



Cognitive Science 45 (2021) e12932

© 2021 Cognitive Science Society, Inc. All rights reserved.

ISSN: 1551-6709 online

DOI: 10.1111/cogs.12932

## Emergent Shared Intentions Support Coordination During Collective Musical Improvisations

Louise Goupil,<sup>a,b</sup>  Thomas Wolf,<sup>c</sup> Pierre Saint-Germier,<sup>a</sup>  
Jean-Julien Aucouturier,<sup>a</sup> Clément Canonne<sup>a</sup>

<sup>a</sup>*Science and Technology of Music and Sound (UMR 9912, IRCAM/CNRS/Sorbonne University)*

<sup>b</sup>*School of Psychology, University of East London*

<sup>c</sup>*Department of Cognitive Science, Central European University*

Received 15 May 2020; received in revised form 26 November 2020; accepted 9 December 2020

---

### Abstract

Human interactions are often improvised rather than scripted, which suggests that efficient coordination can emerge even when collective plans are largely underspecified. One possibility is that such forms of coordination primarily rely on mutual influences between interactive partners, and on perception–action couplings such as entrainment or mimicry. Yet some forms of improvised joint actions appear difficult to explain solely by appealing to these emergent mechanisms. Here, we focus on collective free improvisation, a form of highly unplanned creative practice where both agents' subjective reports and the complexity of their interactions suggest that shared intentions may sometimes emerge to support coordination during the course of the improvisation, even in the absence of verbal communication. In four experiments, we show that shared intentions spontaneously emerge during collective musical improvisations, and that they foster coordination on multiple levels, over and beyond the mere influence of shared information. We also show that musicians deploy communicative strategies to manifest and propagate their intentions within the group, and that this predicts better coordination. Overall, our results suggest that improvised and scripted joint actions are more continuous with one another than it first seems, and that they differ merely in the extent to which they rely on emergent or planned coordination mechanisms.

*Keywords:* Improvisation; Musical performance; Coordination; Joint action; Goal representations; Shared intentions

---

---

Correspondence should be sent to Louise Goupil, School of Psychology, University of East London, Stratford Campus—Water Lane, London E15 4LZ, UK. E-mail: [lougoupil@gmail.com](mailto:lougoupil@gmail.com)

## 1. Introduction

While the ability to plan and to organize our actions accordingly is often considered crucial to collective behavior in humans (Bratman, 2014), a significant part of our interactions seems to take place in the absence of such planification. Sometimes, we have to react to unexpected events, spontaneously adapting our interactions on the fly without having the possibility to rely on pre-established plans (Mendonça & Wallace, 2007). Other times, we simply refuse to commit to a shared plan before engaging in a joint activity, because we trust that it will allow for the emergence of creative or surprising interactions (Sawyer, 2003). Such unplanned joint actions can be referred to as cases of *collective (or joint) improvisations*, and they are encountered in a wide variety of areas (Ingold & Hallam, 2007), from artistic activities (e.g., comedy improv) to work situations (e.g., brainstorming sessions), from day-to-day life (e.g., open-ended conversations) to emergency crisis (e.g., sudden terrorist attacks).

On a general level, collective improvisations can be defined as joint actions in which the precise outcome of the action is not planned ahead, nor is the precise way it will unfold. In such a situation, improvisers must invent ways to coordinate online, as the joint action proceeds, while referring to a joint goal that remains largely under-specified (e.g., “making music together” or “surviving together”) and which, as such, does not entail a given sequence of actions nor a given task distribution. Collective improvisations are thus in stark contrast with *scripted joint actions*, where interacting partners explicitly specify the desired end result (i.e., their joint outcome) beforehand, as well as each agent’s task, and an outline of the steps needed in order to reach this joint outcome. At first sight, scripted and improvised joint actions appear to raise distinct problems of coordination that may be solved by distinct mechanisms. Consequently, research focusing on coordination has mainly studied these two types of joint actions separately.

On the one hand, research on scripted joint actions typically highlights the role of joint planning for coordination (Bratman, 1999; Knoblich, Butterfill, & Sebanz, 2011; Loehr, Kourtis, Vesper, Sebanz, & Knoblich, 2013; Vesper et al., 2017). A central way through which partners are thought to solve coordination problems during scripted joint actions is through the involvement of shared intentions—mental states held by individual agents that represent specific joint outcomes—and specifications of each agent’s tasks, that are common knowledge between them (Bratman, 2014). Beyond abstract, shared intentions, recent evidence suggests that shared goal representations—that have a more concrete, motoric format (Butterfill, 2018)—can also facilitate coordination at shorter time scales (della Gatta, 2017; Kourtis, Woźniak, Sebanz, & Knoblich, 2019; Sacheli, Arcangeli, & Paulesu, 2018). In the following, we refer to processes that involve shared intentions or shared goal representations as *planned coordination* mechanisms (Butterfill, 2018; Knoblich et al., 2011), because they require partners to be jointly oriented toward a given outcome.

On the other hand, most research on improvised joint actions so far has focused on examining embodied and embedded aspects, and describing coordination mechanisms that are thought to operate on short time scales, and to directly arise from dynamic

interactions between partners within a shared environment. Unlike planned coordination, this type of mechanisms does not require that agents hold specific mental representations at the individual level, but primarily rely on agents' dynamic couplings while acting jointly. Here, following previous authors (Butterfill, 2018; Knoblich et al., 2011), we refer to these processes as *emergent coordination* mechanisms. One classic example is the phenomenon of entrainment observed when two agents become more synchronized with one another than expected by chance simply through seeing each other's movements, and even in the absence of, or contrary to, any intention to do so (Issartel, Marin, & Cadopi, 2007; Nessler & Gilliland, 2009; Repp, 2005; Yun, Watanabe, & Shimojo, 2012). Entrainment is often interpreted in the framework of dynamical systems, where it is argued to merely constitute a particular instance of physical coupling that can arise in all (social or nonsocial) kinds of coupled oscillators (Schmidt & Richardson, 2008; Walton et al., 2018). Other studies have documented the role of mimicry, or automatic imitation, showing that individuals often mirror each other's actions, and that such mirroring fosters coordination and acts as social glue by increasing affiliation between individuals (Gueguen, Jacob, & Martin, 2009; Van Baaren, Janssen, Chartrand, & Dijksterhuis, 2009). For instance, one study showed that expert improvisers could smoothly imitate each other's movements while performing a mirror-game task, entering into a state of co-confidence in which each player seems to be both leading and following at the same time (Noy, Dekel, & Alon, 2011). Beyond mirroring, there is some evidence that motor simulation enables observers to predict their partners' actions, which can help them adjust their actions accordingly to improve coordination (Aglioti, Cesari, Romani, & Urgesi, 2008; Novembre, Ticini, Schütz-Bosbach, & Keller, 2014; Noy et al., 2011; Vesper, van der Wel, Knoblich, & Sebanz, 2013). Finally, other research has focused on documenting joint affordances, showing for instance that particularly salient elements present within their environment can constrain improvisers' behavior, leading them to perform actions with a similar functional profile (e.g., changing what they were doing) during the course of the performance (Canonne & Garnier, 2012).

Overall, it seems clear that coordination during collective improvisations heavily relies on the fact that agents' interactions are both embodied and embedded (Linson & Clarke, 2018). Yet, whether emergent mechanisms are sufficient to support coordination in cases of complex and/or temporally extended collective improvisations, without the support of additional (planned) mechanisms, at least punctually, remains far from certain. Indeed, the studies reviewed above document the role of emergent coordination mechanisms in supporting very simple forms of joint action that typically involve agents who perform very similar actions at the same time (e.g., tap in synchrony to the same beat, imitate each other's motion or emotional displays, etc.). A large literature has documented the pervasiveness of these mechanisms, at the behavioral, physiological, and neural levels, and the role they play in coordination from infancy to adulthood (Helm, Miller, Kahle, Troxel, & Hastings, 2018; Wass, Whitehorn, Marriott Haresign, Phillips, & Leong, 2020). Yet how they could account for complex forms of collective improvisations, where each agent has to perform a different type of action, and where no temporal structure is present to support mechanisms such as entrainment, is really unclear. Moreover, these

mechanisms operate on short time scales (seconds, at best minutes), and they are specifically efficient when precision is targeted, while they fall short at explaining how the coordination of complex and flexible behaviors—typical of most creative improvisations—may be achieved (Butterfill, 2018).

Research on *scripted* joint actions generally suggests that both emergent coordination mechanisms and planned coordination mechanisms actually interact to foster coordination, their relative contributions enabling an optimal trade-off between precision and flexibility (Butterfill, 2018). For instance, the fine-tuning of musical expressivity in performing chamber music compositions crucially depends on emergent mechanisms, which regulate the temporal unfolding of performers on very short time scales (D’Ausilio et al., 2012; Keller, 2014). Studies also suggest that when co-agents have a shared intention to synchronize, internal sensorimotor models enable them to predict each other’s timing and to deploy strategies to improve synchrony (Heggli, Konvalinka, Kringelbach, & Vuust, 2019; Vesper, van der Wel, Knoblich, & Sebanz, 2011).

Building upon these studies targeting scripted interactions, here we ask whether such a synergy of planned and emergent coordination mechanisms is also at play during improvised joint actions. More precisely, we test the hypothesis that co-improvisers also coordinate by forming shared intentions that emerge during the course of the interaction. We hypothesize that shared intentions may be particularly crucial to support the most complex and flexible forms of collective improvisations, which require co-agents to perform dissimilar and varied actions that are not necessarily tied to an underlying temporal structure. We thus conducted four experiments using the practice of Collective Free Musical Improvisation (CFI) as an experimental model of improvised joint action.

Collective Free Musical Improvisation constitutes a particularly pure and paradigmatic case of collective improvisation (Bailey, 1992) that is ideal to test our hypotheses for several reasons. First, in CFI, musicians typically do not attribute roles to each other, do not specify melodic or harmonic structures before improvising together, and overall, refuse to specify how the improvisation will unfold. In other words, they refuse to precisely specify their joint outcome and to establish a joint plan beforehand (Pressing, 1984). On a finer level, CFI also crucially differs from more familiar genres of improvised music such as bebop or even free jazz in the sense that it is generally not pulsed and devoid of rhythmical patterns. Free improvisers certainly share a common ground, which imposes nontrivial aesthetical constraints on the group’s performances (e.g., leading musicians to focus on subtle timbral explorations and to avoid conventional rhythmical patterns or chord progressions). However, the issue of how to *temporally* organize the individual and collective musical behaviors on shorter and longer time scales in a given performance remains in its entirety (Canonne, 2018), making CFI an as pure as possible case of real-life improvised joint action (see video and audio examples via this link [[https://osf.io/4pnxh/?view\\_only=75afeb0864964265ab40e29a60895885](https://osf.io/4pnxh/?view_only=75afeb0864964265ab40e29a60895885)]). Second, CFI typically involves a temporally extended situation in which each agent performs highly idiosyncratic, nonimitative actions. This is in sharp contrast with shorter, simpler, and imitation-based forms of improvised interactions used in previous research (Noy et al., 2011), and it makes CFI especially appropriate to track the existence and impact of shared intentions in joint

improvised actions. Finally, like other forms of collective music-making that have been used as a model to investigate joint actions (Aucouturier & Canonne, 2017; D’Ausilio, Novembre, Fadiga, & Keller, 2015; Kirschner & Tomasello, 2010; Michael, 2017), CFI constitutes a model that is ecologically valid, and allows one to measure coordination on multiple levels and to investigate the mechanisms that drive the emergence of shared intentions on the fly, in the absence of verbal communication.

This specific model allows us to ask three questions: Do shared intentions emerge during this complex case of improvised joint actions? If so, how can such shared intentions emerge in the absence of verbal communication? And to which extent does the sharedness of these intentions among co-agents affect coordination? To address these three questions, we focused on a coordination problem that is likely to arise in most—if not all—improvisations: how to collectively *end* the performance.

How and when to end a performance is a coordination problem that is particularly challenging in CFI because musicians do not share a given script nor a repertoire of canonical endings that provide them with clear potential ending points. Even if musicians were to decide to end the piece at the same time, it would still be difficult to do so. Contrary to other musical genres, such as straight-ahead jazz, in which temporal and harmonic structures typically determine specific ending points (e.g., on the beat, or on a closing cadence), provide musicians with the support of a shared entrainment to a beat, or at the least, enable performers to rely on auditory imagery to form precise predictions about what is about to come next (Hadley, Sturt, Moran, & Pickering, 2018; Keller, 2008), in CFI there are no definite structures nor conventional patterns that point to specific ending points. As Alain Savouret—who taught free improvisation at Paris Conservatory for many years—nicely puts it: “If it’s always difficult to start [an improvisation], it’s even harder to finish it” (Savouret, 2010, p. 26). As such, issues of endings are often raised and discussed within CFI classes. At the same time, endings are also moments in which the improvisers’ coordination (or lack thereof) is at its clearest: Musicians (and attuned audience members alike) often speak of “missed endings” when the group members did not “feel” at the same time that the performance was coming to an end or that such or such musical event could act as a good ending point. For these two reasons, endings perfectly encapsulate the coordination problems that are at stake during improvised joint actions. In this regard, they constitute a particularly interesting case to study the role of shared intentions in supporting coordination when multiple agents act in flexible ways. Shared intentions could indeed foster coordination in this context because they would allow improvisers to anticipate that the performance is about to finish, and to plan their actions with respect to this proximate joint outcome, on the basis that their coimprovisers are likely to do the same.

Thus, in Experiments 1 and 2, we invited trios of musicians to a recording studio, where they were asked to perform a series of short improvisations. In Experiment 1, musicians had to perform four improvisations and, while playing, each musician was asked to press a pedal “*as soon as she felt that she was looking for an ending.*” As musicians were playing in separate studio booths, pedal presses were made covertly, with no auditory consequence allowing other musicians to perceive when their partners pressed

the pedal. By testing whether musicians' reports are closer in time to one other than would be predicted by chance, Experiment 1 allowed us to investigate whether shared intentions do emerge during collective improvisations.

In Experiment 2, we tested the extent to which shared intentions actually impact coordination. To do so, we asked the same musicians to perform 12 additional improvisations. We experimentally manipulated musicians' intention to end the piece, by covertly delivering auditory prompts through their headphones. Musicians were prompted with either an *individual*, ME-Goal (i.e., finding a good ending for their own individual parts) or with a *collective*, WE-Goal (i.e., finding a good ending for the group's performance as a whole). We also manipulated the number of musicians who received a prompt ( $N = 1, 2$ , or 3), thereby manipulating the degree of *shared information*. Note that musicians always received the same type of prompt, either ME or WE.

