



**HAL**  
open science

# Musical Friends and Foes: The Social Cognition of Affiliation and Control in Improvised Interactions

Jean-Julien Aucouturier, Clément Canonne

► **To cite this version:**

Jean-Julien Aucouturier, Clément Canonne. Musical Friends and Foes: The Social Cognition of Affiliation and Control in Improvised Interactions. *Cognition*, 2017, 10.1016/j.cognition.2017.01.019 . hal-01451102

**HAL Id: hal-01451102**

**<https://hal.science/hal-01451102>**

Submitted on 31 Jan 2017

**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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



## Original Articles

## Musical friends and foes: The social cognition of affiliation and control in improvised interactions



Jean-Julien Aucouturier\*, Clément Canonne

CNRS Sound and Technology of Music and Sound (UMR9912, CNRS/IRCAM/UPMC), 1 Place Stravinsky, Paris France

## ARTICLE INFO

## Article history:

Received 30 March 2016

Revised 18 January 2017

Accepted 25 January 2017

## Keywords:

Social cognition

Musical improvisation

Interaction

Joint action

Coordination

## ABSTRACT

A recently emerging view in music cognition holds that music is not only social and participatory in its production, but also in its perception, i.e. that music is in fact perceived as the sonic trace of *social relations* between a group of real or virtual agents. While this view appears compatible with a number of intriguing music cognitive phenomena, such as the links between beat entrainment and prosocial behaviour or between strong musical emotions and empathy, direct evidence is lacking that listeners are at all able to use the acoustic features of a musical interaction to infer the affiliatory or controlling nature of an underlying social intention. We created a novel experimental situation in which we asked expert music improvisers to communicate 5 types of non-musical social intentions, such as being domineering, disdainful or conciliatory, to one another solely using musical interaction. Using a combination of decoding studies, computational and psychoacoustical analyses, we show that both musically-trained and non musically-trained listeners can recognize relational intentions encoded in music, and that this social cognitive ability relies, to a sizeable extent, on the information processing of acoustic cues of temporal and harmonic coordination that are not present in any one of the musicians' channels, but emerge from the dynamics of their interaction. By manipulating these cues in two-channel audio recordings and testing their impact on the social judgements of non-musician observers, we finally establish a causal relationship between the affiliation dimension of social behaviour and musical harmonic coordination on the one hand, and between the control dimension and musical temporal coordination on the other hand. These results provide novel mechanistic insights not only into the social cognition of musical interactions, but also into that of non-verbal interactions as a whole.

© 2017 Published by Elsevier B.V.

## 1. Introduction

The study of music cognition, more than solely contributing to the understanding of a specialized form of behaviour, has informed general domains of human cognition for questions as varied as attention (Shamma, Elhilali, & Micheyl, 2011), learning (Bigand & Poulin-Charronnat, 2006), development (Dalla Bella, Peretz, Rousseau, & Gosselin, 2001), sensorimotor integration (Chen, Penhune, & Zatorre, 2008), language (Patel, 2003) or emotions (Zatorre, 2013). In fact, perhaps the most characteristic aspect of how music affects us is its large-scale, concurrent recruitment of such many sensory, cognitive, motor and affective processes (Alluri et al., 2012), a feature which some consider key to its phylogenetic success (Patel, 2010) and others, to its potential for clinical remediation (Wan & Schlaug, 2010).

In the majority of such research, music is treated as an abstract sonic structure, a “sound text” that is received, analysed for syntax and form, and eventually decoded for content and expression. Musical emotions, for instance, are studied as information encoded in sound by a performer, then decoded by the listener (Juslin & Laukka, 2003), using acoustic cues that are often compared to linguistic prosody, e.g. fast pace, high intensity and large pitch variations for happy music (Eerola & Vuoskoski, 2013; Juslin & Laukka, 2003). While this paradigm is appropriate to study information processing in many standard music cognitive tasks, it has come under recent criticism by a number of theorists arguing that musical works are, in fact, more akin to a theatre *script* than a rigid text (Cook, 2013; Maus, 1988): that music is intrinsically performative and participatory in its production (Blacking, 1973; Small, 1998; Turino, 2008) and is thus perceived as the sonic trace of *social relations* between a group of real or virtual agents (Cross, 2014; Levinson, 2004) or between the agents and the self (Evers, 2016; Moran, 2016).

\* Corresponding author.

E-mail address: [aucouturier@gmail.com](mailto:aucouturier@gmail.com) (J.-J. Aucouturier).

Much is already known, of course, about the performative aspects of collective music-making. The perceptual, cognitive and motor processes that enable individuals to coordinate their actions with others have received increasing attention in the last decade (Knoblich, Butterfill, & Sebanz, 2011), and musical joint action, with tasks ranging from finger tapping in dyads to a steady or tempo-varying pulse (Konvalinka, Vuust, Roepstorff, & Frith, 2010; Pecenkova & Keller, 2011) to string quartet performance (Wing, Endo, Bradbury, & Vorberg, 2014), has been no exception. What the above “relational” view of music cognition implies, however, is that music does not only involve social processes in its production, but also in its reception; and that those processes are not restricted to merely retrieving the interpersonal dynamics of the musicians that are responsible for the signal, but also extend to full-fledged social scenes heard by the listener *in the music itself*. “Experiencing music”, writes Elvers (Elvers, 2016), could well “serve as an esthetic surrogate for social interaction”.

