



HAL
open science

Emergent shared intentions support coordination during collective musical improvisations

Louise Goupil, Thomas Wolf, Pierre Saint Germier, Jean-Julien Aucouturier, Clément Canonne

► To cite this version:

Louise Goupil, Thomas Wolf, Pierre Saint Germier, Jean-Julien Aucouturier, Clément Canonne. Emergent shared intentions support coordination during collective musical improvisations. *Cognitive Science*, 2021, 45, 10.1111/cogs.12932 . hal-03088074

HAL Id: hal-03088074

<https://hal.science/hal-03088074>

Submitted on 10 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

Emergent shared intentions support coordination during collective musical improvisations

Louise Goupil^{1,3*}, Thomas Wolf², Pierre Saint-Germier¹, Jean-Julien Aucouturier¹ and Clément
Canonne¹

1- Science and Technology of Music and Sound (UMR 9912, IRCAM/CNRS/Sorbonne
University), Paris, France

2- Department of Cognitive Science, Central European University, Budapest, Hungary

3- School of Psychology, University of East London, London, UK

* corresponding author: lougoupil@gmail.com

Louise Goupil - School of Psychology, University of East London
Stratford Campus – Water Lane, London E15 4LZ

keywords: improvisation; musical performance; coordination; joint action; goal
representations; shared intentions

24 **Abstract**

25

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

42

43 1 Introduction

44

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

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

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

81 On the other hand, most research on improvised joint actions so far has focused on
82 examining embodied and embedded aspects, and describing coordination mechanisms that are
83 thought to operate on short time-scales, and to directly arise from dynamic interactions between
84 partners within a shared environment. Unlike planned coordination, this type of mechanisms
85 does not require that agents hold specific mental representations at the individual level, but
86 primarily rely on agents’ dynamic couplings while acting jointly. Here, following previous
87 authors (Butterfill, 2018; Knoblich et al., 2011), we refer to these processes as *emergent*
88 *coordination* mechanisms. One classic example is the phenomenon of entrainment observed
89 when two agents become more synchronized with one another than expected by chance simply
90 through seeing each other’s movements, and even in the absence of, or contrary to, any intention
91 to do so (Issartel, Marin, & Cadopi, 2007; Nessler & Gilliland, 2009; Repp, 2005; Yun,
92 Watanabe, & Shimojo, 2012). Entrainment is often interpreted in the framework of dynamical

93 systems, where it is argued to merely constitute a particular instance of physical coupling that
94 can arise in all (social or non-social) kinds of coupled oscillators (Schmidt & Richardson, 2008;
95 Walton et al., 2018). Other studies have documented the role of mimicry, or automatic
96 imitation, showing that individuals often mirror each other's actions, and that such mirroring
97 fosters coordination and acts as social glue by increasing affiliation between individuals
98 (Gueguen, Jacob, & Martin, 2009; Van Baaren, Janssen, Chartrand, & Dijksterhuis, 2009). For
99 instance, one study showed that expert improvisers could smoothly imitate each other's
100 movements while performing a mirror-game task, entering into a state of co-confidence in
101 which each player seems to be both leading and following at the same time (Noy, Dekel, &
102 Alon, 2011). Beyond mirroring, there is some evidence that motor simulation enables observers
103 to predict their partners' actions, which can help them adjust their actions accordingly to
104 improve coordination (Aglioti, Cesari, Romani, & Urgesi, 2008; Novembre, Ticini, Schütz-
105 Bosbach, & Keller, 2014; Noy et al., 2011; Vesper, van der Wel, Knoblich, & Sebanz, 2013).
106 Finally, other research has focused on documenting joint affordances, showing for instance that
107 particularly salient elements present within their environment can constrain improvisers'
108 behavior, leading them to perform actions with a similar functional profile (e.g., changing what
109 they were doing) during the course of the performance (Canonne & Garnier, 2012).

110 Overall, it seems clear that coordination during collective improvisations heavily relies
111 on the fact that partners' interactions are both embodied and embedded (Linson & Clarke,
112 2018). Yet, whether emergent mechanisms are sufficient to support coordination in cases of
113 complex and/or temporally-extended collective improvisations, without the support of
114 additional (planned) mechanisms, at least punctually, remains far from certain. Indeed, the
115 studies reviewed above document the role of emergent coordination mechanisms in supporting
116 very simple forms of joint action, that typically involve agents who perform very similar actions
117 at the same time (e.g., tap in synchrony to the same beat, imitate each other's motion or

118 emotional displays, etc.). A large literature has documented the pervasiveness of these
119 mechanisms, at the behavioral, physiological and neural levels, and the role they play in
120 coordination from infancy to adulthood (Helm, Miller, Kahle, Troxel, & Hastings, 2018; Wass,
121 Whitehorn, Marriott Haresign, Phillips, & Leong, 2020). Yet, how they could account for
122 complex forms of collective improvisations, where each agent has to perform a different type
123 of action, and where no temporal structure is present to support mechanisms such as
124 entrainment, is really unclear. Moreover, these mechanisms operate on short time scales
125 (seconds, at best minutes), and are specifically efficient when precision is targeted, while they
126 fall short at explaining how the coordination of complex and flexible behaviors – typical of
127 most creative improvisations – may be achieved (Butterfill, 2018).

128 Research on *scripted* joint actions generally suggests that both emergent coordination
129 mechanisms and planned coordination mechanisms actually interact to foster coordination, their
130 relative contributions enabling an optimal trade-off between precision and flexibility (Butterfill,
131 2018). For instance, the fine-tuning of musical expressivity in performing chamber music
132 compositions crucially depends on emergent mechanisms, which regulate the temporal
133 unfolding of performers on very short time scales (D’Ausilio et al., 2012; Keller, 2014). Studies
134 also suggest that when co-agents have a shared intention to synchronize, internal sensorimotor
135 models enable them to predict each other’s timing, and to deploy strategies to improve
136 synchrony (Heggli, Konvalinka, Kringelbach, & Vuust, 2019; Vesper, van der Wel, Knoblich,
137 & Sebanz, 2011).

138 Building upon these studies targeting scripted interactions, here we ask whether such a
139 synergy of planned and emergent coordination mechanisms is also at play during improvised
140 joint actions. More precisely, we test the hypothesis that co-improvisers also coordinate by
141 forming shared intentions that emerge during the course of the interaction. We hypothesize that
142 shared intentions may be particularly crucial to support the most complex and flexible forms of

143 collective improvisations, that require co-agents to perform dissimilar and varied actions that
144 are not necessarily tied to an underlying temporal structure. We thus conducted four
145 experiments using the practice of Collective Free Musical Improvisation (CFI) as an
146 experimental model of improvised joint action.

147 CFI constitutes a particularly pure and paradigmatic case of collective improvisation
148 (Bailey, 1992) that is ideal to test our hypotheses for several reasons. First, in CFI, musicians
149 typically do not attribute roles to each other, do not specify melodic or harmonic structures
150 before improvising together, and overall, refuse to specify how the improvisation will unfold:
151 they refuse to precisely specify their joint outcome, and to establish a joint plan beforehand
152 (Pressing, 1984). On a finer level, CFI also crucially differs from more familiar genres of
153 improvised music such as bebop or even free jazz in the sense that it is generally not pulsed and
154 devoid of rhythmical patterns. Free improvisers certainly share a common ground, which
155 imposes non-trivial aesthetical constraints on the group's performances (e.g., leading musicians
156 to focus on subtle timbral explorations and to avoid conventional rhythmical patterns or chord
157 progressions). However, the issue of how to *temporally* organize the individual and collective
158 musical behaviors on shorter and longer time scales in a given performance remains in its
159 entirety (Canonne, 2018), making CFI an as pure as possible case of real-life improvised joint
160 action (see video and audio examples via this [link](#)). Second, CFI typically involves a
161 temporally-extended situation in which each agent performs highly idiosyncratic, non-imitative
162 actions. This is in sharp contrast with shorter, simpler, and imitation-based forms of improvised
163 interactions used in previous research (Noy et al., 2011), and makes CFI especially appropriate
164 to track the existence and impact of shared intentions in joint improvised actions. Finally, like
165 other forms of collective music-making that have been used as a model to investigate joint
166 actions (Aucouturier & Canonne, 2017; D'Ausilio, Novembre, Fadiga, & Keller, 2015;
167 Kirschner & Tomasello, 2010; Michael, 2017), CFI constitutes a model that is ecologically

168 valid, and allows to measure coordination on multiple levels and to investigate the mechanisms
169 that drive the emergence of shared intentions on the fly, in the absence of verbal
170 communication.

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

177 How and when to end a performance is a coordination problem that is particularly
178 challenging in CFI because musicians do not share a given script nor a repertoire of canonical
179 endings that provide them with clear potential ending points. Even if musicians were to decide
180 to end the piece at the same time, it would still be difficult to do so. Contrary to other musical
181 genres, such as straight-ahead jazz, in which temporal and harmonic structures typically
182 determine specific ending points (e.g., on the beat, or on a closing cadence), provide musicians
183 with the support of a shared entrainment to a beat, or at the least, enable performers to rely on
184 auditory imagery to form precise predictions about what is about to come next (Hadley, Sturt,
185 Moran, & Pickering, 2018; Keller, 2008), in CFI there are no definite structures nor
186 conventional patterns that point to specific ending points. As Alain Savouret – who taught free
187 improvisation at Paris Conservatory for many years – nicely puts it: “*If it’s always difficult to*
188 *start [an improvisation], it’s even harder to finish it*” (Savouret, 2010, p. 26). As such, issues
189 of endings are often raised and discussed within CFI classes. At the same time, endings are also
190 moments in which the improvisers’ coordination (or lack thereof) is at its clearest: musicians
191 (and attuned audience members alike) often speak of “missed endings” when the group
192 members did not “feel” at the same time that the performance was coming to an end or that

193 such or such musical event could act as a good ending point. For these two reasons, endings
194 perfectly encapsulate the coordination problems that are at stake during improvised joint
195 actions. In this regard, they constitute a particularly interesting case to study the role of shared
196 intentions in supporting coordination when multiple agents act in flexible ways. Shared
197 intentions could indeed foster coordination in this context because they would allow
198 improvisers to anticipate that the performance is about to finish, and to plan their actions with
199 respect to this proximate joint outcome, on the basis that their partners are likely to do the same.

200 Thus, in Experiments 1 and 2, we invited trios of musicians to a recording studio, where
201 they were asked to perform a series of short improvisations. In Experiment 1, musicians had to
202 perform four improvisations and, while playing, each musician was asked to press a pedal “*as*
203 *soon as she felt that she was looking for an ending*”. As musicians were playing in separate
204 studio booths, pedal presses were made covertly, with no auditory consequence allowing other
205 musicians to perceive when their partners pressed the pedal. By testing whether musicians’
206 reports are closer in time to one other than would be predicted by chance, Experiment 1 allowed
207 us to investigate whether shared intentions do emerge during collective improvisations.

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

217

<i>Hypotheses</i>	<i>Predictions</i>	
	<i>Temporal, acoustic and qualitative aspects of musical coordination...</i>	<i>Signaling strategies are...</i>
<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

218 *Table 1. Predictions of the three main hypotheses with respect to the two main aspects examined in this study: 1) coordination,*
219 *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.*
220 *(qualitative aspects of coordination), and 2) signaling strategies (results reported in section 5.2.3.).*

221

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

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

240 Coordination was examined on three levels: 1) by assessing the *temporal coordination*
241 with which musicians stopped playing at the end of the piece; 2) by assessing the musicians’
242 *dynamic, timbral and harmonic coordination* with several acoustical measures and 3) by
243 assessing *qualitative aspects of musical coordination*. Point 3) was achieved by running a
244 follow-up listening experiment (Experiment 3) where a separate group of expert and naive
245 listeners were asked to evaluate the recorded improvisations, in order to assess whether shared
246 intentions impact the aesthetic perception of the joint performance, and some of its qualitative
247 properties corresponding to higher-level aspects of musical coordination that are difficult to
248 capture with acoustic analysis, given the sheer sonic complexity of most CFI performances.
249 Lastly, contrary to the other two hypotheses, the *shared intention* hypothesis also predicts that
250 prompted musicians may engage in signaling strategies to make their intention *manifest* for the
251 group, thereby establishing common knowledge that the piece is about to end, and ensuring the
252 collaboration and commitment of the other performers. Thus, in a fourth experiment with the
253 same listeners involved in the third experiment, we investigated how goals may propagate
254 within the group of improvisers to foster coordination. We examined the possibility that
255 musicians deploy *signaling strategies* to establish common knowledge of their current goal at
256 the level of the group, thereby forming proper shared intentions. To this end, listeners were
257 asked to detect whether they thought individual performers were looking for an end, and to

258 characterize their behavior along several categories. This allowed us to examine whether
259 musicians' intentions to end the piece could be deciphered by listeners, what type of
260 communicative behaviors drive this perception, and how the transparency of performers'
261 intentions relates to coordination.

262

263

264 **2 Experiment 1: Can shared intentions emerge during collective musical** 265 **improvisations?**

266

267 **2.1 Experiment 1 – Methods**

268 *2.1.1 Participants*

269 We invited 21 participants (2 women, age $M = 39.8$ years, $SD = 9.1$ years) to take part
270 in experiments 1 and 2. All were highly-skilled professional musicians actively involved in CFI
271 (average years of experience on their respective instruments $M = 29.2$ years, $SD = 8.3$ years,
272 and number of years of performing CFI $M = 17.3$ years, $SD = 6.8$ years). Participants were
273 grouped into 12 trios, such that no combination of musicians would repeat (see Table S1 for the
274 musical instruments played in each trio). Fifteen of the 21 musicians participated in two
275 different trios. We also tried to minimize the familiarity between musicians, which ensures
276 maximal conditions of free improvisation, and limited the common ground structuring
277 musicians' interactions. We asked musicians to report how much they knew each of the two
278 other musicians on a scale from 1 (not familiar at all) to 7 (very familiar), and how much they
279 enjoyed playing with this trio (1: not at all; 7: very much). Familiarity averaged over the 12
280 trios was $M = 2.6$, $SD = 0.91$, confirming low familiarity overall. Appreciation averaged over
281 the 12 trios was $M = 5.7$, $SD = 1$, suggesting that our procedure was not too invasive and allowed
282 musicians to play together in an ecological fashion. We assessed participants' general empathic

283 traits by using the self-report Basic Empathic Scale in Adults (BESA, Carré, Stefaniak,
284 D’Ambrosio, Bensalah, & Besche-Richard, 2013). Nineteen participants filled in the
285 questionnaire, and 2 musicians refused to do so (including one of the musicians who played
286 twice, leading to 3 missing values). Musicians signed an informed consent and were paid for
287 their contribution.

288

289 **2.1.2 Procedure and Design**

290 The aim of Experiment 1 was to assess whether shared goals spontaneously emerge
291 during improvised joint actions, modeled here with CFI. To this end, we asked each of the 12
292 trios of expert improvisers to perform 4 improvisations of approximately 3-4 minutes (180 –
293 240 seconds). Providing this range was necessary to enable efficient data collection, but the
294 instructions emphasized the fact that this time limit was meant to provide a loose guideline
295 rather than to set a strict boundary. Consistent with these instructions, the durations of the
296 improvisations were widely spread around the recommended time range, effectively extending
297 from 92.8 to 391.3 seconds ($M = 202.8$ seconds, $SD = 52.5$). It should also be noted that agreeing
298 on an approximate duration before the beginning of the improvisation is common practice in
299 this community. For example, trumpet player Axel Dörner states that “*In [one of my trios], we*
300 *say beforehand how long we want to play for. For me, that’s important. When we play a concert,*
301 *we decide how long the concert is going to last and how the concert might be divided into*
302 *pieces. Sometimes we define it closely – longer pieces, shorter pieces or endings. We decide*
303 *together*” (quoted in Denzler & Guionnet, 2020, p. 72). More generally, performing pieces of
304 3-4 minutes is not unheard of for these improvisers, as it corresponds to the typical duration of
305 the “constrained improvisations” they sometimes perform during their working sessions
306 (Canonne, 2018).

307 Musicians were placed in separate studio booths so that they could not see each other,
308 and only heard each other through headphones, as is standard in studio recording practices.
309 Each musician was asked to press a midi pedal (M-Audio SP-2) “*as soon as she felt that she*
310 *was looking for an end to the piece*”. Thus, our focus was on *collective intentions* (i.e., intentions
311 to end the piece that include the group in their contents): for the piece to end, *all* improvisers
312 must stop playing. By testing whether such collective intentions emerge closer to each other
313 than would be expected by chance, we test whether they were shared amongst partners,
314 amounting to *shared intentions*. After each improvisation, musicians were asked to rate on a 7-
315 point Likert scale the extent to which they enjoyed the improvisation, and how much they liked
316 the ending. These ratings suggested that they were not disturbed by having to press the pedal
317 (see section S.1.4 in the supplementary material). They were also asked whether or not they
318 thought that their partners had been looking for an end, and if so why. This experiment was pre-
319 registered at <https://aspredicted.org/k2jf5.pdf>. We note when our analyses departed from the
320 pre-registration. The corpus, data and analysis scripts are available on the Open Science
321 Framework via [this link](#).

322

323 **2.1.3 Data Analysis**

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

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

345

346

347 **2.2 Experiment 1 – Results**

348

349 ***2.2.1 Ending goals emerge in musical improvised interactions, and they are temporally*** 350 ***coordinated***

351 The mean *Temporal Coordination of Endings* for real pairs was $M = 7.74$ seconds, $SD = 4.07$.

352 This was significantly better than the *Temporal Coordination of Endings* calculated for fake

353 pairings ($M = 45.60$ seconds, $SD = 23.88$), $t(11) = 5.152$, $p < .001$, $d = 2.210$. Performances’

354 endings were thus not the mere result of the individual musicians randomly stopping at some

355 point. On the contrary, despite the highly unscripted nature of CFI and the general absence of

356 a shared pulse, it seems that the improvisers were still aiming to achieve some degree of

357 temporal coordination when ending the piece, although it should be noted that 7 seconds is well
358 above the duration that would be expected in a typical, scripted musical performance.
359 The number of Pedal Pressings was 2 or higher in 25 out of the 48 improvisations (see Fig. 1A).
360 The mean *Pedal Pressing Temporal Coordination* was $M = 28.38$ seconds, $SD = 19.97$. To test
361 whether this duration is smaller than what would be expected by chance, we compared it to the
362 temporal coordination of fake pairings ($M = 47.10s$, $SD = 23.51s$). Consistent with our
363 prediction, a paired-sample t-test revealed a significant difference, $t(11) = 2.643$, $p = .025$, $d =$
364 $.797$, with the real *Pedal Pressing Temporal Coordination* being significantly lower than the
365 one for fake pairings. Thus, when two or more musicians pressed their pedals during the
366 performance, those pedal presses were closer in time than would be expected by chance.
367 Additionally, despite the inevitable latency introduced by the experimental setting, pedal
368 pressings were less than 10s apart in 24.3 % of trials (see Fig. 1B), which suggests that, in those
369 cases at least, two or more improvisers were intending to end during the same short time span.
370 Our data reveal that collective intentions can emerge at the same time, and thus be shared by
371 several musicians during improvised interactions.

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

381

382 2.2.2 *Impact of shared intentions on improvised musical coordination*

383 The average *Temporal Coordination of Endings* was $M = 27.38s$ ($SD = 20.57s$) and the
384 average Ending Appreciation was $M = 4.74$ ($SD = 0.99$). Contrary to our predictions, there was
385 no correlation among trials between *Pedal Pressing Temporal Coordination* and the *Temporal*
386 *Coordination of Endings* (Spearman $rs(23) = 3286$, $p = .20$), and no correlation between *Pedal*
387 *Pressing Temporal Coordination* and *Appreciation of Ending* ($rs(23) = 2621.2$, $p = .97$), which
388 we take as a proxy to higher-level aspects of coordination. Thus, there was no evidence that the
389 emergence of shared intentions positively impacted musicians' coordination here. It is worth
390 noting that debriefings with participants revealed that, in some cases, improvisers had forgotten
391 to press their pedal even though they had been actively looking for an end. In the second
392 experiment, which offered a more controlled environment, we investigate the impact of shared
393 intentions on improvised coordination more directly.

394

395 ----- *Insert Fig.1 about here* -----

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

400

401 **3 Experiment 2: Can shared intentions improve coordination during**

402 **collective musical improvisations?**

403

404 Experiment 1 demonstrates that shared intentions to end the joint action can emerge in the
405 course of improvised interactions, even in the absence of verbal communication. In Experiment
406 2, we ask whether these shared intentions actually *impact* coordination. To this end, we
407 experimentally manipulated musicians' intentions: we gave them covert instructions regarding

408 how and when they should start looking for an end to the piece, and measured whether and how
409 these instructions impacted coordination at the level of the group. More precisely, we
410 manipulated both the degree of *shared information* (i.e., the number of musicians receiving
411 instructions) and the *content* of the intention (i.e., whether musicians were supposed to look for
412 an end individually, or collectively). This allowed us to discriminate between the three
413 hypotheses outlined in the introduction, namely, the hypothesis according to which *shared*
414 *information* is crucial to foster coordination, the hypothesis according to which *collective*
415 *intentions* are crucial, and finally, the most demanding hypothesis according to which *shared*
416 *intentions* are crucial.

417

418 **3.1 Experiment 2 – Methods**

419 **3.1.1 Participants and procedure**

420 After completing 4 improvisations for Experiment 1, each of the 12 trios took a short
421 break, before performing 12 additional improvisations for Experiment 2, resulting in a total of
422 144 improvisations. During these 12 additional improvisations, musicians sometimes received
423 covert auditory prompts approximately 2:30 minutes after the beginning of the improvisation
424 (see below for the sampling procedure). Prompts were of two types: upon hearing the keyword
425 “ME”, a musician was asked to “*find a good way for you to stop playing, thus looking for an*
426 *ending for yourself*” (ME-Goal); upon hearing the keyword “WE”, the musicians were asked
427 to “*find a good way for the group to stop playing, thus looking for an ending for the group*”
428 (WE-Goal). Thus, we varied whether musicians had a goal whose content involved the group
429 as a whole (WE-Goal) or only themselves (ME-Goal). In addition, we varied the degree of
430 dissemination of these goals within the group, by prompting either 1, 2 or all 3 musicians. For
431 each improvisation, only one type of prompt could be delivered (i.e., all prompted musicians
432 either received a WE or ME-Goal). Experimental conditions could vary over the 3 Prompt

433 Types (ME-Goal or WE-Goal, or NO-Prompt) and 3 Prompt Numbers (1, 2, 3), resulting in 6
434 experimental conditions at the level of the trio (1 musician with a ME-Goal / 2 non-prompted
435 musicians; 2 musicians with ME-Goals / 1 non-prompted musician; 3 musicians with ME-
436 Goals; 1 musician with a WE-Goal / 2 non-prompted musicians; 2 musicians with a WE-Goal
437 / 1 non-prompted musician; 3 musicians with a We-Goal). Prompt times were semi-randomly
438 sampled from two uniform distributions, one ranging from 2:15 to 2:30 (early prompt) and one
439 ranging from 2:30 to 2:45 minutes (late prompt). Each of the nine conditions had one trial with
440 a time point from the first range and one trial with a time point from the second range. This
441 procedure ensured that the timings of the prompts were not too predictable.

