

LEARNING MUSIC FROM EACH OTHER: SYNCHRONIZATION, TURN-TAKING, OR IMITATION?

ANDREA SCHIAVIO
University of Graz, Graz, Austria

JAN STUPACHER
Aarhus University, Aarhus, Denmark

RICHARD PARNCUTT
University of Graz, Graz, Austria

RENEE TIMMERS
The University of Sheffield, Sheffield, United Kingdom

IN AN EXPERIMENTAL STUDY, WE INVESTIGATED how well novices can learn from each other in situations of technology-aided musical skill acquisition, comparing joint and solo learning, and learning through imitation, synchronization, and turn-taking. Fifty-four participants became familiar, either solo or in pairs, with three short musical melodies and then individually performed each from memory. Each melody was learned in a different way: participants from the solo group were asked via an instructional video to: 1) play in synchrony with the video, 2) take turns with the video, or 3) imitate the video. Participants from the duo group engaged in the same learning trials, but with a partner. Novices in both groups performed more accurately in pitch and time when learning in synchrony and turn-taking than in imitation. No differences were found between solo and joint learning. These results suggest that musical learning benefits from a shared, in-the-moment, musical experience, where responsibilities and cognitive resources are distributed between biological (i.e., peers) and hybrid (i.e., participant(s) and computer) assemblies.

Received: September 12, 2019, accepted March 19, 2020.

Key words: synchronization, interaction, embodied cognition, music education, joint action

JORDAN AND PERRY ARE GOOD FRIENDS. Growing up together, they share a mutual passion for a number of things, including music and music-making. Recently, Jordan came up with the idea of taking music lessons together, about which Perry was

enthusiastic. While they did not know whether this form of learning was appropriate for them, or useful, they nevertheless wanted to give it a try and keep spending time with each other. However, they had little or no money to attend music classes. And even if they did, the only music teacher available lived far away from their homes. A solution, Perry suggested, might involve watching learning videos on the Internet, where expert musicians offer online courses as well as useful tips to help develop one's musical skills more informally. Video after video, Jordan and Perry learned a variety of exercises and techniques that improved their musicianship in a number of ways. However, this kind of training left the two unsatisfied: they were not actually learning together—the process of learning remained inherently individual. Indeed, online and informal teaching resources are mostly designed for individual learners, and often based on imitation and repetition. How can they share their learning experience and acquire new musical skills together? What is the best strategy for novices to learn music from each other while taking advantage of the possibilities offered by technology? How can they optimize the reciprocal exchange of information and enhance their musical experience? And what musical skills are best acquired and developed together?

In this paper, we report data from an original behavioral study that can help advance some preliminary answers. The study addresses the question of how well novices can learn from each other in situations of technology-aided musical skill acquisition. We expected that reciprocal interaction between musically untrained individuals can lead to enhanced learning experiences and musical outcomes when it involves an active, in-the-moment, form of participation. This can contribute to minimize the “cognitive burden of individual responsibility” (El Zein, Bahrami, & Hertwig, 2019), and help learners co-create a “we-space” (Krueger, 2011)—a shared cognitive niche where skills and mental resources are dynamically negotiated and co-determined as the musical activity unfolds. We thus assessed three different forms of technology-aided peer-to-peer musical learning: synchronization (playing the same musical pattern at the same time), turn-taking (playing different musical

patterns one after the other), and imitation (playing the same musical pattern one after the other), and compared them with corresponding tasks based on individual musical learning. The latter involved individual participants learning the melodies in the same conditions (synchronization, turn-taking, and interaction), alone with a computer. As such, three types of novice-novice-computer interactions were tested and compared to analogous novice-computer interactions. This research builds on previous work on collaborative musical pedagogy (see Gaunt & Westerlund, 2013; Hanken, 2016; Nielsen, Johansen, & Jørgensen, 2018), and is inspired by the conceptual tools of ‘4E’ cognitive science - a cross-disciplinary approach that conceives of mental life as Embodied, Embedded, Extended, and Enactive¹ (Menary, 2010; Newen, DeBruin, & Gallagher, 2018). We suggest that putting these two research avenues together can help frame our work within a broader context involving both practice-oriented (i.e., pedagogical), and theoretical interests. This can contribute a new perspective on the interactive roots of musical learning, helping Jordan and Perry find a way to learn music together.

A PROBLEM IN MIND

While there is a long tradition in music scholarship to explore the various forms of participatory music-making across cultures and styles, questions about the capacity of novices to collaboratively generate novel musical skills have rarely been posed—particularly with regard to the psychological mechanisms involved in the process. A reason for this is that the most common learning approaches in Western contexts are based on sessions of master-apprentice tuition and individual training (Hallam, 1998). “[E]stablished as a continuing core activity in Western classical instrumental learning” (Creech & Gaunt, 2018, p. 145), this pedagogical strategy contributed significant benefits for the lives and musical development of the students (see Gaunt, 2008). For instance, individual tuition is particularly

¹ Put simply, Embodiment emphasizes the mutual specifications between a living system’s body and its ability to think, perceive, feel, act, and communicate (e.g., Gallagher, 2005); Embedded cognition focuses on the ways in which the cultural, physical, and social properties of the environment can shape mental life (e.g., Haugeland, 1998); the main concern of Extended cognition is to explore the fluid integration between internal biological resources (e.g., memory) and external tools (e.g., computers), offering insights into how agents offload (parts of) their cognitive domain into the world to serve specific cognitive functions (e.g., Clark & Chalmers, 1998). Finally, Enactive cognitive scientists argue that mind and life compose a structured unity, intimately dependent on the patterns of ecological activity of an organism and its biological complexity (e.g., Thompson, 2007).

useful to foster the creation of personalized learning strategies, monitor the progress of the student over time, and stimulate verbal communication in a rich variety of situations involving close personal interactions (e.g., Barrett & Gromko, 2007). The psychological model adopted to explain the teacher-student interaction often involves a representational, correspondence-based, schema (see van der Schyff, Schiavio, & Elliott, 2016). Here, pre-given cognitive mechanisms are thought to respond to external perturbations (i.e., musical stimuli, verbal instructions from the teacher, etc.) via sub-personal, rule-based computational processes. The internal elaboration of such stimuli, then, gives rise to task-specific mental representations that assist and guide the student in various ways (see Hallam & Bautista, 2018; Lehmann, 1997). In other words, if there is an adequate correspondence between the internal psychological predisposition of the learner and the information being communicated verbally (or transmitted by the structural antecedents intrinsic to the musical stimulus), a representation with a contextual meaning dependent on the learner’s expertise can be formed² (see Lehmann & Jørgensen, 2018).

Novices like Jordan and Perry, however, can hardly provide each other with the information necessary for novel skills to be developed in such a way. They lack the expertise of a teacher, and do not know how to enhance their reciprocal capacity to engage with the musical material to be learned. And even if information is available from external sources (e.g., a teacher, an instructional video, etc.), then it seems implausible that their reciprocal interaction could play any significant role in improving their performance. In fact, one might expect that an interaction between two or more novices while learning would somehow disrupt the linear mechanisms (elaboration of external information followed by the generation of a given behavioral outcome) of the process. Their focus should arguably remain bounded to their *individual* activity, allowing them to accurately represent their learning goal (e.g., playing a chord), and optimize the best behavioral strategy to achieve it (e.g., the most appropriate fingering solution). But when

² The linear model portrayed here displays important similarities with what has been often described as the “sandwich view” of cognition, standing at the core of traditional cognitive psychology (see Hurley, 1998). The metaphor assumes that cognition—the meat—is segregated by two layers of bread, namely perception and action. By this view, the latter two categories remain at unbridgeable removes from each other and end up being only partially involved in what mental life really is: the cognitive “meat.” The same focus on this middle level can arguably be found in the learning approach based on representations described in this section.

looking at concrete musical practices, musical skills appear to be often developed *collaboratively*. And indeed, one can find many music pedagogies specifically oriented toward joint learning:

“These range from well-established music education approaches developed in the 20th century by Kodaly, Orff, Dalcroze, and Gordon to national interpretations and developments of these approaches, traditional, intercultural approaches, and methods developed in the context of psychology research with specific transfer aims in mind. Such methods involve group learning, shared musical experiences, synchronization, imitation, and a range of other socially interactive behaviors that are common to ‘real-world’ social musical experiences, and could perhaps be used effectively and systematically in music intervention research” (Overy, 2012, p. 66).

In such contexts two or more individuals can jointly learn from a teacher, negotiating their skills and expertise as the lesson unfolds. In a recent qualitative study, Schiavio and colleagues (2018) explored the teaching dynamics behind similar forms of collective music lessons, showing how teachers sometimes “step back” to let their students more actively participate in, and guide, their own learning. Such participatory learning dynamics, it is suggested, are functional to the students’ musical development; by being “less present,” teachers aim to foster a shared sense of responsibility in the group. This way, skills and expertise are negotiated in the moment-to-moment contingencies of the continuous interactions among peers. Similar insights are illustrated in well-known work on informal musical learning, where novices are shown to collaboratively acquire and develop musical skills without prescribed instructions (e.g., Green, 2008). Novices are seen here to explore together possibilities, meanings, and challenges that emerge with the collaborative process in itself, rather than through the intrapersonal elaboration of external stimuli. It has been argued that settings where students are asked to actively participate in group musical activities may foster important benefits in terms of negotiating differences and stimulating trust and social understandings (Higgins & Mantie, 2013), improving critical thinking as well as social skills (Gokhale, 1995), and promoting processes of social inclusion (Welch et al., 2014).