As detailed in Table 1, this procedure allowed us to contrast three hypotheses. According to a *shared information* hypothesis, for the presence of goals to impact coordination, agents merely have to represent the same information (i.e., that the piece is about to end). This hypothesis merely predicts tighter coordination as the degree of shared information (i.e., number of prompts) increases. By contrast, according to a *collective intention* hypothesis, what matters is that some agents within the group hold collective intentions, in the sense that they involve the group in their very content. This hypothesis predicts tighter coordination when agents' intentions involve the group (i.e., for WE-Goals) as compared to when agents merely pursue individual goals (i.e., for ME-Goals). Finally, according to a *shared intention* hypothesis, what matters is that agents hold collective intentions, but in addition, that these intentions be shared and common knowledge between them. This hypothesis predicts that the content of the goals (i.e., whether it was

Table 1

Predictions of the three main hypotheses with respect to the two main aspects examined in this study: (a) coordination, assessed at three levels as reported in Sections 3.2.1 (temporal coordination), 3.2.2 (acoustic coordination), and 4.2.1/4.2.2 (qualitative aspects of coordination), and (b) signaling strategies (results reported in Section 5.2.3)

Hypotheses	Predictions	
	Temporal, acoustic, and qualitative aspects of musical coordination...	Signaling strategies are...
<i>Shared information</i>	... improve as the degree of shared information increases <i>main effect of the number of prompts</i>	no specific predictions about signaling strategies
<i>Collective intention</i>	... improve when agents hold collective as compared to individual intentions <i>main effect of prompt type</i>	no specific predictions about signaling strategies
<i>Shared intention</i>	... improve when collective intentions are shared <i>interaction between the number of prompts and the type of prompt</i>	... present in the WE but not in the ME condition, so that collective intentions spread and become common knowledge within the group

an individual ME-goal or a collective WE-goal) should impact coordination over and beyond shared information: We should thus expect tighter coordination when several musicians had the same collective goal of finding a good ending for the group as compared to cases in which the same number of improvisers merely had *parallel* individual goals (i.e., each improviser having the distinct goal of finding a good ending for herself), and this relationship should also vary as a function of the number of prompts (i.e., only one performer having a collective intention may not be enough for coordination to ensue).

Coordination was examined on three levels: (1) by assessing the *temporal coordination* with which musicians stopped playing at the end of the piece; (2) by assessing the musicians' *dynamic, timbral, and harmonic coordination* with several acoustical measures; and (3) by assessing *qualitative aspects of musical coordination*. Point (3) was achieved by running a follow-up listening experiment (Experiment 3) where a separate group of expert and naive listeners were asked to evaluate the recorded improvisations, in order to assess whether shared intentions impact the aesthetic perception of the joint performance, and some of its qualitative properties corresponding to higher level aspects of musical coordination that are difficult to capture with acoustic analysis, given the sheer sonic complexity of most CFI performances. Lastly, contrary to the other two hypotheses, the *shared intention* hypothesis also predicts that prompted musicians may engage in signaling strategies to make their intention *manifest* for the group, thereby establishing common knowledge that the piece is about to end, and ensuring the collaboration and commitment of the other performers. Thus, in a fourth experiment with the same listeners involved in the third experiment, we investigated how goals may propagate within the group of improvisers to foster coordination. We examined the possibility that musicians deploy *signaling strategies* to establish common knowledge of their current goal at the level of the group, thereby forming proper shared intentions. To this end, listeners were asked to detect whether they thought individual performers were looking for an end, and to characterize their behavior along several categories. This allowed us to examine whether musicians' intentions to end the piece could be deciphered by listeners, what type of communicative behaviors drive this perception, and how the transparency of performers' intentions relates to coordination.

## **2. Experiment 1: Can shared intentions emerge during collective musical improvisations?**

### *2.1. Experiment 1—Methods*

#### *2.1.1. Participants*

We invited 21 participants (2 women, age  $M = 39.8$  years,  $SD = 9.1$  years) to take part in Experiments 1 and 2. All were highly skilled professional musicians actively involved in CFI (average years of experience on their respective instruments

$M = 29.2$  years,  $SD = 8.3$  years, and number of years of performing CFI  $M = 17.3$  years,  $SD = 6.8$  years). Participants were grouped into 12 trios, such that no combination of musicians would repeat (see Table S1 for the musical instruments played in each trio). Fifteen of the 21 musicians participated in two different trios. We also tried to minimize the familiarity between musicians, which ensures maximal conditions of free improvisation, and limited the common ground structuring musicians' interactions. We asked musicians to report how much they knew each of the two other musicians on a scale from 1 (not familiar at all) to 7 (very familiar), and how much they enjoyed playing with this trio (1: not at all; 7: very much). Familiarity averaged over the 12 trios was  $M = 2.6$ ,  $SD = 0.91$ , confirming low familiarity overall. Appreciation averaged over the 12 trios was  $M = 5.7$ ,  $SD = 1$ , suggesting that our procedure was not too invasive and allowed musicians to play together in an ecological fashion. We assessed participants' general empathic traits by using the self-report Basic Empathic Scale in Adults (BESA; Carré, Stefaniak, D'Ambrosio, Bensalah, & Besche-Richard, 2013). Nineteen participants filled in the questionnaire, and two musicians refused to do so (including one of the musicians who played twice, leading to three missing values). Musicians signed an informed consent and were paid for their contribution.

### 2.1.2. Procedure and design

The aim of Experiment 1 was to assess whether shared goals spontaneously emerge during improvised joint actions, modeled here with CFI. To this end, we asked each of the 12 trios of expert improvisers to perform four improvisations of approximately 3–4 min (180–240 s). Providing this range was necessary to enable efficient data collection, but the instructions emphasized the fact that this time limit was meant to provide a loose guideline rather than to set a strict boundary. Consistent with these instructions, the durations of the improvisations were widely spread around the recommended time range, effectively extending from 92.8 to 391.3 s ( $M = 202.8$  s,  $SD = 52.5$ ). It should also be noted that agreeing on an approximate duration before the beginning of the improvisation is common practice in this community. For example, trumpet player Axel Dörner states that

In [one of my trios], we say beforehand how long we want to play for. For me, that's important. When we play a concert, we decide how long the concert is going to last and how the concert might be divided into pieces. Sometimes we define it closely—longer pieces, shorter pieces or endings. We decide together. (quoted in Denzler & Guionnet, 2020, p. 72)

More generally, performing pieces of 3–4 min is not unheard of for these improvisers, as it corresponds to the typical duration of the “constrained improvisations” they sometimes perform during their working sessions (Canonne, 2018).

Musicians were placed in separate studio booths so that they could not see each other, and only heard each other through headphones, as is standard in studio recording practices. Each musician was asked to press a midi pedal (M-Audio SP-2) “as soon as she felt that she was looking for an end to the piece.” Thus, our focus was on *collective*

*intentions* (i.e., intentions to end the piece that include the group in their contents): For the piece to end, *all* improvisers must stop playing. By testing whether such collective intentions emerge closer to each other than would be expected by chance, we test whether they were shared among partners, amounting to *shared intentions*. After each improvisation, musicians were asked to rate on a 7-point Likert scale the extent to which they enjoyed the improvisation, and how much they liked the ending. These ratings suggested that they were not disturbed by having to press the pedal (see Section S.1.4 in Appendix S1). They were also asked whether or not they thought that their partners had been looking for an end, and if so why. This experiment was pre-registered at <https://aspredicted.org/k2jf5.pdf>. We note when our analyses departed from the pre-registration. The corpus, data and analysis scripts are available on the Open Science Framework via this link ([https://osf.io/4pnxh/?view\\_only=75afeb0864964265ab40e29a60895885](https://osf.io/4pnxh/?view_only=75afeb0864964265ab40e29a60895885)).

### 2.1.3. Data analysis

Pedal press events were recorded and time stamped. Reports that occurred after the musician actually stopped playing were removed (more on this below). The *Number of Pedal Pressings* per improvisation (0–3) was then computed by summing the number of pedals that were pressed *before* the actual end of the performance. We also computed the *Pedal Pressing Temporal Coordination* for each improvisation, as the absolute time difference between the three possible pairing of events, and took the mean of this value over the whole trio. Note that the *Pedal Pressing Temporal Coordination* could only be computed for improvisations where two or more events were recorded.

To test whether musicians were more temporally coordinated in their intentions to end the improvisation than would be predicted by chance, we also computed temporal coordination between fake pairings of pedal pressings. Fake pairings were defined as pairings of pedal press events from the same trio, but from different improvisations. Theoretically, each pedal pressing could thus be “fakely” paired with six other pedal pressings (i.e., pedal pressings of the two other musicians taken from the three other improvisations performed during the experiment), which would result in 864 possible pairings. In practice, since musicians sometimes did not press the pedal, this step resulted in only 208 fake pairings. We computed the *Temporal Coordination of Endings* in the same way as the *Pedal Pressing Temporal Coordination*, except that we took the time-stamped ending points of each musician’s performance instead of pedal press events. Finally, the *Ending Appreciation* metric was computed based on the appreciation ratings provided by the musicians after each improvisation, by averaging the ratings of all three musicians for each improvisation.

## 2.2. Experiment 1—Results

### 2.2.1. Ending goals emerge in musical improvised interactions, and they are temporally coordinated

The mean *Temporal Coordination of Endings* for real pairs was  $M = 7.74$  s,  $SD = 4.07$ . This was significantly better than the *Temporal Coordination of Endings* calculated

for fake pairings ( $M = 45.60$  s,  $SD = 23.88$ ),  $t(11) = 5.152$ ,  $p < .001$ ,  $d = 2.210$ . Performances' endings were thus not the mere result of the individual musicians randomly stopping at some point. On the contrary, despite the highly unscripted nature of CFI and the general absence of a shared pulse, it seems that the improvisers were still aiming to achieve some degree of temporal coordination when ending the piece, although it should be noted that 7 s is well above the duration that would be expected in a typical, scripted musical performance.

The number of Pedal Pressings was 2 or higher in 25 out of the 48 improvisations (see Fig. 1A). The mean *Pedal Pressing Temporal Coordination* was  $M = 28.38$  s,  $SD = 19.97$ . To test whether this duration is smaller than what would be expected by chance, we compared it to the temporal coordination of fake pairings ( $M = 47.10$  s,  $SD = 23.51$  s). Consistent with our prediction, a paired-sample  $t$  test revealed a significant difference,  $t(11) = 2.643$ ,  $p = .025$ ,  $d = 0.797$ , with the real *Pedal Pressing Temporal Coordination* being significantly lower than the one for fake pairings. Thus, when two or more musicians pressed their pedals during the performance, those pedal presses were closer in time than would be expected by chance. Additionally, despite the inevitable latency introduced by the experimental setting, pedal pressings were  $<10$  s apart in 24.3% of trials (see Fig. 1B), which suggests that, in those cases at least, two or more improvisers were

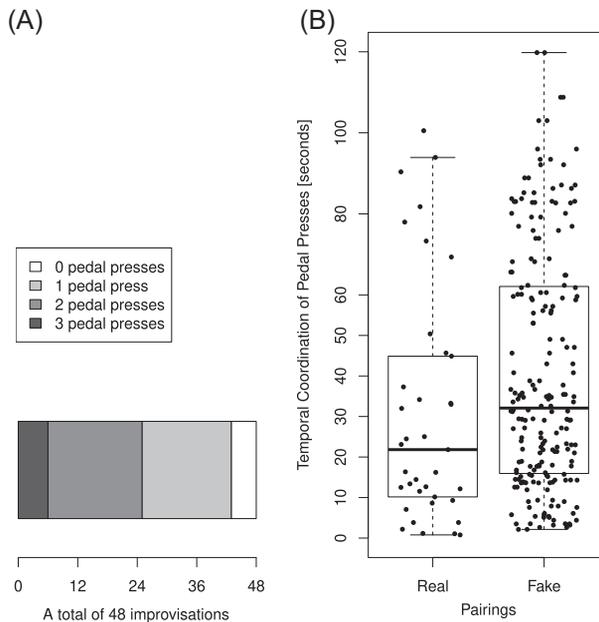


Fig. 1. (A) Percentage of improvisations in which 3, 2, 1, or 0 musicians signaled an intention to end the improvisation by pressing their pedal. (B) Pedal press temporal coordination for real and fake pedal press pairing. Comparing these two conditions allows assessing whether musicians' coordination when pressing the pedal is better than chance. Dots are individual values of temporal coordination between pedal presses occurring in the same improvisation.

intending to end during the same short time span. Our data reveal that collective intentions can emerge at the same time, and thus be shared by several musicians during improvised interactions.

Note that a significant number of pedal presses (22 out of 96) were made after the musician had actually stopped playing. In those cases, it may be that musicians did not have a prior intention to stop playing, or alternatively, that they did not realize that the performance was coming to an end before actually hearing the other musicians stop. Interestingly, however, in 21 of these 22 cases in which one musician pressed her pedal after stopping, at least one of the other musicians had pressed her pedal before her own stopping point. This means that fully “emergent” endings were in fact quite rare, and that the negotiations of endings typically involved a mixture of a short-term micro-planning—including partially or fully shared intentions to end—and emergent reactions to other musicians’ intentions to end the piece.

### 2.2.2. Impact of shared intentions on improvised musical coordination

The average *Temporal Coordination of Endings* was  $M = 27.38$  s ( $SD = 20.57$  s) and the average *Ending Appreciation* was  $M = 4.74$  s ( $SD = 0.99$  s). Contrary to our predictions, there was no correlation among trials between *Pedal Pressing Temporal Coordination* and the *Temporal Coordination of Endings* (Spearman  $r_s(23) = 3,286$ ,  $p = .20$ ), and no correlation between *Pedal Pressing Temporal Coordination* and *Appreciation of Ending* ( $r_s(23) = 2,621.2$ ,  $p = .97$ ), which we take as a proxy to higher level aspects of coordination. Thus, there was no evidence that the emergence of shared intentions positively impacted musicians’ coordination here. It is worth noting that debriefings with participants revealed that, in some cases, improvisers had forgotten to press their pedal even though they had been actively looking for an end. In the second experiment, which offered a more controlled environment, we investigate the impact of shared intentions on improvised coordination more directly.

## 3. Experiment 2: Can shared intentions improve coordination during collective musical improvisations?

Experiment 1 demonstrates that shared intentions to end the joint action can emerge in the course of improvised interactions, even in the absence of verbal communication. In Experiment 2, we ask whether these shared intentions actually *impact* coordination. To this end, we experimentally manipulated musicians’ intentions: We gave them covert instructions regarding how and when they should start looking for an end to the piece, and measured whether and how these instructions impacted coordination at the level of the group. More precisely, we manipulated both the degree of *shared information* (i.e., the number of musicians receiving instructions) and the *content* of the intention (i.e., whether musicians were supposed to look for an end individually, or collectively). This allowed us to discriminate between the three hypotheses outlined in Section 1, namely, the hypothesis according to which *shared information* is crucial to foster coordination,

the hypothesis according to which *collective intentions* are crucial, and finally, the most demanding hypothesis according to which *shared intentions* are crucial.