Many music cognitive phenomena seem amenable to this type of explanation. First, some important characteristics of music are indeed processed by listeners as *relations* between simultaneous parts of the signal, and have been linked to positive emotional or social effects. For instance, cues of harmonic coordination, such as the consonance of two simultaneous musical parts, are associated to increased preference (Zentner & Kagan, 1996) and positive emotional valence (Hevner, 1935). Similarly, simultaneous temporal coordination, such as entrainment to the same beat, are linked to cooperative and prosocial behaviours (Cirelli, Wan, & Trainor, 2014; Wiltermuth & Heath, 2009) and judgements of musical quality (Hagen & Bryant, 2003). However, these effects remain largely indirect and one does not know whether “playing together” in harmony or time reliably signals something social to the listener, and if so, what?

Second, many individual differences in musical processing and enjoyment appear associated with markers of social cognitive abilities. For instance, in a survey of 102 Finnish adults, Eerola, Vuoskoski, and Kautiainen (2016) found that sad responses to music were not correlated to markers of emotional reactivity (such as being prone to absorption or nostalgia) but rather to high trait empathy (see also Egermann & McAdams, 2013). Similarly, in Loersch and Arbuckle (2013), participants’ mood was assessed before and after listening to music, and participants with highest emotional reactivity were also those who ranked high on group motivational attitudes. However, while these capacities linked to evaluating relations between agents and the self appear to modulate our perception and judgements about music, we do not know what exactly in music is the target of such processes.

Finally, both musician discourse on their practice and non-musician reports on their listening experience are replete with social relational metaphors. In jazz and contemporary improvisation (Bailey, 1992; Monson, 1996) but also in classical music (Klorman, 2016), musical phrases are often described as statements that are made and responded to; stereotypical musical behaviours, such as solo-taking or accompaniment, are interpreted as attempts to socially control or affiliate with other musicians; even composers like Messiaen (Healey, 2004) admit to treating some of their rhythmic and melodic motives like “characters”. In non-musicians too, qualitative reports of “strong experiences with music” (Gabrielsson & Bradbury, 2011; Osborne, 1989) show a tendency to construct relational narratives from music, series of “actions and events in a virtual world of structures, space and motion” (Clarke, 2014). However, while abundantly discussed in the anthropological, ethnomusicological and music theoretical literature, we have no mechanistic insight into how these phenomenological qualities may be generated.

In sum, despite all the theoretizing, we have no direct evidence that processing social relations is at all entailed by music cognition.

To demonstrate that human listeners recruit specifically social cognitive concepts (e.g., agency or intentionality) and processes (e.g., simulation or co-representations) to perceive social-relational meaning in music, one would need to show evidence that listeners are able to use acoustic features of a musical interaction to infer that two or more agents are e.g. in an affiliatory (Miles, Nind, & Macrae, 2009) or dominant relationship (Bente, Leuschner, Al Issa, & Blascovich, 2010), or simply exhibiting social contingency (McDonnell, Ennis, Dobbyn, & O’Sullivan, 2009; Neri, Luu, & Levi, 2006). All such evidence exists in the visual domain – sometimes with the mediation of music (Edelman & Harring, 2015; Moran, Hadley, Bader, & Keller, 2015) –, and in non-verbal vocalizations (Bryant et al., 2016), but not in music. Hard data on this is lacking because of the non-referential nature (or “floating intentionality” (Cross, 2014)) of music, which makes it difficult to design a musical social observation task amenable to quantitative psychoacoustical measurements.

Three main features of our experiments made it possible to achieve this goal. First, we created a large dataset of “musical social scenes”, by asking dyads of expert improvisers to use music to communicate a series of relational intentions, such as being domineering or conciliatory, from one to the other. In contrast to other music cognition paradigms which relied almost exclusively on solo monophonic extracts (e.g., 40 out of the 41 studies reviewed in Juslin & Laukka (2003)) or well-known pre-composed pieces (Eerola & Vuoskoski, 2013), this allowed us to show that a variety of relational intentions can be communicated *in music* from one musician to another (Study 1), and that third-party listeners too were able to perceive the social relations of the two interacting musicians (Study 2). Second, by recording these duets in two simultaneous but separate audio channels, we were able to present these channels dichotically to third-party listeners. This allowed us to show that a sizeable share of their recognition accuracy relied on the processing of a series of two-channel “relational” cues that emerged from the interaction and were not present in either one of the players’ behaviour (Study 3 & 4). Finally, using acoustic transformations on the separated channels, we were able to selectively manipulate the apparent level of harmonic and temporal coordination in these recordings (Study 5). This allowed us to establish a causal role for these two types of cues in the perceived level of social affiliation and control (two important dimensions of social behaviour Pincus, Gurtman, & Ruiz, 1998) of the agents involved in these musical interactions.