The same focus on *relationships* is found among the main tenets of 4E cognitive science (Ward & Stapleton, 2012). The label indicates a school of thought that conceives of mental life as Embodied, Embedded, Extended,

and Enactive. Putting together insights from theoretical biology, cognitive neuroscience, artificial intelligence, philosophy of mind, and ecological psychology, this framework trades the traditional view of cognition as information-processing occurring in the head, for a more integrated perspective in which mind is understood as an emerging property of the relationship between brain-body systems and the environment in which they are situated (see Gallagher, 2017; Stewart, Gapenne, & Di Paolo, 2010; Varela, Thompson, & Rosch, 1991). The move points to the constitutive role of (inter)active, sensorimotor experience for the living system’s cognitive economy: as organisms of different biological complexity can establish meaningful relationships with their niche via recursive patterns of action and perception, the boundaries between physiology, world, and mental life become blurred (Colombetti & Thompson, 2008). Such relationships include the exchange of information between agents (e.g., speech), the optimization of certain behavioral dispositions to reach a meaningful objective (e.g., the quest for nutrition for an organism), the coupling between living systems and non-biological devices (such as computers or musical instruments), and many other possibilities. If cognitive states are fluidly integrated with bodily and ecological resources on the basis of these continuous sensorimotor loops, then explanation of the mental should entail a focus on such agent-world dynamical couplings (see Barrett, 2011; Hutto & Myin, 2013; Wilson & Golonka, 2013). Recent scholarship in music cognition and music pedagogy increasingly draws from similar insights, and moved away from brain-centered perspectives to explore the broader network of musical bodies, brains, and niches, from which mental life (and music cognition), emerge (see Bowman, 2004; Clarke, 2005; Krueger, 2014; Schiavio & van der Schyff, 2016, 2018).

In this changing landscape, empirical research inspired by the conceptual resources of the 4E approach is still rare. The present contribution aims to address this gap by quantitatively investigating two aspects of the theory that may have strong implications for musical skill acquisition and development. These are: 1) the primacy of sensorimotor experience, that is, the idea that body and action play a fundamental role in driving cognition, including learning; and 2) the coupling between agents and their environment, namely, the capacity of living systems to functionally integrate social and physical elements of their contingent milieu to achieve a specific (cognitive) task. Before presenting our study, however, we first need to explore in more detail the main features of musical settings based on collective

activity and describe three main learning modalities, which are then tested empirically.

JOINT ACTION IN PRACTICE

In learning settings, joint activities where active music-making is prioritized (being facilitated by a teacher or emerging spontaneously within the group of novices), often involve one or more of the following modalities: *synchronization*, *turn-taking*, and *imitation*.

With regard to *synchronization*, a first thing to notice is how pervasive synchronized behavior is for human and non-human life. Consider, for example, how humans tend to synchronize their gait when walking together (van Ulzen, Lamoth, Daffertshofer, Semin, & Beek, 2008), or how non-human animals can also coordinate with each other synchronically to produce periodic signaling (Patel, 2008, p. 408). Acting in synchrony with another person can increase cohesion and social affiliation among members of a group (Hove & Risen, 2009; Stupacher, Maes, Witte, & Wood, 2017), also inducing a sense of compassion and trust (Launay, Dean, & Bailes, 2013; Valdesolo, Ouyang, & Desteno, 2010). The tendency exhibited by listeners to move together at the same time with a musical beat is considered a human universal, and shared musical experiences can enhance prosocial skills in both children and adults (Kirchner & Tomasello, 2010). Producing synchronous behavior in musical contexts, however, can be complicated. Interesting empirical findings by Endedijk and coworkers (2015) showed that when children between 2- and 4-years old play drums in dyads, older children are able to coordinate and adapt to their partner's drumming significantly better when compared to younger ones, even if all subjects were able to produce a steady beat. This aligns with studies that suggest a facilitation for the productions of synchronous behaviors in individuals with musical experience (e.g., Drake, Jones, & Baruch, 2000). So, while the ability to "lock in" to temporal patterns with other individuals is understood as a natural resource that can be highly useful for the "development of executive functions and far transfer effects" (Miendlarzewska & Trost, 2014), its functional role in novices' peer learning needs to be assessed. Can playing in synchrony with each other lead to enhanced musical outcomes in nonmusicians when compared to other modes of learning?

Joint musical activities, of course, involve more than simply playing in synchrony with others. Rather, such contexts are often based on meaningful exchanges, where implicit rules and skills are negotiated and reciprocally developed during performance. Because of this, training based on co-regulated turn-taking behaviors

may also be beneficial. Like synchronized behavior, *turn-taking* can be found early in life. It has been shown, for example, that young infants are highly sensitive to rhythmic patterns of vocal exchanges based on turn-taking (Murray & Trevarthen, 1985). Trevarthen (1999), among others, suggests that infants and mothers are united by a single rhythm: a turn-taking in a slow Adagio (a beat every 0.9 seconds). In a similar vein, Gratier and colleagues (2015) presented evidence of early turn-taking behaviors in young infants (8 to 21 weeks), suggesting a key role for joint activity and preverbal interaction in the development of socio-cognitive skills, communication, and learning. This is particularly interesting when considering how turn-taking is instead usually associated with the dynamics of spoken conversation (Stivers et al., 2009). This joint activity, indeed, "enables the fluency and continuity of natural conversation and entails regulation of the timings of turns at talk. Over different languages and cultures, the conversation participants share an ability to exchange the speakership in tens of milliseconds and, for most of the time, without overlaps" (Hirvenkari et al., 2013). But can learning modalities based on turn-taking help two nonmusicians improve their musical skills? Are there benefits in taking turns while learning to perform a musical phrase?

Finally, *imitative behaviors* are perhaps the most used tool for learning with a teacher. Here, "[c]ues such as direct eye gaze and pointing signal that the expert is about to communicate something which is learning-relevant and generalizable" (McEllin, Knoblich, & Sebanz, 2018), allowing the novice to parse actions into subunits. Building on previous work by Byrne and Russon (1998), Leman (2008), distinguishes between those imitative processes that copy the organization of actions and their goal(s), and those who copy the kinematics of movements. The former (often called "true" imitation) is particularly suited for learning, as it stimulates the agent to produce an action in function of the goal. Imitating goals might build on initial patterns of copied behavior that are then optimized through repetition and performance. While this solution might work well with expert teachers—who can show what is to be copied with confidence and precision—novices might face more difficulties in imitating each other, struggling to extract the "correct" learning-relevant information.³ Nevertheless, it may still serve as a complementary

³For expert musicians this might involve finding the best fingering solution for a passage they saw performed by a famous guitarist during a concert; for novices, imitating a complex musical goal might be more complicated because of the lack of expertise in their motor repertoire.

resource for elementary goals. By imitating each other, for example, peers can become aware of novel possibilities for action and recursively integrate such expertise with each other.

As we saw, these three modalities of joint action can play an important role in music training, social life, and development more generally, stimulating the emergence of various skills at different levels and timescales. Conceiving of musical learning as an individual process based on the realization of inner representations, however, might play down the moment-to-moment, interactive dynamics of such joint behaviors. Traditional models based on a representationalist doctrine, in other words, might display important limitations when addressing the collective emergence of skills that characterize many participatory forms of musical practices. Inspired by the main insights of 4E cognition, we thus designed an experiment that involves a learning task based on action; examines the reciprocal interaction between co-actors; and includes the manipulation of a non-biological device to aid learning (i.e., a computer). We assessed the three learning modalities discussed here (synchronization, turn-taking, and imitation), and systematically compared collective vs. solo learning with both including computer support. Because 4E accounts help eschew the dichotomy between internal psychological domains and external “objective” environment in favor of a more dynamic story based on the constant interplay of brains, bodies, and world, we expected that active forms of reciprocal participation can determine stronger agent-world couplings, leading to more accurate musical outcomes.

Method

PARTICIPANTS

Fifty-four healthy participants (40 females, 14 males, mean age = 23.1 years, $SD = 4.9$) were recruited for this study by the University of Graz mailing list. Eighteen volunteers were randomly assigned to the solo learning group, while the remaining thirty-six formed eighteen pairs for the duo learning conditions. All participants were nonmusicians: none of them regularly performed music, attended music lessons or music schools, or informally learned to play a musical instrument. They reported to listen to music on average 12.1 hours per week ($SD = 8.3$), with their favorite genre being pop ($n = 17$), rock ($n = 12$), indie ($n = 5$), hip-hop ($n = 3$), electronic music ($n = 2$), R'n'B ($n = 2$), classic ($n = 2$), folk ($n = 2$), tango ($n = 1$), country ($n = 1$), jazz ($n = 1$). The remaining

participants ($n = 6$) reported no preferences for specific genres. The thirty-six participants of the duo group were known to each other for an average of 27.28 months. Here, the high standard deviation ($SD = 56.37$) is due to the fact that two participants were brothers and two participants knew each other for nine years, whereas the remaining participants were unknown to each other ($n = 12$), or knew each other for less than four years ($n = 20$). The study was conducted at the Centre for Systematic Musicology of the University of Graz, Austria. All procedures were approved by the Ethical Committee at the University of Graz and were in accordance with the statements of the Declaration of Helsinki. Participants were monetarily compensated for their involvement in the study and provided written informed consent.

MATERIALS

Three melodies were composed by AS for the purpose of this study. All melodies involved the same limited set of pitches (F#, G#, and A#), in different order, style, and rhythm. Each melody was characterized by a defining feature: *Melody 1* was based on a simple melodic pattern, *Melody 2* included simultaneous tones, whereas *Melody 3* was rhythmically challenging. The purpose of such variety was to keep the task engaging for our participants throughout the experiment. The melodies are shown in Figure 1.

The melodies were performed by RP on a digital piano (*Yamaha Clavinova CLP370*), connected to a computer (*MacBook*) via USB. The audio was recorded with the software *Reaper64*. Videos were recorded with a *JVC Everion* digital camera placed above RP's hand (see Figure 2). The three resulting video recordings (one per melody) were edited with *iMovie* software to assist participants in the different learning conditions. The videos were presented to the participants over a laptop (another *MacBook*) placed over the digital piano. Participants only received instructions using these videos. Music notation was not shown.

EXPERIMENTAL DESIGN

Prior to the testing, participants were assigned to one of two groups: *solo* ($n = 18$) or *duo* ($n = 36$). Participants of the solo group took part in the whole study individually, while participants of the duo group were paired with another participant (see Figure 3). The experiment involved three blocks, each divided into a learning phase and a performance phase. For the entire duration of the experiment, a researcher was present in the experimental room to assist participants, monitor the data



FIGURE 3. Participants learned the melodies either together or individually.