### 3.1. Experiment 2—Methods

#### 3.1.1. Participants and procedure

After completing four improvisations for Experiment 1, each of the 12 trios took a short break, before performing 12 additional improvisations for Experiment 2, resulting in a total of 144 improvisations. During these 12 additional improvisations, musicians sometimes received covert auditory prompts approximately 2:30 min after the beginning of the improvisation (see below for the sampling procedure). Prompts were of two types: Upon hearing the keyword “ME,” a musician was asked to “*find a good way for you to stop playing, thus looking for an ending for yourself*” (ME-Goal); upon hearing the keyword “WE,” the musicians were asked to “*find a good way for the group to stop playing, thus looking for an ending for the group*” (WE-Goal). Thus, we varied whether musicians had a goal whose content involved the group as a whole (WE-Goal) or only themselves (ME-Goal). In addition, we varied the degree of dissemination of these goals within the group, by prompting either one, two, or all three musicians. For each improvisation, only one type of prompt could be delivered (i.e., all prompted musicians either received a WE or ME-Goal). Experimental conditions could vary over the three Prompt Types (ME-Goal or WE-Goal, or NO-Prompt) and three Prompt Numbers (1, 2, 3), resulting in six experimental conditions at the level of the trio (one musician with a ME-Goal/two non-prompted musicians; two musicians with ME-Goals/one non-prompted musician; three musicians with ME-Goals; one musician with a WE-Goal/two non-prompted musicians; two musicians with a WE-Goal/one non-prompted musician; three musicians with a WE-Goal). Prompt times were semi-randomly sampled from two uniform distributions, one ranging from 2:15 to 2:30 (early prompt) and one ranging from 2:30 to 2:45 min (late prompt). Each of the nine conditions had one trial with a time point from the first range and one trial with a time point from the second range. This procedure ensured that the timings of the prompts were not too predictable.

After each improvisation, we asked musicians to rate the extent to which they thought the ending was successful (on a 7-point scale), to justify this judgment with a few words, as well as to guess for each musician whether they had received a prompt, and if so which type of prompt (ME or WE). This allowed us to verify, first, that participants heard the instructions correctly in prompted trials and, second, to assess their ability to “mindread” the intentions of their partners (see Fig. S5).

Auditory prompts were delivered covertly through musicians’ headphones. This solution was preferred over visual prompts because of two practical reasons: (a) Musicians need to wear headphones to hear each other in the studio anyways, and (b) many of them close their eyes when they play, and mostly focus on sounds during the performance. Using auditory prompts thus minimized the risk that musicians would miss the prompts (e.g., due to closed eyes). Despite these precautions, questionnaires revealed that

musicians missed or misheard prompt types on a few occasions ( $N = 32$ , 7.4% of the trials). We excluded eight trials in which two musicians or more made such mistakes and re-coded the other trials to account for what the musician actually perceived. This procedure left a total of 136 improvisations in the dataset. In addition, because of a technical error, the first of the 12 trio only received “ME” prompts.

This experiment was pre-registered at <https://aspredicted.org/k2jf5.pdf>. We note when our analyses departed from the pre-registration. Data and analysis scripts are available via this link ([https://osf.io/4pnxh/?view\\_only=75afeb0864964265ab40e29a60895885](https://osf.io/4pnxh/?view_only=75afeb0864964265ab40e29a60895885)).

### 3.1.2. Data analysis

As in Experiment 1, we computed the *Temporal Coordination of Endings* for each improvisation and trio as the average of the absolute values of each musician’s stopping time minus the timing of the end of the improvisation (i.e., the timing at which the last musician stopped). The smaller the value of this variable, the closer in time the three musicians ended the improvisation. We also computed the unprompted musicians’ *Temporal Coordination with Others*, which reflects the degree to which unprompted musicians coordinated with their (prompted) partners. For each unprompted musician and improvisation, this index was calculated as the absolute value of the difference between the timing at which they stopped and the average of the timings at which their partners stopped. As there were no unprompted musicians in improvisations in which the Prompt Number was three, these trials were not included in this analysis.

### 3.1.3. Acoustic analysis

To investigate whether receiving prompts changed the relationships between the musicians, we conducted an acoustic analysis of musical snippets extracted before and after the prompts. Following previous studies (Pachet, Roy, & Foulon, 2017; Papiotis, Marchini, & Maestre, 2012), we approximated coordination by computing a linear (Pearson correlation) as well as a nonlinear (mutual information) index of dependency for five acoustic features: pitch, volume (RMS), playing time ratio (% of sound), spectral centroid, and harmonic-to-noise ratio (HNR; see below). For each of the five acoustic features and two metrics, we computed values for each pair of musicians, improvisation and timing (before or after the prompt) before averaging the values within the trio for each improvisation and timing. We also estimated the consonance of the music produced at the level of the trio as a measure of harmonic coordination.

For each improvisation and individual musician, pitch, loudness, playing time, spectral centroid, and HNR were estimated in non-overlapping successive time frames of 200 ms in two time windows: (a) in a window starting 1 min before the prompt and ending before the prompt and (b) in a window starting at the prompt and extending until the end of the improvisation ( $M = 54.8$  s,  $SD = 70.7$ ). Pitch was extracted using the *Praat* software (Boersma, 2001). Loudness was approximated as the root-mean square of the amplitude of the sound. Playing time ratio was defined as the ratio of the time spent playing over the total duration of the extract. The HNR was computed following the algorithm described in Boersma (1993). Finally, dissonance/roughness was estimated based on the

algorithm described in Vassilakis (2001) and implemented in the *dissonant* package in *Python*. This method, which is based on a classic model by Sethares (1993), estimates the dissonance/roughness of a sound from the amount of competition between partials (see <https://pypi.org/project/dissonant> for a full detail of this method and formulas). Dissonance is a complex percept that is difficult to capture algorithmically, but listening to a subset of our corpus and comparing values of dissonance obtained by this method confirms that it captures dissonance and/or roughness reliably in our dataset (follow this link ([https://osf.io/4pnxh/?view\\_only=75afeb0864964265ab40e29a60895885](https://osf.io/4pnxh/?view_only=75afeb0864964265ab40e29a60895885)) for sound examples). Takes in which at least 10% of each acoustic feature could be reliably extracted were included in the analysis (this low rate was chosen to allow for the fact that CFI often involves musical textures that do not contain harmonic signal). Pitch, centroid, HNR, and dissonance were only computed in the windows in which the RMS value was above a certain threshold ( $-60$ ), chosen to discriminate between background noise and sound in these recording conditions. To assess changes with respect to the prompt, these values were normalized for each musician and take.

#### 3.1.4. Statistical analysis

Statistical analysis was performed in *R*. We ran *rmANOVAs* whenever possible, and linear mixed regressions with the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2014) when there were missing data, or logistic mixed regressions when the dependent variable was binary. Hierarchical logistic or linear mixed regressions included trios, pairs, or performers as random factors depending on the analysis. We report chi-squares, degrees of freedom, and *p* values for hierarchical nested model comparisons with likelihood ratio tests testing main effects and interactions (Gelman & Hill, 2007), followed by estimates, standard errors, *z* or *t* values and *p* values for model comparisons between factors.

### 3.2. Experiment 2—Results

#### 3.2.1. Impact of the number and type of prompts on temporal coordination

To assess the effect of Prompt Number and Prompt Type on temporal coordination, we ran a linear mixed regression with the *Temporal Coordination of Endings* as a dependent variable, Prompt Number and Prompt Type as independent variables, and Trio as a random factor (see Fig. 2A). This analysis revealed a main effect of Prompt Number ( $\chi^2 = 9.61$ ;  $p = .008$ ), a main effect of Prompt Type ( $\chi^2 = 10.8$ ;  $p = .001$ ), and a significant interaction between the two factors ( $\chi^2 = 8.93$ ;  $p = .011$ ). As predicted by the *shared information* and the *shared intention* hypotheses, temporal coordination improved as the number of prompts increased: It was better when there were three prompts ( $M = 4.5$  s,  $SD = 2.65$ ) as compared to when there was only one ( $M = 8.6$  s,  $SD = 4.46$ ,  $\beta = -3.6$ ,  $SEM = 1.27$ ,  $df = 12$ ,  $t = -2.84$ ,  $p = .014$ ) or two prompts ( $M = 9.9$  s,  $SD = 5.44$ ,  $\beta = -4.8$ ,  $SEM = 1.49$ ,  $df = 12$ ,  $t = -3.22$ ,  $p = .007$ ; the difference between one and two prompts was not significant,  $\beta = -1.2$ ,  $SEM = 1.64$ ,  $df = 11$ ,

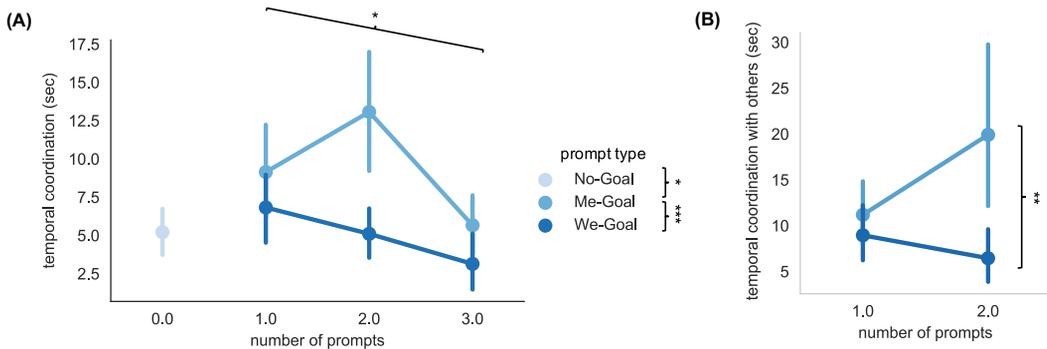


Fig. 2. (A) Temporal coordination of endings averaged per trio depending on prompt type and number. (B) Un-prompted musicians' temporal coordination with other musicians' depending on prompt type and number. \* represents significant outputs of the model with a threshold of  $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ . Error bars show the 95% interval.

$t = -0.74$ ,  $p = .47$ ). Crucially, as predicted by the *collective intention* and the *shared intention* hypotheses, the main effect of Prompt Type was such that musicians exhibited a better temporal coordination in the WE ( $M = 5.25$  s,  $SD = 1.98$ ) as compared to the ME condition ( $M = 10.5$  s,  $SD = 4.73$ ,  $\beta = -4.4$ ,  $SEM = 1.1$ ,  $df = 18$ ,  $t = -3.97$ ,  $p < .001$ ). Thus, the nature of the prompted goals (i.e., collective vs. individual) impacted how well musicians were able to temporally coordinate with each other, which is consistent with the idea that shared information is not the only factor that would impact coordination, but that the content of goals (i.e., whether they involve the individual alone, or the group as a whole) is also crucial.

The interaction between Prompt Type and Prompt Number reflected the fact that the *Temporal Coordination of Endings* significantly improved as the number of prompts increased in the WE condition ( $\chi^2 = 4.96$ ,  $\beta = -1.77$ ,  $SEM = 0.72$ ,  $df = 11$ ,  $t = -2.47$ ,  $p = .03$ ) but not in the ME condition ( $\chi^2 = 2.55$ ,  $\beta = -1.74$ ,  $SEM = 1$ ,  $df = 11$ ,  $t = -1.7$ ,  $p = .11$ ). The *Temporal Coordination of Endings* was significantly smaller in the WE as compared to the ME condition when there were two prompts ( $\beta = -8.54$ ,  $SEM = 1.76$ ,  $df = 58$ ,  $t = -4.84$ ,  $p < .001$ ), but this effect did not reach significance when there was only one prompt ( $\beta = -2.5$ ,  $SEM = 1.69$ ,  $df = 51$ ,  $t = -1.47$ ,  $p = .15$ ), or when there were three prompts ( $\beta = -2.38$ ,  $SEM = 1.8$ ,  $df = 61$ ,  $t = -1.32$ ,  $p = .19$ ). This suggests that the difference between the intentional content of the goals was greatest in situations of partial sharedness, as compared to situation of full sharedness or lack of sharedness. This is not entirely compatible with the *shared intention* hypothesis (and with our pre-registered hypothesis): Although it specifically predicts that temporal coordination should improve with the number of prompts in the WE condition, this hypothesis would also predict that temporal coordination would be maximal in the condition where the three musicians received a WE-Goal. This lack of effect for post hoc comparisons may be due to a lack of power. In any case, the *collective intention* hypothesis does not make specific predictions regarding the impact of the

number of prompts, and the *shared information* hypothesis does not make specific predictions regarding the impact of the type of prompts. Thus, the *shared intention* hypothesis more adequately captures the complexity of the data, in particular since it predicted that there should be an interaction between the number of prompts and prompt type, and that the impact of the number of prompts on temporal coordination should be restricted to the WE condition, as observed here.

Interestingly, the level of temporal coordination measured in the WE condition in Experiment 2 was not different from that measured in Experiment 1 (Experiment 1,  $M = 5.21$ ,  $SD = 2.8$ , linear mixed model comparison:  $\beta = -1.29$ ,  $SEM = 0.84$ ,  $df = 73$ ,  $t = -1.53$ ,  $p = .13$ ). By contrast, temporal coordination was significantly worse in the individual intention condition (ME-Goal) than in Experiment 1 ( $\beta = -1.85$ ,  $SEM = 0.83$ ,  $df = 149$ ,  $t = -2.24$ ,  $p = .027$ ). This is consistent with our observation that in the unconstrained CFI conditions of Experiment 1, the ending goals that spontaneously emerge are likely to be collective intentions rather than individual intentions.

Finally, we computed a linear mixed regression with the unprompted musicians' *Temporal Coordination with Others* as a dependent variable (see Fig. 2B). This analysis revealed a main effect of Prompt Type ( $\chi^2 = 4.31$ ;  $p = .038$ ), no effect of Prompt Number ( $\chi^2 = 0.04$ ;  $p > .5$ ), and a marginal interaction ( $\chi^2 = 3.35$ ;  $p = .07$ ). Unprompted musicians were more temporally coordinated with others in the WE ( $M = 0.75$ ,  $SD = 9.78$ ) condition than in the ME condition ( $M = 10.17$ ,  $SD = 13.58$ , linear mixed comparison:  $\beta = -3$ ,  $SEM = 1.42$ ,  $df = 79$ ,  $t = -2.15$ ,  $p = .035$ ). Thus, the existence of even a partially shared intention within the group was enough to improve the ability of the unprompted musicians to coordinate with others: It not only impacted the performance of prompted musicians, but also the performance of the group as a whole, which is consistent with the *shared intention* hypothesis.