## 2. Study 1: studio improvisations

In this first study, we asked dyads of musicians to communicate a series of non-musical relational intentions (or social attitudes, namely those of being domineering, insolent, disdainful, conciliatory and caring), from one to the other. We then tested their capacity to recognize the intentions of their partner, based solely on their musical interaction.

### 2.1. Methods

**Participants:**  $N = 18$  professional French musicians (male: 13;  $M = 25$ ,  $SD = 3.5$ ), recruited via the collective free improvisation masterclasses of the National Conservatory of Music and Dance in Paris (CNSMD), took part in the studio sessions. All had very substantial musical training (15–20 years of musical practice, 5–10 years of improvisation practice, 2–5 years of free improvisation practice), as well as previous experience playing with one another. Their instruments were saxophone ( $N = 5$ ); piano ( $N = 3$ ); viola ( $N = 2$ ); bassoon, clarinet, double bass, euphonium, flute, guitar, saxhorn, violin ( $N = 1$ ). As a general note, all the

procedures used in this work (Study 1–5) were approved by the Institutional Review Board of the Institut National Supérieur de la Santé et de la Recherche Médicale (INSERM), and the INSEAD/Sorbonne Ethical committee. All participants gave their informed written consent, and were compensated at a standard rate.

**Procedure:** The musicians were paired in 10 random dyads and each dyad recorded 10 short ( $M = 95s$ ) improvised duets. In each duet, one performer (musician A, the “encoder”) was asked to musically express one of 5 social attitudes (see below), while the other (musician B, the “decoder”) was asked to recognize it based on their mutual interaction. At the end of each improvisation, musician A rated how well they thought they had managed to convey their attitude on a 9-point Likert-scale anchored with “completely unsuccessful” to “very successful”. Musician B rated which of the 5 attitudes they thought had been conveyed (forced-choice, 5 response categories). Musicians in a dyad switched role after each improvisation: the decoding musician then became the encoder for another random attitude, which the other musician had to recognize, and so on until each dyad had recorded 5 attitudes each, for a total of 10 duets. Attitudes were presented in random order, so that each musician was both the encoder and the decoder for each of them once. In total, we recorded  $n = 100$  duets (5 attitudes encoded by 2 musicians, in 10 dyads). Musicians played from isolated studio booths without seeing each other, so that communication remained purely acoustical. Each duet was recorded in two separated audio channel, one per musician. A selection of 4 representative interactions can be seen in [SI Video 1](#), and the complete corpus is made available at <https://archive.org/details/socialmusic>.

**Attitudes:** The five social attitudes chosen to prompt the interactions were those of being domineering (DOM; French: *dominant, autoritaire*), insolent (INS; French: *insolent, effronté*), disdainful (DIS; French: *dédaigneux, sans égard*), conciliatory (CON; French: *conciliant, faire un pas vers l'autre*), and caring (CAR; French: *prévenant, e-tre aux petits soins*). These 5 behaviours were selected from the literature to differ along both affiliation (i.e., the degree to which one is inclusive or exclusive towards the other, along which, for instance,  $INS < CAR$ ) and control (i.e., the degree to which one is domineering or submissive towards the other, along which, for instance,  $DIS < DOM$ ), two dimensions which are standardly taken to describe the space of social behaviours (Pincus et al., 1998). In addition, the 5 behaviours considered here correspond to attitudes which cannot exist without the other person being present (Wichmann, 2000) (e.g., one cannot dominate or exclude on one's own). This is in contrast with other intra-personal constructs such as basic emotions (incl. happiness, sadness, anger, fear) or even so-called social emotions like pity, shyness or resentment, which do not require any inter-personal interaction to exist (Johnson-Laird & Oatley, 1989).

The 5 attitudes were presented to the participants using textual definitions ([SI Text 1](#)) and pictures illustrative of various social situations in which they may occur ([SI Fig. 1](#)). The same pictures were used as prompts during the recording of the interactions (see [SI Video 1](#)).

## 2.2. Results

[Table 1](#) gives the hit rates and confusion matrix over all 5 response categories. Overall hit rate from the  $n = 100$  5-forced choice trials was  $H = 64\%$  (45–80%,  $\pi = .77$ –94,  $M = .88$ ), with caring ( $H = 80\%$ ) and domineering ( $H = 63\%$ ) scoring best, and insolent ( $H = 45\%$ ) scoring worst. Misses and false alarms were roughly consistent with an affiliatory vs non-affiliatory dichotomy, with caring and conciliatory on the one hand, and domineering and insolent on the other. The disdainful attitude was equally confused with domineering and conciliatory.

## 2.3. Discussion

We created here a novel experimental situation in which expert improvisers communicated 5 types of relational intentions to one another solely through their musical interaction. The decoding accuracy measured in this task ( $\pi = .90$ ) is relatively high for a music decoding paradigm: for instance, in a meta-study of music emotion decoding tasks, Juslin and Laukka (2003) report  $\pi = .68$ –1.00 ( $M = .85$ ) for happiness and  $\pi = .74$ –1.00 ( $M = .86$ ) for anger. In short, recognizing that a musical partner is trying to dominate, support or scorn us seems a capacity at least as robust as recognizing whether he or she is happy or sad.