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

448 Auditory prompts were delivered covertly through musicians’ headphones. This
449 solution was preferred over visual prompts because of two practical reasons: 1) musicians need
450 to wear headphones to hear each other in the studio anyways, and 2) many of them close their
451 eyes when they play, and mostly focus on sounds during the performance. Using auditory
452 prompts thus minimized the risk that musicians would miss the prompts (e.g., due to closed
453 eyes). Despite these precautions, questionnaires revealed that musicians missed or misheard
454 prompt types on a few occasions ($N = 32$, 7.4% of the trials). We excluded 8 trials in which two
455 musicians or more made such mistakes, and re-coded the other trials to account for what the
456 musician actually perceived. This procedure left a total of 136 improvisations in the dataset. In
457 addition, because of a technical error, the first of the 12 trio only received “ME” prompts.

458 This experiment was pre-registered at <https://aspredicted.org/k2jf5.pdf>. We note when
459 our analyses departed from the pre-registration. Data and analysis scripts are available via [this](#)
460 [link](#).

461

462 **3.1.2 Data analysis**

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

474

475 **3.1.3 Acoustic analysis**

476 To investigate whether receiving prompts changed the relationships between the
477 musicians, we conducted an acoustic analysis of musical snippets extracted before and after the
478 prompts. Following previous studies (Pachet, Roy, & Foulon, 2017; Papiotis, Marchini, &
479 Maestre, 2012), we approximated coordination by computing a linear (Pearson correlation) as
480 well as a non-linear (mutual information) index of dependency for five acoustic features: pitch,
481 volume (RMS), playing time ratio (% of sound), spectral centroid and Harmonic-to-noise ratio
482 (see below). For each of the five acoustic features and two metrics, we computed values for

483 each pair of musicians, improvisation and timing (before or after the prompt) before averaging
484 the values within the trio for each improvisation and timing. We also estimated the consonance
485 of the music produced at the level of the trio as a measure of harmonic coordination.

486 For each improvisation and individual musician, pitch, loudness, playing time, spectral
487 centroid and HNR were estimated in non-overlapping successive time frames of 200ms in two
488 time windows: 1) in a window starting 1 minute before the prompt and ending before the prompt
489 and 2) in a window starting at the prompt and extending until the end of the improvisation (M
490 $= 54.8$ seconds, $SD = 70.7$). Pitch was extracted using the *Praat* software (Boersma, 2001).
491 Loudness was approximated as the root-mean square of the amplitude of the sound. Playing
492 time ratio was defined as the ratio of the time spent playing over the total duration of the extract.
493 The harmonic-to-noise ratio was computed following the algorithm described in (Boersma,
494 1993). Finally, dissonance/roughness was estimated based on the algorithm described in
495 (Vassilakis, 2001), and implemented in the *dissonant* package in *Python*. This method, which
496 is based on a classic model by Sethares (1993), estimates the dissonance/roughness of a sound
497 from the amount of competition between partials (see <https://pypi.org/project/dissonant> for a
498 full detail of this method and formulas). Dissonance is a complex percept that is difficult to
499 capture algorithmically, but listening to a subset of our corpus and comparing values of
500 dissonance obtained by this method confirms that it captures dissonance and/or roughness
501 reliably in our dataset (follow [this link](#) for sound examples). Takes in which at least 10% of
502 each acoustic feature could be reliably extracted were included in the analysis (this low rate
503 was chosen to allow for the fact that CFI often involves musical textures that do not contain
504 harmonic signal). Pitch, centroid, HNR and dissonance were only computed in the windows
505 in which the RMS value was above a certain threshold (-60), chosen to discriminate between
506 background noise and sound in these recording conditions. To assess changes with respect to
507 the prompt, these values were normalized for each musician and take.

508

509 **3.1.4 Statistical analysis**

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

518

519

520 **3.2 Experiment 2 – Results**

521

522 **3.2.1. Impact of the number and type of prompts on temporal coordination**

523 To assess the effect of Prompt Number and Prompt Type on temporal coordination, we
524 ran a linear mixed regression with the *Temporal Coordination of Endings* as a dependent
525 variable, Prompt Number and Prompt Type as independent variables, and Trio as a random
526 factor (see Fig. 2A). This analysis revealed a main effect of Prompt Number ($\chi^2 = 9.61$; $p =$
527 $.008$), a main effect of Prompt Type ($\chi^2 = 10.8$; $p = .001$), and a significant interaction between
528 the two factors ($\chi^2 = 8.93$; $p = .011$). As predicted by the *shared information* and the *shared*
529 *intention* hypotheses, temporal coordination improved as the number of prompts increased: it
530 was better when there were three prompts ($M = 4.5$ seconds, $SD = 2.65$) as compared to when
531 there was only one ($M = 8.6$ seconds, $SD = 4.46$, $beta = -3.6$, $sem = 1.27$, $df = 12$, $t = -2.84$, $p =$
532 $.014$) or two prompts ($M = 9.9$ seconds, $SD = 5.44$, $beta = -4.8$, $sem = 1.49$, $df = 12$, $t = -3.22$,

533 $p = .007$; the difference between one and two prompts was not significant, $beta = -1.2$, $sem =$
534 1.64 , $df = 11$, $t = -0.74$, $p = .47$). Crucially, as predicted by the *collective intention* and the
535 *shared intention* hypotheses, the main effect of Prompt Type was such that musicians exhibited
536 a better temporal coordination in the WE ($M = 5.25$ seconds, $SD = 1.98$) as compared to the
537 ME condition ($M = 10.5$ seconds, $SD = 4.73$, $beta = -4.4$, $sem = 1.1$, $df = 18$, $t = -3.97$, $p < .001$).
538 Thus, the nature of the prompted goals (i.e., collective versus individual) impacted how well
539 musicians were able to temporally coordinate with each other, which is consistent with the idea
540 that shared information is not the only factor that would impact coordination, but that the
541 content of goals (i.e., whether they involve the individual alone, or the group as a whole) is also
542 crucial.

543 The interaction between Prompt Type and Prompt Number reflected the fact that the
544 *Temporal Coordination of Endings* significantly improved as the number of prompts increased
545 in the WE condition ($\chi^2 = 4.96$, $beta = -1.77$, $sem = 0.72$, $df = 11$, $t = -2.47$, $p = .03$) but not in
546 the ME condition ($\chi^2 = 2.55$, $beta = -1.74$, $sem = 1$, $df = 11$, $t = -1.7$, $p = .11$). The *Temporal*
547 *Coordination of Endings* was significantly smaller in the WE as compared to the ME condition
548 when there were 2 prompts ($beta = -8.54$, $sem = 1.76$, $df = 58$, $t = -4.84$, $p < .001$), but this effect
549 did not reach significance when there was only one prompt ($beta = -2.5$, $sem = 1.69$, $df = 51$, t
550 $= -1.47$, $p = .15$), or when there were three prompts ($beta = -2.38$, $sem = 1.8$, $df = 61$, $t = -1.32$,
551 $p = .19$). This suggests that the difference between the intentional content of the goals was
552 greatest in situations of partial sharedness, as compared to situation of full sharedness or lack
553 of sharedness. This is not entirely compatible with the *shared intention* hypothesis (and with
554 our pre-registered hypothesis): although it specifically predicts that temporal coordination
555 should improve with the number of prompts in the WE-condition, this hypothesis would also
556 predict that temporal coordination would be maximal in the condition were the three musicians
557 received a WE-Goal. This lack of effect for post-hoc comparisons may be due to a lack of

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

565 Interestingly, the level of temporal coordination measured in the WE condition in
566 Experiment 2 was not different from that measured in Experiment 1 (Experiment 1, $M = 5.21$,
567 $SD = 2.8$, linear mixed model comparison: $\beta = -1.29$, $sem = 0.84$, $df = 73$, $t = -1.53$, $p = .13$).
568 By contrast, temporal coordination was significantly worse in the individual intention condition
569 (ME-Goal) than in Experiment 1 ($\beta = -1.85$, $sem = 0.83$, $df = 149$, $t = -2.24$, $p = .027$). This
570 is consistent with our observation that in the unconstrained CFI conditions of Experiment 1, the
571 ending goals that spontaneously emerge are likely to be collective intentions rather than
572 individual intentions.

573 Finally, we computed a linear mixed regression with the unprompted musicians'
574 *Temporal Coordination with Others* as a dependent variable (see Fig. 2B). This analysis
575 revealed a main effect of Prompt Type ($\chi^2 = 4.31$; $p = .038$), no effect of Prompt Number ($\chi^2 =$
576 0.04 ; $p > .5$) and a marginal interaction ($\chi^2 = 3.35$; $p = .07$). Unprompted musicians were more
577 temporally coordinated with others in the WE ($M = 0.75$, $SD = 9.78$) condition than in the ME
578 condition ($M = 10.17$, $SD = 13.58$, linear mixed comparison: $\beta = -3$, $sem = 1.42$, $df = 79$, $t =$
579 -2.15 , $p = .035$). Thus, the existence of even a partially shared intention within the group was
580 enough to improve the ability of the unprompted musicians to coordinate with others: it not
581 only impacted the performance of prompted musicians, but also the performance of the group
582 as a whole, which is consistent with the *shared intention* hypothesis.

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

593

594 ----- *Insert Fig.2 about here* -----

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

598

599 ***3.2.2. Impact of the number and type of prompts on dynamic, timbral and harmonic*** 600 ***coordination***

601 To investigate whether receiving prompts changed the relationships between the
602 musicians, we conducted acoustic analysis on musical snippets extracted before and after the
603 prompts. Following previous research (Pachet et al., 2017), we approximated musical
604 coordination by computing a linear (Pearson correlation) as well as a non-linear (mutual
605 information) index of dependency between musicians for five acoustic features: pitch, volume
606 (RMS), playing time ratio (% of sound), spectral centroid and Harmonic-to-noise ratio (see
607 methods).

608 First, and before analyzing how the prompted goals impacted coordination at the
609 acoustic level, we verified that our measures effectively captured some forms of musical
610 coordination. This is non-trivial in our case since, as detailed above, CFI is generally devoid of
611 harmonic and rhythmic structure. To this aim, we simply tested whether the linear correlation
612 between acoustic features across time differed from zero overall. Correlation within trios (i.e.,
613 Pearson's rho averaged for each trio so as to estimate coordination at the level of the group)
614 was significantly higher than chance for 2 of the 5 acoustic features (rms r : $M = .17$, $SD = .06$,
615 $t(11) = 8.9$, $p < .001$; Playing time ratio r : $M = .15$, $SD = .08$, $t(11) = 6.45$, $p < .001$), marginally
616 higher than chance for 2 acoustic features (pitch r : $M = .02$, $SD = .03$, $t(11) = 2.07$, $p = .06$;
617 harmonic-to-noise ratio r : $M = .023$, $SD = .037$, $t(11) = 2.02$, $p = .07$) and did not significantly
618 differ from zero for the spectral centroid ($M = .02$, $SD = .04$, $t(11) = 1.72$, $p = .11$). Thus, four
619 of our five measures captured substantial acoustic coordination. These results – although
620 reflecting rather weak associations – are in fact quite significant when related to the astounding
621 variety and complexity of timbral and instrumental expressions found in CFI, and the fact that
622 previous studies involving jazz musicians and similar measures failed to capture substantial
623 acoustic coordination over and beyond the coordination explained away by the shared musical
624 score (Pachet et al., 2017).

625 With this in mind, we examined our main question of interest, which was to assess
626 whether shared intentions impact musical coordination (see Fig. 3). To assess this, we ran a
627 logistic mixed regression with timing (before or after) as a dependent variable, prompt type,
628 prompt number and acoustic coordination variables (Pearson's rho and MI for the five acoustic
629 dimensions, as well as dissonance) as independent variables, and trio as a random factor. After
630 the prompt, there was a significant increase in mutual information for loudness ($\beta = 4.1$, sem
631 $= 1.1$, $df = 204$, $z = 3.77$, $p < .001$), a significant decrease in mutual information for pitch (β
632 $= -2.8$, $sem = 0.85$, $df = 204$, $z = 3.29$, $p < .005$), as well as a decrease in dissonance ($\beta = -$

633 0.0016, $sem = 0.0005$, $df = 204$, $z = 3$, $p < .005$). Thus, the prompts substantially modified
634 dynamic and harmonic aspects of musical coordination.

635 Over and above these main effects, we also observed that prompt type and number
636 differentially impacted musical coordination, and we break down these effects in Fig.3's
637 caption for each acoustic dimension. For pitch, we found that the decrease in mutual
638 information was actually restricted to the ME condition: there was a significant interaction
639 between timing and prompt type ($beta = -3.9$, $sem = 1.8$, $df = 165$, $z = -2.14$, $p = .03$), and the
640 decrease was significant in the ME ($t(11) = -3.65$, $p = .004$) but not the We ($t(10) = -1.28$, $p >$
641 $.23$) condition. Thus, after hearing a "ME" prompt, the pitch of the music produced by the
642 improvisers became more independent from the pitch produced by other musicians, but this
643 effect was not observed after they heard a "WE" prompt. For loudness, the decrease in mutual
644 information did not significantly interact with prompt type nor number. By contrast, the Pearson
645 correlation was significantly impacted by prompt type ($beta = -6$, $sem = 2.16$, $df = 165$, $z = -$
646 2.79 , $p = .005$): the linear relationships between musicians' volumes significantly increased
647 after WE ($t(10) = 2.26$, $p = .047$) but not ME ($t(11) = 0.5$, $p > .6$) prompts. However, the decrease
648 in dissonance did not significantly interact with prompt type nor number: the music was less
649 dissonant after the prompt both in the ME ($t(11) = -2.9$, $p = .014$) and WE ($t(10) = -3.22$, $p =$
650 $.009$) conditions. Finally, for timbral aspects (centroid and HRN) and the percentage of sound,
651 there were no main effects and no interactions (see Fig.3 caption for details).

652 Overall, these analyses suggest that the presence of goals impacts musical coordination
653 during improvised interactions: even at the basic level captured by our acoustic analysis,
654 prompts had an impact on how improvisers' musical actions related to one another, at least for
655 coordination at the harmonic and dynamic (i.e., loudness) levels. Specifically, when they had a
656 WE-Goal, musicians' productions evolved towards being more consonant, and their loudness
657 was more correlated over time, suggesting tighter musical coordination. When musicians

658 received a ME-Goal, their production also became more consonant but, in addition, the pitches
659 they produced became more independent from one another, and they did not show improved
660 coordination (i.e., tighter correlation) at the level of loudness.

661

662 ----- *Insert Fig.3 about here* -----

663 *Fig.3. Change in dynamic, timbral and harmonic coordination after the prompt depending on prompt type and number. For*
664 *each take, timing (after/before) and each trio, musical coordination was assessed by computing the mutual information or*
665 *Pearson correlation between each pair, between averaging these values within each trio separately depending on prompt type*
666 *and number. We also computed a measure of dissonance over the whole trio for each take and timing, before averaging it*
667 *separately depending on prompt type and number. Black asterisks show main effects of timing (before/after), colored asterisks*
668 *main effects of prompt type. Error bar show the 95% confidence intervals. Significant impacts of prompt type and number on*
669 *the acoustic measures of musical coordination are detailed in the main text. For centroid, there was no main effect of centroid*
670 *on timing, and no interactions with prompt type nor number. It is worth noting however that there was a significant decrease*
671 *in spectral centroid's correlation in the We-3 condition after the prompt ($t(10) = -2.27, p = .046$, all other comparisons n.s.),*
672 *which may reflect an attempt of the musicians to distribute themselves in different parts on the spectrum (i.e., an increase in*
673 *musical coordination). For HNR, there was no main effect of centroid on timing, and no interactions with prompt type nor*
674 *number. Again, it is worth noting nonetheless that there was a significant decrease in HNRs' mutual information in the We-3*
675 *condition after the prompt ($t(10) = -2.87, p = .017$, all other comparisons n.s.), that may reflect an attempt to produce textures*
676 *that are more distinct (i.e., an increase in musical coordination). Percentage of sound: there were no significant effects for this*
677 *measure.*

678

679

680 **4 Experiment 3: Impact of the number and type of prompts on qualitative** 681 **aspects of musical coordination.**

682

683 Next, we wanted to assess whether shared intentions impacted properties of the
684 performance related to higher-level and qualitative aspects of musical coordination, beyond
685 temporal coordination and the relatively low-level acoustical features that we examined in part
686 3.2.1 and 3.2.2. A particularly interesting question is whether the impact of shared intentions

687 on the performance can be perceived by external observers, and reflected in their aesthetic
688 evaluations. Thus, in a third experiment, we asked third-party listeners (both experts and non-
689 experts) to rate the extent to which they thought the ending was successful, and to classify the
690 endings along several categories corresponding to qualitative aspects that are linked to
691 coordination during CFI.

692

693

694 **4.1 Experiment 3 – Methods**

695

696 **4.1.1 Participants**

697 We determined the size of the sample with a power analysis involving musicians’
698 sensitivity in guessing each other’s prompts (Experiment 2, see Fig. S5). To have a power of
699 95% at the 0.05 alpha level, the analysis showed that we should aim to test 23 participants per
700 group. Given scheduling constrains, we finally tested 26 naive listeners (8 women, age $M =$
701 27.4 years, $SD = 8$ years) who were not musicians (mean number of years of instrumental
702 practice: $M = 0.42$, $SD = 1.08$) and had no experience of CFI (mean number of years of CFI
703 practice: $M = 0$, $SD = 0$) and 21 experts (5 women, age $M = 33.9$ years, $SD = 8.8$ years) who
704 were all accomplished musicians (mean number of years of instrumental practice: $M = 23.14$,
705 $SD = 8$) with a strong experience of CFI (mean number of years of CFI practice: $M = 10.7$, SD
706 $= 6.5$). Participants reported having no major hearing or visual impairment, and appropriate
707 corrections allowing them to perceive the stimuli. They signed an informed consent and were
708 compensated financially after the experiment.

709

710 **4.1.2 Stimuli**

711 We selected 24 improvisations pseudo-randomly from those recorded in Experiment 2
712 by ensuring that 1) no trio was over-represented; 2) every trio was included; 3) the main finding

713 were replicated in the subset (i.e., the impact of Prompt Type and Number on the *Temporal*
714 *Coordination of Endings*); 4) half of the improvisations were taken from the ME condition, and
715 half from the WE condition; 5) each individual musician played during at least 19 seconds
716 after prompt delivery (this last condition matters only for the Experiment 4, presented below,
717 which relies on the same subset of improvisations than Experiment 3).

718

719 **4.1.3 Procedure and Data analysis**

720 Listeners heard the last 50 seconds of each of the 24 improvisations, and indicated on a
721 7-points Likert scale whether they thought that what they just heard was a good ending or not.
722 Listeners were also asked in a random order whether the ending was: 1) hierarchical or
723 egalitarian; 2) collective or disjoint; 3) progressive or immediate; 4) predictable or surprising;
724 and 5) timely or not (too late or too early). These five qualitative aspects were derived from
725 musicians' reports during Experiment 2, where their judgements of appreciation were generally
726 related to one or several of these categories. To infer categories from these written reports, three
727 of the authors (L.G., P.S.-G., and C.C.) read all of the reports, and grouped them in several
728 categories. These subjective groupings were quite consistent amongst the three authors, and
729 suggested that the five aspects listed above capture most of the relevant parameters reflecting
730 the success of coordination during CFI. To ensure that all participants understood the five
731 qualitative aspects in a similar fashion, we provided them with a glossary describing the
732 meaning of each label (see glossary in the supplementary materials, section S.3.1.). We
733 analyzed appreciation ratings as a continuous variable, and qualitative ratings were dummy
734 coded as binary variables (e.g., for the hierarchical category, we dummy coded hierarchical
735 responses as 1, and egalitarian as 0). Data, data collection and analysis scripts are available via
736 [this link](#).

737

738

739 4.2 Experiment 3 – Results

740

741 4.2.1 Shared intentions impact the success of endings

742

743 We analyzed the impact of Prompt Type and Prompt Number on listeners' appreciation ratings
744 with a rmANOVA (see Fig. 4A). There was an interaction between Prompt Type and Prompt
745 Number ($F(2,90) = 4, p = .021, \eta p^2 = .04$), a main effect of Prompt Number ($F(2,90) = 11.5, p$
746 $< .001, \eta p^2 = .06$) and no main effect of Prompt Type ($F(1,45) = 0.027, p > .8, \eta p^2 = .00$).
747 Appreciation ratings were highest in the WE-3 condition (ratings were higher in the WE-3
748 condition than in WE-2, $p < .001$; ME-3, $p = .007$, ME-2, $p = .006$, post-hoc Tukey HSD).
749 Listeners' appreciation ratings were thus maximal when performers had a shared intention,
750 which is consistent with our hypothesis that shared intentions help musicians to coordinate and
751 attain a better outcome. We also examined the relationship between appreciation, Prompt Type,
752 Prompt Number and expertise, and report these results in Fig. S6A. Overall, the impact of
753 shared intentions on musical coordination could be perceived independently from expertise,
754 which suggests that even in an avant-garde artistic form like CFI, coordination relies on features
755 that are transparent enough to be accessible to the general population (see Moran, Hadley,
756 Bader, & Keller, 2015 for a similar finding regarding expressive movements).