Overall, the results show that temporal coordination was not only impacted by shared information (i.e., the number of prompts), but also by the collective nature of the intention (i.e., whether it was a WE or a ME-Goal): Crucially, temporal coordination was improved when musicians were asked to look for an end *collectively*. This impact of the collective content of intentions, over and beyond the presence of shared information, shows that the effect of goals on coordination is not only a matter of having *parallel* individual goals (e.g., having the musicians looking to stop their individual parts at the same moment). Rather, having goals that involved the group as a whole—that is, goals whose content can truly be *shared* by different members of the group—made a crucial difference in the temporal coordination of the performers. Taken together, these results favor the *shared intention* hypothesis.

### 3.2.2. *Impact of the number and type of prompts on dynamic, timbral, and harmonic coordination*

To investigate whether receiving prompts changed the relationships between the musicians, we conducted acoustic analysis on musical snippets extracted before and after the prompts. Following previous research (Pachet et al., 2017), we approximated musical coordination by computing a linear (Pearson correlation) as well as a nonlinear (mutual

information) index of dependency between musicians for five acoustic features: pitch, volume (RMS), playing time ratio (% of sound), spectral centroid, and HNR (see methods, section 3.1.3).

First, and before analyzing how the prompted goals impacted coordination at the acoustic level, we verified that our measures effectively captured some forms of musical coordination. This is nontrivial in our case since, as detailed above, CFI is generally devoid of harmonic and rhythmic structure. To this aim, we simply tested whether the linear correlation between acoustic features across time differed from zero overall. Correlation within trios (i.e., Pearson's rho averaged for each trio so as to estimate coordination at the level of the group) was significantly higher than chance for two of the five acoustic features (rms  $r$ :  $M = 0.17$ ,  $SD = 0.06$ ,  $t(11) = 8.9$ ,  $p < .001$ ; playing time ratio  $r$ :  $M = 0.15$ ,  $SD = 0.08$ ,  $t(11) = 6.45$ ,  $p < 0.001$ ), marginally higher than chance for two acoustic features (pitch  $r$ :  $M = 0.02$ ,  $SD = 0.03$ ,  $t(11) = 2.07$ ,  $p = .06$ ; HNR  $r$ :  $M = 0.023$ ,  $SD = 0.037$ ,  $t(11) = 2.02$ ,  $p = .07$ ), and did not significantly differ from zero for the spectral centroid ( $M = 0.02$ ,  $SD = 0.04$ ,  $t(11) = 1.72$ ,  $p = .11$ ). Thus, four of our five measures captured substantial acoustic coordination. These results—although reflecting rather weak associations—are in fact quite significant when related to the astounding variety and complexity of timbral and instrumental expressions found in CFI, and the fact that previous studies involving jazz musicians and similar measures failed to capture substantial acoustic coordination over and beyond the coordination explained away by the shared musical score (Pachet et al., 2017).

With this in mind, we examined our main question of interest, which was to assess whether shared intentions impact musical coordination (see Fig. 3). To assess this, we ran a logistic mixed regression with timing (before or after) as a dependent variable, prompt type, prompt number, and acoustic coordination variables (Pearson's rho and MI for the five acoustic dimensions, as well as dissonance) as independent variables, and trio as a random factor. After the prompt, there was a significant increase in mutual information for loudness ( $beta = 4.1$ ,  $SEM = 1.1$ ,  $df = 204$ ,  $z = 3.77$ ,  $p < .001$ ), a significant decrease in mutual information for pitch ( $beta = -2.8$ ,  $SEM = 0.85$ ,  $df = 204$ ,  $z = 3.29$ ,  $p < .005$ ), as well as a decrease in dissonance ( $beta = -0.0016$ ,  $SEM = 0.0005$ ,  $df = 204$ ,  $z = 3$ ,  $p < .005$ ). Thus, the prompts substantially modified dynamic and harmonic aspects of musical coordination.

Over and above these main effects, we also observed that prompt type and number differentially impacted musical coordination, and we break down these effects in Fig. 3's caption for each acoustic dimension. For pitch, we found that the decrease in mutual information was actually restricted to the ME condition: There was a significant interaction between timing and prompt type ( $beta = -3.9$ ,  $SEM = 1.8$ ,  $df = 165$ ,  $z = -2.14$ ,  $p = .03$ ), and the decrease was significant in the ME ( $t(11) = -3.65$ ,  $p = .004$ ) but not the We ( $t(10) = -1.28$ ,  $p > .23$ ) condition. Thus, after hearing a "ME" prompt, the pitch of the music produced by the improvisers became more independent from the pitch produced by other musicians, but this effect was not observed after they heard a "WE" prompt. For loudness, the decrease in mutual information did not significantly interact with prompt type or number. By contrast, the Pearson correlation was significantly

impacted by prompt type ( $\beta = -6$ ,  $SEM = 2.16$ ,  $df = 165$ ,  $z = -2.79$ ,  $p = .005$ ): The linear relationships between musicians' volumes significantly increased after WE ( $t(10) = 2.26$ ,  $p = .047$ ) but not ME ( $t(11) = 0.5$ ,  $p > .6$ ) prompts. However, the decrease in dissonance did not significantly interact with prompt type or number: The music was less dissonant after the prompt both in the ME ( $t(11) = -2.9$ ,  $p = .014$ ) and WE ( $t(10) = -3.22$ ,  $p = .009$ ) conditions. Finally, for timbral aspects (centroid and HRN) and the percentage of sound, there were no main effects and no interactions (see Fig. 3 caption for details).

Overall, these analyses suggest that the presence of goals impacts musical coordination during improvised interactions: Even at the basic level captured by our acoustic analysis, prompts had an impact on how improvisers' musical actions related to one another, at least for coordination at the harmonic and dynamic (i.e., loudness) levels. Specifically, when they had a WE-Goal, musicians' productions evolved toward being more consonant,

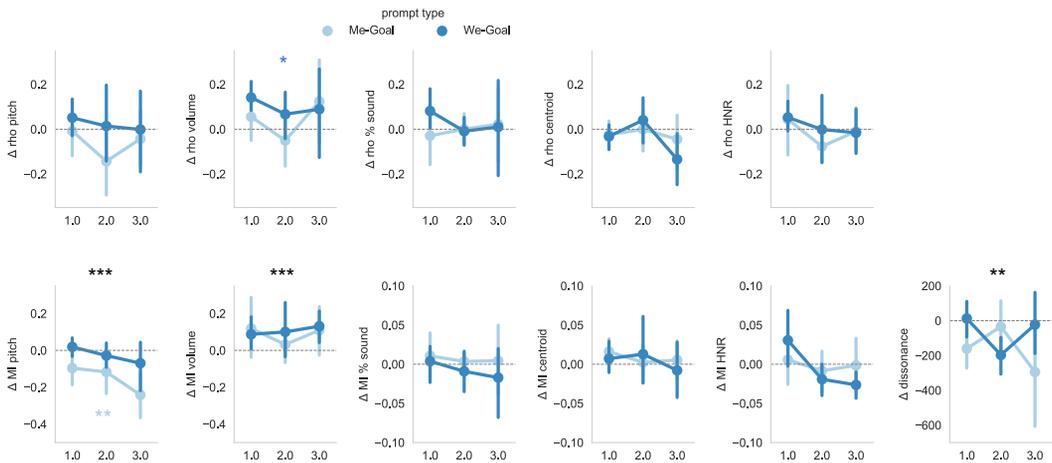


Fig. 3. Change in dynamic, timbral, and harmonic coordination after the prompt depending on prompt type and number. For each take, timing (after/before) and each trio, musical coordination was assessed by computing the mutual information or Pearson correlation between each pair, between averaging these values within each trio separately depending on prompt type and number. We also computed a measure of dissonance over the whole trio for each take and timing, before averaging it separately depending on prompt type and number. Black asterisks show main effects of timing (before/after); colored asterisks show main effects of prompt type. Error bars show the 95% confidence intervals. Significant impacts of prompt type and number on the acoustic measures of musical coordination are detailed in the main text. For centroid, there was no main effect of centroid on timing, and no interactions with prompt type or number. It is worth noting, however, that there was a significant decrease in spectral centroid's correlation in the WE-3 condition after the prompt ( $t(10) = -2.27$ ,  $p = .046$ , all other comparisons n.s.), which may reflect an attempt of the musicians to distribute themselves in different parts on the spectrum (i.e., an increase in musical coordination). For HNR, there was no main effect of centroid on timing, and no interactions with prompt type or number. Again, it is worth noting nonetheless that there was a significant decrease in HNRs' mutual information in the WE-3 condition after the prompt ( $t(10) = -2.87$ ,  $p = .017$ , all other comparisons n.s.), which may reflect an attempt to produce textures that are more distinct (i.e., an increase in musical coordination). Percentage of sound: There were no significant effects for this measure.

and their loudness was more correlated over time, suggesting tighter musical coordination. When musicians received a ME-Goal, their production also became more consonant but, in addition, the pitches they produced became more independent from one another, and they did not show improved coordination (i.e., tighter correlation) at the level of loudness.

#### **4. Experiment 3: Impact of the number and type of prompts on qualitative aspects of musical coordination**

Next, we wanted to assess whether shared intentions impacted properties of the performance related to higher level and qualitative aspects of musical coordination, beyond temporal coordination and the relatively low-level acoustical features that we examined in Sections 3.2.1 and 3.2.2. A particularly interesting question is whether the impact of shared intentions on the performance can be perceived by external observers and reflected in their aesthetic evaluations. Thus, in a third experiment, we asked third-party listeners (both experts and nonexperts) to rate the extent to which they thought the ending was successful and to classify the endings along several categories corresponding to qualitative aspects that are linked to coordination during CFI.

##### *4.1. Experiment 3—Methods*

###### *4.1.1. Participants*

We determined the size of the sample with a power analysis involving musicians' sensitivity in guessing each other's prompts (Experiment 2, see Fig. S5). To have a power of 95% at the 0.05 alpha level, the analysis showed that we should aim to test 23 participants per group. Given scheduling constraints, we finally tested 26 naive listeners (8 women, age  $M = 27.4$  years,  $SD = 8$  years) who were not musicians (mean number of years of instrumental practice:  $M = 0.42$ ,  $SD = 1.08$ ) and had no experience of CFI (mean number of years of CFI practice:  $M = 0$ ,  $SD = 0$ ) and 21 experts (5 women, age  $M = 33.9$  years,  $SD = 8.8$  years) who were all accomplished musicians (mean number of years of instrumental practice:  $M = 23.14$ ,  $SD = 8$ ) with a strong experience of CFI (mean number of years of CFI practice:  $M = 10.7$ ,  $SD = 6.5$ ). Participants reported having no major hearing or visual impairment, and appropriate corrections allowing them to perceive the stimuli. They signed an informed consent and were compensated financially after the experiment.

###### *4.1.2. Stimuli*

We selected 24 improvisations pseudo-randomly from those recorded in Experiment 2 by ensuring that (a) no trio was over-represented; (b) every trio was included; (c) the main findings were replicated in the subset (i.e., the impact of Prompt Type and Number on the *Temporal Coordination of Endings*); (d) half of the improvisations were taken from the ME

condition, and half from the WE condition; (e) each individual musician played during at least 19 s after prompt delivery (this last condition matters only for Experiment 4, presented below, which relies on the same subset of improvisations than Experiment 3).

#### 4.1.3. Procedure and data analysis

Listeners heard the last 50 s of each of the 24 improvisations and indicated on a 7-point Likert scale whether they thought that what they just heard was a good ending or not. Listeners were also asked in a random order whether the ending was (a) hierarchical or egalitarian; (b) collective or disjoint; (c) progressive or immediate; (d) predictable or surprising; and (e) timely or not (too late or too early). These five qualitative aspects were derived from musicians' reports during Experiment 2, where their judgments of appreciation were generally related to one or several of these categories. To infer categories from these written reports, three of the authors (L.G., P.S.-G., and C.C.) read all of the reports and grouped them in several categories. These subjective groupings were quite consistent among the three authors and suggested that the five aspects listed above capture most of the relevant parameters reflecting the success of coordination during CFI. To ensure that all participants understood the five qualitative aspects in a similar fashion, we provided them with a glossary describing the meaning of each label (see Section S.3.1 in Appendix S1). We analyzed appreciation ratings as a continuous variable, and qualitative ratings were dummy coded as binary variables (e.g., for the hierarchical category, we dummy coded hierarchical responses as 1, and egalitarian as 0). Data, data collection, and analysis scripts are available via this link ([https://osf.io/4pnxh/?view\\_only=75afeb0864964265ab40e29a60895885](https://osf.io/4pnxh/?view_only=75afeb0864964265ab40e29a60895885)).

## 4.2. Experiment 3—Results

### 4.2.1. Shared intentions impact the success of endings

We analyzed the impact of Prompt Type and Prompt Number on listeners' appreciation ratings with an rmANOVA (see Fig. 4A). There was an interaction between Prompt Type and Prompt Number ( $F(2, 90) = 4, p = .021, \eta_p^2 = 0.04$ ), a main effect of Prompt Number ( $F(2, 90) = 11.5, p < .001, \eta_p^2 = 0.06$ ), and no main effect of Prompt Type ( $F(1, 45) = 0.027, p > .8, \eta_p^2 = .00$ ). Appreciation ratings were highest in the WE-3 condition (ratings were higher in the WE-3 condition than in WE-2,  $p < .001$ ; ME-3,  $p = .007$ , ME-2,  $p = .006$ , post hoc Tukey HSD). Listeners' appreciation ratings were thus maximal when performers had a shared intention, which is consistent with our hypothesis that shared intentions help musicians to coordinate and attain a better outcome. We also examined the relationship between appreciation, Prompt Type, Prompt Number, and expertise, and report these results in Fig. S6A. Overall, the impact of shared intentions on musical coordination could be perceived independently from expertise, which suggests that even in an avant-garde artistic form like CFI, coordination relies on features that are transparent enough to be accessible to the general population (see Moran, Hadley, Bader, & Keller, 2015, for a similar finding regarding expressive movements).