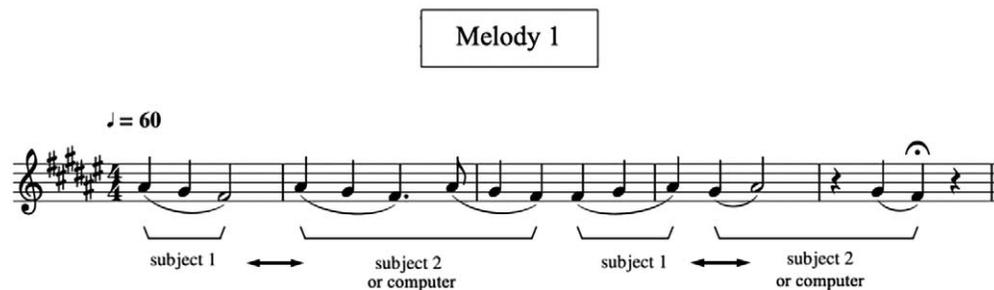


FIGURE 4. Example of segmentation of a melody for turn-taking and imitation condition. Here, Melody 1 was divided into four different sections. In turn-taking conditions, if subject 1 plays the first bit, then either subject 2 (in the duo group) or the computer (in the solo group) plays the following one, and so on for six minutes. After this phase, subject 1 plays the part previously played by the machine or by subject 2. In the imitation condition: after each section, a pause in the video allowed participants to imitate the given musical fragment (in this example, the first three notes). Participants were thus instructed not to play along with the video, but rather to imitate what was seen in the video and either show it to another participant (who then has to play the section herself), or play it individually in the solo condition.

and their association with the melodies were pseudorandomized to avoid order artifacts. All learning phases were video recorded with a *JVC Everion* digital camera placed at around two meters on the back of the participant(s). Videos were coded offline by a coder blind to the study. A representation of all learning phases is illustrated in Table 1. The coder watched all videos and reported the time spent by participants: 1) talking to each other (duo) or to the researcher (solo); 2) playing according to the learning condition; 3) interacting with the computer (e.g., stopping the video, replaying a specific passage, etc.); 4) observing the instructional videos; and 5) resting (e.g., periods of silence or inaction, where participants tried to remember the motor sequence).

Performance phase. Each learning phase was followed by a performance phase, in which the participants' ability to perform the learned melodies was assessed.

Participants of both groups individually performed the learned melody without any visual or acoustic information—in the same octave register as in the learning phase—in the original tempo (60 BPM), and in a faster tempo (90 BPM). Participants had the chance to perform each melody three times in each tempo, and then chose the two best performances (one in the original tempo, one in the fast tempo). Best performances were chosen to give participants the best opportunity of success, reducing the inclusion of noisy data due to performance errors. An additional benefit is a likely reduction of anxiety to perform accurately. The faster tempo was chosen to challenge participants and test whether learning was robust and adaptable. Playing at a faster tempo may help to bring out weaknesses in learning, more so than when exactly the same task was repeated. The tempo was given by a metronome prior to the performances.

PROCEDURE

Having obtained written informed consent, participants (individuals or pairs) were invited to enter the experimental room and were instructed about the aims of the study (i.e., they were told that their ability to perform different melodies after three learning phases was tested). Participants were comfortably seated in front of the digital piano, above which the laptop was placed. Depending on melody and condition, the corresponding instructional video was looped on the laptop for a total of twelve minutes. In the *solo* group, participants played the exact same notes as in the original melody. In the *duo* group, participants played one octave lower or higher, depending on where they were seated. During the learning phase, participants were instructed to interact with the laptop as they like in order to optimize their learning. For example, they were free to stop the video and focus on a particularly difficult passage, or talk to

their partner about a given issue. If they felt ready before the twelve minutes ended, they asked the experimenter to terminate the learning phase. After the third, final, performance phase, each subject completed a questionnaire involving a total of 13 items (*solo*) or 16 items (*duo*). These comprised an initial section on the background and preferences in music listening of participants, followed by a series of Likert-scale items dedicated to assess the felt quality of their learning experience. The questionnaire for duos also featured the 7-point single-item “Inclusion of Other in the Self” (IOS) scale, originally developed by Aron and colleagues (1992), and since then adopted to measure how close respondents feel to another individual or social group (e.g., Himberg et al., 2018; Stupacher et al., 2017). A representation of the experimental procedure is depicted in Figure 5.

TABLE 1. Representation of all Learning Phases

Learning phase	Description
Imitation	Alternation between absorption through listening and self-driven performance.
Turn-taking	As imitation, but requires active readiness to continue, strengthening the need of anticipating the other’s performance.
Synchronization	Requires online listening and adaptation, affords continuous coupling, risks interference through mistakes.

SOLO

Learning phase	Part 1	Part 2
Imitation	Imitating all 4 sections one after the other (12’ max)	–
Turn-taking	Playing sections 1 and 3 (6’ max)	Playing sections 2 and 4 (6’ max)
Synchronization	Playing the entire melody (12’ max)	–

DUO

Learning phase	Part 1		Part 2	
	Subject 1	Subject 2	Subject 1	Subject 2
Imitation	Playing sections 1 and 2 (6’ max)	Imitating section 1 and 2 (6’ max)	Imitating section 3 and 4 (6’ max)	Playing sections 3 and 4 (6’ max)
Turn-taking	Playing sections 1 and 3 (6’ max)	Playing sections 2 and 4 (6’ max)	Playing sections 2 and 4 (6’ max)	Playing sections 1 and 3 (6’ max)
Synchronization	Playing the entire melody (12’ max)	Playing the entire melody (12’ max)	–	–

DEPENDENT VARIABLES

Two measures of accuracy of the performances of the melodies were employed as dependent variables, which captured pitch similarity and temporal similarity.

Pitch similarity. The similarity of notes played by the participants in the final performances and the original melodies was calculated with the *edr* (edit distance on real signals) function in MATLAB (MathWorks, Natick, MA). Notes were represented as MIDI numbers. The *edr* function returned the number of notes that a participant played wrong, missed, or added in comparison to the original melodies. The *edr* outcome was divided by the total number of notes in the original melodies, to normalize pitch similarity. In the *duo* group, the *edr* function was additionally used to assess the pitch similarity of the two performances of the paired participants. Here, the mean of the number of notes played by the participants was used to normalize the *edr* outcome. As Melody 2 mostly consisted of chords, notes played in a time window of 60 ms were considered as part of a chord and the order of these notes was irrelevant. To give an example, if the chord of the original

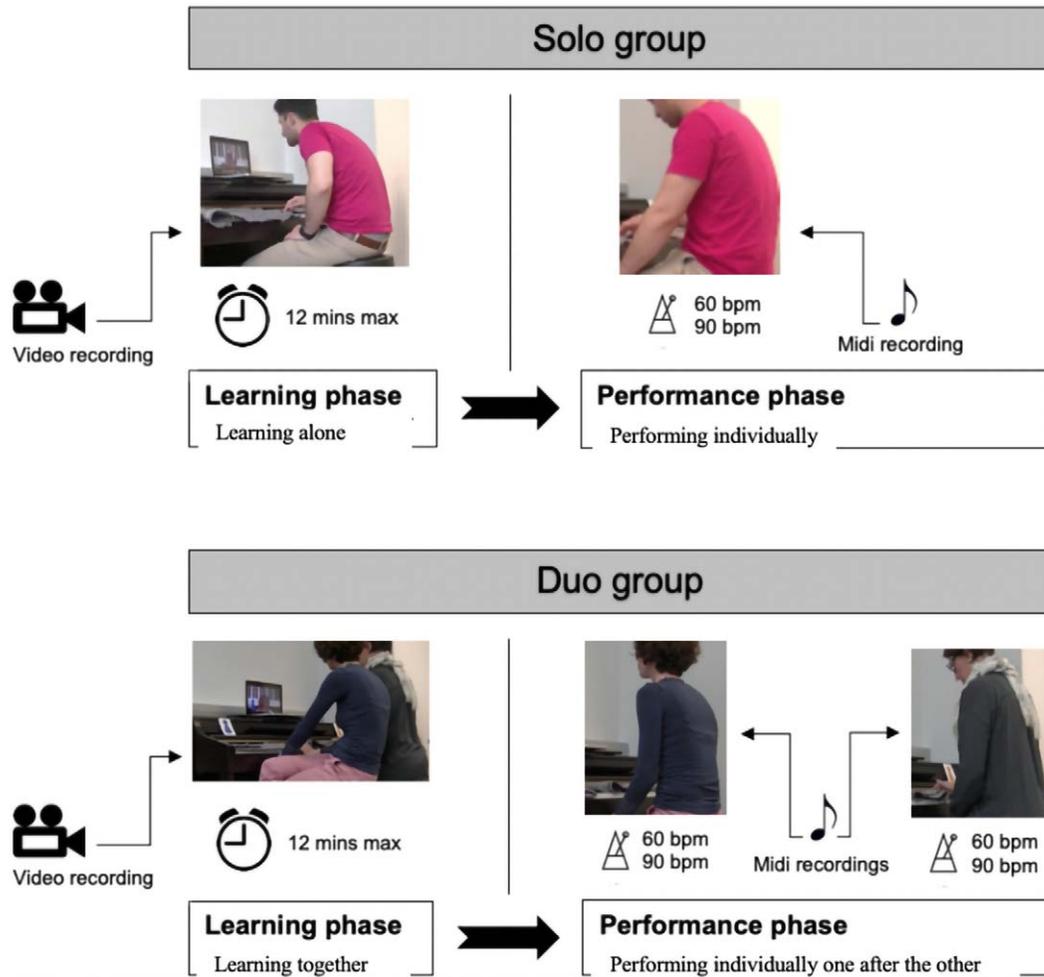


FIGURE 5. Experimental procedure per group. Each block is repeated three times with changes in melody and learning condition.

melody consisted of the MIDI notes 66, 68, and 70 being played at the same time, and the participant played the notes 68, 70, and 66 within 60 ms, the order of these notes was marked as correct.

Temporal similarity. The similarity of the time series in performed and original melodies was calculated with the *dtw* (distance between signals using dynamic time warping) function in MATLAB. The function stretched the performed and original melodies such that the sum of all asynchronies between the two melodies was the smallest. This was done to account for performances that were played too fast or too slow. The outcome was the sum of all asynchronies between the performed and original melody in milliseconds. To normalize temporal similarity, the sum of the asynchronies was divided by the absolute time of the original melody. In the *duo*

group, the *dtw* function was additionally used to assess the temporal similarity of the two performances of the paired participants. Here, the mean of the absolute time of performances was used to normalize the *dtw* outcome. For the calculation of temporal similarity between the original stimulus and the faster performance (count in metronome with 90 BPM), the time series of the original stimulus was multiplied by 0.6667 before using the *dtw* function.