So far, empirical evidence for the possibility to induce or regulate social behaviours with music had been mostly indirect, and always suggestive of affiliatory behaviours. For instance, collective singing or music-making was reported to lead both adults and children to be more cooperative (Wiltermuth & Heath, 2009), empathetic (Rabinowitch, Cross, & Burnard, 2013) or trusting of each other (Anshel & Kipper, 1988). However, because these behaviours can all be mediated by other non-social effects of music, such as its inducing basic emotions or relaxation, it had remained unknown whether music could directly communicate (i.e., encode and be decoded as) a relational intention. We found here that it could, both for affiliatory and non-affiliatory behaviours and with unprecedented complex nested levels of intentionality (e.g., consider discriminating domineering: “*I want you to understand that I want you to do as I want*” vs insolence: “*I want you to understand that I do not wish to do what you want*”). That it is possible for a musical interaction to communicate such a large array of interpersonal relations gives interesting support to the widespread use of musical improvisation in music therapy, in which collective improvisation is often practiced as a way to undergo rich social processes and explore inter-personal communication without involving verbal exchanges (MacDonald & Wilson, 2014).

From a cognitive mechanistic point of view, however, results from Study 1 remain nondescript. First, little is known about the processes with which the interacting musicians convey and recognize these attitudes. In particular, decoding musicians can rely on two types of cues: single-channel prosodic cues in the expression of the encoding musician (e.g. loud, fast music to convey domination or insolence) and dual-channel, emerging properties of their joint action with the encoder (e.g. synchronization, mutual adaptation, etc.). In this task, the decoding musicians' gathering of socially relevant information is embedded in their ongoing musical interaction, and it is impossible to distinguish the contributions of both types of cues. Second, the experimental paradigm in which each musician encoded and decoded each attitude exactly once does not provide a well-controlled basis for psychophysical analysis. In particular, because participants knew that each response category had only one exemplar, various strategic processes could occur between trials (e.g., if *that* was caring, then *this* must be conciliatory). This probably lead to an over-estimation of response accuracy, and makes it difficult to statistically test for difference from chance performance. In the following study, we used this corpus of improvisations to collect more rigorous decoding performance data in the lab, on third-party observers.

## 3. Study 2: decoding study

In this study, we took the preceding decoding task offline and measured decoding accuracy for a selection of the best-rated duets recorded in Study 1. Because we recorded these duets in two simultaneous but separate audio channels, we were able to present these channels dichotically to third-party listeners, and thus sepa-

**Table 1**

Decoding confusion matrix for Study 1. Hit rates  $H$  computed over all decoding trials (collapsed over all dyads). Proportion indices  $\pi$  express the raw hit rate  $H$  transformed to a standard scale where a score of 0.5 equals chance performance and a score of 1.0 represents a decoding accuracy of 100 percent (Rosenthal & Rubin, 1989). Abbreviations: CAR: caring; CON: conciliatory; DIS: disdainful; DOM: domineering; INS: insolent.

Encoded as	Decoded as					Total
	CAR	CON	DIS	DOM	INS	
CAR	16	3	0	0	1	20
CON	5	12	1	0	2	20
DIS	1	3	13	3	0	20
DOM	1	0	3	14	2	20
INS	0	2	4	5	9	20
Total	23	20	21	22	14	100
$H$	80%	60%	62%	63%	45%	64%
$\pi$	.94	.86	.88	.90	.77	.90

rate the contribution of acoustical cues coming from one channel from those only available when processing both channels simultaneously. Finally, to test for an effect of musical expertise on decoding accuracy, we collected data for both musicians and non-musicians.

### 3.1. Methods

**Stimuli:** We selected the  $n = 64$  duets that were found to be decoded correctly by the musicians participating in Study 1 (domineering: 14, insolent: 9, disdainful: 13, conciliatory: 12, caring: 16). We then selected the  $n = 25$  best-rated duets among them (5 per attitude), with an effort to counterbalance the types of musical instruments appearing in each attitude.

**Procedure:** Two groups of  $N = 20$  musicians and two groups of  $N = 20$  non-musicians were then tasked to listen to either a dichotically-presented stereo version (musician A, the “encoder”, in the left ear, musician B, the “decoder”, in the right ear), or a diotically-presented mono version (musician A only, in both ears) of the  $n = 25$  duets. Each trial included visual information to help participants keep track of the relation between both channels: in both conditions, the name and an illustration of instrument A was presented in the left side of the trial presentation screen. In the stereo condition, the name and an illustration of instrument B was also presented in the right side. An arrow pointing from the left instrument to the right instrument (in stereo condition) or from the left instrument outwards (in the mono condition) also served to illustrate the direction of the relation that had to be judged. In each trial, participants made a forced-choice decision of which of the 5 relational intentions best described the behaviour of musician A (with respect to musician B in the stereo condition; with respect to an imaginary partner in the mono condition). Trials were presented in random order, and condition (stereo/mono) was varied between-subject. Participants could listen to each recording as many times as needed to respond.