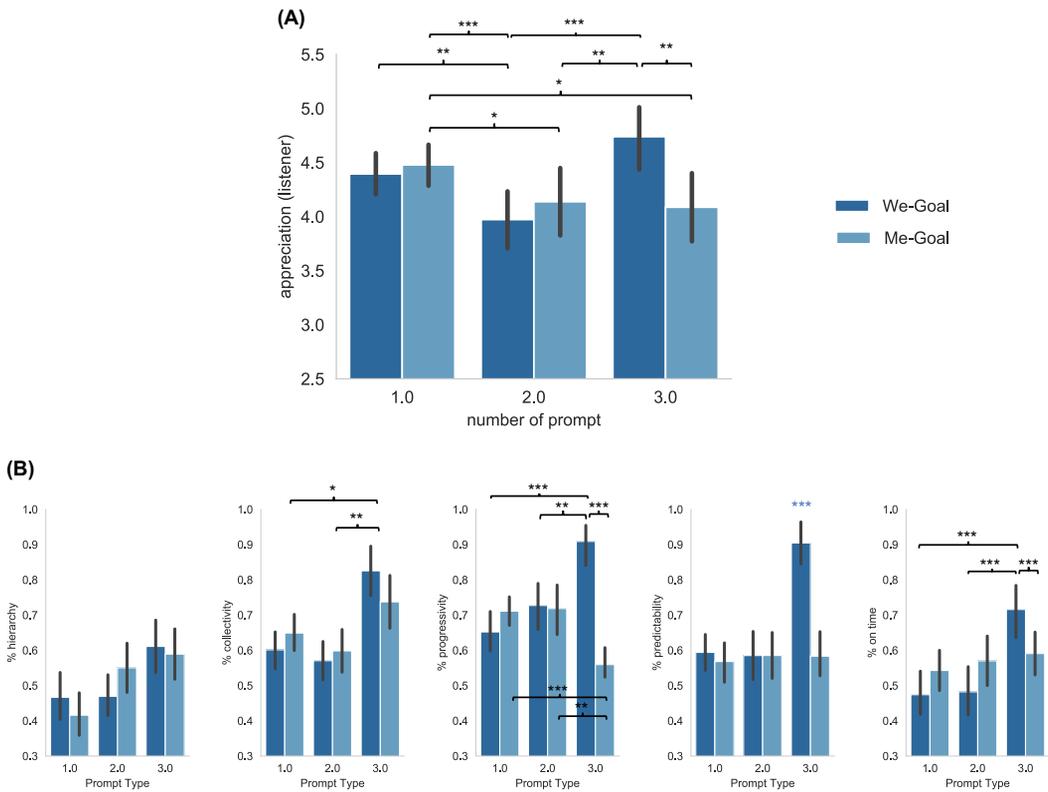


Fig. 4. Mains results of Experiment 3. (A) Expert and naive listeners’ appreciation ratings were averaged separately for each participant, prompt number, and prompt type, before being averaged in the group. Black asterisks show post hoc Tukey HSD comparisons. As reported in the main text, appreciation ratings were highest in the shared goal (WE-3). Participants also preferred the ME-1 condition over the ME-2 ( $p = .04$ ), ME-3 ( $p = .04$ ), and WE-2 ( $p = .001$ ) conditions. Similarly, they preferred the WE-1 condition over the WE-2 condition ( $p = .007$ , all other comparisons were nonsignificant). Thus, listeners also preferred conditions in which fewer prompts were present (WE-1 and ME-1 conditions did not differ  $p > .5$ ). This may be due to the fact that these interactions are less artificial than the others (i.e., only one of the musicians receives a prompt while the other musicians remain unconstrained). Note that musicians in these more natural conditions may also spontaneously form shared intentions, as suggested by the results observed in the first experiment. (B) The percentage of hierarchical, collective, progressive, predictable, and on-time assessment was computed for each of the five qualitative questions, separately for each participant, prompt number, and prompt type, before being averaged in the group. Black asterisks show the logistic regression model comparisons, and the blue asterisk represents the fact that all comparisons were significant with respect to the indicated condition.  $p < .05$ ;  $**p < .01$ ;  $***p < .001$ . Error bars show 95% confidence intervals.

#### 4.2.2. Shared intentions impact qualitative aspects of endings

To measure the impact of goals on the characteristics of the improvised joint action, we ran logistic mixed regressions for each of the five qualitative aspects (i.e., Hierarchy, Collectivity, Progressivity, Predictability, Timing), with Prompt Type and Prompt Number as independent variables, and listener as a random factor (see Fig. 4B).

For Collectivity, there was a significant effect of Prompt Number ( $\chi^2 = 9.5, p = .009$ ), in which listeners perceived endings to be more collective when the three musicians received a prompt than when only one musician received a prompt (model comparison between 3 vs. 1 prompt:  $beta = 0.37, SEM = 0.18, df = 1053, z = 1.99, p = .047$ ) and when two musicians received a prompt (3 vs. 2 prompts:  $beta = 0.6, SEM = 0.2, df = 1,053, z = 3, p = .002$ ). While this is consistent with both the *shared information* hypothesis and the *shared intention* hypothesis, the results are more clearly in favor of the *shared intention* hypothesis for the remaining aspects.

For Progressivity, there was a significant interaction ( $\chi^2 = 32, p < .001$ ). Listeners judged endings to be more progressive when the three musicians received a WE-Goal (model comparison between 3 vs. 1 prompt:  $beta = 1.32, SEM = 0.33, df = 1053, z = 3.9, p < .001$ ; 3 vs. 2 prompts:  $beta = 1.11, SEM = 0.35, df = 1053, z = 3.15, p = .002$ ), and less progressive when the three musicians received a ME-Goal (3 vs. 1 prompt:  $beta = -0.9, SEM = 0.25, df = 1053, z = -3.63, p < .001$ ; 3 vs. 2 prompts:  $beta = -0.86, SEM = 0.29, df = 1053, z = -3, p = .003$ ; comparison between WE-3 and ME-3:  $beta = 1.94, SEM = 0.38, df = 1053, z = 5, p < .001$ ) as compared to the other conditions.

For Predictability, there was a significant effect of Prompt Type ( $\chi^2 = 9, p = .003$ ) and a significant interaction between the two factors ( $\chi^2 = 21.57, p < .001$ ). Listeners judged endings to be more predictable when the three musicians had received a WE-Goal as compared to the other conditions (all comparisons between the WE-3 condition and the other conditions were highly significant, and none of the other comparisons were significant).

Finally, and crucially, regarding Timing, there was a significant interaction ( $\chi^2 = 16.52, p < .001$ ). While in the ME condition no significant differences were observed depending on prompt number (all  $ps > .07$ ), listeners in the WE condition judged endings to be timelier when the three musicians had received a prompt (3 vs. 1:  $beta = 0.96, SEM = 0.26, df = 1053, z = 3.7, p < .001$ ; 3 vs. 2:  $beta = 1.06, SEM = 0.27, df = 1,053, z = 3.86, p < .001$ ). In addition, listeners judged endings to be significantly timelier in the WE as compared to the ME condition when there were three prompts ( $beta = 1.13, SEM = 0.31, df = 1053, z = 3.6, p < .001$ ), but not two prompts ( $beta = 0.12, SEM = 0.23, df = 1053, z = 0.6, p > .5$ ) or one prompt ( $beta = 0.28, SEM = 0.17, df = 1,053, z = 1.7, p > .09$ ).

In other words, for Progressivity, Predictability, and Timing, there was a specific impact of shared intentions over and beyond shared information. These results complement the findings above and confirm that shared intentions impact not only temporal and acoustic coordination, but also higher level qualitative properties of the joint improvisation that can be perceived by expert and naive listeners alike.

## 5. Experiment 4: How do improvisers' goals propagate?

A remaining question concerns how goals propagate within the group, and whether they can be perceived from the music alone. In a last experiment, we wanted to test the

claim that transparent goals (i.e., goals that are easier to detect) have a more positive impact on coordination. This is a specific prediction of the *shared intention* hypothesis, according to which improvisers may coordinate through forming collective intentions that are shared and common knowledge between them. To this aim, we asked naive and expert listeners to try and detect whether individual musicians had an intention to end the performance. We also examined the relationship between listeners' detections of goals and temporal coordination, to see whether transparent goals corresponded to better temporal coordination. Finally, we wanted to try and assess how goals may be manifested, and thus effectively propagate within the group. To examine this issue, listeners were also asked to characterize performers' behaviors along four qualitative aspects. They were asked whether they thought that the musicians' behavior was: (a) descending or not descending (i.e., ascending, constant or without direction); (b) repetitive or varied; (c) predictable or surprising; and (d) confident or hesitant. This also allowed us to examine whether specific behaviors are associated with better temporal coordination and/or shared intentions, suggesting that they may be used by the performers as coordination smoothers or communicative signals (Vesper et al., 2017).

## 5.1. Experiment 4—Methods

### 5.1.1. Stimuli

Stimuli were 72 audio extracts from the three individual performances in each of the 24 improvisations used in Experiment 3. All stimuli were 17 s long, extracted either 17 s before the prompt (Before condition,  $N = 18$  extracts) or 17 s after the prompt, either in trials in which the musician heard a ME-Prompt (ME-Goal condition,  $N = 18$  extracts), a WE-Prompt (WE-Goal condition,  $N = 18$  extracts), or did not hear a prompt (No-Prompt condition,  $N = 18$  extracts). None of the extracts included the actual ending of the piece (i.e., in all of these takes, every musician stopped at least 19 s after hearing the prompt).

### 5.1.2. Procedure and design

Participants were the same as for Experiment 3. They were told that in about half of the musical extracts, musicians were looking for an ending and were about to stop playing, while in the other half they were not looking for an ending. They were asked to report—via a key press (left or right arrow, counterbalanced between participants)—whether the musician was about to stop playing (i.e., to detect ending goals). Participants then provided a confidence rating in their answer on a scale from 1 to 4 and categorized the musician's behavior by responding to four questions presented in a random order. For each category, participants were presented with several alternatives (direction: ascending/descending/constant/none; repetition: repetitive/varied; prevision: predictable/surprising; assurance: confident/hesitant) and asked to select one of them by pressing one of the arrows on the keyboard. These categories were derived from the musicians' reports during Experiment 1, where decisions about their partners' intentions were reported to be caused by one or several of these behaviors (see Section S.1.2 in Appendix S1, for a few

examples and details of the procedure that allowed us to extract these categories from musicians' written reports about how they detected their partners' intention to end during experiment 1). Listeners were provided a glossary to make sure that all of them understood these categories in the same way (see Section S.4.1 in Appendix S1).

### 5.1.3. Data analysis

We computed a measure of sensitivity based on signal detection theory ( $d'$ , Green & Swets, 1966) for each participant and condition, taking tracks extracted after the prompt (NO/ME/WE) as targets, and tracks extracted before the prompt (Before) as non-target. For each participant and condition (NO/ME/WE), the hit rate was computed as: (the number of positive responses for extracts taken after the prompt for that condition/the total number of extracts taken after the prompt for that condition); and the false alarm rate as: (the number of positive responses for extracts taken before the prompt/the total number of extracts taken before the prompt). Note that, although we treated the NO-Goal condition like the WE and ME-Goal conditions to compute  $d'$  here, so as to allow direct comparison between the three conditions, detecting an ending in this condition is not necessarily a “wrong” response: The unprompted musician may or may not have an intention to end depending on whether the goal propagated in the group or not. Data, data collection, and analysis scripts are available via this link ([https://osf.io/4pnxh/?view\\_only=75afeb0864964265ab40e29a60895885](https://osf.io/4pnxh/?view_only=75afeb0864964265ab40e29a60895885)).

## 5.2. Experiment 4—Results

### 5.2.1. Third-party listeners can detect improvisers' goals

Average sensitivity ( $d'$ ) was  $M = 0.37$ ,  $SD = 0.56$ , which was significantly above chance level ( $t(46) = 4.48$ ,  $p < .001$ ). An rmANOVA revealed a main effect of Prompt Type (NO/ME/WE:  $F(2, 90) = 30.8$ ,  $p < .001$ ) on sensitivity, and an interaction between Expertise and Prompt Type ( $F(2, 90) = 3.14$ ,  $p = .048$ ). As can be seen in Fig. 5A, both experts ( $d'$ :  $M = 0.71$ ,  $SD = 0.74$ ) and naive listeners ( $d'$ :  $M = 0.55$ ,  $SD = 0.46$ ) achieved above chance sensitivity in the ME condition (musicians:  $t(20) = 4.28$ ,  $p < .001$ ; non-musicians:  $t(25) = 6$ ,  $p < .001$ ), and there was no difference between the two groups in this condition (post hoc Tukey HSD:  $p = .32$ ). By contrast, sensitivity in the WE condition varied with expertise: While experts achieved above chance sensitivity ( $M = 0.72$ ,  $SD = 0.64$ ,  $t(20) = 5$ ,  $p < .001$ ), naive listeners' sensitivity did not significantly differ from chance ( $M = 0.21$ ,  $SD = 0.67$ ,  $t(25) = 1.56$ ,  $p = .13$ ; group difference:  $p = .002$ ).

Thus, ME-Goals could be perceived from musicians' behavior independently from listener's expertise, while the detection of WE-goals depended on expertise. This suggests that WE-goals—that is, goals whose content refer to the group's performance as a whole—may be characterized by specific features that are only accessible to expert listeners. One could argue that this impact of expertise is due to musicians' better auditory processing capacities that would enable them to attend to finer acoustic cues which carry this information. Yet this interpretation is not compatible with the lack of difference between

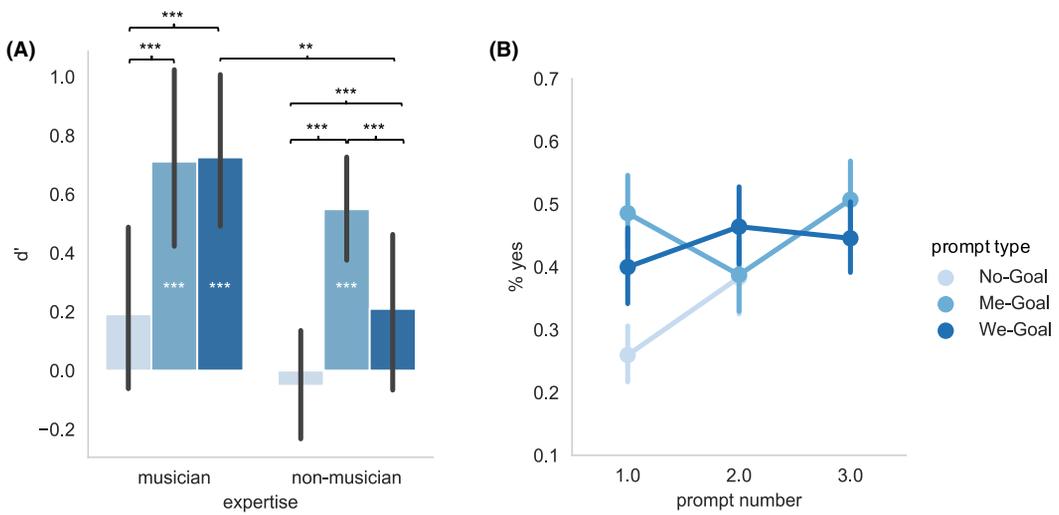


Fig. 5. (A) Participants sensibility ( $d'$ ) was assessed by computing for each condition and participant the hit rate (number of positive responses for snippets extracted after the prompt/number of snippets extracted after the prompt) and false alarm rate (number of positive responses for snippets extracted before the prompt/number of snippets extracted before the prompt). White asterisks show  $p$  values for one-sample  $t$  tests against chance level; black asterisks show post hoc Tukey HSD for between-group or conditions comparisons.  $***p < .001$ ;  $**p < .01$ . (B) The percentage of positive responses (i.e., “Yes, I think the performer is looking for an end”) was computed separately for each participant depending on prompt type and number, before being averaged in the group. A logistic mixed regression with responses (yes/no) as a dependent variable revealed that when only one of the performers had a Goal, listeners detected an intention to end less often when listening to the unprompted performer as compared to when both other performers had a ME-Goal ( $\beta = 1.19$ ,  $SEM = 0.16$ ,  $df = 5,025$ ,  $z = 7.4$ ,  $p < .001$ ) or a WE-Goal ( $\beta = 0.53$ ,  $SEM = 0.16$ ,  $df = 5,025$ ,  $z = 3.26$ ,  $p = .001$ ). Listeners also reported an intention to end more often when the performer was the only one having a ME- as compared to a WE-Goal ( $\beta = 0.65$ ,  $SEM = 0.18$ ,  $df = 5,025$ ,  $z = 3.63$ ,  $p < .001$ ). Error bars show 95% confidence intervals.

the two groups in the ME condition. More interestingly, it could be that WE-Goals depend on conventional behaviors that are only accessible to listeners possessing the same cultural background as the performers. We come back to this issue below. Notwithstanding, the results show that improvisers’ goals have some degree of transparency, and that they are manifested in the performance in ways that allow performers and external listeners to detect them.