STATISTICAL ANALYSIS

The distributions of pitch similarity outcomes (i.e., the normalized edit distances) were heavily right-skewed. This is because many comparisons of performance and original melody revealed no wrong notes (i.e., a value of zero). Thus, Wilcoxon rank tests for paired samples were used to assess differences between the three

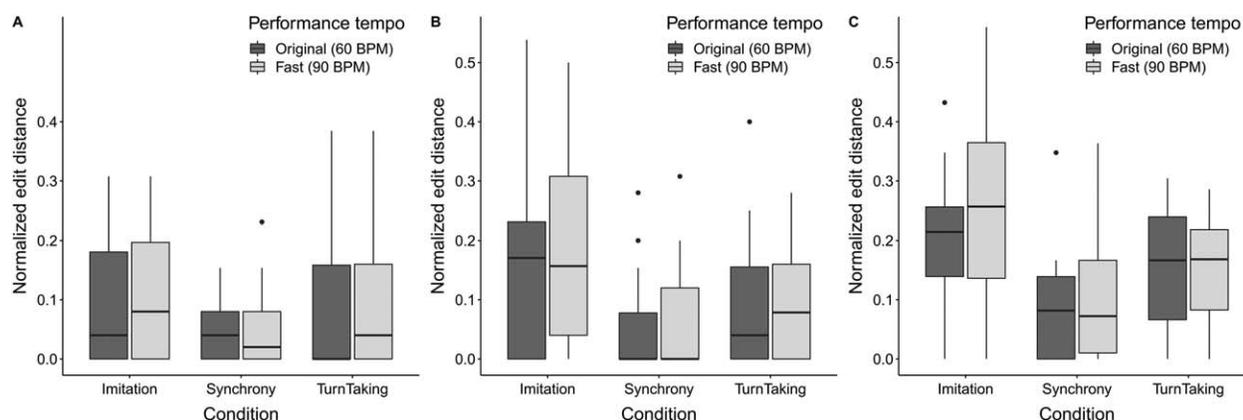


FIGURE 6. Tukey boxplots visualizing pitch similarity operationalized as the normalized edit distance between two vectors including MIDI pitch values. A) Solo group: Normalized edit distances between performances and original stimulus; B) Duo group: Normalized edit distances between performances and original stimulus; C) Duo group: Normalized edit distances between the two performances. Whiskers represent lowest and highest values within 1.5 * interquartile range (IQR) of the lower and upper quartiles, respectively. Dots represent values outside 1.5 * IQR.

learning conditions. As a visual data inspection of the boxplots of normalized edit distances suggested no interaction between condition and performance tempo (Figure 6), comparisons between the conditions were based on both performance tempi. The effect size is computed as $r = Z/\sqrt{N}$. The temporal similarity outcomes (i.e., the normalized Euclidean distance) were ln-transformed to meet the assumptions of normality and entered into repeated measures ANOVAs with the factors Condition (imitation, synchrony, turn-taking) and Performance tempo (original, fast). Post hoc comparisons were Bonferroni corrected.

Results

PITCH SIMILARITY

Solo group - Comparison of performance and original stimulus (Figure 6A). In the imitation condition, participants played more wrong notes than in the synchrony condition ($Z = 2.03, p = .042, r = .34$). The differences between imitation and turn-taking ($Z = 0.47, p = .646, r = .08$), and synchrony and turn-taking ($Z = -1.13, p = .264, r = .19$) conditions, were nonsignificant.

Duo group - Comparison of performance and original stimulus (Figure 6B). In the imitation condition, participants played more wrong notes than in the synchrony condition ($Z = 5.22, p < .001, r = .87$) and the turn-taking condition ($Z = 3.51, p < .001, r = .59$). In the synchrony condition, participants played fewer wrong notes than in the turn-taking condition ($Z = -2.90, p = .003, r = .48$).

Duo group - Comparison of performances in duos (Figure 6C). In the imitation condition, participants played more wrong notes than in the synchrony condition ($Z = 4.08, p < .001, r = .68$) and the turn-taking condition ($Z = 2.85, p = .004, r = .48$). In the synchrony condition, participants played fewer wrong notes than in the turn-taking condition ($Z = -2.11, p = .034, r = .35$).

Solo vs. duo group. Three individual Mann-Whitney U tests on the normalized edit distances (one per learning condition) revealed no significant difference between the solo and the duo performances (all $p > .13$).

TEMPORAL SIMILARITY

Solo group - Comparison of performance and original stimulus (Figure 7A). An ANOVA on the ln-transformed normalized Euclidean distances revealed no significant main effect of condition, $F(2, 34) = 1.33, p = .278, \eta^2 = .07$, or performance tempo, $F(1, 17) = 3.66, p = .073, \eta^2 = .18$, and no interaction between the two factors, $F(2, 34) = 0.80, p = .804, \eta^2 = .05$.

Duo group - Comparison of performance and original stimulus (Figure 7B). An ANOVA on the ln-transformed normalized Euclidean distances revealed a significant main effect of condition, $F(2, 70) = 8.47, p < .001, \eta^2 = .20$, with larger distances for imitation than for synchrony ($p = .001, d = .65$) and turn-taking ($p = .041, d = .43$) and no significant difference between synchrony and turn-taking conditions ($p = .361, d = .27$). Euclidean distances also differed significantly between performance tempi, $F(1, 35) = 6.61, p = .015, \eta^2 = .16$, with smaller distances for the original

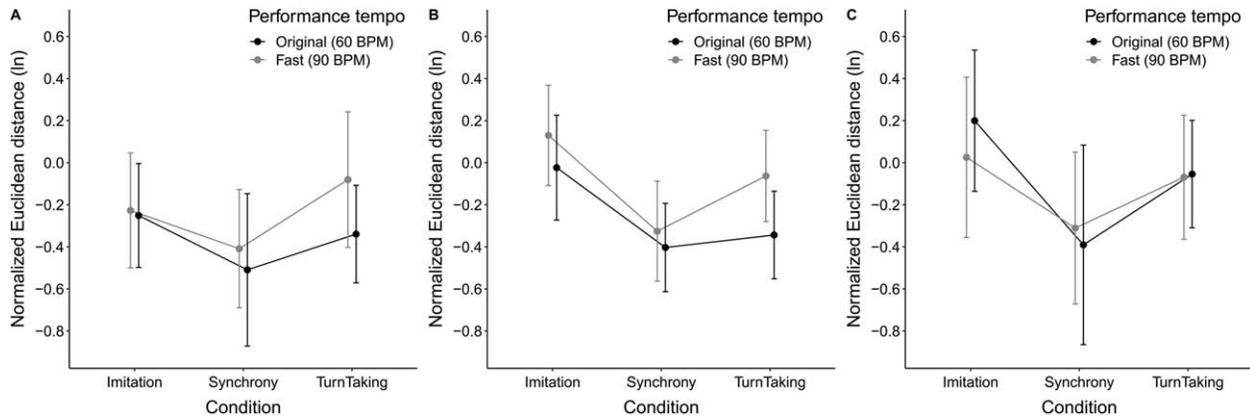


FIGURE 7. Temporal similarity measures operationalized as the normalized Euclidean distance between two time series. A) Solo group: In-transformed sum of Euclidean distances between performances and the original stimulus; B) Duo group: In-transformed sum of Euclidean distances between performances and original stimulus; C) Duo group: In-transformed sum of Euclidean distances between the two performances. Error bars: 95% CI.

performance tempo compared to the faster performance tempo ($p = .015$, $d = .43$). The interaction between the two factors was nonsignificant, $F(2, 70) = 1.55$, $p = .220$, $\eta^2 = .04$.

Duo group - Comparison of performances in duos (Figure 7C). An ANOVA on the ln-transformed normalized Euclidean distances revealed a significant main effect of condition, $F(2, 34) = 4.63$, $p = .017$, $\eta^2 = .21$, with larger distances for imitation than for synchrony ($p = .006$, $d = .86$). The other two post hoc comparisons were nonsignificant (both $p > .30$). No significant difference was found for performance tempo, $F(1,17) = 0.11$, $p = .743$, $\eta^2 = .01$, and the two factors showed no significant interaction ($F(2,34) = 0.54$, $p = .588$, $\eta^2 = .03$).

Solo vs. duo group. An ANOVA on the averaged ln-transformed normalized Euclidean distances of original and fast tempo performances with the within-subjects factor learning condition and the between-subjects factor group (solo vs. duo) revealed a significant main effect of condition, $F(2, 104) = 6.00$, $p = .003$, $\eta^2 = .10$, but no significant main effect of group, $F(1, 52) = 1.05$, $p = .311$, $\eta^2 = .02$, and no significant interaction between condition and group, $F(2, 104) = 1.23$, $p = .297$, $\eta^2 = .02$. Distances were smaller for synchrony compared to imitation ($p = .002$, $d = .49$). The other two post hoc comparisons were nonsignificant (imitation vs. turn-taking, $p = .169$, $d = .27$; synchrony vs. turn-taking, $p = .084$, $d = .31$).

VIDEO CODING

Learning durations. On average, participants of the duo group engaged in learning phases based on

synchronization for 9 min 40 s ($SD = 2.48$); imitation for 9 min 55 s ($SD = 2.41$); and turn-taking for 10 min 53 s ($SD = 2.52$). In the solo group, participants engaged with synchronous learning for an average of 9 min 37 s ($SD = 3.09$); learning based on imitation for 8 min 40 s ($SD = 2.84$); and learning based on turn-taking for 10 min 16 s ($SD = 2.43$).