757

758

----- Insert Fig.4 about here -----

759

760

761

762

763

764

Fig.4. Mains results of Experiment 3. A) Expert and naïve 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 non-significant). Thus, listeners also preferred conditions in which less prompts were present (WE-1 and ME-1 conditions did not differ $p > .5$). This may be due to

765 *the fact that these interactions are less artificial than the others (i.e., only one of the musicians receives a prompt and the others*
766 *playing remains unconstrained). Note that musicians in these more natural conditions may also spontaneously form shared*
767 *intentions, as suggested by the results observed in the first experiment. B) The percentage of hierarchical, collective,*
768 *progressive, predictable, and on time assessment was computed for each of the five qualitative questions, separately for each*
769 *participant, prompt number and prompt type, before being averaged in the group. Black asterisks show the logistic regression*
770 *model comparisons, and the blue asterisk represents the fact that all comparisons were significant with respect to the indicated*
771 *condition. $p < .05$; **: $p < .01$; ***: $p < .001$. Error bars show the 95% confidence intervals.*

772

773 **4.2.2 Shared intentions impact qualitative aspects of endings**

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

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

785 For Progressivity, there was a significant interaction ($\chi^2 = 32, p < .001$). Listeners
786 judged endings to be more progressive when the three musicians received a WE-Goal (model
787 comparison between 3 vs. 1 prompt: $beta = 1.32, sem = 0.33, df = 1053, z = 3.9, p < .001$; 3 vs.
788 2 prompt: $beta = 1.11, sem = 0.35, df = 1053, z = 3.15, p = .002$), and less progressive when the
789 three musicians received a ME-Goal (3 vs. 1 prompt: $beta = -0.9, sem = 0.25, df = 1053, z = -$
790 $3.63, p < .001$; 3 vs. 2 prompt: $beta = -0.86, sem = 0.29, df = 1053, z = -3, p = .003$; comparison

791 between WE-3 and ME-3: $\beta = 1.94$, $se = 0.38$, $df = 1053$, $z = 5$, $p < .001$) as compared to
792 the other conditions.

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

798 Finally, and crucially, regarding Timing, there was a significant interaction ($\chi^2 = 16.52$, $p <$
799 $.001$). While in the ME condition no significant differences were observed depending on prompt
800 number (all $ps > .07$), listeners in the WE condition judged endings to be timelier when the
801 three musicians had received a prompt (3 vs. 1: $\beta = 0.96$, $se = 0.26$, $df = 1053$, $z = 3.7$, $p <$
802 $.001$; 3 vs. 2: $\beta = 1.06$, $se = 0.27$, $df = 1053$, $z = 3.86$, $p < .001$). In addition, listeners judged
803 endings to be significantly timelier in the WE as compared to the ME condition when there
804 were 3 prompts ($\beta = 1.13$, $se = 0.31$, $df = 1053$, $z = 3.6$, $p < .001$), but not 2 prompts (β
805 $= 0.12$, $se = 0.23$, $df = 1053$, $z = 0.6$, $p > .5$) nor 1 prompt ($\beta = 0.28$, $se = 0.17$, $df = 1053$,
806 $z = 1.7$, $p > .09$).

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

812

813

814 **5 Experiment 4: How do improvisers' goals propagate?**

815

816 A remaining question concerns how goals propagate within the group, and whether they can be
817 perceived from the music alone. In a last experiment, we wanted to test the claim that
818 transparent goals (i.e., goals that are easier to detect) have a more positive impact on
819 coordination. This is a specific prediction of the *shared intention* hypothesis, according to
820 which improvisers may coordinate through forming collective intentions that are shared and
821 common knowledge between them. To this aim, we asked naive and expert listeners to try and
822 detect whether individual musicians had an intention to end the performance. We also examined
823 the relationship between listeners' detections of goals and temporal coordination, to see whether
824 transparent goals corresponded to better temporal coordination. Finally, we wanted to try and
825 assess how goals may be manifested, and thus effectively propagate within the group. To
826 examine this issue, listeners were also asked to characterize performers' behaviors along four
827 qualitative aspects. They were asked whether they thought that the musicians' behavior was: 1)
828 descending or not descending (i.e., ascending, constant or without direction); 2) repetitive or
829 varied; 3) predictable or surprising; and 4) confident or hesitant. This also allowed us to
830 examine whether specific behaviors are associated with better temporal coordination and/or
831 shared intentions, suggesting that they may be used by the performers as coordination
832 smoothers or communicative signals (Vesper et al., 2017).

833

834

835 **5.1 Experiment 4 – Methods**

836 **5.1.1 Stimuli**

837 Stimuli were 72 audio extracts from the three individual performances in each of the 24
838 improvisations used in Experiment 3. All stimuli were 17 seconds long, extracted either 17
839 seconds before the prompt (Before condition, $N = 18$ extracts) or 17 seconds after the prompt,
840 either in trials in which the musician heard a ME-Prompt (ME-Goal condition, $N = 18$ extracts),

841 a WE-Prompt (WE-Goal condition, $N = 18$ extracts), or did not hear a prompt (No-Prompt
842 condition, $N = 18$ extracts). None of the extracts included the actual ending of the piece (i.e., in
843 all of these takes, every musician stopped at least 19 seconds after hearing the prompt).

844

845 *5.1.2 Procedure and Design*

846 Participants were the same as for Experiment 3. They were told that in about half of the
847 musical extracts, musicians were looking for an ending and were about to stop playing, while
848 in the other half they weren't looking for an ending. They were asked to report – via a key press
849 (left or right arrow, counterbalanced between participants) – whether the musician was about
850 to stop playing (i.e., to detect ending goals). Participants then provided a confidence rating in
851 their answer on a scale from 1 to 4, and categorized the musician's behavior by responding to
852 four questions presented in a random order. For each category, participants were presented with
853 several alternatives (direction: ascending / descending / constant / none; repetition: repetitive /
854 varied; prevision: predictable / surprising; assurance: confident / hesitant) and asked to select
855 one of them by pressing one of the arrows on the keyboard. These categories were derived from
856 the musicians' reports during Experiment 1, where decisions about their partners' intentions
857 were reported to be caused by one or several of these behaviors (see supplementary materials,
858 section S.1.2., for a few examples and details of the procedure that allowed us to extract these
859 categories from musicians' written reports about how they detected their partners' intention to
860 end during experiment 1). Listeners were provided a glossary to make sure that all of them
861 understood these categories in the same way (see glossary in the supplementary materials,
862 section S.4.1.).

863

864 *5.1.3 Data Analysis*

865 We computed a measure of sensitivity based on signal detection theory (d' , Green &
866 Swets, 1966) for each participant and condition, taking tracks extracted after the prompt
867 (NO/ME/WE) as targets, and tracks extracted before the prompt (Before) as non-target. For each
868 participant and condition (NO/ME/WE), the hit rate was computed as: (the number of positive
869 responses for extracts taken after the prompt for that condition / the total number of extracts taken
870 after the prompt for that condition); and the false alarm rate as: (the number of positive responses
871 for extracts taken before the prompt / the total number of extracts taken before the prompt). Note
872 that, although we treated the NO-Goal condition like the WE and ME-Goal conditions to compute
873 d' here, so as to allow direct comparison between the three conditions, detecting an ending in
874 this condition is not necessarily a “wrong” response: the unprompted musician may or may not
875 have an intention to end depending on whether the goal propagated in the group or not. Data,
876 data collection and analysis scripts are available via [this link](#).

877

878 **5.2 Experiment 4 – Results**

879

880 ***5.2.1. Third-party listeners can detect improvisers’ goals***

881 Average sensitivity (d') was $M = 0.37$, $SD = 0.56$, which was significantly above chance
882 level ($t(46) = 4.48$, $p < .001$). A rmANOVA revealed a main effect of Prompt Type (NO / ME
883 / WE: $F(2,90) = 30.8$, $p < .001$) on sensitivity, and an interaction between Expertise and Prompt
884 Type ($F(2,90) = 3.14$, $p = .048$). As can be seen in Fig. 5A, both experts (d' : $M = 0.71$, $SD =$
885 0.74) and naive listeners (d' : $M = 0.55$, $SD = 0.46$) achieved above chance sensitivity in the ME
886 condition (musicians: $t(20) = 4.28$, $p < .001$; non-musicians: $t(25) = 6$, $p < .001$), and there was
887 no difference between the two groups in this condition (post-hoc Tukey HSD: $p = .32$). By
888 contrast, sensitivity in the WE condition varied with expertise: while experts achieved above
889 chance sensitivity ($M = 0.72$, $SD = 0.64$, $t(20) = 5$, $p < .001$), naive listeners’ sensitivity did not

890 significantly differ from chance ($M = 0.21$, $SD = 0.67$, $t(25) = 1.56$, $p = .13$; group difference:
891 $p = .002$).

892 Thus, ME-Goals could be perceived from musicians' behavior independently from
893 listener's expertise, while the detection of WE-goals depended on expertise. This suggests that
894 WE-goals – i.e., goals whose content refer to the group's performance as a whole – may be
895 characterized by specific features that are only accessible to expert listeners. One could argue
896 that this impact of expertise is due to musicians' better auditory processing capacities, that
897 would enable them to attend to finer acoustic cues which carry this information. Yet, this
898 interpretation is not compatible with the lack of difference between the two groups in the ME
899 condition. More interestingly, it could be that WE-Goals depend on conventional behaviors that
900 are only accessible to listeners possessing the same cultural background as the performers. We
901 come back to this issue below. Notwithstanding, the results show that improvisers' goals have
902 some degree of transparency, and that they are manifested in the performance in ways that allow
903 performers and external listeners to detect them.

904

905 ----- *Insert Fig.5 about here* -----

906 *Fig.5. A) Participants sensibility (d') was assessed by computing for each condition and participant the hit rate (number of*
907 *positive responses for snippets extracted after the prompt / number of snippets extracted after the prompt)) and false alarm*
908 *rate (number of positive responses for snippets extracted before the prompt / number of snippets extracted before the prompt).*
909 *White asterisks show p -values for one-sample t -tests against chance level, black asterisks show post-hoc Tukey HSD for*
910 *between group or conditions comparisons. ***: $p < .001$; **: $p < .01$. B) The percentage of positive responses (i.e., "Yes, I*
911 *think the performer is looking for an end") was computed separately for each participant depending on prompt type and*
912 *number, before being averaged in the group. A logistic mixed regression with responses (yes/no) as a dependent variable*
913 *revealed that when only one of the performers had a Goal, listeners detected an intention to end less often when listening to*
914 *the unprompted performer as compared to when both other performers had a ME-Goal (beta = 1.19, sem = 0.16, $df = 5025$, z*
915 *= 7.4, $p < .001$) or a WE-Goal (beta = 0.53, sem = 0.16, $df = 5025$, $z = 3.26$, $p = .001$). Listeners also reported an intention*
916 *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*
917 *= 5025, $z = 3.63$, $p < .001$). Error bars show the 95% confidence interval.*

918

919 **5.2.2. Goal propagation: shared intentions impact how listeners perceive unprompted**
920 **musicians' goals**

921

922 In the NO-Prompt condition, sensitivity did not differ from chance level in any of the groups
923 (musicians d' : $M = 0.19$, $SD = 0.66$, $t(20) = 1.29$, $p = .2$; non-musicians d' : $M = -0.05$, $SD =$
924 0.49 , $t(25) = -0.52$, $p = .6$). Thus, overall, listeners did not perceive ending goals when
925 performers did not receive a prompt themselves. This may suggest that the behavior of
926 unprompted performers did not reflect an intention to end after one or both of their co-
927 performers were prompted. Yet, it remains possible that it *only* did when *both* of their co-
928 performers were prompted.

929 To examine this possibility, we examined how detection responses (yes/no) depended
930 on Prompt Type and Number (Fig. 5B). We ran a mixed logistic regression with detection
931 response as a dependent variable, and prompt type and number as independent variables. There
932 was a main effect of Prompt Number ($\chi^2 = 6.17$; $p = .046$), a main effect of Prompt Type ($\chi^2 =$
933 61.17 ; $p < .001$), as well as an interaction between Prompt Number and Prompt Type ($\chi^2 =$
934 27.93 ; $p < .001$). Post-hoc test revealed that when both of an unprompted performer's co-
935 performers had a goal, listeners reported that the unprompted performer had an intention to end
936 as often as they did when listening to prompted performers that had a ME-Goal ($beta = 0.018$,
937 $sem = 0.18$, $df = 5025$, $z = 0.09$, $p = .92$), but less often as compared to prompted performers
938 who had a WE-Goal ($beta = 0.41$, $sem = 0.18$, $df = 5025$, $z = 2.26$, $p = .024$). In addition,
939 listeners reported that unprompted performers had an intention to end more often when both of
940 their partners had an intention to end as compared to when only one of their partners had an
941 intention to end (1 vs. 2 in the NO-Prompt condition: $beta = 0.47$, $sem = 0.17$, $df = 5025$, $z =$
942 2.8 , $p = .005$, see Fig. 5B for a full output of the model).

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

950

951

952 ***5.2.3. Improvisers adopt signaling strategies to communicate their goals***

953 How may such goal propagation occur? To examine whether musicians deployed
954 particular strategies to signal their intentions to end, we assessed the impact of our experimental
955 conditions on how listeners described the musicians' behaviors. We ran a linear regression
956 including percentage of response as a dependent variable, Condition (Before-Prompt / NO-
957 Prompt / ME-Goal / WE-Goal), Category (descending / repetitive / predictable / confident) and
958 Expertise (naive / expert) as independent variables, and listener as a random factor. There was
959 a main effect of Condition ($\chi^2 = 21, p < .001$), a main effect of Category ($\chi^2 = 647, p < .001$)
960 and, more importantly, a significant interaction between Condition and Category ($\chi^2 = 50, p <$
961 $.001$), which revealed that listeners judgments about performers behaviors along each Category
962 varied differently depending on Condition (see Fig. 6B). There was no additional interaction
963 with Expertise ($p > .14$), so we collapsed the data for the two groups of listeners for the
964 remaining analyses.

965 Regarding direction, listeners responded that the musician's behavior was *descending*
966 significantly more often when they heard prompted musicians (ME: $M = 0.24, SD = 0.12$; WE:
967 $M = 0.28, SD = 0.15$) than un-prompted musicians ($M = 0.16, SD = 0.1$; post-hoc Tukey HSD

968 No versus ME: $p < .001$; NO versus WE: $p < .001$) or extracts taken before the prompt ($M =$
969 0.19 , $SD = 0.12$; Before versus ME: $p = .001$; Before versus WE: $p < .001$; no significant
970 difference between Before and NO-Prompt: $p = .26$). Interestingly, there were no significant
971 difference between the rate of descending responses in the ME and WE condition ($p = .08$),
972 which rules out the possibility that WE-Goals simply foster coordination because performers
973 rely on decrescendos to drive the improvisation towards the end (also see the acoustic analysis
974 presented in Fig. S10).

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

980 Finally, and more importantly, listeners responded that behaviors were *predictable* and
981 *repetitive* significantly more often when the performer had a WE-Goal ($M = 0.66/0.72$, $SD =$
982 $0.15/0.17$), as compared to when the performer had a ME-Goal ($M = 0.58/0.61$, $SD = 0.16/0.16$,
983 $ps < .001$), was not prompted ($M = 0.57/0.62$, $SD = 0.14/0.14$, $ps < .001$) or for extracts taken
984 before the prompt ($M = 0.48/0.5$, $SD = 0.15/0.16$, $ps < .001$). Listeners also perceived behaviors
985 to be more predictable/repetitive when performers had a ME-Goal ($ps < .001$) or NO-Goals (ps
986 $< .001$), as compared to the extracts taken before the prompt.

987 The crucial finding here is that musicians relied on more predictable and repetitive
988 behaviors when they had a WE-Goal, presumably to allow their partners to coordinate with
989 them. These repetitive/predictable behaviors could be due to performers playing the same
990 complex pattern over and over again or holding a single tone, but were not necessarily related
991 to performers playing a regular pulse (see Fig. S9 and glossary in the supplementary materials
992 section S.4.1). This finding is consistent with previous research emphasizing the role of

993 predictability and repetitive actions for coordination (Vesper et al., 2011) and emerging
994 communication systems in the visual modality (Scott-Phillips, Kirby, & Ritchie, 2009), and
995 shows that improvisers used basic signaling strategies to help establish common ground when
996 they have to reach a joint outcome with their fellow improvisers.

997

998

----- *Insert Fig.6 about here* -----

999 *Fig.6. Musicians behavior. A) The temporal coordination of endings in improvisations corresponding to the snippets heard by*
1000 *the participants was averaged separately for each listener, prompt type and response type (yes/no), before being averaged in*
1001 *the group. *** show the significance of paired t-tests with a threshold of $p < .001$. B) We show the percentage of descending,*
1002 *repetitive, predictable and confident responses computed for each condition and listener, before being averaged separately in*
1003 *the group of experts (plain line) and naive listeners (dashed line). Error bars show the 95% CI.*

1004

1005

1006 **5.2.4. Goal transparency predicts better temporal coordination**

1007

1008 Finally, we wanted to test the claim that transparent goals (i.e., goals that are easier to detect)
1009 foster coordination. To this aim, we examined the relationship between listeners' goal detection
1010 and subsequent temporal coordination at the end of the piece (which was not presented to the
1011 participants). In a linear mixed regression restricted to judgements made on extracts taken after
1012 the prompt, and including listener and trio as random factors, listeners' detection choices (yes
1013 versus no) significantly predicted the subsequent *Temporal Coordination of Endings* ($\beta =$
1014 0.4 , $sem = 0.07$, $df = 5068$, $t = 6$, $\chi^2 = 36.7$, $p < .001$). On average, performers were more
1015 temporally coordinated in musical extracts where listeners' detected an intention to end ($M =$
1016 5.68 , $SD = 1.19$) as compared to when they did not ($M = 6.97$, $SD = 0.62$, $t(46) = 5.15$, $p < .001$,
1017 see Fig. 6A). This was true for both ME ($t(46) = 3.72$, $p < .001$) and WE-Goals ($t(45) = -5.07$,
1018 $p < .001$), and also after accounting for the effect of Prompt Number on temporal coordination

1019 ($\beta = 0.24$, $sem = 0.06$, $df = 5068$, $t = 3.8$, $p = .007$). This result is therefore consistent with
1020 the idea that goal transparency helps coordination, and that making one's goal easier to detect
1021 by fellow improvisers might be key to coordination during improvised interactions.

1022

1023

1024 **6 Discussion**

1025

1026 Despite being an integral part of our social lives, joint improvised actions have been
1027 understudied to date, and the mechanisms that allow agents to coordinate in complex and
1028 temporally extended forms of collective improvisation remain elusive. The experiments
1029 reported here shed a new light on these mechanisms in the context of collective free musical
1030 improvisations: in Experiment 1, we show that shared intentions emerge on the fly during
1031 collective musical improvisations; in Experiment 2, we show that the presence of such shared
1032 intentions fosters temporal and acoustic coordination; in Experiment 3, we show that shared
1033 intentions also have an effect on qualitative properties of the performance that reflect higher-
1034 level aspects of musical coordination (such as the endings being rated as more successful,
1035 timelier and more progressive); finally, in Experiment 4, we show that improvisers' goals can
1036 be inferred by third-party listeners from their musical behavior and that, strikingly, unprompted
1037 musicians may come to reflect the behaviors of their prompted co-improvisers. The results also
1038 show that improvisers adopt signaling strategies when they have to communicate their goals to
1039 reach a joint outcome collectively, which explains how collective intentions can propagate,
1040 become common knowledge, and improve musical coordination.

1041 Overall, the results are compatible with the hypothesis that *shared intention* foster
1042 coordination during improvised musical joint actions, over and beyond the role of mere *shared*
1043 *information* and of the isolated formation of *collective intentions* in individual musicians. This

1044 demonstrates that the synergy between planned and emergent coordination mechanisms that
1045 had so far been considered exclusively in scripted joint actions is also at play in improvised
1046 joint actions. While our results are in line with the idea that shared intentions support
1047 coordination over long as well as short time scales (Vesper, Butterfill, Knoblich, & Sebanz,
1048 2010), they – perhaps counterintuitively – extend its relevance to the case of collective musical
1049 improvisations.

1050 An important theoretical consequence of our study is that it gives some additional ground
1051 to the idea that shared intentions do not intrinsically depend on verbal communication for their
1052 existence: we show that shared intentions can emerge when agents are freely and spontaneously
1053 interacting within a medium that is semantically underspecified (i.e., music), and that they play
1054 a key role in supporting coordination. The condition of common knowledge, where agents are
1055 not only geared towards a joint outcome, but also represent that this state of affair is publicly
1056 accessible to all members of the group, is generally taken to be one of the crucial features of
1057 shared intentions (Bratman, 2014). Now, in Experiment 2, the WE-goals were communicated
1058 covertly to each musician, apparently violating the requirement of common knowledge.
1059 However, this does not mean that such common knowledge status could not emerge in the
1060 course of the performance, *after* the musicians were prompted, using joint affordances (i.e.
1061 events that afford actions or gestures for the group as a whole, Knoblich et al., 2011), signaling
1062 strategies that trigger “*distinctive cognitive states, corresponding to the sense that something is*
1063 *public and unignorable*” (De Freitas, Thomas, DeScioli, & Pinker, 2019), and focal points that
1064 act as points of converging expectations for the improvisers (Canonne, 2013). Several aspects
1065 of our results are consistent with this possibility.

1066 Results from Experiment 4 (Fig. 5) show that third-party listeners were able to infer WE-
1067 Goals from musicians’ behavior, demonstrating that these goal representations are indeed
1068 manifest, and publicly observable. Results from Experiment 4 further suggest that particular

1069 communicative behaviors (e.g., repetitions) may be especially efficient to signal an intention to
1070 end the piece. Lastly, we saw in Experiment 2 that both ME-Goals and WE-Goals were
1071 detectable by co-agents (see Fig. S5). However, WE-Goals and ME-Goals did not differ in
1072 terms of their directionality (i.e., both were perceived as “descending”) (see Fig. 6B). This rules
1073 out the possibility that improvisers merely detect teleological aspects such as a directionality in
1074 the joint action (e.g., decrescendos) without representing the mental states that may underlie
1075 this directionality in their co-agents, analogously to two-year-old children who engage
1076 successfully in joint action before they have a full understanding of folk psychological concepts
1077 such as *intention* (Butterfill, 2013; Butterfill & Apperly, 2013). On the contrary, the findings
1078 suggest that improvisers considered additional cues, beyond the mere sonic target, and engaged
1079 into some form of mentalizing to discriminate between the two types of goals. These elements
1080 indicate that musicians’ goals were both *manifest* and *mentally represented* by their co-
1081 improvisers. As such, they had the potential to become common knowledge between
1082 improvisers, and to amount to full-fledged shared intentions.