### 5.2.2. Goal propagation: Shared intentions impact how listeners perceive unprompted musicians’ goals

In the NO-Prompt condition, sensitivity did not differ from chance level in any of the groups (musicians  $d'$ :  $M = 0.19$ ,  $SD = 0.66$ ,  $t(20) = 1.29$ ,  $p = .2$ ; non-musicians  $d'$ :  $M = -0.05$ ,  $SD = 0.49$ ,  $t(25) = -0.52$ ,  $p = .6$ ). Thus, overall, listeners did not perceive ending goals when performers did not receive a prompt themselves. This may suggest that the behavior of unprompted performers did not reflect an intention to end after one

or both of their co-performers were prompted. Yet it remains possible that it *only* did when *both* of their co-performers were prompted.

To examine this possibility, we examined how detection responses (yes/no) depended on Prompt Type and Number (Fig. 5B). We ran a mixed logistic regression with detection response as a dependent variable, and prompt type and number as independent variables. There was a main effect of Prompt Number ( $\chi^2 = 6.17$ ;  $p = .046$ ), a main effect of Prompt Type ( $\chi^2 = 61.17$ ;  $p < .001$ ), as well as an interaction between Prompt Number and Prompt Type ( $\chi^2 = 27.93$ ;  $p < .001$ ). A post hoc test revealed that when both of an unprompted performer's co-performers had a goal, listeners reported that the unprompted performer had an intention to end as often as they did when listening to prompted performers who had a ME-Goal ( $beta = 0.018$ ,  $SEM = 0.18$ ,  $df = 5,025$ ,  $z = 0.09$ ,  $p = .92$ ), but less often as compared to prompted performers who had a WE-Goal ( $beta = 0.41$ ,  $SEM = 0.18$ ,  $df = 5,025$ ,  $z = 2.26$ ,  $p = .024$ ). In addition, listeners reported that unprompted performers had an intention to end more often when both of their partners had an intention to end as compared to when only one of their partners had an intention to end (1 vs. 2 in the NO-Prompt condition:  $beta = 0.47$ ,  $SEM = 0.17$ ,  $df = 5,025$ ,  $z = 2.8$ ,  $p = .005$ , see Fig. 5B for a full output of the model).

Thus, unprompted performers' behavior did reflect their co-performers' goals to some extent, when those goals were shared by both co-performers. In line with the results of Experiment 2, this suggests that once goals are partially shared within the group, some form of goal propagation is taking place in the direction of the remaining individuals, with unprompted musicians behaving as if they had themselves received a prompt to find an end. Musicians may thus deploy communicative strategies to establish shared intentionality when their aim is to find an end to the piece collectively.

### 5.2.3. Improvisers adopt signaling strategies to communicate their goals

How may such goal propagation occur? To examine whether musicians deployed particular strategies to signal their intentions to end, we assessed the impact of our experimental conditions on how listeners described the musicians' behaviors. We ran a linear regression including percentage of response as a dependent variable, Condition (Before-Prompt/NO-Prompt/ME-Goal/WE-Goal), Category (descending/repetitive/predictable/confident), and Expertise (naive/expert) as independent variables, and listener as a random factor. There was a main effect of Condition ( $\chi^2 = 21$ ,  $p < .001$ ), a main effect of Category ( $\chi^2 = 647$ ,  $p < .001$ ) and, more importantly, a significant interaction between Condition and Category ( $\chi^2 = 50$ ,  $p < .001$ ), which revealed that listeners' judgments about performers' behaviors along each Category varied differently depending on Condition (see Fig. 6B). There was no additional interaction with Expertise ( $p > .14$ ), so we collapsed the data for the two groups of listeners for the remaining analyses.

Regarding direction, listeners responded that the musician's behavior was *descending* significantly more often when they heard prompted musicians (ME:  $M = 0.24$ ,  $SD = 0.12$ ; WE:  $M = 0.28$ ,  $SD = 0.15$ ) than unprompted musicians ( $M = 0.16$ ,  $SD = 0.1$ ; post hoc Tukey HSD No vs. ME:  $p < .001$ ; NO vs. WE:  $p < .001$ ) or extracts taken before the prompt ( $M = 0.19$ ,  $SD = 0.12$ ; Before vs. ME:  $p = .001$ ; Before vs. WE:

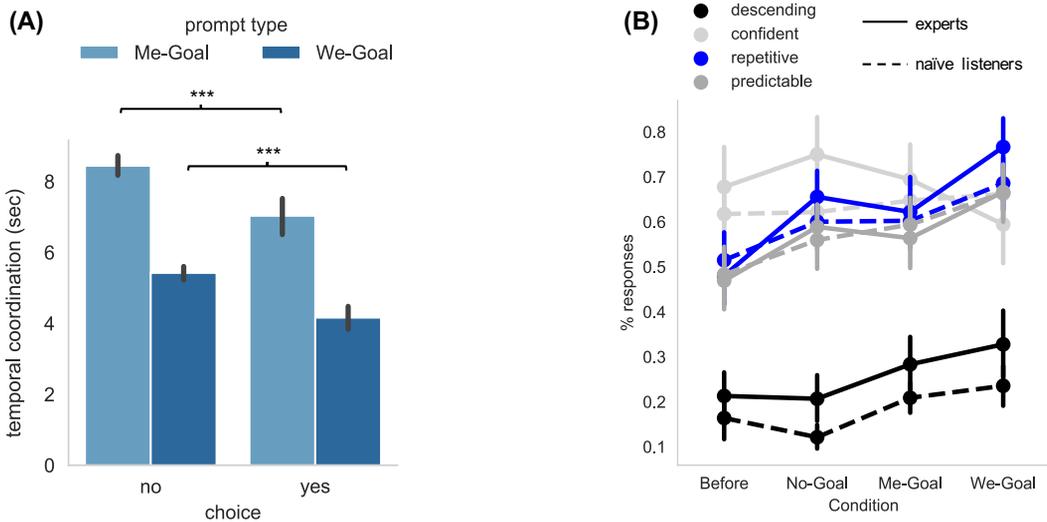


Fig. 6. Musicians' behavior. (A) The temporal coordination of endings in improvisations corresponding to the snippets heard by the participants was averaged separately for each listener, prompt type, and response type (yes/no), before being averaged in the group. \*\*\* show the significance of paired  $t$  tests with a threshold of  $p < .001$ . (B) We show the percentage of descending, repetitive, predictable, and confident responses computed for each condition and listener, before being averaged separately in the group of experts (plain line) and naive listeners (dashed line). Error bars show 95% confidence intervals.

$p < .001$ ; no significant difference between Before and NO-Prompt:  $p = .26$ ). Interestingly, there were no significant differences between the rate of descending responses in the ME and WE condition ( $p = .08$ ), which rules out the possibility that WE-Goals simply foster coordination because performers rely on decrescendos to drive the improvisation toward the end (also see the acoustic analysis presented in Fig. S10).

Listeners also perceived musicians to be less *confident* in the WE ( $M = 0.63$ ,  $SD = 0.18$ ) as compared to the NO-Prompt condition ( $M = 0.68$ ,  $SD = 0.19$ ,  $p = .026$ ) and, marginally, than in the ME condition ( $M = 0.67$ ,  $SD = 0.19$ ,  $p = .053$ ; comparison with Before condition:  $p = .5$ , all other comparisons nonsignificant,  $p > .1$ ). Thus, it seems that WE-Goals lead performers to be more hesitant, perhaps reflecting that they were “waiting for each other.”

Finally, and more importantly, listeners responded that behaviors were *predictable* and *repetitive* significantly more often when the performer had a WE-Goal ( $M = 0.66/0.72$ ,  $SD = 0.15/0.17$ ), as compared to when the performer had a ME-Goal ( $M = 0.58/0.61$ ,  $SD = 0.16/0.16$ ,  $ps < .001$ ), was not prompted ( $M = 0.57/0.62$ ,  $SD = 0.14/0.14$ ,  $ps < .001$ ), or for extracts taken before the prompt ( $M = 0.48/0.5$ ,  $SD = 0.15/0.16$ ,  $ps < .001$ ). Listeners also perceived behaviors to be more predictable/repetitive when performers had a ME-Goal ( $ps < .001$ ) or NO-Goals ( $ps < .001$ ), as compared to the extracts taken before the prompt.

The crucial finding here is that musicians relied on more predictable and repetitive behaviors when they had a WE-Goal, presumably to allow their partners to coordinate with them. These repetitive/predictable behaviors could be due to performers playing the same complex pattern over and over again or holding a single tone, but they were not necessarily related to performers playing a regular pulse (see Fig. S9 and Section S.4.1 in Appendix S1). This finding is consistent with previous research emphasizing the role of predictability and repetitive actions for coordination (Vesper et al., 2011) and emerging communication systems in the visual modality (Scott-Phillips, Kirby, & Ritchie, 2009), and it shows that improvisers used basic signaling strategies to help establish common ground when they have to reach a joint outcome with their fellow improvisers.

#### 5.2.4. Goal transparency predicts better temporal coordination

Finally, we wanted to test the claim that transparent goals (i.e., goals that are easier to detect) foster coordination. To this aim, we examined the relationship between listeners' goal detection and subsequent temporal coordination at the end of the piece (which was not presented to the participants). In a linear mixed regression restricted to judgments made on extracts taken after the prompt, and including listener and trio as random factors, listeners' detection choices (yes vs. no) significantly predicted the subsequent *Temporal Coordination of Endings* ( $\beta = 0.4$ ,  $SEM = 0.07$ ,  $df = 5068$ ,  $t = 6$ ,  $\chi^2 = 36.7$ ,  $p < .001$ ). On average, performers were more temporally coordinated in musical extracts where listeners detected an intention to end ( $M = 5.68$ ,  $SD = 1.19$ ) as compared to when they did not ( $M = 6.97$ ,  $SD = 0.62$ ,  $t(46) = 5.15$ ,  $p < .001$ , see Fig. 6A). This was true for both ME ( $t(46) = 3.72$ ,  $p < .001$ ) and WE-Goals ( $t(45) = -5.07$ ,  $p < .001$ ), and also after accounting for the effect of Prompt Number on temporal coordination ( $\beta = 0.24$ ,  $SEM = 0.06$ ,  $df = 5068$ ,  $t = 3.8$ ,  $p = .007$ ). This result is therefore consistent with the idea that goal transparency helps coordination, and that making one's goal easier to detect by fellow improvisers might be key to coordination during improvised interactions.

## 6. Discussion

Despite being an integral part of our social lives, joint improvised actions have been understudied to date, and the mechanisms that allow agents to coordinate in complex and temporally extended forms of collective improvisation remain elusive. The experiments reported here shed a new light on these mechanisms in the context of CFIs: In Experiment 1, we show that shared intentions emerge on the fly during collective musical improvisations; in Experiment 2, we show that the presence of such shared intentions fosters temporal and acoustic coordination; in Experiment 3, we show that shared intentions also have an effect on qualitative properties of the performance that reflect higher level aspects of musical coordination (such as the endings being rated as more successful, timelier, and more progressive); finally, in Experiment 4, we show that improvisers' goals can be inferred by third-party listeners from their musical behavior and that, strikingly, unprompted musicians may come to reflect the behaviors of their prompted co-

improvisers. The results also show that improvisers adopt signaling strategies when they have to communicate their goals to reach a joint outcome collectively, which explains how collective intentions can propagate, become common knowledge, and improve musical coordination.

Overall, the results are compatible with the hypothesis that *shared intention* fosters coordination during improvised musical joint actions, over and beyond the role of mere *shared information* and of the isolated formation of *collective intentions* in individual musicians. This demonstrates that the synergy between planned and emergent coordination mechanisms that had so far been considered exclusively in scripted joint actions is also at play in improvised joint actions. While our results are in line with the idea that shared intentions support coordination over long as well as short time scales (Vesper, Butterfill, Knoblich, & Sebanz, 2010), they—perhaps counterintuitively—extend its relevance to the case of collective musical improvisations.

An important theoretical consequence of our study is that it gives some additional ground to the idea that shared intentions do not intrinsically depend on verbal communication for their existence: We show that shared intentions can emerge when agents are freely and spontaneously interacting within a medium that is semantically underspecified (i.e., music), and that they play a key role in supporting coordination. The condition of common knowledge, where agents are not only geared toward a joint outcome, but also represent that this state of affair is publicly accessible to all members of the group, is generally taken to be one of the crucial features of shared intentions (Bratman, 2014). Now, in Experiment 2, the WE-goals were communicated covertly to each musician, apparently violating the requirement of common knowledge. However, this does not mean that such common knowledge status could not emerge in the course of the performance, *after* the musicians were prompted, using joint affordances (i.e., events that afford actions or gestures for the group as a whole, Knoblich et al., 2011), signaling strategies that trigger “*distinctive cognitive states, corresponding to the sense that something is public and unignorable*” (De Freitas, Thomas, DeScioli, & Pinker, 2019), and focal points that act as points of converging expectations for the improvisers (Canonne, 2013). Several aspects of our results are consistent with this possibility.