Activities during learning phase. Time spent playing piano: The analyses of the video coding data (Table 2) indicate that in the solo as well as in the duo group, participants spent less time playing in the imitation condition (Solo: $M = 2$ min 44 s, Duo: $M = 4$ min 58 s) than in the synchrony (Solo: $M = 6$ min 12 s, Duo: $M = 6$ min 44 s) and turn-taking (Solo: $M = 3$ min 13 s, Duo: $M = 7$ min 13 s) conditions. In the solo group participants also spent less time playing in the turn-taking condition compared to the synchrony condition. Within each learning condition, the time spent playing was not significantly correlated with pitch or temporal performance measures (Table 3). *Time spent talking:* The only significant difference of talking time occurred in the comparison of synchrony and turn-taking conditions in the duo group ($M = 2$ min 12 s, and $M = 1$ min 30 s, respectively). *Time spent interacting with the computer:* In the duo group, participants spent more time manipulating the video on the computer (for example, repeating a specific section of the video, or stopping the loop) in the synchrony ($M = 18$ s) compared to the imitation ($M = 5$ s) and turn-taking ($M = 3$ s) conditions. In the solo group, participants spent more time interacting with the computer in the synchrony ($M = 21$ s) compared to the turn-taking ($M = 2$ s) condition,

TABLE 2. Comparison of the Time Spent for Specific Activities Between Learning Conditions in Solo and Duo Groups

	Solo group			Duo group		
	Imitation vs. synchrony	Imitation vs. turn-taking	Synchrony vs. turn-taking	Imitation vs. synchrony	Imitation vs. turn-taking	Synchrony vs. turn-taking
Time playing	$t(17) = -8.92,$ $p < .001$	$t(17) = -3.35,$ $p = .004$	$t(17) = 7.01,$ $p < .001$	$t(17) = -3.87,$ $p = .001$	$t(17) = -6.00,$ $p < .001$	$t(17) = -1.01,$ $p = .328$
Time talking ¹	$Z = 0.97,$ $p = .343$	$Z = 0.55,$ $p = .607$	$Z = -0.46,$ $p = .669$	$Z = -1.55,$ $p = .130$	$Z = 0.76,$ $p = .468$	$Z = 2.11,$ $p = .034$
Time interacting with computer	$Z = -1.60,$ $p = .116$	$Z = 1.02,$ $p = .326$	$Z = 2.60,$ $p = .007$	$Z = -2.14,$ $p = .032$	$Z = 0.52,$ $p = .628$	$Z = 2.78,$ $p = .004$
Time watching the video	$Z = -3.59,$ $p < .001$	$Z = -3.72,$ $p < .001$	$Z = 0.37,$ $p = .734$	$Z = -3.29,$ $p < .001$	$Z = -3.72,$ $p < .001$	$Z = -2.29,$ $p = .021$

¹ Solo group: talking with experimenter; Duo group: talking with each other.

TABLE 3. Spearman Correlations for Pitch Similarity Measures and Pearson Correlations for Temporal Similarity Measures Between the Time Spent Playing in One Condition and the Related Performance Measures

		Solo group: Time spent playing ($N = 18$)			Duo group: Time spent playing ($N = 36$)		
		Imitation	Synchrony	Turn-Taking	Imitation	Synchrony	Turn-Taking
Pitch similarity	Imitation	$\rho = .29,$ $p = .236$	—	—	$\rho = -.27,$ $p = .113$	—	—
	Synchrony	—	$\rho = .33,$ $p = .186$	—	—	$\rho = -.14,$ $p = .405$	—
	Turn-Taking	—	—	$\rho = -.16,$ $p = .519$	—	—	$\rho = .03,$ $p = .871$
Temporal similarity	Imitation	$r = .19,$ $p = .463$	—	—	$r = .02,$ $p = .903$	—	—
	Synchrony	—	$r = -.45,$ $p = .059$	—	—	$r = -.12,$ $p = .481$	—
	Turn-Taking	—	—	$r = -.04,$ $p = .865$	—	—	$r = -.09,$ $p = .602$

Note: Mean values for performances in original tempo and fast tempo.

whereas imitation ($M = 8$ s) did not differ significantly from the other two conditions. *Time spent watching the video*: In the duo group, participants spent most time watching the instruction video in the turn-taking condition ($M = 10$ min 11 s), followed by synchrony ($M = 8$ min 28 s) and imitation ($M = 5$ min 1 s). In the solo group, participants spent more time watching the video in turn-taking ($M = 9$ min 31 s) and synchrony ($M = 9$ min 24 s) compared to imitation ($M = 5$ min 30 s). We should note here that one could be engaged in two or more of such activities at the same time: e.g., one can talk while playing and watching the video.

QUESTIONNAIRES

Subjective ratings of learning and performance. Participants of both groups completed a written post-test questionnaire based on a series of Likert-scale items dedicated to assess the felt quality of their learning

experience (range 1–7). On average, the overall enjoyment of learning was ranked 6.25 ($SD = 0.84$), and satisfaction with their own learning 4.81 ($SD = 1.38$). Confidence about learning was 4.88 ($SD = 1.65$), while feeling of being nervous was 4.22 ($SD = 1.72$). Confidence of performance was 4.81 ($SD = 1.49$), and nervousness while performing 4.20 ($SD = 1.89$). Finally, an average of 5.96 ($SD = 1.03$) was reported with regards to motivation to learn music in the future. These ratings are averaged from both groups, and did not significantly differ between the solo and duo group (p values ranging from .14 to .99). Participants of the duo group reported a level of satisfaction with their partner's learning ability of 5.88 ($SD = 0.69$) and ranked the quality of their reciprocal interaction as 5.80 ($SD = 0.84$). The Inclusion of Other in the Self (IOS; Aron et al., 1992) in the duo group was rated as $M = 4.38$ ($SD = 1.49$) on a scale from 1 (*no connectedness*) to 7 (*very high*

connectedness). IOS ratings were not significantly correlated with the performance measures (pitch similarity and temporal similarity). All participants were also asked to indicate the learning condition they preferred: synchrony was the top choice ($n = 24$), followed by imitation ($n = 16$) and turn-taking ($n = 14$).

Discussion and Conclusion

Our study addressed the question of whether technology-aided collaborations rooted in reciprocal interactions can help enhance musical learning in nonmusicians. We operationalized collaborative learning as three main learning conditions based on synchronization, imitation, and turn-taking; we then compared participants involved in joint learning with subjects involved in corresponding tasks based on individual learning. In assessing post-training individual performances, we found that pitch and temporal cues of the newly learned musical excerpts were generally more accurate when participants engaged in synchronous learning and turn-taking, over imitation. As no significant differences of performance measures (i.e., correct pitch and timing) were found between solo and duo groups, our results suggest that novices can maximize their learning in both individual and collective settings when they *actively (co-)participate* in the generation of musical material, rather than imitating external stimuli. The lack of significant differences between groups might be associated to the key role of computers in the learning process. Because of the intimate coupling developed between learners and the device they interact with, the functional role of the latter during individual training becomes comparable to that played by peers. While this does not imply that, phenomenologically speaking, peers and technology are “present” to each other in the same way, the proposed insight is consistent with the idea of fluid integration of biological and non-biological systems at the core of the extended mind thesis. A suggestive implication of this scenario points to an original account of how learning responsibilities are distributed within the large cognitive assembly: as computers arguably become part of the learner’s cognitive ecology, the responsibilities peers can develop for each other during collaborative tasks might be efficiently offloaded into the device in the absence of the partner; this allows the hybrid network to be equally conducive to optimal learning when compared to peer-to-peer occurrences, adequately compensating for the “loss” of one of the peers.

JOINT LEARNING

In the performance phase, participants of the duo group played more right notes when the learning was done in

synchrony with others, when compared to learning based on turn-taking and imitation. Looking at temporal accuracy, we found that participants were more precise when the melodies were learned in synchrony or turn-taking, when compared to imitation. We suggest that the preference for contextual collaborations based on a moment-to-moment interaction, points to a sense of shared responsibility⁴: when engaging in synchronous or turn-taking behaviors with a partner, a single mistake can immediately compromise the shared outcome of a given musical passage. Conversely, in imitative tasks, one could make mistakes, and repeat the pattern to be copied by the other without too much worry. Surprisingly, though, the latter makes the actors less prone to learn effectively. With regard to this point, one could ask whether our results (with imitation being the less efficient learning condition overall) could not be determined by the level of expertise of peers: as they are both novices, it is not surprising to see no real benefit when it comes to imitative learning. You simply *cannot* show to your partner how to play a melody on the piano correctly if you do not know how to do it. If there was an expert, instead, the quality of the information being transmitted from one another would have been certainly better, allowing the imitator to learn efficiently without any need to play in synchrony or in turn-taking with the other. If this was the case, however, we would have likely witnessed data pointing to the benefits of imitation in individual settings, where the novice can freely observe the video and learn from the expert playing in it. Instead, our data show that also participants of the solo group did not benefit from learning based on imitation. Another possible confound is that synchronization might be regarded as the condition with the most reinforcement of the instructional video, which may have led to more effective learning. However, in both collective and individual learning, participants spent most time watching the instruction video in the turn-taking condition, suggesting that even in situations where one does not have to play continuously, technologies can provide additional audio-visual support to enhance learning. It should also be noted that participants in the duo group (alike those in the solo group reported below) spent more time playing the piano during learning phases based on turn-taking and synchronization when compared to imitation. However, this potential confound was arguably mitigated by the possible flexibility exhibited by imitative behaviors

⁴ Here, shared responsibility refers to a condition where the success of the duo (learning the melodies) is equally distributed between co-actors, and not concentrated in a single individual.

involving two participants (when compared to learning based on synchronization and turn-taking, as well as to imitations based on repeating visual information delivered via computer, as those involved in individual learning). Moreover, the total exposure to the melodies was not reduced during imitation, as the latter included being visually informed about aspects of the stimulus that may not be easily captured by playing only.

INDIVIDUAL LEARNING

Participants of the solo group performed more wrong notes during performance done after imitative training, when compared to synchronous learning; and although no significant differences emerged when comparing temporal accuracy across learning conditions, our results remain somewhat surprising given that much musical practices in Western tradition entails forms of imitative behavior, where experts produce and show an exemplary model to novices, who are then asked to imitate. A factor that could have possibly contributed to this result is that learning phases based on imitation and turn taking (as a direct consequence of the experimental design) involved less actual playing time than learning based on synchronous playing. However, the segmentation of the musical material into chunks, central to imitative learning and turn-taking, allowed our participants to focus on specific issues of the musical passages being performed, which could have not been addressed within other settings. For instance, in light of said focus on chunks rather than overall phrases, in imitative settings one could more easily discover and repeat a particularly challenging fingering configuration. This might have compensated for the difference in time spent playing. Moreover, individual performances measured after learning based on turn-taking (where learning is also based on different chunks rather than continuous playing) were generally more accurate than those analyzed after imitative learning; this suggests that what matters is the quality (i.e., the condition) of learning, rather than time spent on the task (Williamson & Valentine, 2000).