1083 Now, even when musicians collectively hold a shared intention to end the performance,
1084 how and when the performance will actually end still remains poorly specified: such abstract
1085 goals do not specify precise temporal or harmonic structures allowing the musicians to
1086 coordinate on fine time scales. In other words, even if musicians manage to form a shared
1087 intention to end the performance, the ending will still have to be spontaneously and collectively
1088 negotiated in a matter of seconds, without the support of a shared entrainment to a beat. How
1089 do such abstract intentions can support coordination in cases where the outcome remains highly
1090 undetermined? Several non-exclusive explanations might be provided here.

1091 A first possibility is that once it is common ground for co-improvisers that there is a
1092 shared intention to X (e.g., “to end the performance together”), they can coordinate by relying
1093 on interconnected planning. That is, they can form compatible sub-plans that are constrained

1094 by their shared intention to look for an end to the performance together (Bratman, 2014). This
1095 is not to say that each agent necessarily represents the other agents' part precisely (Vesper et
1096 al., 2010). Still, once a shared intention is established, performers can monitor and predict their
1097 co-performers actions more finely, and adjust their own behavior accordingly, because the
1098 shared intention constrains the range of possible interpretations of partners' behaviors, as well
1099 as each agent's action repertoire. This being said, although a minimal representation of one's
1100 own task and of the group's shared intention may suffice to finely coordinate in scripted joint
1101 actions that involve predetermined outcomes (Vesper et al., 2017), it is difficult to see how
1102 these mechanisms could allow musicians to precisely coordinate in the case of collective
1103 improvisations. Motor simulation is thought to be one of the crucial mechanisms that enable
1104 co-agents to predict each other's actions and coordinate on short time scales (Knoblich et al.,
1105 2011; Novembre et al., 2014; Vesper et al., 2013). Here however, it is unlikely that musicians
1106 simply rely on their motor system, given that they play on different instruments (Bishop &
1107 Goebel, 2014), and that they use idiosyncratic instrumental techniques. This is not to say that
1108 they cannot rely on action prediction at all. For instance, both expert and naive listeners
1109 perceived an intention to end in conjunction with decrescendos (see Fig. S8 and S11), which
1110 can be argued to be an index with a teleological origin (i.e., "descending" actions typically
1111 precede endings).

1112 A second way in which shared intentions may foster coordination is by enabling
1113 behavioral strategies designed to help coordination (Vesper et al., 2017). For instance, we found
1114 some evidence that musicians' behavior tended to be more repetitive and predictable in the WE-
1115 Goal condition (see Fig. 6B). One interpretation of this result is that, in improvised joint actions,
1116 agents use repetitive actions and other predictable behaviors not only as signals but also as
1117 "coordination smoothers" (Vesper et al., 2017), to help other improvisers predict and coordinate

1118 with them. In favor of this interpretation, we also found that predictability was associated with
1119 better temporal coordination (Fig. S8).

1120 Lastly, at shorter time scales (i.e., a few seconds), it is possible that shared intentions
1121 regulate the emergent mechanisms that are at play to support fine grained coordination. For
1122 instance, when musicians had shared intentions, the dynamics of their amplitude variations were
1123 more tightly coupled (see Fig. 3). This being said, the role of emergent coordination
1124 mechanisms is probably less crucial here than in other types of improvisations involving
1125 imitations such as the mirror game (Noy et al., 2011), because CFI is generally devoid of regular
1126 rhythmic pulsations, and straightforward imitations are often frowned upon amongst free
1127 improvisers. The acoustical analysis presented in Fig. S10, that shows little mimicry in
1128 unprompted musicians, is consistent with this idea: there was little to no evidence in favor of
1129 the idea that un-prompted musicians adapt their behavior by simply mimicking prompted
1130 musicians (e.g., by playing decrescendos).

1131 On the other hand, our results do not imply that agents engaged in collective musical
1132 improvisation always have a shared intention in mind, nor that they systematically need to. It
1133 is likely that musical improvisers oscillate between phases where they unreflectively “go with
1134 the flow”, and phases in which they are more self-conscious and engage in deliberate planning
1135 of their actions, and mindreading (Canonne & Garnier, 2012; Denzler & Guionnet, 2020). In
1136 that perspective, the fact that in Experiment 1 some musicians pressed their pedals after they
1137 had actually stopped playing suggests that musicians can be as surprised as audience members
1138 by the unfolding of their own performance. More generally, it is likely that shared intentions,
1139 to the extent that they are present, are of a rather punctual, short-term nature, emerging when
1140 acute coordination problems, such as endings or the consolidation of a new attractor (Borgo,
1141 2005), arise. Further studies could examine in a more systematic fashion the temporal dynamics
1142 of the kinds of abstract, shared intentions we evidenced here.

1143 On the methodological side, our study shows that collective musical improvisation
1144 constitutes an interesting case study to examine how individuals coordinate in the absence of
1145 scripts, and to investigate coordination dynamics in improvised interactions over an extended
1146 time span. When interactions between individuals are mediated by a pre-existing script, even
1147 loose ones such as the lead sheet of a jazz standard or conversational guidelines, it can be
1148 difficult to tease apart actual inter-personal interactions from individual’s isolated interactions
1149 with the script they all share (Pachet et al., 2017). CFI does not involve such referents and, as
1150 such, it allows a direct, unmediated examination of inter-personal interactions.

1151 Another interest of our approach is that it allows comparing expert and naive listeners, and
1152 makes it possible to uncover the (cultural) knowledge that mediates coordination during joint
1153 actions, an aspect often neglected in cognitive science (Vesper et al., 2017). While CFI is clearly
1154 a highly unplanned form of joint action, it does not happen in a cultural vacuum. Free
1155 improvisers spend many hours developing idiosyncratic instrumental techniques and a repertoire
1156 of distinctive musical materials (Arthurs, 2016). According to MacDonald & Wilson (2020,
1157 p. 115) though, “*particular knowledge or skills (...) are not in themselves a measure of the*
1158 *broader capacity to improvise*”. As such, an important part of free improvisers’ training,
1159 whether formal – through Conservatories classes – or informal – through listening to and
1160 playing with other improvisers – consists in developing broader coordination and
1161 communication skills, as well as highly general attributes, such as “*confidence in exercising*
1162 *choice in real time*”, “*discrimination and discernment of emerging performed material*”
1163 and “*facility in accommodating and responding to unprecedented or unexpected events*” that
1164 are “*transferable across genres or settings in ways that some other musical attributes are not*”
1165 (Wilson & MacDonald, 2020, p. 116-119). As Pelz-Sherman (1998, p. 127) puts it, learning to
1166 “*convey the semantic intent of their own musical ideas to other performers in real time*” and

1167 “to make accurate judgments in real time about the semantic intent of each performer” is
1168 crucial.

1169 Overall, expertise in CFI seems to largely rely on social cognition, contextual
1170 attunement, and interpersonal coordination, as is the case for other forms of freely improvised
1171 practices such as comedy improv (Walsh, Roberts, & Besser, 2013) or contact improvisation
1172 (De Spain, 2014). Thus, it is possible that the signaling mechanisms used by free improvisers
1173 were only accessible to expert listeners in our study, not necessarily because they rely on group-
1174 specific or genre-specific expertise and conventions, but perhaps because they require a high
1175 level of social attunement to the behavior of the improvisers. In other words, it might just be
1176 that our expert listeners, being also expert free improvisers, were more used to face improvised
1177 coordination problems, and thus simply better at abstracting signaling strategies from subtle
1178 variations in the performers’ behaviors. The wide variety of ending behaviors found in our
1179 corpus suggests that the signaling strategies used by the improvisers were not tied to precise
1180 instrumental or musical patterns, but were rather of a very abstract nature (e.g., decrease in
1181 energy, use of salient events, repetition, etc.), and thus possibly independent from the sonic and
1182 aesthetic specificities of CFI as a genre. Further experiments could directly test this hypothesis
1183 by assessing whether expert improvisers from another domain (e.g., improv theater) are able to
1184 detect our musicians’ intentions, despite their being unfamiliar with the genre of CFI. Note also
1185 that, consistently with this last hypothesis, some of our results tend to downplay the importance
1186 of group-specific stylistic conventions in the emergence of shared intentions amongst free
1187 improvisers. In particular, a high degree of familiarity between the musicians (and the implicit
1188 conventions that are likely to come with it) did not seem to give them any advantage in
1189 negotiating their joint endings: familiarity did not correlate with how temporally coordinated
1190 they were in pressing the pedals (experiment 1, Spearman’s rho between pedal pressing
1191 temporal coordination and familiarity scores: $rs(10) = -.38, p = .240$), how well musicians

1192 coordinated at the end of the piece (experiment 2, Spearman's rho between the temporal
1193 coordination of endings and familiarity, $rs(10) = .11, p = .730$), and nor even with how much
1194 they enjoyed playing together overall during experiment 1 and 2 (Spearman's rho between
1195 global appreciation and familiarity scores, $rs(10) = 0.4, p = .200$).

1196 If expertise in collective improvisation is mainly a matter of being able to attune oneself
1197 to the specificities of a given social setting, then the fact that improvisers interact in a shared
1198 environment should play a key role in the emergence of locally shared intentions. In particular,
1199 it is likely that salient features of the improvisers' sonic environment (e.g., a clear pitch in an
1200 otherwise noisy texture or simultaneous impacts in an otherwise asynchronous sequence)
1201 provide the improvisers with the opportunity to adopt similar local goals (such as changing the
1202 musical direction, performing a collective crescendo or accelerando, developing a given idea,
1203 or ending the performance). And similarly, it is likely that the improvisers' active engagement
1204 in embodied interactions – the fact that they could continuously feel each other's actions and
1205 reactions on a fine-grained scale – played a significant role in the remarkable understanding of
1206 each other's intentions they displayed (Michael, 2011). In that sense, emphasizing the
1207 supporting role of shared intentions in an explanation of coordination in complex improvised
1208 actions does not necessarily undermine the role played by interactional or contextual factors;
1209 on the contrary, it is precisely *because* collective improvisations are embodied and embedded
1210 interactions – because improvisers both co-construct and explore their shared sonic
1211 environment *through* their bodily interactions – that local shared intentions can emerge.

1212 An important question is whether and how our findings may generalize to other forms
1213 of collective improvisations. Here, we used CFI as a paradigm for studying joint improvised
1214 action, but we should emphasize that every instance of collective improvisation is not akin to
1215 CFI, since collective improvisations greatly vary on at least three dimensions. First,
1216 improvisation comes in degrees (Nettl, 1974): some collective improvisations are highly

1217 unplanned, others only allow for circumscribed spontaneous decisions within a more or less
1218 loose script. In those latter cases, the role of local shared intentions may be less crucial, as
1219 coordination is then typically supported by a broad script that is common knowledge among
1220 improvisers (think of the role played by standards such as *My Funny Valentine* in jazz
1221 improvisation). Second, some collective improvisations aim at creative and unprecedented
1222 results, while others are more concerned with efficiency in spontaneously achieving a clear goal
1223 (e.g. unarming a terrorist). Again, it is likely that local shared intentions are especially important
1224 in the first case, as they can be seen as compensating the absence of a clear overarching goal.
1225 Third, collective improvisations differ in terms of the medium in which the interaction between
1226 the agents take place. Here, it is obvious that the specificities of our musical paradigm impacted
1227 the resources that our participants were able to use to communicate with each other, and more
1228 generally, the processes through which shared intentions could emerge. But importantly, it did
1229 so mainly by *depriving* them of key coordination resources, most notably verbal communication
1230 (which facilitates the spread of local shared intentions within the group and the emergence of
1231 common knowledge) and physical co-localization (which facilitate the triggering of joint
1232 attention, joint affordances, and more generally emergent coordination mechanisms).
1233 Musicians thus had to rely on resources that were both more abstract and more indeterminate.
1234 If locally shared intentions could emerge to support the improvisers' coordination in such bare-
1235 bone situations, then there is no reason to think that they would not in "richer", more favorable
1236 contexts, in which improvisers are also engaged in highly unplanned and creative joint actions,
1237 but have in addition access to verbal communication and are co-located in the same physical
1238 environment. While the overall context in which the collective improvisation takes place, the
1239 nature of the improvisers' shared environment (sonic, audio-visual, or haptic?), the structure of
1240 the interactions (organized in turn-takings or simultaneous?) and the modes of communication
1241 (non-verbal or verbal?) necessarily impact how improvisers coordinate, we believe that our core

1242 finding that locally shared intentions can support improvisers' coordination should extend to
1243 other kinds of complex improvised joint actions.

1244 In particular, the ending goals we studied here are paradigmatic of the kind of local,
1245 shared intentions that are likely to emerge in complex and temporally extended joint
1246 improvisations – intentions that are abstract enough to be plausibly shared by several
1247 improvisers at a given point of the joint improvised action, while still retaining enough
1248 specificity to constrain the temporal and interactional dynamics at play. For example, the
1249 spontaneous tactics in which team players engage in collective sports such as basketball
1250 (Bourbousson, Poizat, Saury, & Seve, 2010) might be precisely analyzed in terms of the
1251 emergence of such local, shared intentions (e.g., preparing a shooting possibility for the team),
1252 beyond the primary, overarching shared goal of scoring baskets (Steiner, Macquet, & Seiler,
1253 2017). Local shared intentions may also explain temporal coordination (i.e., the smooth
1254 switching between speaker and listener roles) (Corps, Gambi, & Pickering, 2018), and content-
1255 based coordination (i.e., negotiating the actual question under discussion) during open-ended
1256 conversations (Beaver, Roberts, Simons, & Tonhauser, 2017). Because these shared intentions
1257 do not specify the details of the improvisers' contribution, they are likely to allow partners to
1258 act with the high degree of flexibility required by the unpredictable dynamics of an improvised
1259 interaction, while maintaining a minimal level of precision in the agents' coordination by
1260 providing them with a shared directionality (e.g., continuing or changing). In that sense, shared
1261 intentions are perhaps especially important to facilitate coordination when joint outcomes are
1262 underdetermined. To further test this idea, future work could manipulate joint outcomes'
1263 determinacy and measure the rate and level of abstraction of the shared intentions that emerge
1264 in these situations. Another important venue for future research will be to apply our design to
1265 other forms of collective improvisations (e.g., open-ended conversations), and to precisely
1266 examine how shared intentions may emerge from non-verbal (musical) interactions. Finally,

1267 our method makes it possible to ask whether coordination can occur at all when partners hold
1268 incongruent intentions simultaneously (e.g., what happens when some improvisers want to
1269 change the music while others wish to maintain the music?).

1270 Improvisation has once been defined as the “*coordination and concatenation of actions*
1271 *over time by means other than planning*” (Preston, 2013, p. 63). At core, improvisation is the
1272 way we have of navigating our social lives when we cannot or do not want to engage in
1273 extensive planning. But this does not mean that improvisers are locked in an eternal present,
1274 only able to blindly interact without any foresight of what is to come next. By highlighting the
1275 role played by shared intentions in joint improvised actions, our study opens up new avenues
1276 to explore the many ways we have to engage with the future while acting jointly.

1277

1278 ***Acknowledgments.*** This work was funded by ANR MICA (ANR-17-CE27-0021, to C.C.),
1279 ERC StG CREAM 335536 (to J.J.A.), a H2020-MSCA-IF-2018 grant (JDIL-845859 to L.G.),
1280 and partially funded by ERC SOMICS 609819, and ERC JAXPERTISE 616072 and the Central
1281 European University Foundation, Budapest (CEUBPF; partially funding T.W.). The theses
1282 explained herein are representing the own ideas of the authors, and do not necessarily reflect
1283 the opinion of CEUBPF. The authors thank the musicians and sound engineers for participating
1284 in, and recording the music involved in study 1 and 2 at the Aeronef Studio, Paris, France.
1285 Ethical approval was obtained, and experimental data for study 3 and 4 were collected at
1286 INSEAD/ Sorbonne University Center for Behavioural Science, Paris, France.

1287

1288 ***Authors contributions.*** C.C., T.W., L.G., and J.-J.A. designed the experiment. L.G., T.W., and
1289 C.C., collected the data. L.G. and T.W. analyzed the data. L.G. and C.C. wrote the paper with
1290 comments from P.S.-G., T.W. and J.-J.A.

1291 The authors declare that there is no conflict of interest regarding the publication of this article.

1292 **References**

- 1293 Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor
1294 resonance in elite basketball players. *Nature Neuroscience*.
1295 <https://doi.org/10.1038/nn.2182>
- 1296 Arthurs, T. (2016). *Secret gardeners: An ethnography of improvised music in Berlin (2012-*
1297 *13)*. University of Edinburgh.
- 1298 Aucouturier, J. J., & Canonne, C. (2017). Musical friends and foes: The social cognition of
1299 affiliation and control in improvised interactions. *Cognition*, *161*, 94–108.
- 1300 Bailey, D. (1992). *Improvisation: Its Nature and Practice in Music*. Da Capo Press.
- 1301 Beaver, D. I., Roberts, C., Simons, M., & Tonhauser, J. (2017). Questions Under Discussion:
1302 Where Information Structure Meets Projective Content. *Annual Review of Linguistics*.
1303 <https://doi.org/10.1146/annurev-linguistics-011516-033952>
- 1304 Bishop, L., & Goebel, W. (2014). Context-specific effects of musical expertise on audiovisual
1305 integration. *Frontiers in Psychology*, *5*, 1123. <https://doi.org/10.3389/fpsyg.2014.01123>
- 1306 Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the
1307 harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic*
1308 *Sciences*.
- 1309 Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*.
- 1310 Borgo, D. (2005). *Sync or swarm: Improvising music in a complex age*. A&C Black.
- 1311 Bourbousson, J., Poizat, G., Saury, J., & Seve, C. (2010). Team coordination in basketball:
1312 Description of the cognitive connections among teammates. *Journal of Applied Sport*
1313 *Psychology*. <https://doi.org/10.1080/10413201003664657>
- 1314 Bratman, M. E. (1999). Shared Intention. In *Faces of Intention: selected essays on intention*
1315 *and agency* (pp. 109–129).
- 1316 Bratman, M. E. (2014). *Shared agency: A planning theory of acting together*. Oxford

- 1317 University Press.
- 1318 Butterfill, S. A. (2013). Interacting mindreaders. *Philosophical Studies*, 165(3), 841–863.
- 1319 <https://doi.org/10.1007/s11098-012-9980-x>
- 1320 Butterfill, S. A. (2018). Coordinating Joint Action. In *The Routledge Handbook of Collective*
- 1321 *Intentionality*. <https://doi.org/10.4324/9781315768571-8>
- 1322 Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind*
- 1323 *and Language*, 28(5), 606–637. <https://doi.org/10.1111/mila.12036>
- 1324 Canonne, C. (2013). Focal Points in Collective Free Improvisation. *Perspectives of New*
- 1325 *Music*. <https://doi.org/10.7757/persnewmusi.51.1.0040>
- 1326 Canonne, C. (2018). Rehearsing free improvisation? An ethnographic study of free
- 1327 improvisers at work. *Music Theory Online*. <https://doi.org/10.30535/mto.24.4.1>
- 1328 Canonne, C., & Garnier, N. (2012). Cognition and segmentation in collective free
- 1329 improvisation: An exploratory study. In *Proceedings of the 12th international conference*
- 1330 *on music perception and cognition and 8th triennial conference of the european society*
- 1331 *for the cognitive sciences of music* (pp. 197–204).
- 1332 Carré, A., Stefaniak, N., D’Ambrosio, F., Bensalah, L., & Besche-Richard, C. (2013). The
- 1333 basic empathy scale in adults (BES-A): Factor structure of a revised form. *Psychological*
- 1334 *Assessment*.
- 1335 Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating Utterances During Turn-
- 1336 Taking: The Role of Prediction, Response Preparation, and Articulation. *Discourse*
- 1337 *Processes*. <https://doi.org/10.1080/0163853X.2017.1330031>
- 1338 D’Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., ... Fadiga, L. (2012).
- 1339 Leadership in Orchestra Emerges from the Causal Relationships of Movement
- 1340 Kinematics. *PLoS ONE*, 7(5), e35757. <https://doi.org/10.1371/journal.pone.0035757>
- 1341 D’Ausilio, A., Novembre, G., Fadiga, L., & Keller, P. E. (2015). What can music tell us about

1342 social interaction? *Trends in Cognitive Sciences*.

1343 <https://doi.org/10.1016/j.tics.2015.01.005>

1344 De Freitas, J., Thomas, K., DeScioli, P., & Pinker, S. (2019). Common knowledge,
1345 coordination, and strategic mentalizing in human social life. *Proceedings of the National*
1346 *Academy of Sciences of the United States of America*, 201905518.

1347 <https://doi.org/10.1073/pnas.1905518116>

1348 De Spain, K. (2014). *Landscape of the Now: a Topography of Movement Improvisation*.
1349 Oxford University Press.

1350 della Gatta, F., Garbarini, F., Rabuffetti, M., Viganò, L., Butterfill, S. A., & Sinigaglia, C.
1351 (2017). Drawn together: When motor representations ground joint actions. *Cognition*,
1352 165, 53–60. <https://doi.org/10.1016/J.COGNITION.2017.04.008>

1353 Denzler, B., & Guionnet, J.-L. (2020). *The Practice of Musical Improvisation: dialogues with*
1354 *contemporary musical improvisers*. New York: Bloomsbury Academic.

1355 Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical*
1356 *models. Policy Analysis*. Cambridge University Press.

1357 Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York
1358 Wiley (Vol. 4054).

1359 Gueguen, N., Jacob, C., & Martin, A. (2009). Mimicry in social interaction: Its effect on
1360 human judgment and behavior. *European Journal of Social Sciences*.

1361 Hadley, L. V., Sturt, P., Moran, N., & Pickering, M. J. (2018). Determining the end of a
1362 musical turn: Effects of tonal cues. *Acta Psychologica*.