Results from Experiment 4 (Fig. 5) show that third-party listeners were able to infer WE-Goals from musicians’ behavior, demonstrating that these goal representations are indeed manifest and publicly observable. Results from Experiment 4 further suggest that particular communicative behaviors (e.g., repetitions) may be especially efficient to signal an intention to end the piece. Lastly, we saw in Experiment 2 that both ME-Goals and WE-Goals were detectable by co-agents (see Fig. S5). However, WE-Goals and ME-Goals did not differ in terms of their directionality (i.e., both were perceived as “descending”; see Fig. 6B). This rules out the possibility that improvisers merely detect teleological aspects such as a directionality in the joint action (e.g., decrescendos) without representing the mental states that may underlie this directionality in their co-agents, analogously to 2-year-old children who engage successfully in joint action before they have a full understanding of folk psychological concepts such as *intention* (Butterfill, 2013; Butterfill & Apperly, 2013). On the contrary, the findings suggest that improvisers considered

additional cues, beyond the mere sonic target, and engaged in some form of mentalizing to discriminate between the two types of goals. These elements indicate that musicians' goals were both *manifest* and *mentally represented* by their co-improvisers. As such, they had the potential to become common knowledge between improvisers and to amount to full-fledged shared intentions.

Now, even when musicians collectively hold a shared intention to end the performance, how and when the performance will actually end still remains poorly specified: Such abstract goals do not specify precise temporal or harmonic structures, allowing the musicians to coordinate on fine time scales. In other words, even if musicians manage to form a shared intention to end the performance, the ending will still have to be spontaneously and collectively negotiated in a matter of seconds, without the support of a shared entrainment to a beat. How do such abstract intentions support coordination in cases where the outcome remains highly undetermined? Several nonexclusive explanations might be provided here.

A first possibility is that once it is common ground for co-improvisers that there is a shared intention to *X* (e.g., “to end the performance together”), they can coordinate by relying on interconnected planning. That is, they can form compatible sub-plans that are constrained by their shared intention to look for an end to the performance together (Bratman, 2014). This is not to say that each agent necessarily represents the other agents' part precisely (Vesper et al., 2010). Still, once a shared intention is established, performers can monitor and predict their co-performers' actions more finely, and adjust their own behavior accordingly, because the shared intention constrains the range of possible interpretations of partners' behaviors, as well as each agent's action repertoire. This being said, although a minimal representation of one's own task and of the group's shared intention may suffice to finely coordinate in scripted joint actions that involve predetermined outcomes (Vesper et al., 2017), it is difficult to see how these mechanisms could allow musicians to precisely coordinate in the case of collective improvisations. Motor simulation is thought to be one of the crucial mechanisms that enable co-agents to predict each other's actions and coordinate on short time scales (Knoblich et al., 2011; Novembre et al., 2014; Vesper et al., 2013). Here, however, it is unlikely that musicians simply rely on their motor system, given that they play on different instruments (Bishop & Goebel, 2014), and that they use idiosyncratic instrumental techniques. But this does not mean that they cannot rely on action prediction at all. For instance, both expert and naive listeners perceived an intention to end in conjunction with decrescendos (see Figs. S8 and S11), which can be argued to be an index with a teleological origin (i.e., “descending” actions typically precede endings).

A second way in which shared intentions may foster coordination is by enabling behavioral strategies designed to help coordination (Vesper et al., 2017). For instance, we found some evidence that musicians' behavior tended to be more repetitive and predictable in the WE-Goal condition (see Fig. 6B). One interpretation of this result is that, in improvised joint actions, agents use repetitive actions and other predictable behaviors not only as signals but also as “coordination smoothers” (Vesper et al., 2017), to help

other improvisers predict and coordinate with them. In favor of this interpretation, we also found that predictability was associated with better temporal coordination (Fig. S8).

Lastly, at shorter time scales (i.e., a few seconds), it is possible that shared intentions regulate the emergent mechanisms that are at play to support fine-grained coordination. For instance, when musicians had shared intentions, the dynamics of their amplitude variations were more tightly coupled (see Fig. 3). This being said, the role of emergent coordination mechanisms is probably less crucial here than in other types of improvisations involving imitations such as the mirror game (Noy et al., 2011), because CFI is generally devoid of regular rhythmic pulsations, and straightforward imitations are often frowned upon among free improvisers. The acoustical analysis presented in Fig. S10, which shows little mimicry in unprompted musicians, is consistent with this idea: There was little to no evidence in favor of the idea that unprompted musicians adapt their behavior by simply mimicking prompted musicians (e.g., by playing decrescendos).

On the other hand, our results do not imply that agents engaged in collective musical improvisation always have a shared intention in mind, nor that they systematically need to. It is likely that musical improvisers oscillate between phases where they unreflectively “go with the flow” and phases in which they are more self-conscious and engage in deliberate planning of their actions and mind reading (Canonne & Garnier, 2012; Denzler & Guionnet, 2020). In that perspective, the fact that in Experiment 1 some musicians pressed their pedals after they had actually stopped playing suggests that musicians can be as surprised as audience members by the unfolding of their own performance. More generally, it is likely that shared intentions, to the extent that they are present, are of a rather punctual, short-term nature, emerging when acute coordination problems, such as endings or the consolidation of a new attractor (Borgo, 2005), arise. Further studies could examine in a more systematic fashion the temporal dynamics of the kinds of abstract, shared intentions we evidenced here.

On the methodological side, our study shows that collective musical improvisation constitutes an interesting case study to examine how individuals coordinate in the absence of scripts, and to investigate coordination dynamics in improvised interactions over an extended time span. When interactions between individuals are mediated by a pre-existing script, even loose ones such as the lead sheet of a jazz standard or conversational guidelines, it can be difficult to tease apart actual interpersonal interactions from an individual’s isolated interactions with the script they all share (Pachet et al., 2017). CFI does not involve such referents and, as such, it allows a direct, unmediated examination of interpersonal interactions.

Another interest of our approach is that it allows comparing expert and naive listeners, and makes it possible to uncover the (cultural) knowledge that mediates coordination during joint actions, an aspect often neglected in cognitive science (Vesper et al., 2017). While CFI is clearly a highly unplanned form of joint action, it does not happen in a cultural vacuum. Free improvisers spend many hours developing idiosyncratic instrumental techniques and a repertoire of distinctive musical materials (Arthurs, 2016). According to MacDonald and Wilson (2020, p. 115) though, “*particular knowledge or skills (...) are not in themselves a measure of the broader capacity to improvise.*” As such, an important

part of free improvisers' training, whether formal—through Conservatories classes—or informal—through listening to and playing with other improvisers—consists in developing broader coordination and communication skills, as well as highly general attributes, such as “*confidence in exercising choice in real time*,” “*discrimination and discernment of emerging performed material*,” and “*facility in accommodating and responding to unprecedented or unexpected events*” that are “*transferable across genres or settings in ways that some other musical attributes are not*” (MacDonald & Wilson, 2020, pp. 116–119). As Pelz-Sherman (1998, p. 127) puts it, learning to “*convey the semantic intent of their own musical ideas to other performers in real time*” and “*to make accurate judgments in real time about the semantic intent of each performer*” is crucial.

Overall, expertise in CFI seems to largely rely on social cognition, contextual attunement, and interpersonal coordination, as is the case for other forms of freely improvised practices such as comedy improv (Walsh, Roberts, & Besser, 2013) or contact improvisation (De Spain, 2014). Thus, it is possible that the signaling mechanisms used by free improvisers were only accessible to expert listeners in our study, not necessarily because they rely on group-specific or genre-specific expertise and conventions, but perhaps because they require a high level of social attunement to the behavior of the improvisers. In other words, it might just be that our expert listeners, being also expert free improvisers, were more used to face improvised coordination problems, and thus simply better at abstracting signaling strategies from subtle variations in the performers' behaviors. The wide variety of ending behaviors found in our corpus suggests that the signaling strategies used by the improvisers were not tied to precise instrumental or musical patterns, but were rather of a very abstract nature (e.g., decrease in energy, use of salient events, repetition, etc.), and thus possibly independent from the sonic and aesthetic specificities of CFI as a genre. Further experiments could directly test this hypothesis by assessing whether expert improvisers from another domain (e.g., comedy improv) are able to detect our musicians' intentions, despite the fact that they are unfamiliar with the genre of CFI. Note also that, consistent with this last hypothesis, some of our results tend to downplay the importance of group-specific stylistic conventions in the emergence of shared intentions among free improvisers. In particular, a high degree of familiarity between the musicians (and the implicit conventions that are likely to come with it) did not seem to give them any advantage in negotiating their joint endings: Familiarity did not correlate with how temporally coordinated they were in pressing the pedals (Experiment 1, Spearman's rho between pedal pressing temporal coordination and familiarity scores:  $r_s(10) = -.38, p = .240$ ), how well musicians coordinated at the end of the piece (Experiment 2, Spearman's rho between the temporal coordination of endings and familiarity,  $r_s(10) = .11, p = .730$ ), and nor even with how much they enjoyed playing together overall during Experiments 1 and 2 (Spearman's rho between global appreciation and familiarity scores,  $r_s(10) = 0.4, p = .200$ ).

If expertise in collective improvisation is mainly a matter of being able to attune oneself to the specificities of a given social setting, then the fact that improvisers interact in a shared environment should play a key role in the emergence of locally shared intentions. In particular, it is likely that salient features of the improvisers' sonic environment

(e.g., a clear pitch in an otherwise noisy texture or simultaneous impacts in an otherwise asynchronous sequence) provide the improvisers with the opportunity to adopt similar local goals (such as changing the musical direction, performing a collective crescendo or accelerando, developing a given idea, or ending the performance). And similarly, it is likely that the improvisers' active engagement in embodied interactions—the fact that they could continuously feel each other's actions and reactions on a fine-grained scale—played a significant role in the remarkable understanding of each other's intentions they displayed (Michael, 2011). In that sense, emphasizing the supporting role of shared intentions in an explanation of coordination in complex improvised actions does not necessarily undermine the role played by interactional or contextual factors; on the contrary, it is precisely *because* collective improvisations are embodied and embedded interactions—because improvisers both co-construct and explore their shared sonic environment *through* their bodily interactions—that local shared intentions can emerge.

An important question is whether and how our findings may generalize to other forms of collective improvisations. Here, we used CFI as a paradigm for studying joint improvised action, but we should emphasize that every instance of collective improvisation is not akin to CFI, since collective improvisations greatly vary on at least three dimensions. First, improvisation comes in degrees (Nettl, 1974): Some collective improvisations are highly unplanned, and others only allow for circumscribed spontaneous decisions within a more or less loose script. In those latter cases, the role of local shared intentions may be less crucial, as coordination is then typically supported by a broad script that is common knowledge among improvisers (think of the role played by standards such as *My Funny Valentine* in jazz improvisation). Second, some collective improvisations aim at creative and unprecedented results, while others are more concerned with efficiency in spontaneously achieving a clear goal (e.g., unarming a terrorist). Again, it is likely that local shared intentions are especially important in the first case, as they can be seen as compensating the absence of a clear overarching goal. Third, collective improvisations differ in terms of the medium in which the interaction between agents takes place. Here, it is obvious that the specificities of our musical paradigm impacted the resources that our participants were able to use to communicate with each other, and more generally, the processes through which shared intentions could emerge. But importantly, it did so mainly by *depriving* them of key coordination resources, most notably verbal communication (which facilitates the spread of local shared intentions within the group and the emergence of common knowledge) and physical co-localization (which facilitate the triggering of joint attention, joint affordances, and more generally emergent coordination mechanisms). Musicians thus had to rely on resources that were both more abstract and more indeterminate. If locally shared intentions could emerge to support the improvisers' coordination in such bare-bones situations, then there is no reason to think that they would not in “richer,” more favorable contexts, in which improvisers are also engaged in highly unplanned and creative joint actions, but have in addition access to verbal communication and are co-located in the same physical environment. While the overall context in which the collective improvisation takes place, the nature of the improvisers' shared environment (sonic, audio-visual, or haptic?), the structure of the interactions (organized in turn-

takings or simultaneous?) and the modes of communication (nonverbal or verbal?) necessarily impact how improvisers coordinate, we believe that our core finding that locally shared intentions can support improvisers' coordination should extend to other kinds of complex improvised joint actions.

In particular, the ending goals we studied here are paradigmatic of the kind of local, shared intentions that are likely to emerge in complex and temporally extended joint improvisations—intentions that are abstract enough to be plausibly shared by several improvisers at a given point of the joint improvised action, while still retaining enough specificity to constrain the temporal and interactional dynamics at play. For example, the spontaneous tactics in which team players engage in collective sports such as basketball (Bourbousson, Poizat, Saury, & Seve, 2010) might be precisely analyzed in terms of the emergence of such local, shared intentions (e.g., preparing a shooting possibility for the team), beyond the primary, overarching shared goal of scoring baskets (Steiner, Macquet, & Seiler, 2017). Local shared intentions may also explain temporal coordination (i.e., the smooth switching between speaker and listener roles; Corps, Gambi, & Pickering, 2018) and content-based coordination (i.e., negotiating the actual question under discussion) during open-ended conversations (Beaver, Roberts, Simons, & Tonhauser, 2017). Because these shared intentions do not specify the details of the improvisers' contribution, they are likely to allow coagents to act with the high degree of flexibility required by the unpredictable dynamics of an improvised interaction, while maintaining a minimal level of precision in their coordination by providing them with a shared directionality (e.g., continuing or changing). In that sense, shared intentions are perhaps especially important to facilitate coordination when joint outcomes are underdetermined. To further test this idea, future work could manipulate joint outcomes' determinacy and measure the rate and level of abstraction of the shared intentions that emerge in these situations. Another important venue for future research will be to apply our design to other forms of collective improvisations (e.g., open-ended conversations), and to precisely examine how shared intentions may emerge from nonverbal (musical) interactions. Finally, our method makes it possible to ask whether coordination can occur at all when agents hold incongruent intentions simultaneously (e.g., what happens when some improvisers want to change the music while others wish to maintain the music?).

Improvisation has once been defined as the “*coordination and concatenation of actions over time by means other than planning*” (Preston, 2013, p. 63). At core, improvisation is the way we have of navigating our social lives when we cannot or do not want to engage in extensive planning. But this does not mean that improvisers are locked in an eternal present, only able to blindly interact without any foresight of what is to come next. By highlighting the role played by shared intentions in joint improvised actions, our study opens up new avenues to explore the many ways we have to engage with the future while acting jointly.

## Acknowledgments

This work was funded by ANR MICA (ANR-17-CE27-0021, to C.C.), ERC StG CREAM 335536 (to J.-J.A.), a H2020-MSCA-IF-2018 grant (JDIL-845859 to L.G.), and

partially funded by ERC SOMICS 609819, and ERC JAXPERTISE 616072 and the Central European University Foundation, Budapest (CEUBPF; partially funding T.W.). The theses explained herein represent the own ideas of the authors and do not necessarily reflect the opinion of CEUBPF. The authors thank the musicians and sound engineers for participating in and recording the music involved in study 1 and 2 at the Aeronef Studio, Paris, France. Ethical approval was obtained, and experimental data for study 3 and 4 were collected at INSEAD/ Sorbonne University Center for Behavioural Science, Paris, France.