**Participants:** Two groups of musicians participated in the study in both the stereo ( $N = 19$ , male: 12,  $M = 28.4$ yo,  $SD = 5.8$ ) and mono conditions ( $N = 20$ , male: 15,  $M = 24.1$ yo,  $SD = 4.6$ ). They were students recruited from the improvisation and composition degrees (“classe d’improvisation générative” and “classe de composition”) of the Paris and Lyon National Music Conservatories (CNSMD), from the same population as the improvisers recruited for Study 1.

Two groups of non-musicians also participated in both the stereo ( $N = 20$ , male: 10,  $M = 21.9$ yo,  $SD = 1.9$ ) and mono conditions ( $N = 20$ , male: 11,  $M = 24.4$ yo;  $SD = 5.0$ ). They were undergraduate students recruited via the CNRS’s Cognitive Science Network (RISC). They were self-declared non-musicians, and we verified that none had had regular musical practice since at least 5 years before the study.

**Statistical analysis:** For result presentation purposes, hit rates were calculated as ratios of the total number of trials in each condition ( $25 \text{ trials} \times 20 \text{ participants} = 500$ ). For hypothesis testing, unbiased hit rates (Wagner, 1993) were calculated in each of the response categories for each participant, and compared either to their theoretical chance level using uncorrected paired  $t$ -tests or to one another between conditions using uncorrected independent-sample  $t$ -tests.

### 3.2. Results

Fig. 1 gives hit rates for all 5 response categories, in both conditions for both participant groups. The mean hit rate for musician listeners was  $M = 54.3\%$ , a 10% decrease from the decoding accuracy of the similarly experienced musicians performing in the online task of Study 1, confirming the greater difficulty of the present task. While still significantly above chance level, hit rates on stereo stimuli were 20.3% lower for non-musician than musician listeners ( $M = 34\% < 54.3\%$ ), a significant decrease in all 5 attitudes.

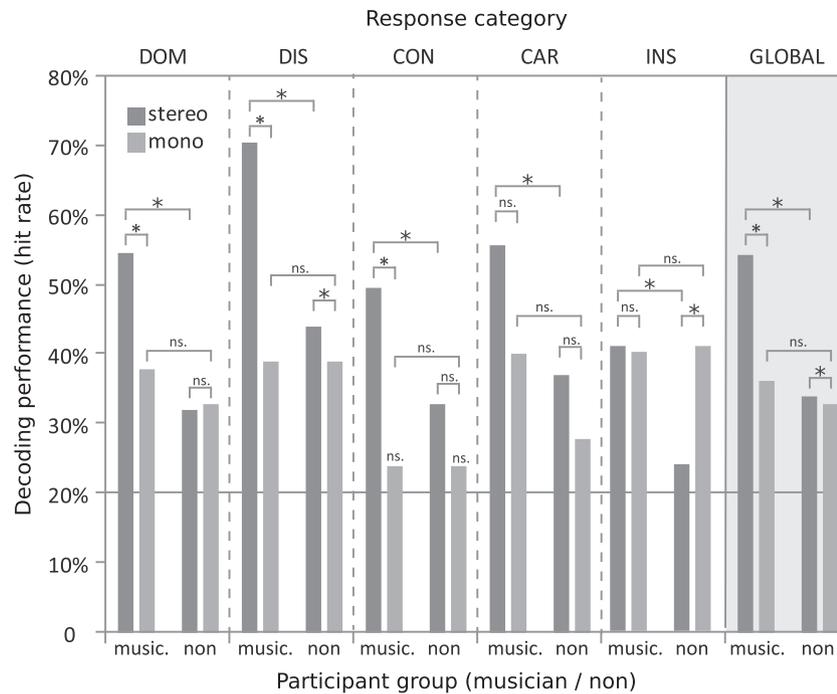
For musician listeners, hit rates were degraded by a mean 18.1% (stereo:  $M = 54.3\% > \text{mono}: 36.2\%$ ) when only mono was available. This decrease was significant for domineering ( $t(37) = 2.52$ ,  $p = .02$ ), disdainful ( $t(37) = 5.61$ ,  $p = .00$ ), insolent ( $t(37) = 2.01$ ,  $p = .05$ ) and conciliatory ( $t(37) = 3.34$ ,  $p = .002$ ), the latter of which degraded to chance level ( $t(19) = 1.93$ ,  $p = .07$ ). Because stimuli in the mono condition lacked information about the decoder’s interplay with the encoder, this shows that musician listeners relied for an important part on dual-channel coordination cues to correctly identify the type of social behaviour expressed in the duets.

Strikingly, for non-musician listeners, no further degradation was incurred by relying only on mono stimuli ( $M = 33\%$ ; even a +17% increase for INS). This suggests that non-musician listeners, who had the same performance as musicians with mono, were not able to interpret the dual-channel coordination cues that let musician listeners perform 18.1% better in stereo than mono.