CONSTITUTING COGNITIVE ASSEMBLIES FOR MUSICAL ACTION AND LEARNING

When looking at both individual and collective learning (with the notable exception of temporal accuracy in the solo group), it appears that musical skills are better acquired when the target melodies are performed collaboratively with the video or with the other peer, rather than linearly observed, internalized, and transformed into behavioral outcomes as in the imitation condition. We argue that this resonates with existing literature that

stresses the role of collaboration, interaction, and co-presence for cognitive life more generally (see De Jaegher & Di Paolo, 2007; De Jaegher, Di Paolo, & Gallagher, 2010; Di Paolo, Cuffari, & De Jaegher, 2018). Along these lines, it has been argued that information concerning jointly performed activities might be processed more efficiently when compared to other information, resulting in enhanced accuracy for remembering and performing a given task (see Mesoudi, Whiten, & Dunbar, 2006; Smith & Semin, 2004). This might depend on the kind of co-regulation occurring “when both partners are responsive to mutual influence, resulting in the emergence of new information not previously available to participants prior to their joint engagement” (Krueger, 2011, p. 645). We suggest synchronous learning and turn-taking to be good examples of such dynamical interplay rooted in bodily action, and achieved via back and forth processes of reciprocal causation (Clark, 1997).

This is consistent with the main insights from 4E cognitive science discussed above: as living systems create their own experience through action rather than through representational recovery, much of their cognitive life is dynamically looped with social and physical resources of their bodies and surrounding environments (Chemero, 2009; Thompson, 2007; see also Malafouris, 2013, 2015). The intersection of gestures and actions emerging from learning settings based on synchrony and turn-taking appears to facilitate the collective and individual acquisition and development of skill (see Schiavio, Gesbert, Reybrouck, Hauw, & Parncutt, 2019), being implemented by “ongoing feedback loops that transform our cognitive profile in real-time and help us negotiate complex cognitive tasks” (Krueger, 2019, p. 48). This is particularly interesting from an “extended mind” perspective, which is increasingly applied to musical contexts (Cochrane, 2008; Kersten, 2017; Ryan & Schiavio, 2019): the theory holds that under certain conditions, living systems can realize part of their cognitive processes thanks to various forms of external scaffolding involving physical and social elements (Clark & Chalmers, 1998). Consider how we use musical notation to facilitate the performance of a piece that might be too demanding to be remembered. It is argued that because we use the information stored on the musical score as if the same information were stored in our biological memory, there is a functional similarity between “internal” and “external” domains, which can enrich our cognitive capacity (e.g., remembering a non-written fingering configuration while playing music written on the score). As such, the achievement of a cognitive task can often depend on the fluid integration of

biological and non-biological resources forming a hybrid cognitive assembly.⁵ In social situations, this might involve the complex musical interplay brought forth by groups such as orchestras or ensembles, where the reciprocal interaction of all performers constitutes the final musical outcome. With regards to learning music, our study shows that the continuous integration of social and physical resources is more effective when it involves a moment-to-moment, online participation, in which action is prioritized over observation and repetition. This gives rise to the constitution of a distributed cognitive system where biological (peers) and non-biological (computers) resources are functionally coupled, possibly leading to a reduced cognitive load. If such resources are partially distributed in a shared space for musical action, tasks can be achieved more easily, resulting in enhanced musical outcomes. Such insights align with pedagogical perspectives based on an “action-first” philosophy, which put major emphasis on learning settings where students are encouraged to *act* musically together—and explore their creativity—from the very beginning of their musical journey (see e.g., Borgo, 2005, 2007; Schiavio, van der Schyff, Gande, & Kruse-Weber, 2019).

Finally, our study provides further empirical grounding to existing cross-disciplinary line of research that stresses the importance of synchrony for human artistic and social behavior (see e.g., Bowling, Herbst, & Fitch, 2013; Merker, 2000). In evolutionary terms, for example, it is argued that the “prosocial consequences of interpersonal synchrony conferred a fitness advantage on individuals in groups that practiced music” (Ravignani, Bowling, & Fitch, 2014). The importance of synchrony has been also addressed from an ontogenetical standpoint: as reported by Baimel and co-workers (2015), in learning how to synchronize with others or through spatiotemporal coordination to an external stimulus, young children can successfully (re)define the boundaries between self and other, “while simultaneously allowing for effective navigation of those boundaries in fostering efficient interpersonal

coordination” (Baimel, Severson, Baron, & Birch, 2015). In all, it can be argued that synchrony contributes to the development of a shared space for action, a performative niche in which agents can co-create a unique experience, and consequently learn new skills. In inter-subjective contexts, this aligns with experimental work from Paladino and colleagues (2010) on self-other merging, where it was found that the ability to distinguish one’s face from those of whom a subject has been synched up with, is affected by synchronous multisensory stimulation. As distinctions from self to other are blurred, cognitive resources can be distributed across agents more fluidly, giving rise to enhanced behavioral outcomes. Overy and Molnar-Szackacs (2009), among others, argue that multiagent synchrony can constitute a shared musical environment where the prediction of others’ behavior is facilitated (see also Schaefer & Overy, 2015). Here music can convey a sense of agency: a feeling of presence of the other that can help anticipate and understand each other. And indeed, “the experience of being synchronized together in time, and yet with a musical, human flexibility and variety creates a powerful sense of togetherness [. . .]. The emerging sound is a group sound, almost ‘larger than life,’ created by a sense of shared purpose” (Overy & Molnar Szackacs, 2009, p. 495).

We should note here that our data failed to show a relation between performance quality (most importantly temporal synchrony) and interpersonal connectedness as measured by the IOS scale. A potential reason for this might involve the focus on individual performance throughout the study. While participants in the duo group learned together, they also performed individually: such discontinuity might have contributed to prevent social connectedness in the reported form. Having participants performing together, however, might have introduced additional confounding effects, such as nervousness of playing in front of the peer. This brings us to one of the main limitations of the study, that is, a lack of focus concerning the emotional aspects involved in the dynamics of learning and interacting. This is indeed a fruitful research avenue, which features a growing number of studies in the cognitive psychology and neuroscience of social collaboration. Here, as reported by Clark and Dumas (2015), recent work in the field provides additional evidence that concerns the recruitment of the mesolimbic dopamine reward system during peer-interaction, pointing to a mutual feeling of fulfillment and pleasure in shared activity (Pfeiffer et al., 2014; Redcay et al., 2010). As a number of recent studies show how emotionality contributes to the development of the motivational cues adopted to facilitate coordination (see

⁵ It should be noted that some scholars remain skeptical of the explanatory power and logical consistency of the extended mind approach. Among others, Adams & Aizawa (2010) argue that the theory operates under a “coupling-constitution fallacy,” namely it confuses the causal relationship between mind and environment with a constitutive one. Put simply, while it can be true that living systems are closely coupled with external tools, it would be a mistake to conceive of the latter as actual “parts” of the mind (see also Ross & Ladyman, 2010). According to Clark (2010), however, such criticism ultimately fails to appreciate the functional significance of various forms of coupling for large-scale systems, leaving the core thesis fundamentally untouched.

Michael, 2011), new empirical work might address this point more clearly, exploring how musical emotions play a role in driving learning processes individually and collaboratively, and how they contribute to form cognitive assemblies with biological and non-biological tools from the environment (see also Colombetti, 2014). This can complement the focus of the present study on action and interaction, and offer novel fascinating possibilities for concrete musical settings and learning contexts. For example, more attention dedicated to the emotional aspects involved in the process might provide future studies with different options to avoid stress-reduction in participants (in the present study, as reported above, we let participants chose their best performance out of three possibilities). Another possible limitation involves the differences in time spent playing by participants during the three learning conditions. These differences are a result of the design of the learning conditions and often occur in real-world practice; playing in synchrony will usually result in more spent playing time than imitating or taking turns. However, we demonstrated that within each learning condition, the time spent playing was not associated with better or worse performance measures. Moreover, our results indicate that active playing time may be more beneficial than total exposure time that partially includes active playing. To better account for this difference, future research might develop experimental paradigms that more specifically control for actual playing time during each learning trial. This work could also explore whether the key role of synchronization and turn-taking in a novice' learning experiences is also present in expert musicians. Comparing novices and experts might offer valuable insights into the cognitive mechanisms associated with musical learning, and shed new light on their coupling with the social and physical elements of the environment in which they are embedded.

CONCLUSION

We examined performance accuracy of novices who learned to play short musical pieces on the piano through individual or joint learning. We contrasted methods based on synchronization, turn-taking, and imitation (all assisted by instructional videos on a computer). We found that in both solo and duo settings, imitation is the least efficient condition overall, with learning settings based on synchrony and turn-taking leading to better pitch and tempo accuracy. These results are consistent with cross-disciplinary literature that highlights the role of co-presence and embodied action for musical learning, and that calls into question too rigid separations between the social, the individual, and the non-biological in the study of human cognition

and skill acquisition. Novices like Jordan and Perry, to come back to our initial vignette, might benefit from technology-aided learning settings rooted in reciprocal interaction as much as their friends who learn music individually. As long as active musicking and moment-to-moment participation are prioritized, novices can share their musical journey from their very first lesson, and participating in each other's learning. This opens up fascinating possibilities for music educators interested in developing novel teaching resources. For example, future computer-assisted music pedagogies might include instrumental courses specifically dedicated to groups of novices. This way, those who prefer learning music with friends from the beginning—as Jordan and Perry—can find opportunities and resources similar to those offered to more individually inclined learners. This scenario seems to be particularly suited for genres such as rock, jazz, or blues, where playing together is the norm; however, it might as well be applied to Western classical repertoires. This possibility for learning resonates with a more general trend in music scholarship that increasingly places emphasis on a variety of multi-agent contexts. In doing so, recent research explores creativity as a distributed phenomenon (see Bishop, 2018; Sawyer, 2006; van der Schyff, Schiavio, Walton, Velardo, & Chemero, 2018); considers musical agency as multiply constituted and inter-personal (Ryan & Schiavio, 2019; Schiavio & Cummins, 2015); and emphasizes the “hidden socialities” that dwell in individual musicking (Cook, 2018) without posing fundamental distinctions between genres or styles. Openly engaging with these “socialities” in joint and computer-based learning settings might help individual performers develop richer expressive nuances as well as communicate their musical intentions and ideas more clearly. An exploration of such dimensions goes well beyond the scope of the present study, but will hopefully find home in future research studies. While the precise relationship between 4E cognition and the acquisition of musical skills, as well as the nature of the cognitive assemblies formed during musical learning, remain to be further addressed, an understanding of the musical mind as fluidly distributed across social and physical resources continues to be developed.