1363 <https://doi.org/10.1016/j.actpsy.2017.11.001>

1364 Heggli, O. A., Konvalinka, I., Kringelbach, M. L., & Vuust, P. (2019). Musical interaction is
1365 influenced by underlying predictive models and musical expertise. *Scientific Reports*,
1366 9(1). <https://doi.org/10.1038/s41598-019-47471-3>

- 1367 Helm, J. L., Miller, J. G., Kahle, S., Troxel, N. R., & Hastings, P. D. (2018). On Measuring
1368 and Modeling Physiological Synchrony in Dyads. *Multivariate Behavioral Research*.
1369 <https://doi.org/10.1080/00273171.2018.1459292>
- 1370 Ingold, T., & Hallam, E. (2007). Creativity and Cultural Improvisation: An Introduction. In E.
1371 Hallam & T. Ingold (Eds.), *Creativity and Cultural Improvisation* (pp. 1–24). Berg.
1372 <https://doi.org/10.1017/S1537781415000316>
- 1373 Issartel, J., Marin, L., & Cadopi, M. (2007). Unintended interpersonal co-ordination: “can we
1374 march to the beat of our own drum?” *Neuroscience Letters*, *411*(3), 174–179.
1375 <https://doi.org/10.1016/J.NEULET.2006.09.086>
- 1376 Keller, P. E. (2008). Joint action in music performance. *Emerging Communication: Studies in*
1377 *New Technologies and Practices in Communication*.
- 1378 Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression.
1379 In *Expressiveness in music performance: Empirical approaches across styles and*
1380 *cultures*. <https://doi.org/10.1093/acprof:oso/9780199659647.001.0001>
- 1381 Kirschner, S., & Tomasello, M. (2010). Joint music making promotes prosocial behavior in 4-
1382 year-old children. *Evolution and Human Behavior*.
1383 <https://doi.org/10.1016/j.evolhumbehav.2010.04.004>
- 1384 Knoblich, G., Butterfill, S., & Sebanz, N. (2011). *Psychological Research on Joint Action.*
1385 *Theory and Data. Psychology of Learning and Motivation - Advances in Research and*
1386 *Theory*. <https://doi.org/10.1016/B978-0-12-385527-5.00003-6>
- 1387 Kourtis, D., Woźniak, M., Sebanz, N., & Knoblich, G. (2019). Evidence for we-
1388 representations during joint action planning. *Neuropsychologia*, *131*, 73–83.
1389 <https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2019.05.029>
- 1390 Kuznetsova, A., Brockhoff, P. B., & Christensen, H. B. (2014). lmerTest: Tests for random
1391 and fixed effects for linear mixed effect models (lmer objects of lme4 package). *R*.

- 1392 Linson, A., & Clarke, E. F. (2018). Distributed cognition, ecological theory, and group
1393 improvisation. In E. Clarke & M. Doffman (Eds.), *Distributed creativity: Collaboration*
1394 *and improvisation in contemporary music* (pp. 52–69). Oxford, UK: Oxford University
1395 Press.
- 1396 Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring
1397 Individual and Joint Action Outcomes in Duet Music Performance. *Journal of Cognitive*
1398 *Neuroscience*, 25(7), 1049–1061. https://doi.org/10.1162/jocn_a_00388
- 1399 MacDonald, R. A. R., & Wilson, G. B. (2020). *The Art of Becoming: How Group*
1400 *Improvisation Works*. OUP USA.
- 1401 Mendonça, D. J., & Wallace, W. Al. (2007). A cognitive model of improvisation in
1402 emergency management. *IEEE Transactions on Systems, Man, and Cybernetics Part*
1403 *A: Systems and Humans*. <https://doi.org/10.1109/TSMCA.2007.897581>
- 1404 Michael, J. (2011). Interactionism and Mindreading. *Review of Philosophy and Psychology*,
1405 2(3), 559. <https://doi.org/10.1007/s13164-011-0066-z>
- 1406 Michael, J. (2017). Music Performance as Joint Action.pdf. In *The Routledge Companion to*
1407 *Embodied Music Interaction*.
- 1408 Moran, N., Hadley, L. V., Bader, M., & Keller, P. E. (2015). Perception of “back-channeling”
1409 nonverbal feedback in musical duo improvisation. *PLoS ONE*.
1410 <https://doi.org/10.1371/journal.pone.0130070>
- 1411 Nessler, J. A., & Gilliland, S. J. (2009). Interpersonal synchronization during side by side
1412 treadmill walking is influenced by leg length differential and altered sensory feedback.
1413 *Human Movement Science*, 28(6), 772–785.
1414 <https://doi.org/10.1016/J.HUMOV.2009.04.007>
- 1415 Nettl, B. (1974). Thoughts on improvisation: A comparative approach. *The Musical Quarterly*,
1416 60(1), 1–19.

- 1417 Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2014). Motor simulation
1418 and the coordination of self and other in real-time joint action. *Social Cognitive and*
1419 *Affective Neuroscience*, 9(8), 1062–1068. <https://doi.org/10.1093/scan/nst086>
- 1420 Noy, L., Dekel, E., & Alon, U. (2011). The mirror game as a paradigm for studying the
1421 dynamics of two people improvising motion together. *Proceedings of the National*
1422 *Academy of Sciences of the United States of America*, 108(52), 20947–20952.
1423 <https://doi.org/10.1073/pnas.1108155108>
- 1424 Pachet, F., Roy, P., & Foulon, R. (2017). Do Jazz Improvisers Really Interact? In *The*
1425 *Routledge Companion to Embodied Music Interaction* (pp. 167–176). New York ;
1426 London : Routledge, 2017.: Routledge. <https://doi.org/10.4324/9781315621364-19>
- 1427 Papiotis, P., Marchini, M., & Maestre, E. (2012). Computational Analysis of Solo Versus
1428 Ensemble Performance in String Quartets: Intonation and DynamicsNo Title. In
1429 *Proceedings of the 12th International Conference on Music Perception and Cognition*.
- 1430 Pelz-Sherman, M. (1998). *A Framework for the Analysis of Performer Interactions in Western*
1431 *Improvised Contemporary Art Music*. University of California.
- 1432 Pressing, J. (1984). Cognitive Processes in Improvisation. *Advances in Psychology*.
1433 [https://doi.org/10.1016/S0166-4115\(08\)62358-4](https://doi.org/10.1016/S0166-4115(08)62358-4)
- 1434 Preston, B. (2013). *A philosophy of material culture: Action, function, and mind*. Routledge.
1435 <https://doi.org/10.4324/9780203069844>
- 1436 Repp, B. H. (2005). Sensorimotor synchronization: A review of the tapping literature.
1437 *Psychonomic Bulletin and Review*. <https://doi.org/10.3758/BF03206433>
- 1438 Sacheli, L. M., Arcangeli, E., & Paulesu, E. (2018). Evidence for a dyadic motor plan in joint
1439 action. *Scientific Reports*. <https://doi.org/10.1038/s41598-018-23275-9>
- 1440 Savouret, A. (2010). Introduction à un solfège de l’audible. L’improvisation libre comme
1441 outil pratique. *Symétrie*.

- 1442 Sawyer, R. K. (2003). Group Creativity: Music. In *Theater, collaboration*. Erlbaum.
- 1443 Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of interpersonal coordination. In
1444 *Coordination: Neural, behavioural and social dynamics* (pp. 281–308). Springer Berlin
1445 Heidelberg. https://doi.org/10.1007/978-3-540-74479-5_14
- 1446 Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. S. (2009). Signalling signalhood and the
1447 emergence of communication. *Cognition*.
1448 <https://doi.org/10.1016/j.cognition.2009.08.009>
- 1449 Sethares, W. A. (1993). Local consonance and the relationship between timbre and scale.
1450 *Journal of the Acoustical Society of America*. <https://doi.org/10.1121/1.408175>
- 1451 Steiner, S., Macquet, A. C., & Seiler, R. (2017). An integrative perspective on interpersonal
1452 coordination in interactive team sports. *Frontiers in Psychology*.
1453 <https://doi.org/10.3389/fpsyg.2017.01440>
- 1454 Van Baaren, R., Janssen, L., Chartrand, T. L., & Dijksterhuis, A. (2009). Where is the love?
1455 The social aspects of mimicry. *Philosophical Transactions of the Royal Society B:*
1456 *Biological Sciences*. <https://doi.org/10.1098/rstb.2009.0057>
- 1457 Vesper, C., Abramova, E., Bütelage, J., Ciardo, F., Crossey, B., Effenberg, A., ... Wahn, B.
1458 (2017). Joint Action: Mental Representations, Shared Information and General
1459 Mechanisms for Coordinating with Others. *Frontiers in Psychology*, 07, 2039.
1460 <https://doi.org/10.3389/fpsyg.2016.02039>
- 1461 Vesper, C., Butterfill, S., Knoblich, G., & Sebanz, N. (2010). A minimal architecture for joint
1462 action. *Neural Networks*, 23(8–9), 998–1003.
1463 <https://doi.org/10.1016/J.NEUNET.2010.06.002>
- 1464 Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2011). Making oneself
1465 predictable: reduced temporal variability facilitates joint action coordination.
1466 *Experimental Brain Research*, 211(3–4), 517–530. <https://doi.org/10.1007/s00221-011->

1467 2706-z

1468 Vesper, C., van der Wel, R. P. R. D., Knoblich, G., & Sebanz, N. (2013). Are you ready to
1469 jump? Predictive mechanisms in interpersonal coordination. *Journal of Experimental*
1470 *Psychology: Human Perception and Performance*. <https://doi.org/10.1037/a0028066>

1471 Walsh, M., Roberts, I., & Besser, M. (2013). *Upright Citizens Brigade Comedy Improvisation*
1472 *Manual*. Comedy Council of Nicea, LLC.

1473 Walton, A. E., Washburn, A., Langland-Hassan, P., Chemero, A., Kloos, H., & Richardson,
1474 M. J. (2018). Creating Time: Social Collaboration in Music Improvisation. *Topics in*
1475 *Cognitive Science*, 10(1), 95–119. <https://doi.org/10.1111/tops.12306>

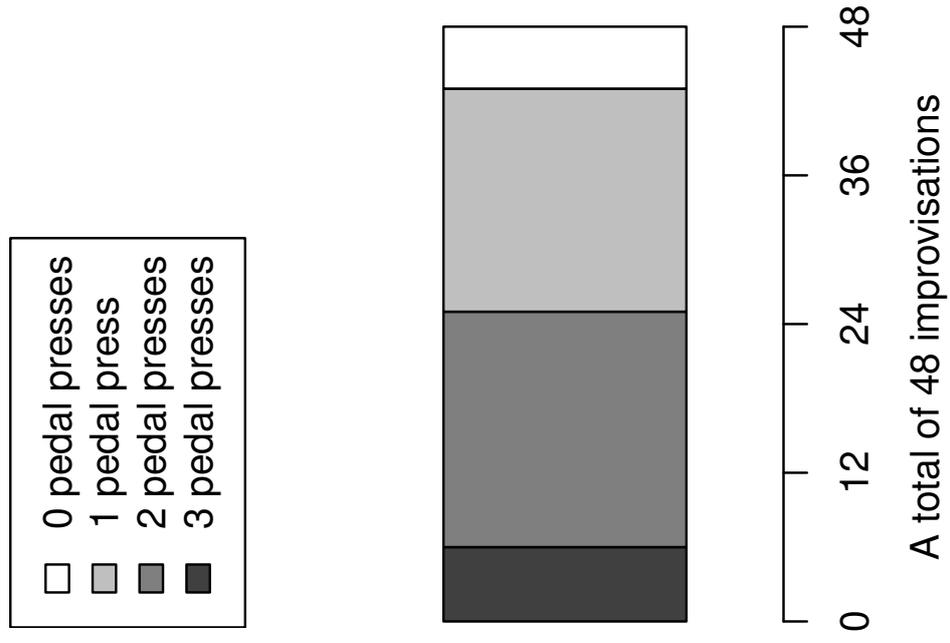
1476 Wass, S. V., Whitehorn, M., Marriott Haresign, I., Phillips, E., & Leong, V. (2020).
1477 Interpersonal Neural Entrainment during Early Social Interaction. *Trends in Cognitive*
1478 *Sciences*. <https://doi.org/10.1016/j.tics.2020.01.006>

1479 Yun, K., Watanabe, K., & Shimojo, S. (2012). Interpersonal body and neural synchronization
1480 as a marker of implicit social interaction. *Scientific Reports*.
1481 <https://doi.org/10.1038/srep00959>

1482

Figure 1

A



B

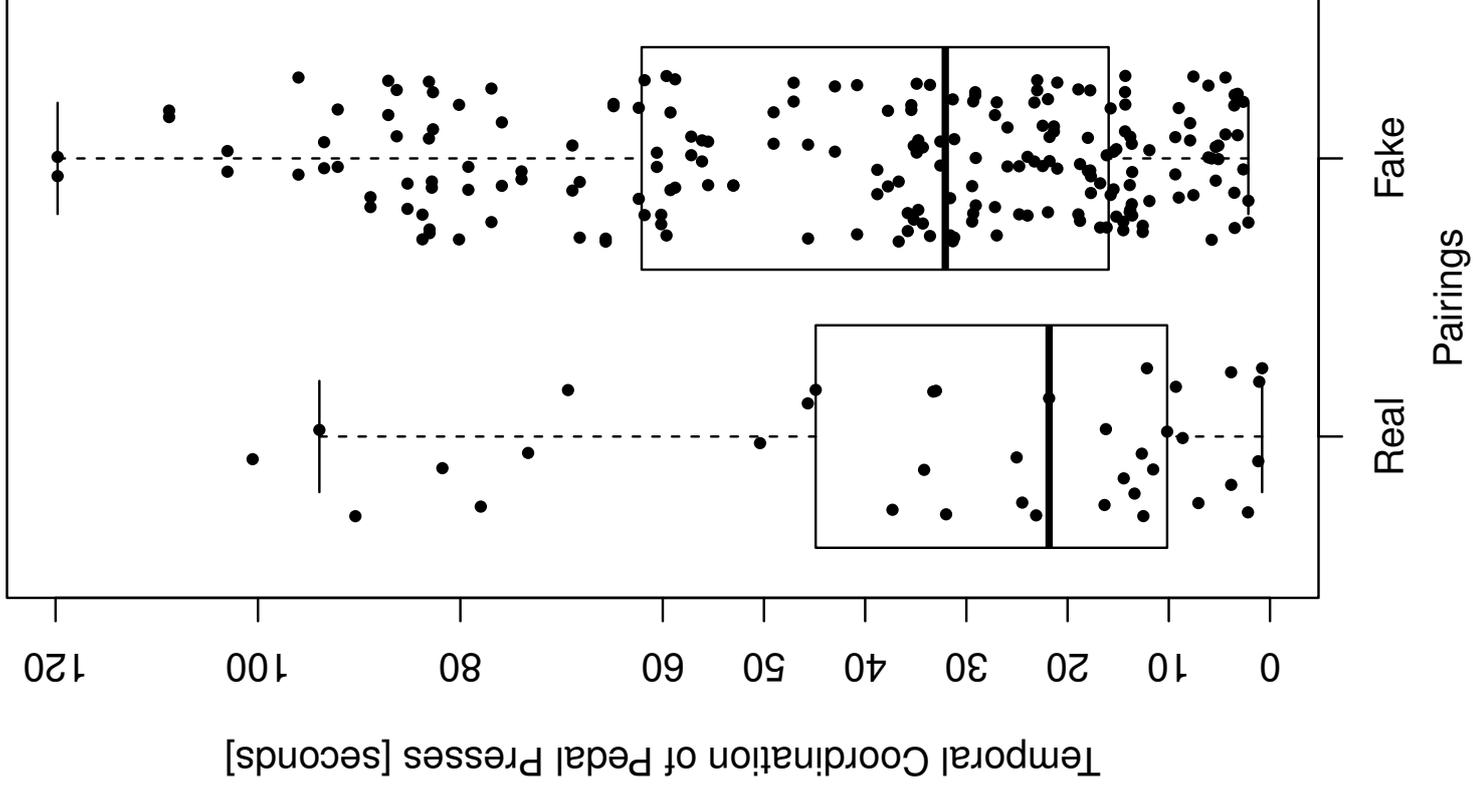


Figure 2

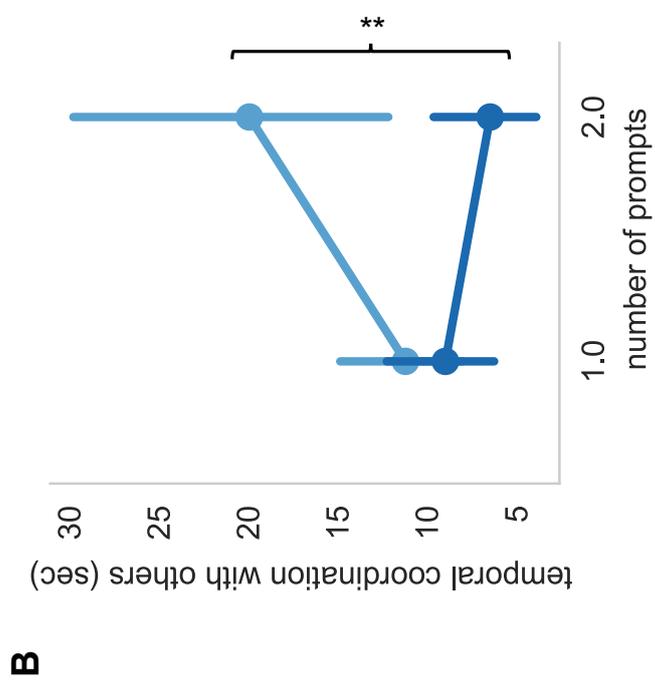
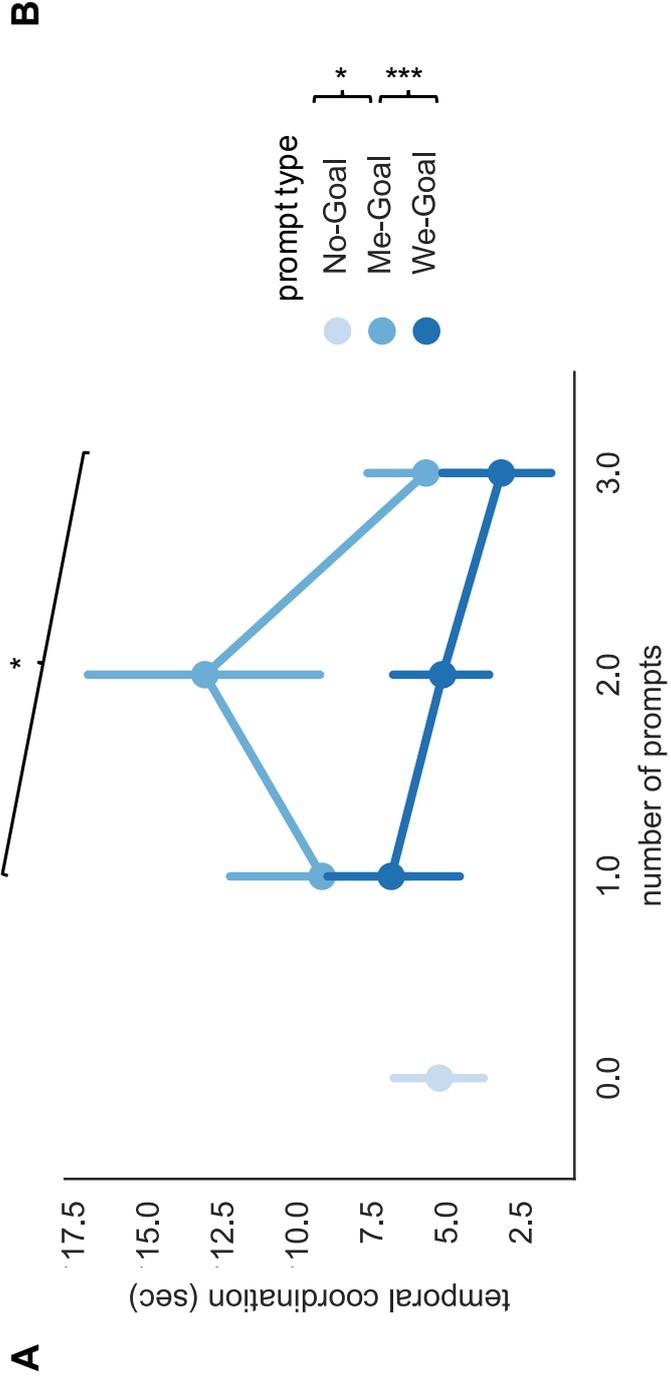