## Authors' contributions

C.C., T.W., L.G., and J.-J.A. designed the experiment. L.G., T.W., and C.C. collected the data. L.G. and T.W. analyzed the data. L.G. and C.C. wrote the paper with comments from P.S.-G., T.W., and J.-J.A. The authors declare that there is no conflict of interest regarding the publication of this article.

## Open Research badges



This article has earned Open Data and Open Materials badges. Data and materials are available at <https://doi.org/10.17605/OSF.IO/4PNXH>.

## References

- Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*, 1109–1116. <https://doi.org/10.1038/nn.2182>
- Arthurs, T. (2016). *Secret gardeners: An ethnography of improvised music in Berlin (2012–13)*. Edinburgh: University of Edinburgh.
- Aucouturier, J. J., & Canonne, C. (2017). Musical friends and foes: The social cognition of affiliation and control in improvised interactions. *Cognition*, *161*, 94–108.
- Bailey, D. (1992). *Improvisation: Its nature and practice in music*. New York: Da Capo Press.
- Beaver, D. I., Roberts, C., Simons, M., & Tonhauser, J. (2017). Questions under discussion: Where information structure meets projective content. *Annual Review of Linguistics*, *3*, 265–284. <https://doi.org/10.1146/annurev-linguistics-011516-033952>
- Bishop, L., & Goebel, W. (2014). Context-specific effects of musical expertise on audiovisual integration. *Frontiers in Psychology*, *5*, 1123. <https://doi.org/10.3389/fpsyg.2014.01123>
- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences*, *17*, 97–110.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*, *5*(9/10), 341–345.
- Borgo, D. (2005). *Sync or swarm: Improvising music in a complex age*. London: A&C Black.
- Bourbousson, J., Poizat, G., Saury, J., & Seve, C. (2010). Team coordination in basketball: Description of the cognitive connections among teammates. *Journal of Applied Sport Psychology*, *22*, 150–166. <https://doi.org/10.1080/10413201003664657>

- Bratman, M. E. (1999). Shared intention. In E. Sosa (Ed.), *Faces of intention: Selected essays on intention and agency* (pp. 109–129). Cambridge: Cambridge University Press.
- Bratman, M. E. (2014). *Shared agency: A planning theory of acting together*. Oxford: Oxford University Press.
- Butterfill, S. A. (2013). Interacting mindreaders. *Philosophical Studies*, 165(3), 841–863. <https://doi.org/10.1007/s11098-012-9980-x>
- Butterfill, S. A. (2018). Coordinating joint action. In M. Jankovic & K. Ludwig (Eds.), *The Routledge handbook of collective intentionality* (pp. 68–82). New York: Routledge. <https://doi.org/10.4324/9781315768571-8>
- Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind and Language*, 28(5), 606–637. <https://doi.org/10.1111/mila.12036>
- Canonne, C. (2013). Focal points in collective free improvisation. *Perspectives of New Music*, 51, e40. <https://doi.org/10.7757/persnewmusi.51.1.0040>
- Canonne, C. (2018). Rehearsing free improvisation? An ethnographic study of free improvisers at work. *Music Theory Online*, 24(4). <https://doi.org/10.30535/mto.24.4.1>
- Canonne, C., & Garnier, N. (2012). Cognition and segmentation in collective free improvisation: An exploratory study. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pasiadis (Eds.), *Proceedings of the 12th international conference on music perception and cognition and 8th triennial conference of the European Society for the Cognitive Sciences of Music* (pp. 197–204). Thessaloniki: Aristotle University of Thessaloniki.
- Carré, A., Stefaniak, N., D'Ambrosio, F., Bensalah, L., & Besche-Richard, C. (2013). The basic empathy scale in adults (BES-A): Factor structure of a revised form. *Psychological Assessment*, 25(3), 679–691.
- Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. *Discourse Processes*, 55(2), 230–240. <https://doi.org/10.1080/0163853X.2017.1330031>.
- D'Ausilio, A., Novembre, G., Fadiga, L., & Keller, P. E. (2015). What can music tell us about social interaction? *Trends in Cognitive Sciences*, 19(3), 111–114. <https://doi.org/10.1016/j.tics.2015.01.005>
- D'Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., Aloimonos, Y., & Fadiga, L. (2012). Leadership in orchestra emerges from the causal relationships of movement kinematics. *PLoS One*, 7(5), e35757. <https://doi.org/10.1371/journal.pone.0035757>
- De Freitas, J., Thomas, K., DeScioli, P., & Pinker, S. (2019). Common knowledge, coordination, and strategic mentalizing in human social life. *Proceedings of the National Academy of Sciences of the United States of America*, 116(28), 13751–13758. <https://doi.org/10.1073/pnas.1905518116>
- De Spain, K. (2014). *Landscape of the now: A topography of movement improvisation*. Oxford: Oxford University Press.
- della Gatta, F., Garbarini, F., Rabuffetti, M., Viganò, L., Butterfill, S. A., & Sinigaglia, C. (2017). Drawn together: When motor representations ground joint actions. *Cognition*, 165, 53–60. <https://doi.org/10.1016/J.COGNITION.2017.04.008>
- Denzler, B., & Guionnet, J.-L. (2020). *The practice of musical improvisation: Dialogues with contemporary musical improvisers*. New York: Bloomsbury Academic.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models. Policy analysis*. Cambridge: Cambridge University Press.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 4054). New York: Wiley.
- Gueguen, N., Jacob, C., & Martin, A. (2009). Mimicry in social interaction: Its effect on human judgment and behavior. *European Journal of Social Sciences*, 8(2), 253–259.
- Hadley, L. V., Sturt, P., Moran, N., & Pickering, M. J. (2018). Determining the end of a musical turn: Effects of tonal cues. *Acta Psychologica*, 182, 89–193. <https://doi.org/10.1016/j.actpsy.2017.11.001>

- 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*(1). <https://doi.org/10.1038/s41598-019-47471-3>
- Helm, J. L., Miller, J. G., Kahle, S., Troxel, N. R., & Hastings, P. D. (2018). On measuring and modeling physiological synchrony in dyads. *Multivariate Behavioral Research*, *53*(4), 521–543. <https://doi.org/10.1080/00273171.2018.1459292>
- Ingold, T., & Hallam, E. (2007). Creativity and cultural improvisation: An introduction. In E. Hallam & T. Ingold (Eds.), *Creativity and cultural improvisation* (pp. 1–24). New York: Berg. <https://doi.org/10.1017/S1537781415000316>
- Issartel, J., Marin, L., & Cadopi, M. (2007). Unintended interpersonal co-ordination: “Can we march to the beat of our own drum?” *Neuroscience Letters*, *411*(3), 174–179. <https://doi.org/10.1016/J.NEULET.2006.09.086>
- Keller, P. E. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Emerging communication: Studies on new technologies and practices in communication: Vol. 10. Enacting intersubjectivity: A cognitive and social perspective on the study of interactions* (pp. 205–211). Amsterdam: IOS Press.
- Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 260–282). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199659647.001.0001>
- Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-year-old children. *Evolution and Human Behavior*, *31*(5), 354–364. <https://doi.org/10.1016/j.evolhumbehav.2010.04.004>
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. *Psychology of Learning and Motivation—Advances in Research and Theory*, *54*, 59–101. <https://doi.org/10.1016/B978-0-12-385527-5.00003-6>
- Kourtis, D., Woźniak, M., Sebanz, N., & Knoblich, G. (2019). Evidence for we-representations during joint action planning. *Neuropsychologia*, *131*, 73–83. <https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2019.05.029>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, H. B. (2014). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R package 2.0-11.
- Linson, A., & Clarke, E. F. (2018). Distributed cognition, ecological theory, and group improvisation. In E. Clarke & M. Doffman (Eds.), *Distributed creativity: Collaboration and improvisation in contemporary music* (pp. 52–69). Oxford, UK: Oxford University Press.
- Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of Cognitive Neuroscience*, *25*(7), 1049–1061. [https://doi.org/10.1162/jocn\\_a\\_00388](https://doi.org/10.1162/jocn_a_00388)
- MacDonald, R. A. R., & Wilson, G. B. (2020). *The art of becoming: How group improvisation works*. Oxford: Oxford University Press.
- Mendonça, D. J., & Wallace, W. A. (2007). A cognitive model of improvisation in emergency management. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, *37*(4): 547–561. <https://doi.org/10.1109/TSMCA.2007.897581>
- Michael, J. (2011). Interactionism and mindreading. *Review of Philosophy and Psychology*, *2*(3), 559. <https://doi.org/10.1007/s13164-011-0066-z>
- Michael, J. (2017). Music performance as joint action.pdf. In M. Lesaffre, P.-J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 160–166). New York: Routledge.
- Moran, N., Hadley, L. V., Bader, M., & Keller, P. E. (2015). Perception of “back-channeling” nonverbal feedback in musical duo improvisation. *PLoS One*, *10*(6), e0130070. <https://doi.org/10.1371/journal.pone.0130070>

- Nessler, J. A., & Gilliland, S. J. (2009). Interpersonal synchronization during side by side treadmill walking is influenced by leg length differential and altered sensory feedback. *Human Movement Science*, 28(6), 772–785. <https://doi.org/10.1016/J.HUMOV.2009.04.007>
- Nettl, B. (1974). Thoughts on improvisation: A comparative approach. *The Musical Quarterly*, 60(1), 1–19.
- Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2014). Motor simulation and the coordination of self and other in real-time joint action. *Social Cognitive and Affective Neuroscience*, 9(8), 1062–1068. <https://doi.org/10.1093/scan/nst086>
- Noy, L., Dekel, E., & Alon, U. (2011). The mirror game as a paradigm for studying the dynamics of two people improvising motion together. *Proceedings of the National Academy of Sciences of the United States of America*, 108(52), 20947–20952. <https://doi.org/10.1073/pnas.1108155108>
- Pachet, F., Roy, P., & Foulon, R. (2017). Do jazz improvisers really interact?. In M Lesaffre, P. J. Maes, & M. Leman (Eds.), *The Routledge companion to embodied music interaction* (pp. 167–176). New York: Routledge. <https://doi.org/10.4324/9781315621364-19>
- Papiotis, P., Marchini, M., & Maestre, E. (2012). Computational analysis of solo versus ensemble performance in string quartets: Intonation and dynamics. *Proceedings of the 12th international conference on music perception and cognition*, Thessaloniki, Greece.
- Pelz-Sherman, M. (1998). *A framework for the analysis of performer interactions in western improvised contemporary art music*. San Diego: University of California.
- Pressing, J. (1984). Cognitive processes in improvisation. *Advances in Psychology*, 19, 345–363. [https://doi.org/10.1016/S0166-4115\(08\)62358-4](https://doi.org/10.1016/S0166-4115(08)62358-4)
- Preston, B. (2013). *A philosophy of material culture: Action, function, and mind*. New York: Routledge. <https://doi.org/10.4324/9780203069844>
- Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature. *Psychonomic Bulletin and Review*, 12(6), 969–992. <https://doi.org/10.3758/BF03206433>
- Sacheli, L. M., Arcangeli, E., & Paulesu, E. (2018). Evidence for a dyadic motor plan in joint action. *Scientific Reports*, 8, 5027. <https://doi.org/10.1038/s41598-018-23275-9>
- Savouret, A. (2010). Introduction à un solfège de l'audible. L'improvisation libre comme outil pratique. *Symétrie*.
- Sawyer, R. K. (2003). Group creativity: Music. In *Theater, collaboration*. Mahwah: Erlbaum.
- Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In A. Fuchs, & V. K. Jirsa (Eds.), *Coordination: Neural, behavioural and social dynamics* (pp. 281–308). Berlin: Springer, Springer Nature. Coordination: Neural, behavioral and social dynamics. [https://doi.org/10.1007/978-3-540-74479-5\\_14](https://doi.org/10.1007/978-3-540-74479-5_14)
- Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. S. (2009). Signalling signalhood and the emergence of communication. *Cognition*, 113(2), 226–233. <https://doi.org/10.1016/j.cognition.2009.08.009>
- Sethares, W. A. (1993). Local consonance and the relationship between timbre and scale. *Journal of the Acoustical Society of America*, 94, 1218. <https://doi.org/10.1121/1.408175>
- Steiner, S., Macquet, A.-C., & Seiler, R. (2017). An integrative perspective on interpersonal coordination in interactive team sports. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.01440>
- Van Baaren, R., Janssen, L., Chartrand, T. L., & Dijksterhuis, A. (2009). Where is the love? The social aspects of mimicry. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 364(1528), 2381–2389. <https://doi.org/10.1098/rstb.2009.0057>
- Vassilakis, P. N. (2001). *Perceptual and physical properties of amplitude fluctuation and their musical significance*. PhD thesis. University of California.
- Vesper, C., Abramova, E., Bütepage, J., Ciardo, F., Crossey, B., Effenberg, A., Hristova, D., Karlinsky, A., McEllin, L., Nijssen, S. R. R., Schmitz, L., & Wahn, B. (2017). Joint action: Mental representations, shared information and general mechanisms for coordinating with others. *Frontiers in Psychology*, 07, 2039. <https://doi.org/10.3389/fpsyg.2016.02039>
- Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint action. *Neural Networks*, 23(8–9), 998–1003. <https://doi.org/10.1016/J.NEUNET.2010.06.002>

- Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2011). Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research*, 211(3–4), 517–530. <https://doi.org/10.1007/s00221-011-2706-z>
- Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2013). Are you ready to jump? Predictive mechanisms in interpersonal coordination. *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), 48–61. <https://doi.org/10.1037/a0028066>
- Walsh, M., Roberts, I., & Besser, M. (2013). *Upright citizens brigade comedy improvisation manual*. Comedy Council of Nicea LLC.
- Walton, A. E., Washburn, A., Langland-Hassan, P., Chemero, A., Kloos, H., & Richardson, M. J. (2018). Creating time: Social collaboration in music improvisation. *Topics in Cognitive Science*, 10(1), 95–119. <https://doi.org/10.1111/tops.12306>
- Wass, S. V., Whitehorn, M., Marriott Haresign, I., Phillips, E., & Leong, V. (2020). Interpersonal neural entrainment during early social interaction. *Trends in Cognitive Sciences*, 24(4), 329–342. <https://doi.org/10.1016/j.tics.2020.01.006>
- Yun, K., Watanabe, K., & Shimojo, S. (2012). Interpersonal body and neural synchronization as a marker of implicit social interaction. *Scientific Reports*, 2, 959. <https://doi.org/10.1038/srep00959>

### Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article:

**Appendix S1:** Materials and results.