### 3.3. Discussion

Study 2 extended results from Study 1 by establishing a more conservative measure of decoding accuracy ( $H = 54\%$ ,  $\pi = .82$  for the musician group), and confirming that both musician and non-musician listeners could recognize directed relational intentions in musical interactions above chance level for the 5 response categories.

In addition, by comparing stereo and mono conditions for musician participants, Study 2 showed that this ability relied, to a sizeable extent (18.1% recognition accuracy), on dual-channel cues which were not present in any one of the performers’ channels. This result suggests that musical joint action has specific communicative content, and that the way two musicians “play together”



**Fig. 1.** (Study 2) Decoding performance (hit rates) by musician and non-musician listeners for the task of recognizing 5 types of relational intentions conveyed by musician A to musician B. In the stereo condition, participants could listen to the simultaneous recording of both musicians. In the mono condition, participants only listened to musician A. Reported hit rates were calculated as ratios of the total number of trials in each condition. Hypothesis tests of difference between samples were conducted on the corresponding unbiased hit rates, with \* indicating significance at the  $p < .05$  level. Abbreviations: CAR: caring; CON: conciliatory; DIS: disdainful; DOM: domineering; INS: insolent; music.: musicians; non: non-musicians.

conveys information that is greater than, or complementary to, its separated parts.

Several factors can explain that musicians from the same population as those tested in Study 1 had a 10% lower decoding accuracy in the offline task than in Study 1. First, as already noted, task characteristics were more demanding in Study 2: stimuli were more varied (several performers, several instruments) and more numerous both in total ( $n = 25$  vs  $n = 5$ ) and per response category ( $n = 5$  vs  $n = 1$ ). This did not allow participants to rely on strong trial-to-trial response strategies. Second, musician listeners in Study 2 had a more limited emotional engagement with the task, having not participated in the music they were asked to evaluate. Finally, and perhaps most interestingly, it could be that when interacting, participants in Study 1 have had access to more information about the behaviour of their partner than participants in Study 2 had as mere observers of the interaction. The cognitive mechanisms behind this augmentation of social capacities in the online case (Schilbach, 2014) remain poorly understood: in particular, debate is ongoing whether it involves certain forms of social knowing or “we-mode” representations primed within the individual by the context of the interaction (Gallotti & Frith, 2013), or whether the added social knowledge directly resides in the dynamics of the interaction (De Jaegher, Di Paolo, & Gallagher, 2010). The performance discrepancy between the Study 1 and Study 2 suggests that variations around the same paradigm provide a way to study differences between such interaction-based and observation-based processes (Schilbach, 2014).

For third-party observers, the information processing of dual-channel cues suggests a capacity for processing two simultaneous stimuli associatively. This capacity does not seem of the same nature as the information-processing of the single-channel expressive cues usually described in the music cognition literature, but rather seems related to that of co-representation (the capacity of people to simultaneously keep track of the actions and perspectives of two interacting agents Gallotti & Frith, 2013). In the visual domain,

co-representational processes occur implicitly at the earliest stages of sensory processing (Neri et al., 2006; Samson, Apperly, Braithwaite, Andrews, & Bodley Scott, 2010) and are believed to be an essential part in many social-specific processes, such as imitation, simulation (Korman, Voiklis, & Malle, 2015), sharing of action repertoire and team-reasoning (Gallotti & Frith, 2013). In the music domain however, evidence for co-representations of two real or virtual agents has so far been mostly indirect, with music mediating the co-representation of visual stimuli: in Moran et al. (2015), observers were able to discriminate whether wire-frame videos of two musicians corresponded to a real, or simulated, interaction by using the music as a cue to detect social contingency; similarly, in Edelman and Harring (2015), the presence of background music lead participants to judge agents in a video-taped group to be more affiliative. In contrast, the present data establishes a causal role for co-representation to not only detect social contingency, but to qualify a variety of social relations, and this based on the sole musical signal without the mediation of any visual stimuli. Our results therefore suggest that co-representational processes, and with them the ability to cognize social relations between agents, are entailed by music cognition in general, and not only by specialized social tasks that *happen to involve* music. The question remains, however, of what exactly in music is the target of such processes. The following two studies examine two aspects of the musical signal that appear to be involved: temporal coordination (Study 3) and vertical/harmonic coordination (Study 4).

Study 2 also compared musically- and non musically-trained listeners and found a robust effect of musical expertise: only musicians were able to utilize dichotomically-presented joint action cues to interpret social intent from musical interactions; non-musicians here essentially listened to duets as if they were soli. Several factors may account for this result. First, it is possible that musicians and non-musicians differed in socio-educational background, general cognitive or personality characteristics, and that

musical expertise covaried with interactive skills unlinked to musical ability. Second, it is possible that the capacity to decode social cues from acoustically-expressed coordination of non-verbal material is acquired culturally, as part of a general set of musical skills. Finally, it is possible that the single-channel cues involved in the stimuli used in Study 2 were so salient for non-musicians that there was too much competition to recruit the more effortful co-representational processes necessary to assess dual-channel cues. If so, it is possible that, when placed in a situation where they can compare stimuli with identical single-channel cues which only differed by dual-channel cues, even non-musicians could demonstrate evidence of co-representational processing (for a similar argument in the visual domain, see Samson et al., 2010). Study 5 will later test this hypothesis.