Figure 3

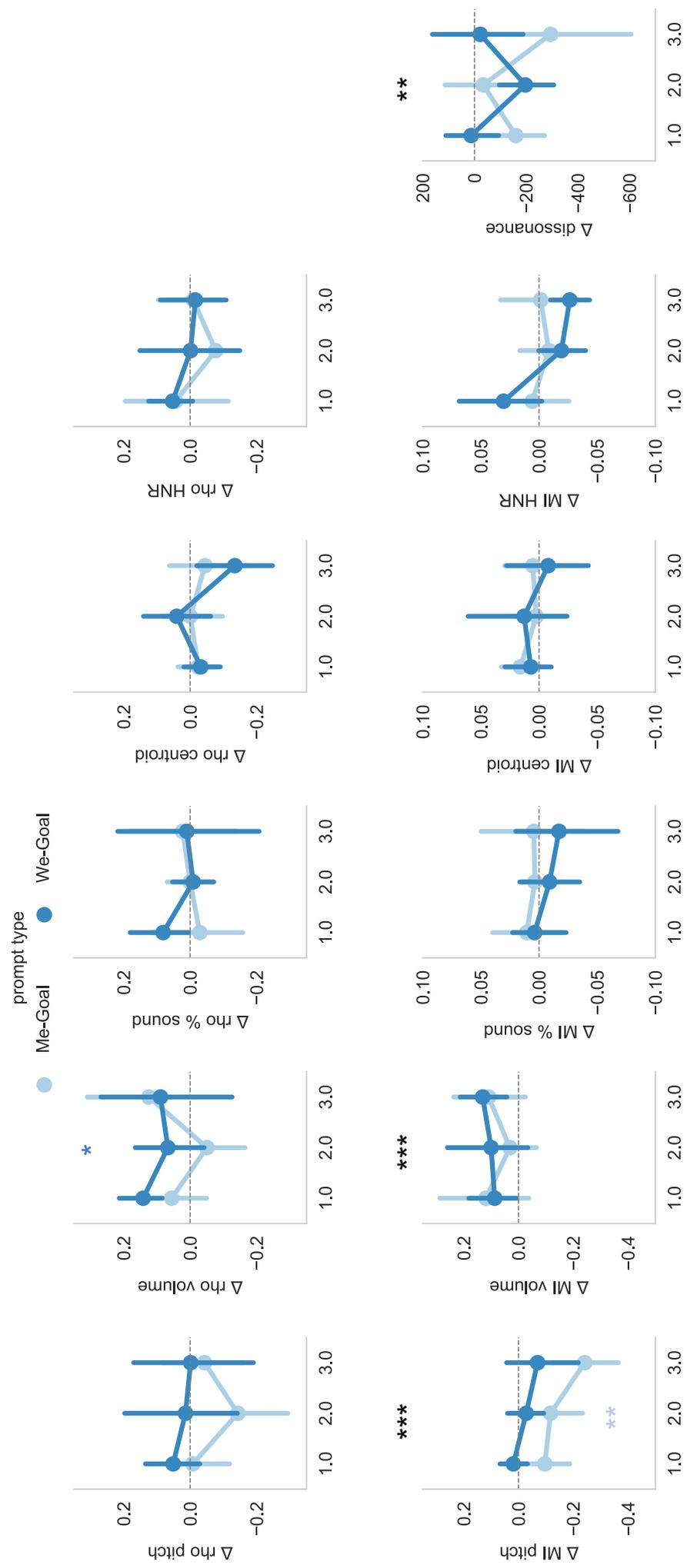


Figure 4

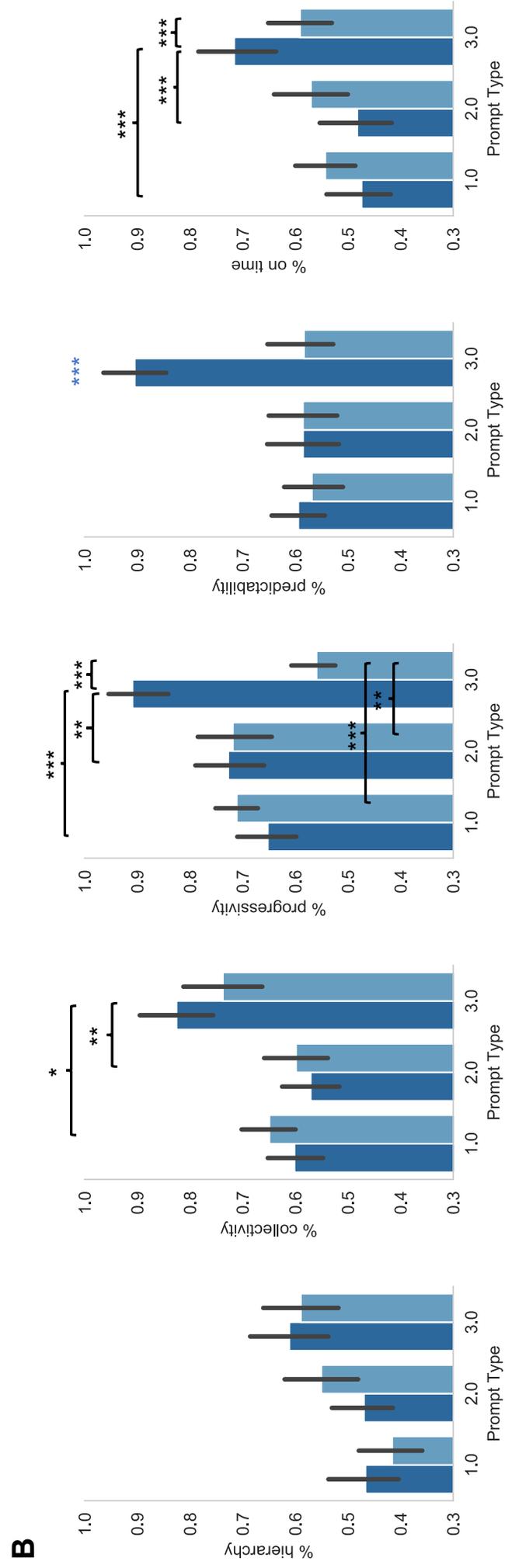
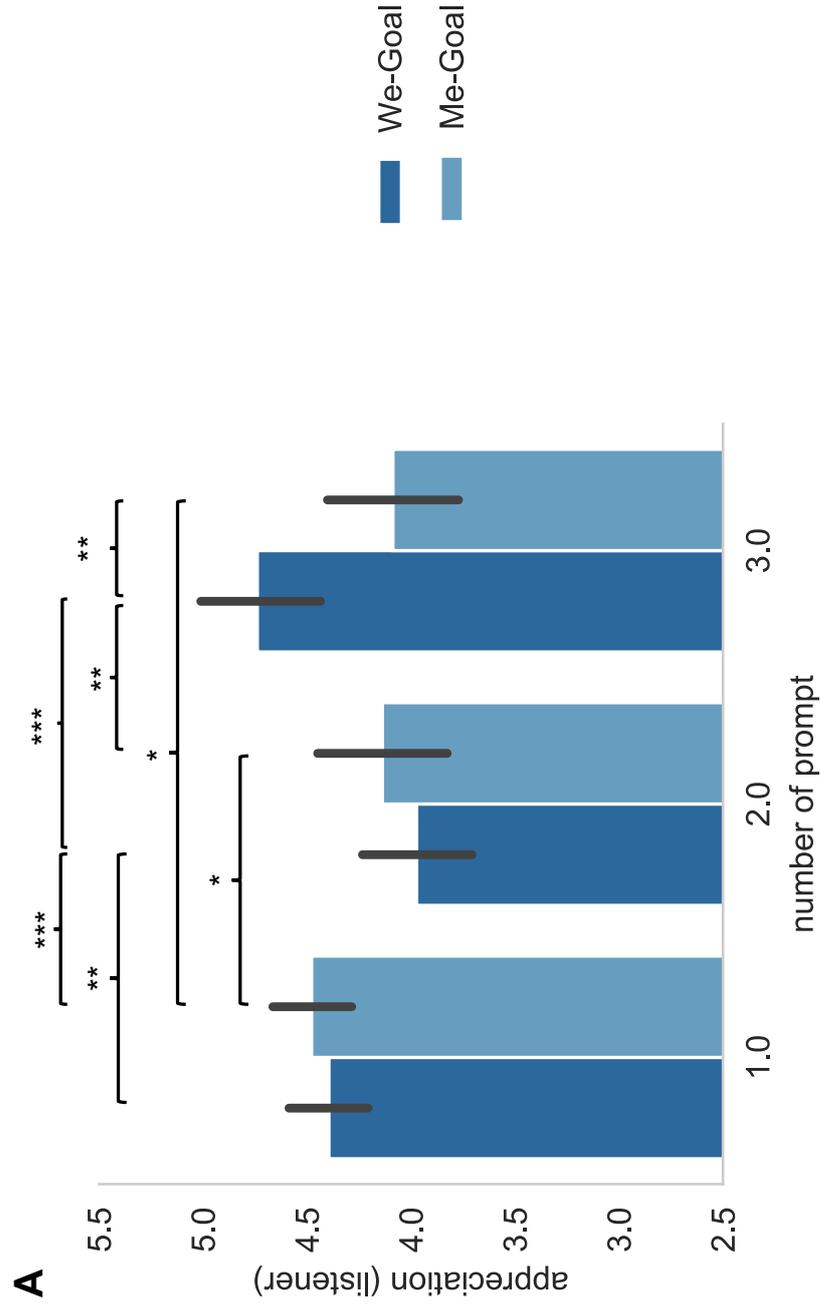


Figure 5

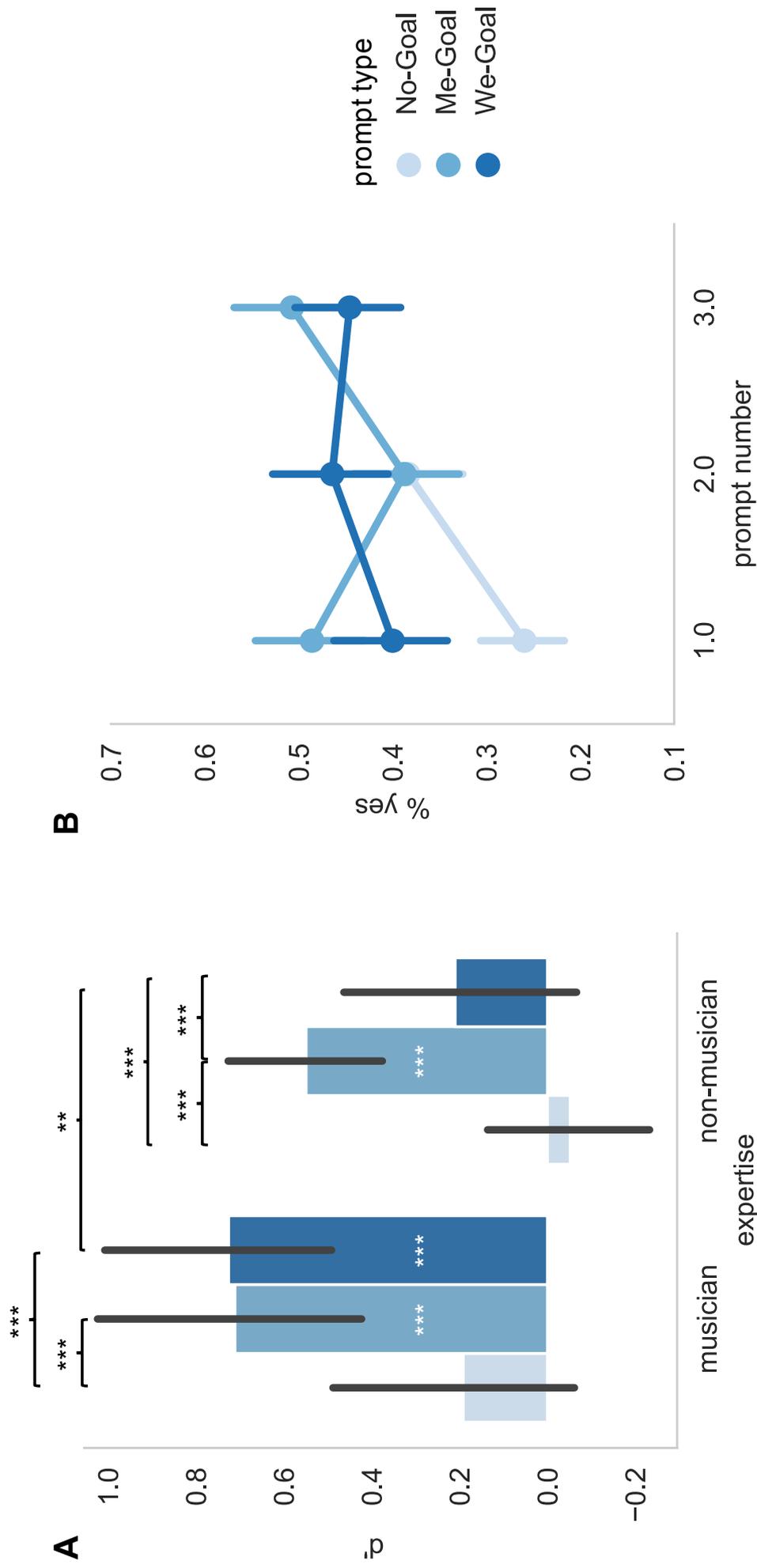
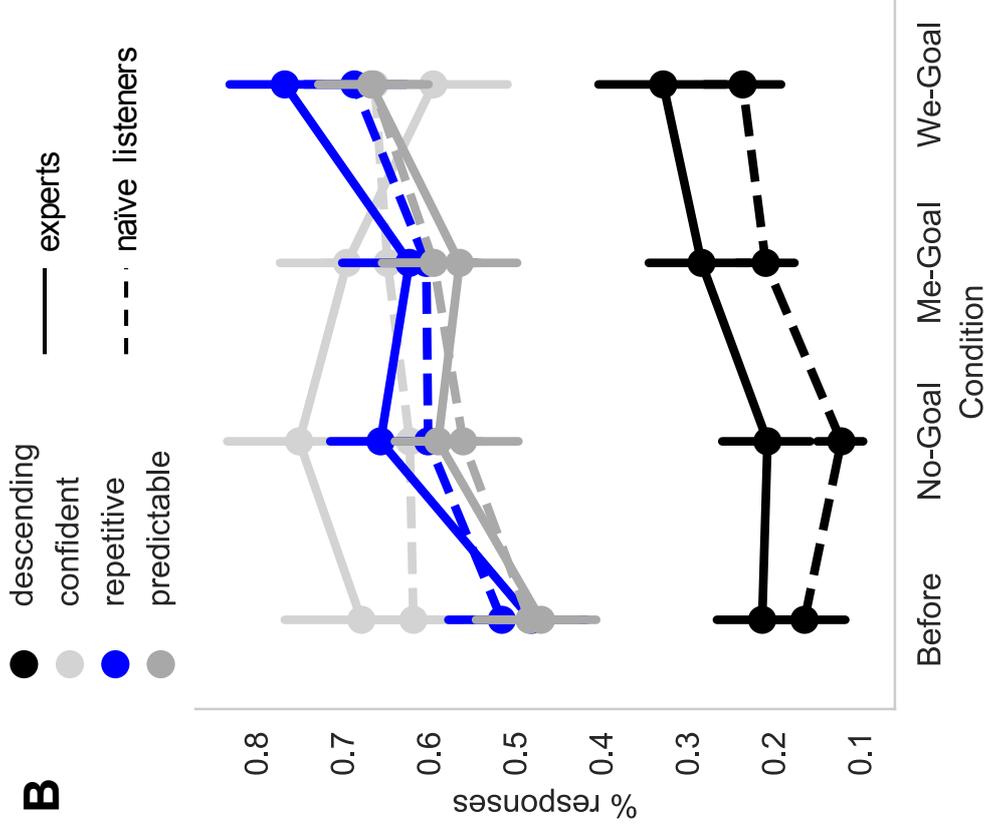
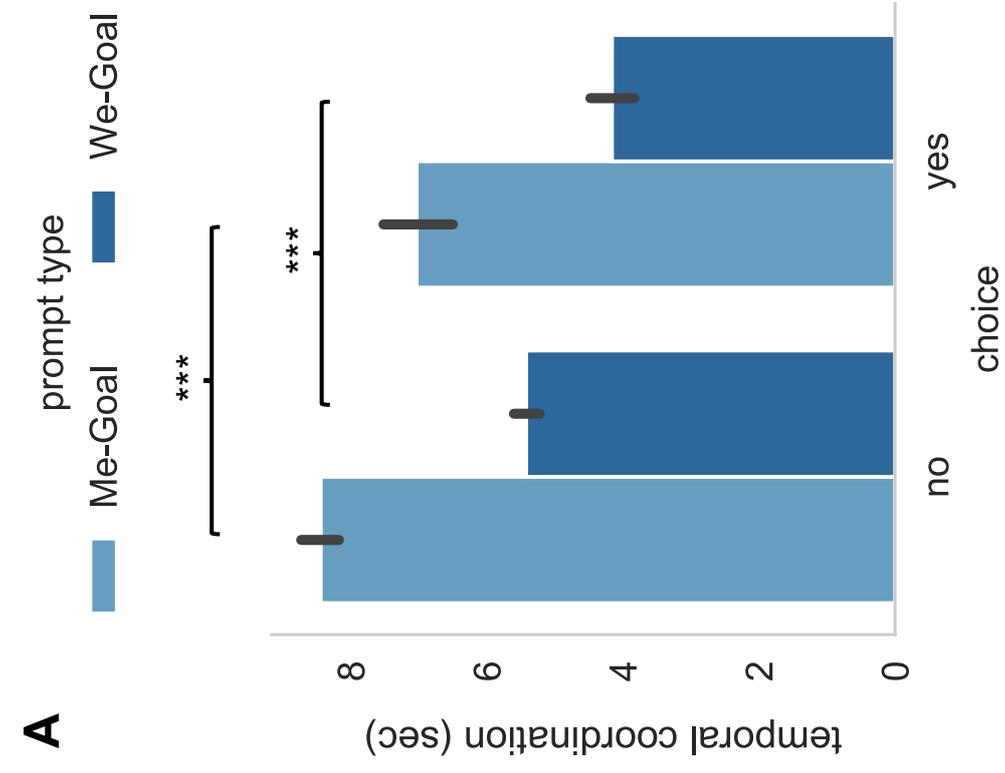


Figure 6



Supplementary Materials and Results

S.1. Experiment 1

S.1.1. Composition of the trios.

Trio n°	Instruments
1	Bass clarinet (musician 1); alto saxophone (musician 2); drums (musician 3)
2	Voice-clarinet (musician 4); prepared piano (musician 5); alto saxophone (musician 2)
3	Tenor saxophone (musician 6); Voice-clarinet (musician 4); drums (musician 3)
4	Guitar (musician 7); electronic (musician 8); drums (musician 9)
5	Drums (musician 9); prepared piano (musician 5); trumpet (musician 10)
6	Guitar (musician 7); Tenor saxophone (musician 6); drums (musician 11)
7	Electronic (musician 8); prepared piano (musician 12); Baryton saxophone / duduk (musician 13)
8	Baryton saxophone / duduk (musician 13); Alto saxophone- piano (musician 14); Alto saxophone (musician 15)
9	Alto saxophone - piano (musician 14); Double bass (musician 16); Alto saxophone (musician 17)
10	Flute (musician 18); Soprano saxophone (musician 19); Double bass (musician 20)
11	Bass clarinet (musician 1); Trumpet (musician 21); Double bass (musician 20)
12	Drums (musician 11); Alto saxophone (musician 17); Flute (musician 18)

Table S1. Composition of the trios.

S.1.2. Information used by improvisers to detect their partners' intention to end the piece

Following the improvisation, musicians were asked whether they thought that their partners had been looking for an end, and if so, why. These post-improvisation reports can provide some hints regarding which indices musicians use to attribute goals to each other during

CFI. Musicians often referenced aspects such as “descending” behaviors (e.g., “*decrease in energy*”, “*decrescendo*”, “*slowing*” ...), changes in structures (e.g., “*change in the structure of the piece*”, “*he played a conclusive note*”, “*new texture*”, “*held a long note*”...), and importantly, perception of changing intentions in others (“*more tension in the listening*”, “*he believed there was an end*”, “*he thought I was stopping*”, ...). These reports were used to construct the categories used in the fourth Experiment with independent listeners (see Glossary in section S.5.1.). Three of the authors (L.G., P.S.G., C.C.) read all of the reports, and grouped them in several subjective categories. These categories were then compared and discussed, which led to reducing them to four categories:

- *direction*, related to the directionality of the performer’s musical actions, either ascending, descending, constant or not perceptible),
- *repetition*, related to the tendency of the performer to engage in repetitive actions such as holding a note, repeating a pattern, etc, or varied actions)
- *prevision*, related to the predictability of the musician’s actions (predictable or surprising)
- *assurance*, related to the confidence with which the musician seemed to perform musical propositions.

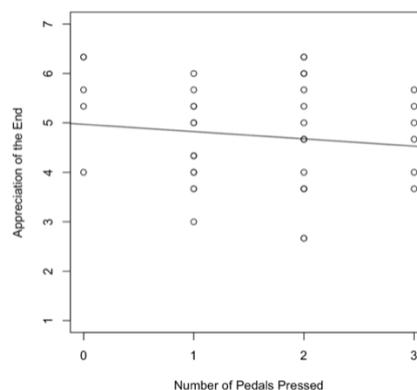
S.1.3. Musicians appreciation ratings and reports

Musicians were asked to rate the endings of their performance after each improvisation, on a scale from 1 to 7. Although we originally planned to use these ratings as a proxy for coordination success, we realized that this measure was not appropriate for several reasons. First, the debriefings we had with the musicians revealed a general discomfort with post-hoc evaluations, given the subjectivity and partiality of such evaluations, especially in a context where the variety of aesthetics and individual preference is highly valued (Bailey, 1992), and the only limited and partial access they could have to the overall result as participants deeply

immersed in the performance. Second, and most importantly, musician appreciation judgements appeared to be very influenced by the prompts here. This is clear when examining their responses to the question “why? (did you like/dislike this ending)”: out of the 432 reports, 32 (7%) explicitly comported a reference to the prompt.

Thus, probably unsurprisingly given the elements mentioned above, we did not find any significant correlations between our experimental conditions and the improvisers’ evaluations (see Fig. S4B below). However, such absence of correlation does not necessarily mean that shared intentions did not impact the success of the performances’ endings. To approach this question, we decided, in Experiment 3, to turn to third party listeners who could have an external overview on the performance and were thus in a better position to evaluate the success of the performances’ endings.

S.1.4. Musicians appreciation ratings and pedal presses



We checked that having to report their intentions to end did not significantly perturb the performance by examining whether appreciation ratings decreased with the number of pedals pressed. In a linear mixed regression, we found no significant effect of the number of pedal pressed on appreciation (beta = -0.15 +/- 0.17 se, $t = -0.86$, $p = 0.39$, suggesting that takes in which musicians pressed the pedal more often could still be enjoyed to the same extent as performances in which the pedal was not used (either because musicians did not form an intention to end, or because they forgot to report it).