Author Note

We would like to thank Elli Xypolitaki for coding the videos and Nils Meyer-Kahlen for technical support during data collection. We are grateful to the action editor and three anonymous reviewers for their helpful suggestions. AS was supported by a Lise Meitner

Postdoctoral Fellowship from the Austrian Science Fund (FWF): project number M2148. JS was supported by an Erwin Schrödinger Fellowship from the Austrian Science Fund (FWF): project number J4288.

Correspondence concerning this article should be addressed to Andrea Schiavio, Centre for Systematic Musicology, University of Graz, Merangasse 70, 8010, Graz, Austria, andrea.schiavio@uni-graz.at

References

- ADAMS, F., & AIZAWA, K. (2010). Defending the bonds of cognition. In R. Menary (Ed.), *The extended mind* (pp. 67–80). Cambridge, MA: MIT Press.
- ARON, A., ARON, E. N., & SMOLLAN, D. (1992). Inclusion of other in the self-scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, 63(4), 596–612
- BAIMEL, A., SEVERSON, R. L., BARON, A. S., & BIRCH, S. A. J. (2015). Enhancing “theory of mind” through behavioral synchrony. *Frontiers in Psychology*, 6, 870. DOI: 10.3389/fpsyg.2015.00870
- BARRETT, L. (2011). *Beyond the brain: How body and environment shape animal and human minds*. Princeton, NJ: Princeton University Press.
- BARRETT, M. S., & GROMKO, J. E. (2007). Provoking the muse: A case study of teaching and learning in music composition. *Psychology of Music*, 35(2), 213–230. DOI: 10.1177/0305735607070305
- BISHOP, L. (2018) Collaborative musical creativity: How ensembles coordinate spontaneity. *Frontiers in Psychology*, 9, 1285. DOI: 10.3389/fpsyg.2018.01285
- BORGO, D. (2005). *Sync or swarm: Improvising music in a complex age*. New York: Continuum.
- BORGO, D. (2007). Free jazz in the classroom: An ecological approach to music education. *Jazz Perspectives*, 1, 61–88. DOI: 10.1080/17494060601061030.
- BOWLING, D. L., HERBST, C. T., & FITCH, W. T. (2013). Social origins of rhythm? Synchrony and temporal regularity in human vocalization. *PLoS ONE*, 8, e80402 10.1371/journal.pone.0080402
- BOWMAN, W. (2004). Cognition and the body: Perspectives from music education. In L. Bresler (Ed.), *Knowing bodies, moving minds: Toward embodied teaching and learning* (pp. 29–50). Dordrecht, Netherlands: Kluwer Academic Press.
- BYRNE, R. W., & RUSSON, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and Brain Sciences*, 21(5), 667–684.
- CHEMERO, A. (2009). *Radical embodied cognitive science*. Cambridge, MA: MIT Press.
- CLARK, A. (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.
- CLARK A. (2010). Coupling, constitution, and the cognitive kind: A reply to Adams and Aizawa. In R. Menary (Ed.), *The extended mind* (pp. 81–99). Cambridge, MA: MIT Press.
- CLARK, A., & CHALMERS, D. (1998). The extended mind. *Analysis* 58(1), 7–19.
- CLARK, I., & DUMAS, G. (2015). Toward a neural basis for peer interaction: What makes peer learning tick? *Frontiers in Psychology*, 6, 28. DOI: 10.3389/fpsyg.2015.00028
- CLARKE, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. New York: Oxford University Press.
- COCHRANE, T. (2008). Expression and extended cognition. *Journal of Aesthetics and Art Criticism*, 66, 59–73.
- COLOMBETTI, G. (2014). *The feeling body: Affective science meets the enactive mind*. Cambridge, MA: MIT Press.
- COLOMBETTI, G., & THOMPSON, E. (2008). The feeling body: Towards an enactive approach to emotion. In W. F. Overton, U. Müller, & J. L. Newman (Eds.), *Developmental perspectives on embodiment and consciousness* (pp. 45–68). New York: Erlbaum.
- COOK, N. (2018). *Music as creative practice*. New York: Oxford University Press.
- CREECH, A., & GAUNT, H. (2018). The changing face of individual instrumental tuition: Value, purpose, and potential. In G. McPherson, & G. F. Welch, (Eds.), *Vocal, instrumental, and ensemble learning. An Oxford handbook of music education* (Vol. 3, pp. 145–164). New York: Oxford University Press.
- DE JAEGER, H., & DI PAOLO, E. (2007) Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the Cognitive Sciences*, 6, 485–507.
- DE JAEGER, H., DI PAOLO, E., & GALLAGHER, S. (2010). Can social interaction constitute social cognition? *Trends in Cognitive Sciences*, 14, 441–447.
- DI PAOLO, E., CUFFARI, E. C., & DE JAEGER, H. (2018). *Linguistic bodies. The continuity between life and language*. Cambridge, MA: MIT Press.
- DRAKE, C., JONES, M. R., & BARUCH, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, focal attending. *Cognition*, 77, 251–288.
- EL ZEIN, M., BAHRAMI, B., & HERTWIG, R. (2019). Shared responsibility in collective decisions. *Nature Human Behaviour*, 3(6), 554–559. DOI: 10.1038/s41562-019-0596-4
- ENDEDIJK, H. M., RAMENZONI, V. C. O., COX, R. F. A., CILLESSEN, A. H. N., BEKKERING, H., & HUNNIUS, S. (2015). Development of interpersonal coordination between peers during a drumming task. *Developmental Psychology*, 51(5), 714–721.

- GALLAGHER, S. (2017). *Enactivist interventions: Rethinking the mind*. Oxford, UK: Oxford University Press.
- GALLAGHER, S. (2005). *How the body shapes the mind*. Oxford, UK: Oxford University Press.
- GAUNT, H. (2008). One-to-one tuition in a conservatoire: The perceptions of instrumental and vocal teachers. *Psychology of Music*, 36(2), 215–245.
- GAUNT, H., & WESTERLUND, H. (Eds.) (2013). *Collaborative learning in higher music education*. Farnham, UK: Ashgate.
- GOKHALE, A. (1995). Collaborative learning enhances critical thinking. *Journal of Technology Education*, 7, 22–30.
- GRATIER, M., DEVOUCHE, E., GUELLAI, B., INFANTI, R., YILMAZ, E., & PARLATO-OLIVEIRA, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology*, 6, 1167. DOI:10.3389/fpsyg.2015.01167
- GREEN, L. (2008). *Music, informal learning and the school: A new classroom pedagogy*. London, UK: Ashgate Press.
- HALLAM, S. (1998). *Instrumental teaching*. Oxford, UK: Heinemann.
- HALLAM, S., & BAUTISTA, A. (2018). Processes of instrumental learning: the development of musical expertise. In G. McPherson, & G. F. Welch (Eds.), *Vocal, instrumental, and ensemble learning. An Oxford handbook of music education* (Vol. 3, pp. 108–125). New York: Oxford University Press.
- HANKEN, I. M. (2016). Peer learning in specialist higher music education. *Arts and Humanities in Higher Education*, 15(3–4), 364–375.
- HAUGELAND, J. (1998). Mind embodied and embedded. In J. Haugeland (Ed.), *Having thought: Essays in the metaphysics of mind*. (pp. 207–227) Cambridge, MA: Harvard University Press.
- HIGGINS, L., & MANTIE, R. (2013). Improvisation as ability, culture, and experience. *Music Educators Journal*, 100(2), 38–44. <https://doi.org/10.1177/0027432113498097>
- HIMBERG, T., LAROCHE, J., BIGÉ, R., BUCHKOWSKI, M., & BACHRACH, A. (2018). Coordinated interpersonal behaviour in collective dance improvisation: The aesthetics of kinaesthetic togetherness. *Behavioral Sciences*, 8, 23. DOI: 10.3390/bs8020023
- HIRVENKARI, L., RUUSUVUORI, J., SAARINEN, V.-M., KIVIOJA, M., PERÄKYLÄ, A., & HARI, R. (2013) Influence of turn-taking in a two-person conversation on the gaze of a viewer. *PLoS ONE* 8(8), e71569. <https://doi.org/10.1371/journal.pone.0071569>
- HOVE, M. J., & RISEN, J. L. (2009). It's all in the timing: Interpersonal synchrony increases affiliation. *Social Cognition*, 27, 949–960. DOI: 10.1521/soco.2009.27.6.949
- HURLEY, S. (1998). *Consciousness in action*. Cambridge MA: Harvard University Press.
- HUTTO, D., & MYIN, E. (2013). *Radicalizing enactivism: Basic minds without content*. Cambridge, MA: MIT Press.
- KERSTEN, L. (2017). Extended music cognition. *Philosophical Psychology*, 30, 1078–1103.
- KIRSCHNER S., & TOMASELLO M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behaviour*, 31, 354–364. DOI: 10.1016/j.evolhumbehav.2010.04.004
- KRUEGER, J. (2019). Music as affective scaffolding. In D. Clarke, R. Herbert, & E. Clarke (Eds.), *Music and consciousness II* (pp. 48–63). Oxford, UK: Oxford University Press.
- KRUEGER, J. (2014). Affordances and the musically extended mind. *Frontiers in Psychology*, 4, 1003. DOI:10.3389/fpsyg.2013.01003
- KRUEGER, J. (2011). Extended cognition and the space of social interaction. *Consciousness and Cognition*, 20(3), 643–657.
- LAUNAY, J., DEAN, R. T., & BAILES, F. (2013). Synchronization can influence trust following virtual interaction. *Experimental Psychology*, 60, 53–63. DOI: 10.1027/1618-3169/a000173
- LEHMANN, A. (1997). Acquired mental representations in musical performance. Anecdotal and preliminary empirical evidence. In H. Jørgensen & A. Lehmann (Eds.), *Does practice make perfect? Current theory and research on instrumental music practice* (pp. 55–67). Oslo, NO: Norges musikkhøgskole.
- LEHMANN, A., & JØRGENSEN, H. (2018). Practice. In G. McPherson, & G. F. Welch (Eds.), *Vocal, instrumental, and ensemble learning. An Oxford handbook of music education* (Vol. 3, pp. 126–144). New York: Oxford University Press.
- LEMAN, M. (2008). *Embodied music cognition and mediation technology*. Cambridge, MA: MIT Press.
- MALAFOURIS, L. (2013). *How things shape the mind: A theory of material engagement*. Cambridge, MA: MIT Press.
- MALAFOURIS, L. (2015). Metaplasticity and the primacy of material engagement. *Time and Mind*, 8, 351–371. DOI: 10.1080/1751696X.2015.1111564
- MCELLIN, L., KNOBLICH, G. O., & SEBANZ, N. (2018). Imitation from a joint action perspective. *Mind and Language*, 33, 342–354. <https://doi.org/10.1111/mila.12188>
- MENARY, R. (2010). Introduction to the special issue on 4E cognition. *Phenomenology and the Cognitive Sciences*, 9, 459. <https://doi.org/10.1007/s11097-010-9187-6>
- MERKER, B. (2000). Synchronous chorusing and the origins of music. *Musicae Scientiae*, 3(1) 59–73.
- MESOUDI, A., WHITEN, A., & DUNBAR, R. (2006) A bias for social information in human cultural transmission. *British Journal of Psychology*, 97(3), 405–423.
- MICHAEL, J. (2011). Shared emotions and joint action. *Review of Philosophy and Psychology*, 2, 355–373. DOI: 10.1007/s13164-011-0055-2
- MIENDLARZEWSKA, E. A., & TROST, W. J. (2014) How musical training affects cognitive development: Rhythm, reward and other modulating variables. *Frontiers in Neuroscience*, 7, 279. DOI: 10.3389/fnins.2013.00279