#### 4. Study 3: acoustical analyses (temporal coordination)

To examine whether dual-channel cues of temporal coordination (simultaneous playing and leader/follower patterns) covaried with the types of social relations encoded in the interactions, we subjected the stimuli of Study 1 to acoustical analyses. The results of Study 3 will be discussed jointly with Study 4.

##### 4.1. Methods

**Stimuli:** We analysed the  $n = 64$  duets that were decoded correctly by the musicians participating in Study 1 (domineering: 14, insolent: 9, disdainful: 13, conciliatory: 12, caring: 16).

**Simultaneous playing time:** In each duet, we measured the amount of simultaneous playing time between musician A (the “encoder”) and musician B (the “decoder”). In both channels, we calculated Root-mean-Square (RMS) energy on successive 10-ms windows. Simultaneous playing time was estimated as the number of frames with RMS energy larger or equal to 0.2 ( $z$ -score) in both channels.

**Granger causality:** In each duet, we measured the pair-wise conditional Granger causality between the two musicians’ time-aligned series of RMS energy, as a way to estimate leader/follower situations. Prior to Granger causality estimation, time-series of 10-ms RMS values were  $z$ -scored, median-filtered to 100 ms time-windows and derivated to ensure covariance-stationarity. Granger causality was estimated using the GGCA toolbox (Seth, 2010). All series passed an augmented Dickey-Fuller unit-root test of covariance-stationarity at the .05 significance level. Series were modelled with autoregressive models with AIC order estimation (mean order  $M = 7$ , i.e. 700 ms), and Granger causalities were estimated with pairwise  $G$ -magnitudes and  $F$ -values. For analysis, duets were considered to have “encoder-to-decoder” causality either if only the A-to-B  $F$ -value reached .05 significance level or if both A-to-B and B-to-A  $F$ -values reached significance and A-to-B  $G$ -magnitude was larger or equal than B-to-A  $G$ -magnitude (converse criteria for “decoder-to-encoder”); duets were labeled “no causality” if neither directional  $F$ -values reached significance).

**Statistical analysis:** Difference of simultaneous playing time across attitudes was tested with a one-way ANOVA (5 levels). The distribution of the three categories of pair-wise conditional causality (encoder-to-decoder, decoder-to-encoder, no-causality) was tested for goodness of fit to a uniform distribution across the 5 attitudes using Pearson’s  $\chi^2$ .

##### 4.2. Results

Fig. 2a shows an analysis of simultaneous playing time in a prototypical disdainful interaction, and Fig. 2b summarizes the amount of simultaneous playing time in all 5 attitudes. Simultane-

ous play between musician A and musician B ( $M = 39\%$ ,  $SD = 17\%$ ) differed with the encoded attitude, with caring and domineering prompting longer overlap than insolent, disdainful and conciliatory ( $F(4,59) = 2.56$ ,  $p = .05$ ). Sustained playing together with the interlocutor appeared to convey supportive behaviours and interventions restricted to the other’s silent parts suggested contradiction, or active avoidance (see Fig. 2a).

Fig. 3a shows an analysis of RMS Granger-causality from musician A to musician B in a prototypical domineering interaction, and Fig. 3b summarizes how different types of causality were distributed across the attitudes. These were non uniformly distributed across attitudes ( $\chi^2(8) = 19.21$ ,  $p = .01$ ). Domineering interactions were characterized by predominant encoder-to-decoder causalities (7/14), indicating that musician A tended to precede, and musician B submit, during simultaneous play (Fig. 3a); in contrast, caring interactions were mostly decoder-to-encoder (9/16), suggesting musician A’s tendency to follow/support rather than lead. In addition, musician A allowed almost no causal role for musician B in either the insolent (1/16) and disdainful (0/13) interactions.

#### 5. Study 4: psychoacoustical analysis (spectral/harmonic coordination)

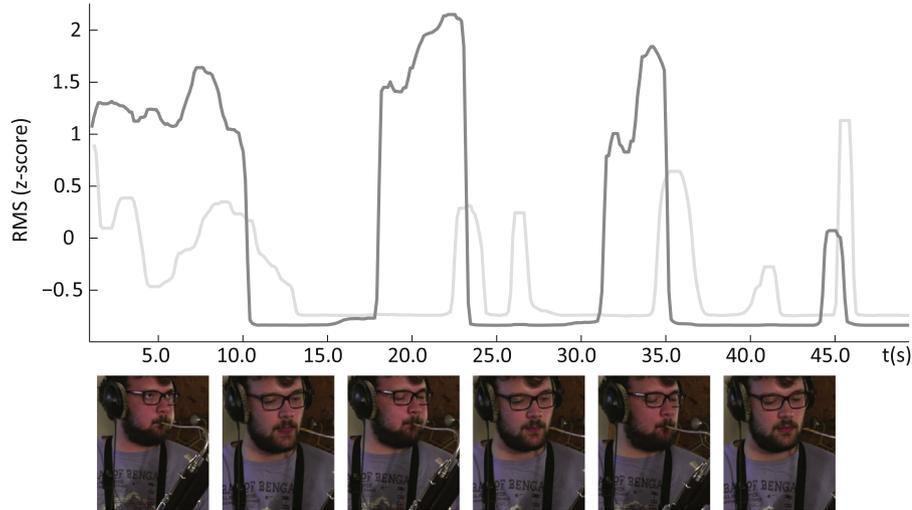
To examine whether dual-channel cues of vertical coordination (consonance/dissonance between the two musicians, and more generally spectral/harmonic patterns of coordination) covaried with the types of social relations encoded in the interactions, we conducted a free-sorting psychoacoustical study of the interactions with an additional group of musician listeners.