S.2. Experiment 2

S.2.1. Supplementary results

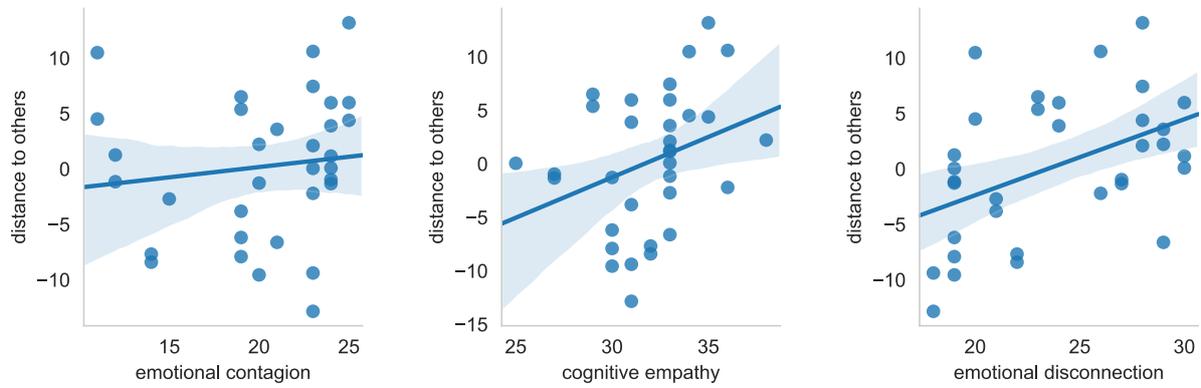


Figure S1. Experiment 2. Correlation between musicians' cognitive empathy and synchrony. Musicians filled in the BESA, an empathy questionnaire that measures three components of empathy: emotional contagion, cognitive empathy and disconnection (which involves the ability to engage cognition to regulate affects) (Carré, Stefaniak, D'Ambrosio, Bensalah, & Besche-Richard, 2013). Interestingly, we found that the two sub-scales related to cognitive empathy correlated with musicians' ability to synchronize with others (i.e., distance to others variable, see methods; cognitive empathy: Spearman's $\rho = 0.39$, $p = .023$; disconnection: $\rho = 0.48$, $p = .005$), while the emotional contagion sub-scale did not correlate with musicians' ability to synchronize with others ($\rho = 0.24$, $p > .17$). These results complement previous reports showing that empathic perspective taking promotes synchrony in non-musicians (Novembre, Mitsopoulos, & Keller, 2019).

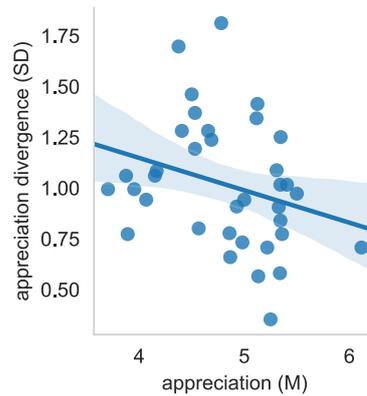


Figure S2. Experiment 2. Correlation between pairs' overall appreciation and divergence in appreciation. Judgements of appreciations were highly correlated amongst pairs of musicians (spearman's $\rho = 0.2$, $p < .001$, not shown). More interestingly, the degree of divergence in a pair's judgements of appreciations (standard deviation computed across each pair for each improvisation, averaged for each pair) was negatively correlated with its global level of appreciation (average computed across each pair for each improvisation, averaged for each pair; spearman's $\rho = -0.33$, $p = .05$). This suggests that appreciation alignment leads to better coordination. Each dot represents one of the thirty-six pairs of musicians, and the line shows the best regression fit with 95% confidence intervals.

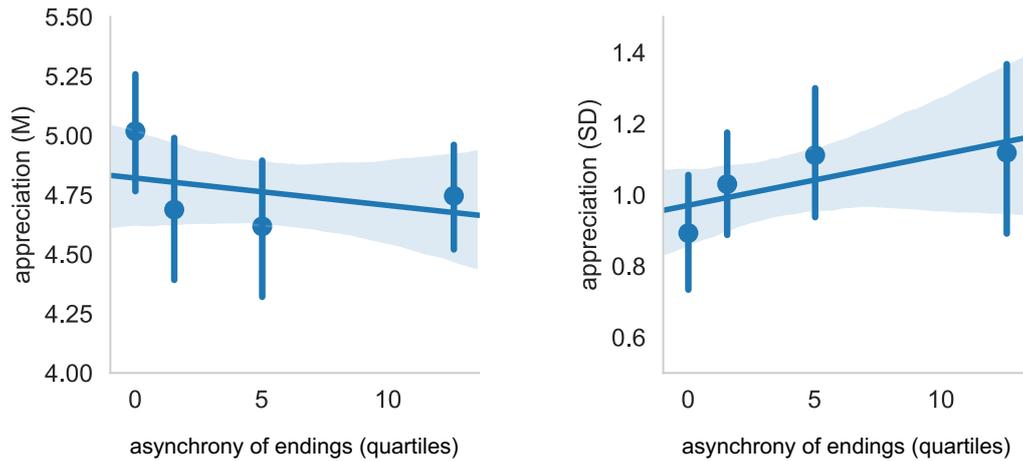


Figure S3. Experiment 2. Correlation between appreciation, asynchrony of endings and divergence in appreciation at the level of the improvisation for each pair. A) Pairs' mean appreciation reports did not systematically co-vary with asynchrony from one improvisation to the next: a hierarchical mixed linear regression (number of observations = 550, see methods for details) showed no significant effect of mean appreciation on asynchrony ($\beta = 0.33 \pm 0.62$ sem, $t = 0.5$, $p > .6$). A similar result was obtained when examining listeners' ratings: there was no significant linear relationship between appreciation and asynchrony ($\beta = 0.01 \pm 0.009$ sem, $t = 1$, $p > .28$). B) By contrast, divergence significantly related to asynchrony ($\beta = 1.9 \pm 0.8$ sem, $t = 2.4$, $p = .017$). For visualization purposes, mean appreciation (A), and standard deviation appreciation (B) were averaged in four separate bins depending on the asynchrony of endings quartiles for each pair, before being averaged across the group. The line shows the best regression fit. Shaded area and error bars show 95% confidence intervals. We present the data for pairs for precision, but similar results were obtained at the level of the trios.

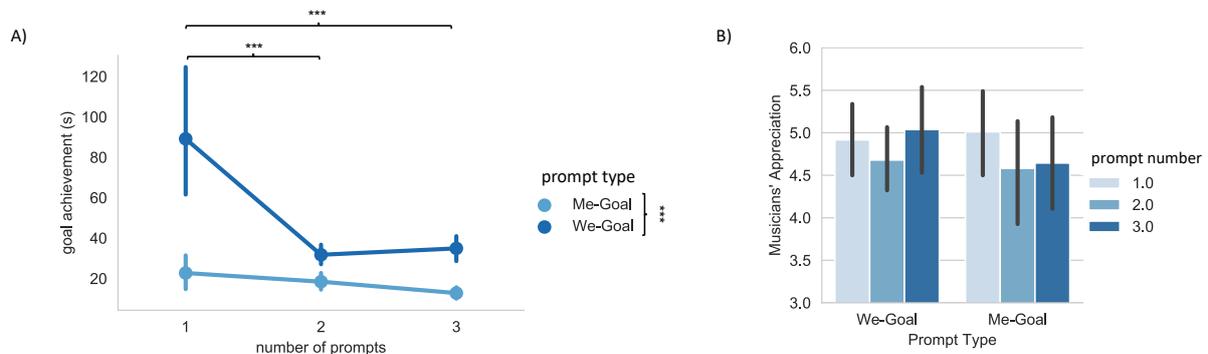


Figure S4. Experiment 2. A) Goal achievement time averaged per trio depending on prompt type and number. In the WE condition, goal achievement time was computed as the time between the end of the improvisation (latest note played) and prompt time. In the ME condition, goal achievement time corresponded to the time between the prompted musician's ending time and prompt time. A linear mixed regression with achievement time as a dependent variable, prompt type and number as independent variables, and trio as a random factor, revealed a main effect of prompt sharedness ($\chi^2 = 7.25$; $p = .027$), a main effect of prompt type ($\chi^2 = 15$; $p < .001$), and a significant interaction between the two predictors ($\chi^2 = 38.3$; $p < .001$). The main effect of prompt type reflected the fact that goal achievement time was longer in the WE as compared to the ME condition at each level of prompt number (1: $\beta = 65$, sem = 7.35, $t = 8.965$, $p < .001$; 2: $\beta = 13$, sem = 5.7, $t = 2.28$, $p = .0287$; 3: $\beta = 18.82$, sem = 5.19, $t = 3.625$, $p = .004$). The effect of prompt sharedness and the interaction reflected the fact that, although in the ME condition there was no impact of prompt number on goal achievement time (all p -values $> .2$), in the WE condition goal achievement time was significantly longer when only one of the performers was prompted as compared to the other conditions (1 vs. 2: $\beta = 56$, sem = 10, $t = 5.4$, $p < .001$; 1 vs. 3: $\beta = 54.8$, sem = 9.82, $t = 5.58$, $p < .001$). B) Self-appreciation ratings. There was no effect of Prompt Type or Number and no interaction (all p -values $> .16$).

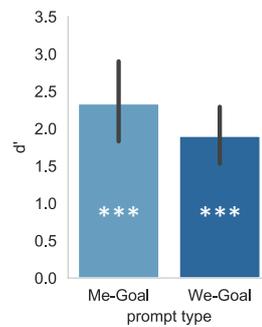


Figure S5. Goal transparency (Experiment 2). After each improvisation we asked performers to guess whether they thought their partners had received a prompt. Performers' ability to guess their partners prompts was assessed with signal detection theory by computing a d' (Green & Swets, 1966) for each participant and condition from the hit rate (number of prompts reported for prompted trials / total number of prompts) and false alarm rate (number of prompts reported for un-prompted trials / total number of un-prompted trials). Musicians' sensitivity in guessing their partners' goals significantly differed from chance level for both ME (mean $d' = 2.33 \pm 1.56$ SD, $t(31) = 8.17$, $p < .001$) and WE-Goals (mean $d' = 1.9 \pm 1$ SD, $t(31) = 9.81$, $p < .001$) condition, and there was no difference in performance between the two prompt types ($t(27) = 1$, $p = .31$). Trials where musicians were prompted were excluded from this analysis. The difference in performances in between the two conditions could only be computed for 28 musicians out of 36 (the three musicians of trio 1 did not have data for the WE condition, and five musicians did not respond to the mindreading question for one or the other condition). Thus, it appeared that goals were transparent for the other members of the group. Even though these estimations were made post-hoc, and may thus rely on offline reasoning, such goal transparency may form the basis of a mechanism through which goals propagate within a group. White asterisks show the significance of one-sample t-tests against chance.

S.3. Experiment 3

S.3.1. Glossary

Listeners were provided a glossary to define the categories along which they were asked to rate the endings as follows:

Collective: Musicians seem to agree about how they should end the improvisation.

Disjoint: Musicians seem to disagree about how they should end the improvisation.

Hierarchical: The ending seems to be conducted by one of the musicians, and/or forced by one of the musicians upon the others.

Egalitarian: The different musicians seem to contribute more or less equally to the end of the improvisation.

Progressive: The ending comes in a gradual fashion.

Immediate: The ending comes suddenly.

Predictable: You were expecting that the ending would to happen this way.

Surprising: You were not expecting that the ending would happen this way.

Too early: The ending seems to have happened a bit too early.

Timely: The ending seems to have happened at the right time.

Too late: The ending seems to have happened a bit too late.

S.3.2. Supplementary results

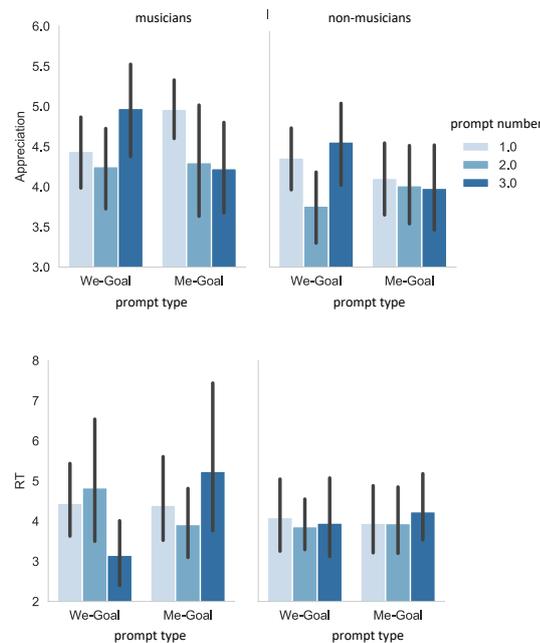


Figure S6. Mains results of Experiment 3 depending on expertise. A) Listeners' appreciation ratings and **B)** response times were averaged separately for each participant, prompt number and prompt type, before being averaged in the group. **A)** Expectedly, we found a main effect of expertise such that musicians had higher appreciation ratings overall ($F(1,44) = 5, p = .29$), but no interaction between expertise and prompt number ($F < 0.4, p > .6$), no significant interaction between expertise and prompt type ($F(1,44) = 2.66, p = .11$), and a marginal triple interaction between the three factors ($F(2,88) = 2.9, p = .06$). **B)** We examined listeners response times in providing appreciation ratings, in order to get a hint at how difficult the task was depending on prompt type and number. There was an interaction between Prompt Type and Prompt Number ($F(2,90) = 5.25, p = .007, \eta p^2 = .02$), reflecting the fact that listeners tended to respond faster when performers had a shared goal (i.e., in the WE-3 condition, difference between WE-3 and ME-3: $p = .01$) and slower when all performers had an individual goal (ME-3 versus ME-2: $p = .049$, all other comparisons non-significant). In addition, there was a significant triple interaction between Prompt Type, Prompt Number and expertise ($F(2,88) = 4.135, p = .019$) and an interaction between Prompt Type and Number ($F(2,88) = 4.7, p = .011$). This interaction reflected the fact that expert listeners responded faster when performers had a shared goal (i.e., difference between WE-3 and ME-3: $p = .001$; WE-3 and WE-2: $p = .002$; WE-3 and WE-1: $p = .011$; WE-3 and ME-1: $p = .013$; post-hoc Tukey HSD) and slower when each individual musician had an individual goal (ME-3 versus ME-2: $p = .02$). Thus, expert listeners had more difficulties when evaluating interactions where musicians did not share a goal, as compared to improvisations where they did. None of these comparisons reached significance in naive listeners (all p -values $> .4$). The fact that experts took longer to evaluate interactions where musicians did not share a WE-goal as compared to improvisations where they did, while naive listeners remain unaffected may suggest that endings in which musicians were sharing a goal appeared to expert musicians as more similar to what they would expect in a natural CFI setting. Error bars show the 95% confidence intervals.

S.4. Experiment 4

S.4.1. Glossary

Listeners were provided a glossary to define the categories along which they were asked to rate the musician's behavior as follows:

Ascending: The musician globally follows an ascending trajectory: playing with increasing loudness, increasing density, or increasing pitch etc...

Descending: The musician globally follows a descending trajectory: playing with decreasing loudness, decreasing density, or decreasing pitch etc...

Constant: The musician globally maintains the same trajectory: playing with a stable loudness, stable density, or stable pitch etc...

Without direction: The musician does not follow a specific trajectory.

Repetitive: The musician repeats more or less the same sound, the same rhythm or the same sentence; there is little variation across the musical extract.

Varied: The musician frequently varies, often changing sounds, rhythm or sentences across the musical extract.

Predictable: The musician follows a path that is quite predictable, there are no surprising events across the duration of the musical extract.

Surprising: One or several remarkable or unpredictable events happened during this musical extract.

Confident: The musician seems franc, direct, without hesitations.

Hesitant: The musician seems retained, lacks confidence.

S.4.2. Supplementary results

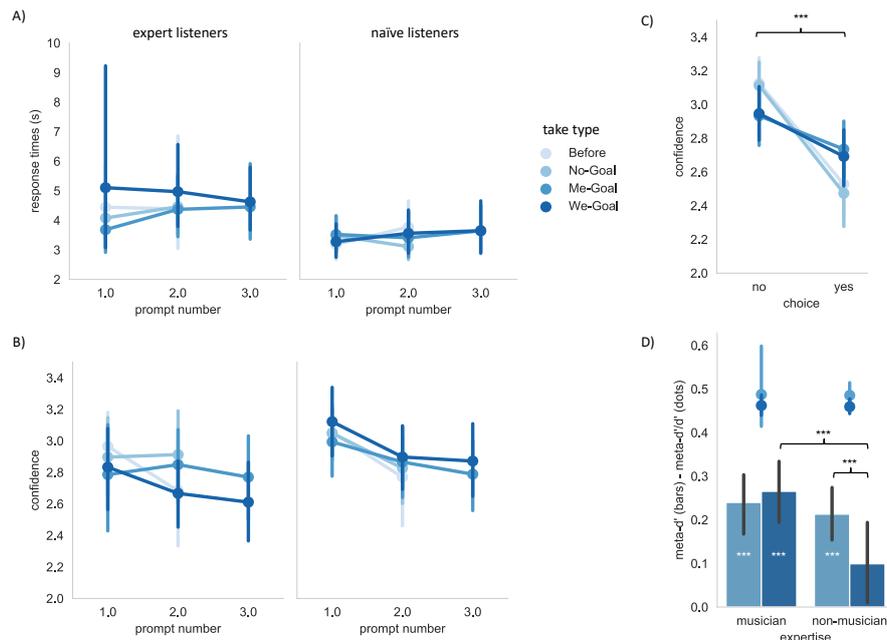


Figure S7. Confidence and response times Experiment 4. **A)** We examined response times to see whether listeners engaged more cognitive processes to respond depending on the conditions and expertise. There was a main effect of expertise ($F(1,45) = 5, p = .029$): naïve listeners responded faster than musicians (mean RT musicians: 4.43 seconds \pm 1.72 SD; non-musicians 3.41 \pm 1.18 SD, $t(45) = 2.33, p = .024$), suggesting that experts deliberated more than naïve participants in this task. There was also a marginal interaction between expertise, prompt type, and prompt number ($F(3,135) = 2.34, p = .076$), but no effect of prompt number ($F(1,45) = 0.5, p > .47$), and no effect of prompt type ($F(3,135) = 1.18, p = .18$). **B)** Regarding confidence, there was a main effect of take type ($F(3,135) = 4.53, p = .005$), a main effect of prompt number ($F(1,45) = 22.44, p < .001$), and a marginal interaction between the two factors ($F(3,135) = 2.6, p = .054$). There was no significant effect of expertise ($F(1,45) = 0.9, p > .3$) and no interactions between expertise and the other factors (all p -values $> .05$). These effects reflected the fact that participants were more confident in their responses when only one of the musicians had been prompted (post-hoc Tukey HSD, 1 vs. 2: $p < .001$; 1 vs. 3: $p < .001$, 2 vs. 3: $p = .2$). There were also more confident for the Before and NO-Prompt conditions as compared to both the ME-Goal ($p = .003 / 0.018$) and WE-Goal conditions ($p = .003 / 0.019$, comparison between Before and NO-Prompt: $p = .53$; WE-Goal and ME-Goal: $p = .97$). Note that this effect of prompt type is not very informative however, without considering the choice made by the participants. Thus **(C)**, we also ran a $rmANOVA$ including choice, prompt type and expertise. Here again, there was no effect interaction with expertise ($p > .9$). On top of the main effect of take type ($F(3,127) = 4.95, p = .003$), there was also a main effect of choice ($F(1,39) = 52.4, p < .001$) and an interaction ($F(3,129) = 12.9, p < .001$). Overall, listeners were more confident when responding that the performer was not looking for an end (post-hoc Tukey HSD: $p < .001$). This effect varied with condition however: as expected if listeners had some metacognitive access to their performances, they were more confident when correctly responding that the performer was not looking for an end in the Before condition as compared to when they incorrectly responded that the performer was looking for an end ($p < .001$). The same effect was found in the NO-Prompt condition ($p < .001$). Yet, as mentioned above, participants had a large metacognitive bias: they tended to be more confident when responding negatively. Thus, they were also more confident when incorrectly responding that the performer was not looking for an end in the WE-Goal and ME-Goal conditions as compared to when they correctly responded that the performer was looking for an end (Me: $p < .008$; We: $p < .001$; there were no differences in between the Before and NO-Prompt conditions: all p -values $> .6$; and We versus ME-Goals: all p -values $> .5$). Thus, participants had poor metacognitive sensitivity in this task, due to a strong metacognitive bias. Still, participants were more confident for positive responses given in the ME- and WE-Goal conditions as compared to positive responses given in the No- and Before conditions (all p -values $> .01$), and vice versa for negative responses (all p -values $> .0006$). **D)** Participants metacognitive sensitivity ($meta-d'$, bars) and metacognitive efficiency ($meta-d'/d'$, dots) was assessed for the Me and We conditions separately. To assess whether participants' confidence judgements still tracked performances despite this large bias, we computed a $meta-d'$, which is a measure of metacognitive sensitivity that, like d' for discrimination measures, relies on signal detection theory to measure sensitivity independently from bias (Fleming, 2017). Participants metacognitive sensitivity was significantly above chance overall (mean $meta-d' = 0.2 \pm 0.17$ SD; $t(46) = 7.9, p < .001$). Yet, a $rmANOVA$ revealed that, like for d' , there was an interaction between expertise and condition ($F(1,45) = 6.13, p = .017$) reflecting the fact that while expert listeners achieved above chance metacognitive sensitivity in both conditions (mean $meta-d'$ in the ME condition: $0.24 \pm 0.17, t(20) = 6.44, p < .001$; WE condition: $0.26 \pm 0.17, t(20) = 7.11, p < .001$), naïve listeners only achieved above chance metacognitive sensitivity in the ME condition ($0.21 \pm 0.15, t(25) = 7.06, p < .001$) but not in the WE condition ($0.1 \pm 0.24, t(25) = 2, p = .056$, post-hoc Tukey HSD comparison between the two conditions: $p < .001$). Moreover, naïve listeners' metacognitive sensitivity was significantly worse than experts' metacognitive sensitivity in the WE ($p < .001$) but not the ME ($p = .58$) conditions. Thus, despite their strong bias, listeners' confidence was still reflecting their performances to a weak extent, and showed better than chance metacognitive sensitivity when this bias is considered.