- MURRAY, L., & TREVARTHEN, C. (1985). Emotional regulation of interaction between two-month-olds and their mothers. In T. M. Field, & N. A. Fox (Eds.), *Social perception in infants* (pp. 177–198). Norwood, NJ: Ablex Publishing Corporation.
- NEWEN, A., DE BRUIN, L., & GALLAGHER, S. (EDS.) (2018). *The Oxford handbook of 4E cognition*. Oxford, UK: Oxford University Press.
- NIELSEN, S., JOHANSEN, G., & JØRGENSEN, H. (2018). Peer learning in instrumental practicing. *Frontiers in Psychology, 9*, 339. DOI: 10.3389/fpsyg.2018.00339
- OVERY, K. (2012). Making music in a group: Synchronization and shared experience. *Annals of the New York Academy of Sciences, 1252*, 65–68. DOI: 10.1111/j.1749-6632.2012.06530.x
- OVERY, K., & MOLNAR-SZAKACS, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception, 26*, 489–504.
- PALADINO, M.-P., MAZZUREGA, M., PAVANI, F., & SCHUBERT, T. W. (2010). Synchronous multisensory stimulation blurs self-other boundaries. *Psychological Science, 21*(9), 1202–1207. <https://doi.org/10.1177/0956797610379234>
- PATEL A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.
- PFEIFFER, U. J., SCHILBACH, L., TIMMERMANS, B., KUZMANOVIC, B., GEORGESCU, A. L., BENTE, G., ET AL. (2014). Why we interact: On the functional role of the striatum in the subjective experience of social interaction. *Neuroimage, 101*, 124–137. DOI: 10.1016/j.neuroimage.2014.06.061
- RAVIGNANI, A., BOWLING, D. L., & FITCH, W. T. (2014). Chorusing, synchrony, and the evolutionary functions of rhythm. *Frontiers in Psychology, 5*, 1118. DOI: 10.3389/fpsyg.2014.01118
- REDCAY, E., DODELL-FEDER, D., PEARROW, M. J., MAVROS, P. L., KLEINER, M., GABRIELI, J. D., & SAXE, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *NeuroImage, 50*(4), 1639–1647. DOI: 10.1016/j.neuroimage.2010.01.052
- ROSS, D., & LADYMAN, L. (2010). The alleged coupling-constitution fallacy and the mature sciences. In R. Menary (Ed.) *The extended mind* (pp. 155–166). Cambridge, MA: MIT Press.
- RYAN, K., & SCHIAVIO, A. (2019) Extended musicking, extended mind, extended agency. Notes on the third wave. *New Ideas in Psychology, 55*, 8–17. DOI: 10.1016/j.newideapsych.2019.03.001
- SAWYER, R. K. (2006). Group creativity: Musical performance and collaboration. *Psychology of Music, 34* (2), 148–165.
- SCHAEFER, R. S., & OVERY, K. (2015), Motor responses to a steady beat. *Annals of the New York Academy of Sciences, 1337*, 40–44. DOI: 10.1111/nyas.12717.
- SCHIAVIO, A., BIASUTTI, M., VAN DER SCHYFF, D., & PARNCUTT, R. (2018). A matter of presence. A qualitative study on teaching individual and collective music classes. *Musicae Scientiae, Online first*. DOI: 10.1177/1029864918808833
- SCHIAVIO, A., & CUMMINS, F. (2015). An inter(en)active approach to musical agency and learning. In R. Timmers, N. Dibben, Z. Eitan, R. Granot, T. Metcalfe, A. Schiavio, & V. Williamson (Eds.), *Proceedings of the International Conference on the Multimodal Experience of Music*. Sheffield, UK: HRI Online Publications.
- SCHIAVIO, A., GESBERT, V., REYBROUCK, M., HAUW, D., & PARNCUTT, R. (2019). Optimizing performative skills in social interaction. Insights from embodied cognition, music education, and sport psychology. *Frontiers in Psychology, 10*, 1542. DOI: 10.3389/fpsyg.2019.01542
- SCHIAVIO, A., & VAN DER SCHYFF, D. (2016). Beyond musical qualia. Reflecting on the concept of experience. *Psychomusicology: Music, Mind, and Brain, 26*(4), 366–378.
- SCHIAVIO, A., & VAN DER SCHYFF, D. (2018). 4E music pedagogy and the principles of self-organization. *Behavioral Sciences, 8*(8), 72. DOI: 10.3390/bs8080072.
- SCHIAVIO, A., VAN DER SCHYFF, D., GANDE, A., & KRUSE-WEBER, S. (2019). Negotiating individuality and collectivity in community music. A qualitative case study. *Psychology of Music, 47*(5), 706–721. DOI: 10.1177/0305735618775806
- SMITH, E. R., & SEMIN, G. R. (2004). Socially situated cognition: Cognition in its social context. *Advances in Experimental Social Psychology, 36*, 53–117.
- STEWART, J., GAPENNE, O., & DI PAOLO, E. A. (Eds.). (2010). *Enaction: Toward a new paradigm for cognitive science*. Cambridge, MA: MIT Press.
- STIVERS, T., ENFIELD, N. J., BROWN, P., ENGLERT, C. HAYASHI, M., HEINEMANN, T., ET AL. (2009) Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences USA, 106*, 10587–10592.
- STUPACHER, J., MAES, P.-J., WITTE, M., & WOOD, G. (2017). Music strengthens prosocial effects of interpersonal synchronization – If you move in time with the beat. *Journal of Experimental Social Psychology, 72*, 39–44. DOI: 10.1016/j.jesp.2017.04.007
- THOMPSON, E. (2007). *Mind in life: Biology, phenomenology, and the sciences of mind*. Cambridge, MA: Harvard University Press.
- TREVARTHEN, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. In I. Deliege (Ed.) *Rhythms, musical narrative, and the origins of human communication. Musicae Scientiae, Special Issue*, (pp. 157–213). Liege: European Society for the Cognitive Sciences of Music.
- VALDESOLO, P., OUYANG, J., & DESTENO, D. (2010). The rhythm of joint action: Synchrony promotes cooperative ability. *Journal of Experimental Social Psychology, 46*, 693–695. DOI: 10.1016/j.jesp.2010.03.004
- VAN DER SCHYFF, D., SCHIAVIO, A., & ELLIOTT, D. (2016). Critical ontology for an enactive music pedagogy. *Action, Criticism and Theory for Music Education, 15*(5), 81–121. DOI: 10.22176/act15.5.81

- VAN DER SCHYFF, D., SCHIAVIO, A., WALTON, A., VELARDO, V., & CHEMERO, T. (2018). Musical creativity and the embodied mind. Exploring the possibilities of 4E cognition and dynamical systems theory. *Music and Science, 1*. DOI: 10.1177/2059204318792319
- VAN ULZEN, N. R., LAMOTH, C. J. C., DAFFERTSHOFER, A., SEMIN, G. R., & BEEK, P. J. (2008). Characteristics of instructed and uninstructed interpersonal coordination while walking side-by-side. *Neuroscience Letters, 432*, 88–93. DOI: 10.1016/j.neulet.2007.11.070
- VARELA, F., THOMPSON, E., & ROSCH, E. (1991). *The embodied mind: Cognitive science and human experience*. Cambridge MA: MIT Press.
- WARD, D., & STAPLETON, M. (2012). Es are good: Cognition as enacted, embodied, embedded, affective and extended. In F. Paglieri (Ed.), *Consciousness in interaction: The role of the natural and social context in shaping consciousness* (pp. 89–105). Amsterdam, The Netherlands: John Benjamins.
- WELCH, G. F., HIMONIDES, E., SAUNDERS, J., PAPAGEORGI, I., & SARAZIN, M. (2014). Singing and social inclusion. *Frontiers in Psychology, 5*, 803. DOI: 10.3389/fpsyg.2014.00803
- WILLIAMON, A., & VALENTINE, E. (2000). Quantity and quality of musical practice as predictors of performance quality. *British Journal of Psychology, 91*, 353–376. DOI: 10.1348/000712600161871
- WILSON, A. D., & GOLONKA, S. (2013). Embodied cognition is not what you think it is. *Frontiers in Psychology, 4*, 58. DOI: 10.3389/fpsyg.2013.00058