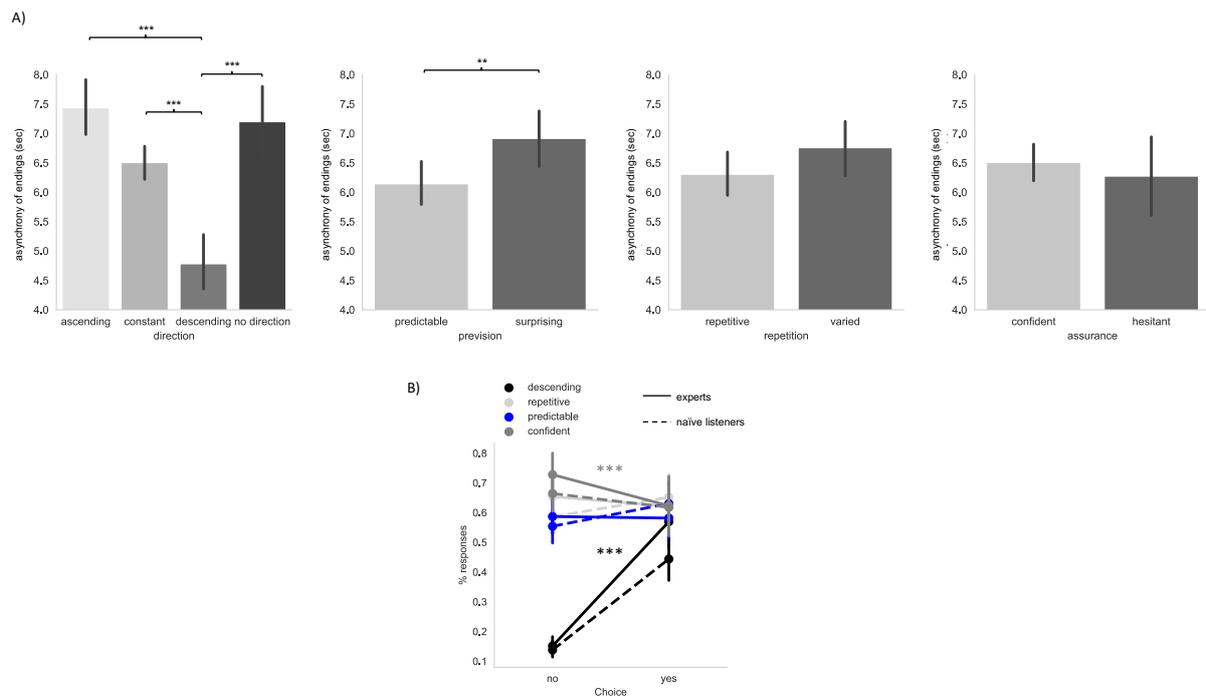


Figure S8. Do repetitions act as coordination smoothers or as communicative signals? **A)** To try and investigate whether repetitions can act as coordination smoothers here, we examined how behavioral categories related to the asynchrony of the trio. In a linear mixed regression with trios' asynchrony of endings as a dependent variable, listeners' judgements about musicians behavior (directionality, predictability, variety and assurance) as independent variables, and performer and listener as random factors, we found that predictability was significantly related to asynchrony ($\beta = -0.21$, $sem = 0.1$, $t = -1.97$, $p = 0.009$), as well as direction (descending versus ascending: $\beta = -1.42$, $sem = 0.29$, $t = -4.82$, $p < .001$; constant: $\beta = -1.14$, $sem = 0.25$, $t = -4.52$, $p < .001$; no direction: $\beta = -1.56$, $sem = 0.29$, $t = -5.51$, $p < .001$; assurance and repetition were not significantly associated with asynchrony when taking the other factors into account, $p > .6$). Thus, the more musicians were judged to have a descending and predictable behavior following the prompt, the more synchronized the trio. This is consistent with the idea that performers relied on descending and predictable behaviors as coordination smoothers. Asterisks show significant model comparisons with ** representing $p < .01$; ***: $p < .001$. **B)** The percentage of descending, repetitive, predictable and confident responses was computed separately for yes and no detection responses for each listener, before being averaged in the group. Another possibility is that musicians use repetitions as communicative signals. If such was the case, we may expect that expert listeners should report an intention to end more often in improvisations with repetitions, which would suggest that they generally interpret repetitive behaviors as a signaling strategy that, in the appropriate musical context, could be construed as an intention to end the performance. To assess this claim, we examined how listeners' judgements about performers' behaviors related to judgements about intentions to end the performance by running a logistic mixed regression with detection choice (yes / no) as a dependent variable, behavioral category (direction / repetition / prediction / confidence) as an independent variable, and listener and performer as random factors. Direction and assurance significantly predicted detection choices, but not judgements about repetition and prediction ($p > .2$ and 0.6 respectively). Confident behaviors were associated with less positive (i.e., ending) responses ($\beta = -0.23 \pm 0.047$ sem, $z = -5.1$, $p < .001$). On the opposite, descending behaviors were associated with more positive responses (descending vs. ascending: $\beta = 2.6 \pm 0.14$ sem, $z = 18.2$, $p < .001$; descending vs. constant: $\beta = 2.1 \pm 0.12$ sem, $z = 17.7$, $p < .001$; descending vs. no direction: $\beta = 2 \pm 0.13$ sem, $z = 15$, $p < .001$). Note that there were no significant interactions between musicians' behavior, prompt type and number, and expertise. Thus, listeners associated descending and hesitant behaviors with intentions to end, but did not systematically report an intention to end when they heard repetitions. Taken together with the results presented above, this may suggest that repetitions were used as coordination smoothers rather than communicative signal here. Alternatively, it may still be that performers used repetitions as communicative signals, but that context was crucial for their interpretation: given that our analysis does not take the context into account, we may miss the information that potentially enable musicians to pragmatically infer intentions to end from repetitions. Asterisks show significant model comparisons with *** representing $p < .001$.

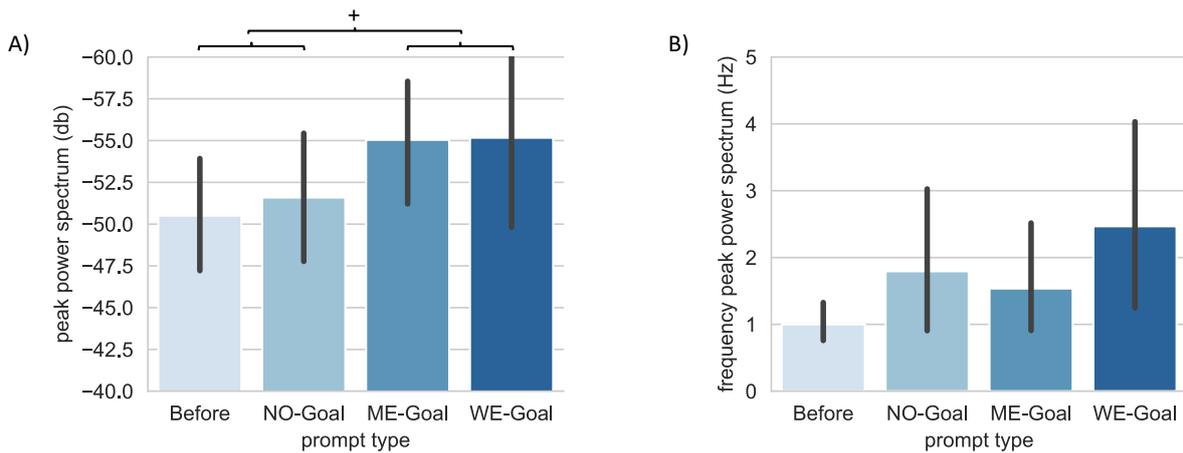


Figure S9. Rhythmic content estimation. Listeners perceived performers' musical actions to be more repetitive and predictable in the WE condition as compared to the other conditions (see Figure 6). This may be because performers produced more rhythmical patterns in this condition, which would be a good strategy to make themselves more predictable and help other musicians synchronize with them. To evaluate this possibility, we estimated the extent to which musicians produced rhythmical actions, as well as the main frequency at which they produced them. For each musical extract, we: 1) extracted its amplitude envelope with the Hilbert transform (using the `signal` function of the `scipy` package in python); 2) down-sampled this envelope to 32 Hz, a resolution allowing to capture beats while ignoring faster variations that may be related to vibrato or instruments' resonances; 3) computing the power spectrum of the envelope with a fast Fourier transform (using the `fft` function of the `numpy` package in python); 4) keeping only the values that corresponded to frequencies that are known to induce perceptions of rhythms in humans (between 0.5 to 16 Hz, according to London, 2012). **A)** The extent to which the musical extract contained rhythmical content was then estimated as the peak value of the power spectrum in this restricted range, and averaged separately for each condition. An ANOVA revealed no main effect of condition on the peak value of the power spectrum ($F(3,68) = 1.21, p > 0.31$). There was a marginal difference between the peak value of the power spectrum for unprompted (Before and NO-Goal conditions) and prompted (WE and ME-Goal conditions) trials ($t(70) = 1.9, p = 0.06$; represented by the cross), suggesting that if something, performers tended to play more rhythmical contents following a prompt, although they did so to a similar extent in the ME and WE conditions. **B)** The dominating tempo was estimated as the frequency at which the peak value was observed in the power spectrum. An ANOVA revealed no main effect of condition on tempo ($F(3,68) = 1.36, p > 0.26$).

S.5. Acoustic Analysis.

S.5.1. Supplementary results.

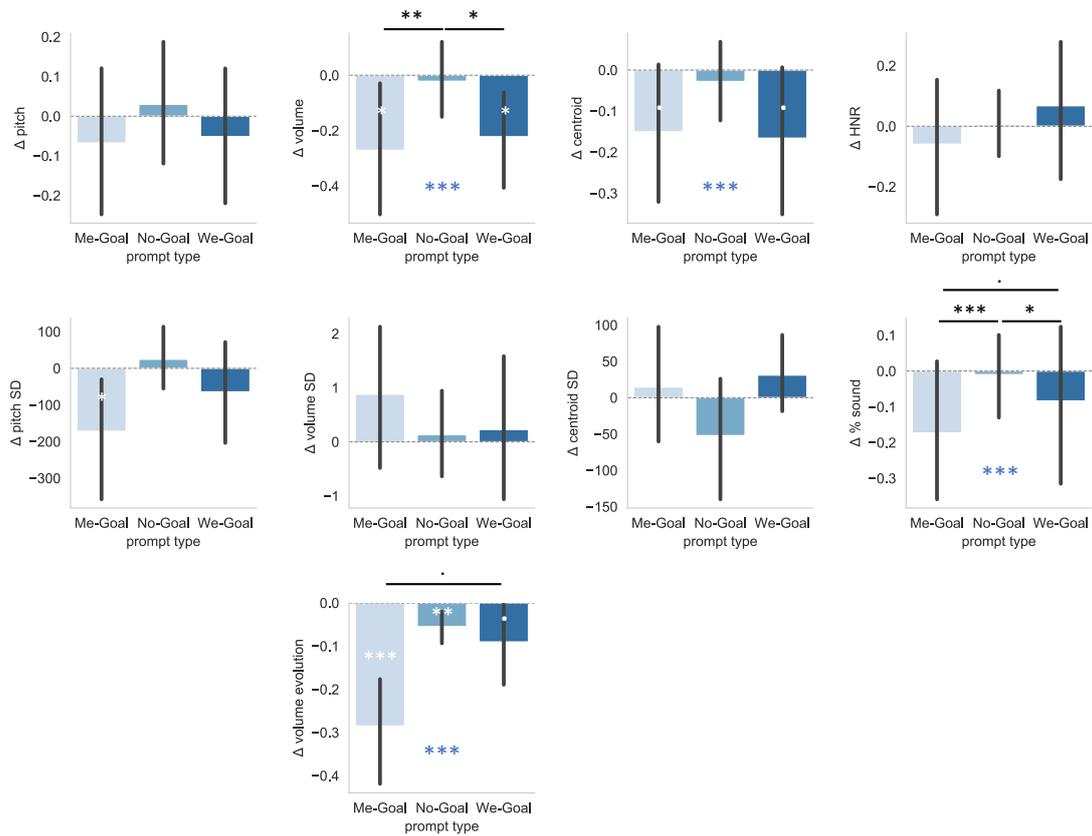


Figure S10. Impact of goals on acoustic features. For each take and musician, we computed nine acoustic features: pitch, volume (RMS), spectral centroid, harmonic to noise ratio (HNR), pitch, RMS, and centroid variability (approximated by the standard deviation), the percentage of sound (i.e., time played / time played + silence), and volume evolution (defined as the slope of the RMS in each snippet). Values were z-scored to allow comparison in between musicians / instruments and minimize the impact of recording, mixing, etc. On top of the main effects of volume, spectral centroid, evolution of the volume and % of sound on timing, and the interaction between prompt type and volume reported in the main text, we observed three other noteworthy interactions. First, there was a significant interaction between prompt type and density ($\chi^2 = 10$, $p = .007$), reflecting the fact that the decrease in density observed after the prompt was more important in the Me ($\beta = 5.46$, $sem = 1.63$, $z = 3.34$, $p < .001$) and We ($\beta = 3.16$, $sem = 1.55$, $z = 2$, $p = .04$) conditions as compared the NO-Prompt condition. The decrease in density was also marginally more pronounced in the Me as compared to the We condition ($\beta = -2.3$, $sem = 1.38$, $z = -1.7$, $p = .096$). Second, there was a marginal interaction between volume evolution and prompt type ($\chi^2 = 5.9$, $p = .053$) reflecting the fact that musicians performed more pronounced decrescendos in the Me condition as compared to the We condition ($\beta = -1.43$, $sem = 0.8$, $z = -1.76$, $p = .078$). Third, regarding timber, it also seemed to be the case that spectral centroid shifted to lower values mostly for the Me and We condition, but the interaction was not significant ($p > .6$). Blue asterisks within graphs show the results of logistic regressions testing the predictive effect of acoustic features on timing (after/before), and black asterisks represent pairwise model comparisons for significant interactions with prompt type. White asterisks represent one-sample t-tests against chance-level; $. p < .1$; $* p < .05$; $** p < .01$; $*** p < .001$.

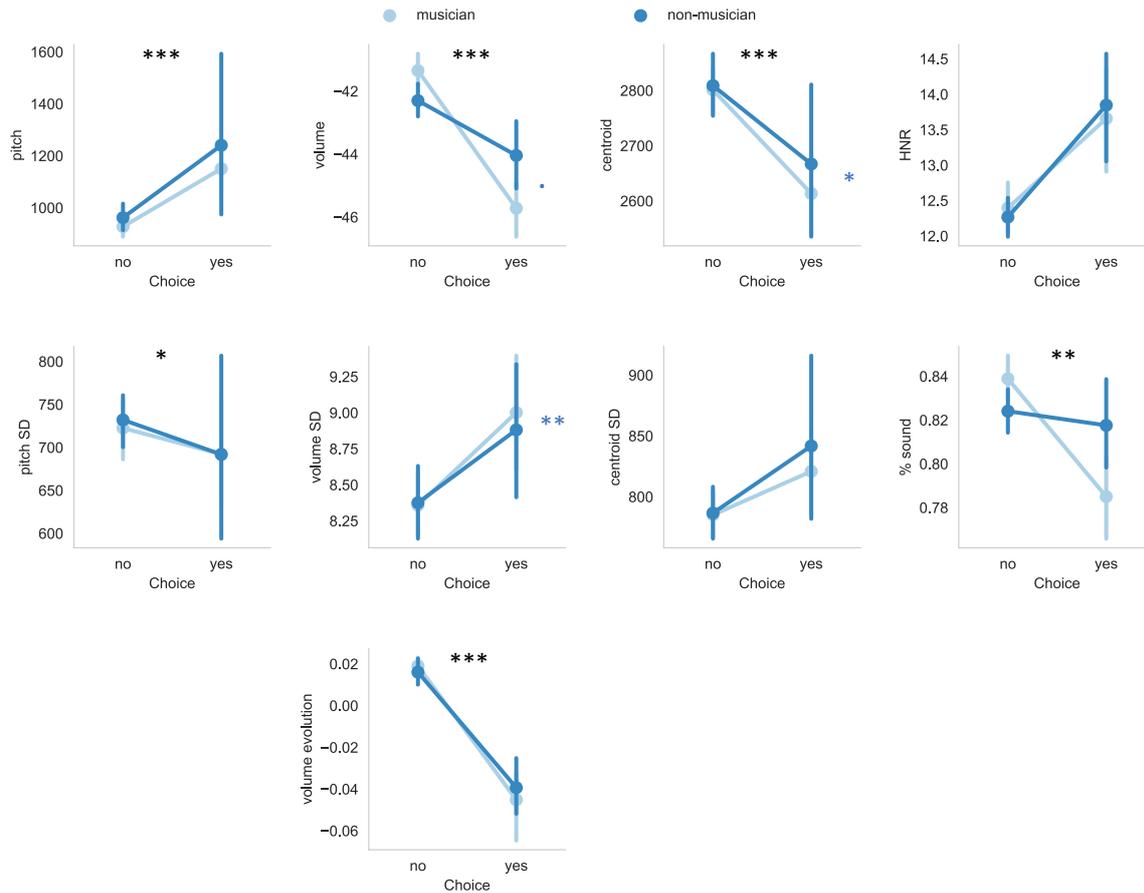


Figure S11. Impact of acoustic features on listeners' detection of an end. To assess this, we examined how acoustic features predicted listeners' detection responses. We ran a logistic mixed regression with choice (yes/no) as a dependent variable, the nine acoustic features and expertise as independent variables, and listener and musician as random factors. The probability to detect an end significantly increased with pitch ($\beta = 0.0003$, $sem = 0.00006$, $z = 6.1$, $p < .001$), and decreased with volume ($\beta = -0.063$, $sem = 0.01$, $z = -6.6$, $p < .001$), centroid ($\beta = -0.0004$, $sem = 0.00006$, $z = -5.98$, $p < .001$), volume evolution ($\beta = -1.7$, $sem = 0.38$, $z = -4.42$, $p < .001$) and density ($\beta = -1.7$, $sem = 0.58$, $z = -3$, $p = .002$). In addition, there was an interaction between expertise and centroid ($\beta = -0.0001$, $sem = 0.00005$, $z = -2$, $p = .045$), suggesting that experts relied on timber more than naïve listeners. There was also a significant interaction between expertise and volume variability ($\beta = -0.03$, $sem = 0.01$, $z = -2.62$, $p = .009$) and a marginal interaction with volume ($\beta = -0.012$, $sem = 0.007$, $z = -1.7$, $p < .09$), which reflected the fact that musicians tended to detect an end when the volume was more variable and negative more than naïve listeners. Black asterisks within graphs show the results of logistic regressions testing the predictive effect of acoustic features on choice. Blue asterisks show significant interactions with expertise; $.p < .08$; $* p < .05$; $** p < .01$; $*** p < .001$.

S.5.2. Goal propagation does not reduce to simple imitations.

If emergent mechanisms such as mimicry or entrainment were driving the improvement in coordination that we observed in the WE condition, we may see some evidence for low-level adaptations, or perception-action matchings after the prompt in non-prompted musicians. For instance, we may see that they directly mimic decrescendos. To examine this possibility, we extracted nine acoustic features from the snippets heard by the listeners before or after the

prompt: the mean pitch (fundamental frequency), the mean volume (RMS), the spectral centroid, the harmonic to noise ratio (HNR), the variability of the pitch, volume and spectral centroid, the percentage of time that the musician spent playing rather than remaining silent (% sound), and finally, the volume evolution (difference between the second and the first part of the sound: negative values reflect decrescendos).

We assessed how these acoustic features changed after the prompt depending on prompt type, by running a logistic mixed regression with the timing of the snippet (before or after the prompt) as a dependent variable, prompt type and the nine acoustic features as independent variables, and musician as a random factor (see Fig. S9). The model revealed that, following the prompt, there was a significant decrease in volume (beta = -2.13, sem = 0.48, $z = 4.42$, $\chi^2 = 37.2$, $p < .001$) and spectral centroid (beta = -1, sem = 0.3, $z = 3.6$, $\chi^2 = 14.1$, $p < .001$), as well as a more negative evolution of the volume (beta = -1.2, sem = 0.31, $z = 3.8$, $\chi^2 = 40$, $p < .001$) and a decrease in the % of sound (beta = -1.37, sem = 0.55, $z = 2.5$, $\chi^2 = 19.3$, $p < .001$). Thus, after the prompt musicians tended to play less, more quietly, and with a darker sound.

We also observed some interactions, reflecting the fact that certain acoustic features changed differently as a function of prompt type. In particular, there was an interaction between prompt type and volume ($\chi^2 = 8.8$, $p = .013$), reflecting the fact that the decrease in volume observed after the prompt was more important in the Me (beta = 4.03, sem = 1.46, $z = 2.75$, $p = .006$) and We (beta = 3.14, sem = 1.39, $z = 2.25$, $p = .024$) conditions as compared the No-Goal condition. Furthermore, pairwise comparisons showed that the decrease in volume was significant in the Me ($t(35) = -2.17$, $p = .037$) and We conditions ($t(32) = -2.34$, $p = .025$), but not in the No-Goal condition ($t(35) = -0.3$, $p = .76$). Thus, it was not the case that unprompted musicians' behavior directly reflected the behavior of their prompted partners at the level of volume (see Fig. S9 caption for the full analysis that more generally supports this conclusion).

Overall, there was little to no evidence in favor of the idea that un-prompted musicians adapt their behavior by directly mimicking prompted musicians for the acoustic features and the time window under scrutiny (note that equivalent results were obtained in a bigger window of 29 seconds, corresponding to the average time that musicians took to stop after hearing the prompt). Thus, it does not seem to be the case that musicians engage in simple perception-action matching, unless it occurs at some other level that we do not capture here.

References

- Bailey, D. (1992). *Improvisation: Its Nature and Practice in Music*. Da Capo Press.
- Canonne, C. (2019). Listening to Improvisation. *Empirical Musicology Review*.
<https://doi.org/10.18061/emr.v13i1-2.6118>
- Canonne, & Garnier. (2012). Cognition and segmentation in collective free improvisation: An exploratory study. *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*.
- 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*.
- Corbett, J. (2016). *A listener's guide to free improvisation*. The University of Chicago press.
- Denzler, B., & Guionnet, J.-L. (2020). *The Practice of Musical Improvisation: dialogues with contemporary musical improvisers*. New York: Bloomsbury Academic.
- Fleming, S. M. (2017). HMeta-d: hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, 2017(1).
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York Wiley (Vol. 4054).

London, J. (2012). *Hearing in Time: Psychological Aspects of Musical Meter*. *Hearing in Time: Psychological Aspects of Musical Meter*.

<https://doi.org/10.1093/acprof:oso/9780199744374.001.0001>

Michaelsen, G. (2019). Making “Anti-Music”: Divergent Interactional Strategies in the Miles Davis Quintet’s *The Complete Live at the Plugged Nickel 1965*. *Music Theory Online*, 25(3).

Novembre, G., Mitsopoulos, Z., & Keller, P. E. (2019). Empathic perspective taking promotes interpersonal coordination through music. *Scientific Reports*.

<https://doi.org/10.1038/s41598-019-48556-9>