##### 5.1. Methods

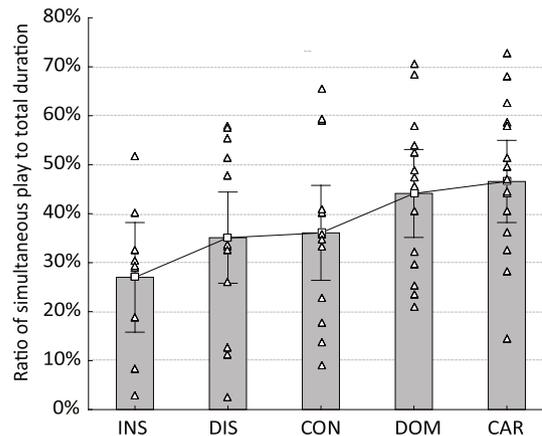
**Stimuli:** In order to let participants judge the vertical aspects of coordination in the interactions, we masked the temporal dynamics of the musical signal using a shuffling procedure. Audio montages of the same  $n = 64$  duets as Study 3 were generated by extracting sections of simultaneous playing time and cutting them into 500-ms windows with 250-ms overlap; all windows were then randomly permuted in time; the resulting “shuffled” version of each duet was then reconstructed by overlap-and-add, and cut to include only the first 10 s of material. The resulting audio preserved the vertical relation between channels at every frame, but lacked the cues of temporal coordination already analysed in Study 3. One duet (4\_9) was removed because it did not include sufficient simultaneous playing.

**Procedure:** The remaining  $n = 63$  were presented dichotically (musician A, the “encoder”, on the left; musician B, the “decoder”, on the right) to  $N = 10$  expert music listeners, who were tasked to categorize them with a free sorting procedure according to how well the two channels harmonized with one another. Groupings and annotations for the groupings were collected using the Tcl-LabX platform (Gaillard, 2010). Extracts were represented by randomly numbered on-screen icons, and could be heard by clicking on the icon. Icons were initially presented aligned at the top of the screen. Participants were instructed to first listen to the stimuli in sequential random order, and then to click and drag the icons on-screen so as to organize them into as many groups as they wished. Participants were told that the stimuli were electroacoustic compositions composed of two simultaneous tracks, and were asked to organize them into groups so as to reflect their “degree of vertical organization in the spectral/harmonic space”. Groups had to contain at least one sound. Throughout the procedure, participants were free to listen to the stimuli individually as many times as needed. It

(a) Time-series of RMS energy in a prototypical DIS interaction, with time-aligned photographs of the encoder. Dark grey : encoder (bassoon) ; light grey : decoder (alto saxophone).



(b) Proportion of simultaneous play (as a ratio of total duration), categorized by encoded intent



**Fig. 2.** (Study 3) Acoustical analysis of proportion of simultaneous playing time. In prototypical disdainful interactions, encoders stopped playing during the decoder's interventions to suggest active avoidance (a). Simultaneous play was maximum in caring (i.e., supporting) and domineering (i.e., monopolizing) interactions (b). Abbreviations: CAR: caring; CON: conciliatory; DIS: disdainful; DOM: domineering; INS: insolent.

took each participant approximately 20 min to do the groupings. After the participants validated their groupings, they were asked to type a verbal description of each group, in the form of a free annotation. Participants were left unaware that the stimuli were in fact dyadic improvisations, and that they corresponded to specific social attitudes.

**Participants:** Musicians who participated in the study ( $N = 10$ , male: 9,  $M = 30.1$ yo,  $SD = 2.6$ ) were students recruited from the composition degree ("classe de composition") of the Paris and Lyon CNSMD.

**Analysis:** Duets groupings for each participant were analysed as 63-dimension binary co-occurrence matrices. All  $N = 10$  matrices were averaged into one, which was subjected to complete-linkage hierarchical clustering, yielding a dendrogram that was cut at the top-most 3-cluster level. Grouping annotations for each participant were encoded manually by extracting and normalizing all adjectives, adverbs and names describing the harmonic coordination of the stimuli, yielding 100 distinct tags for which we counted the number of occurrences in each of the 3 clusters. The most discriminative tags were then selected to describe each cluster according to their TF/IDF score.

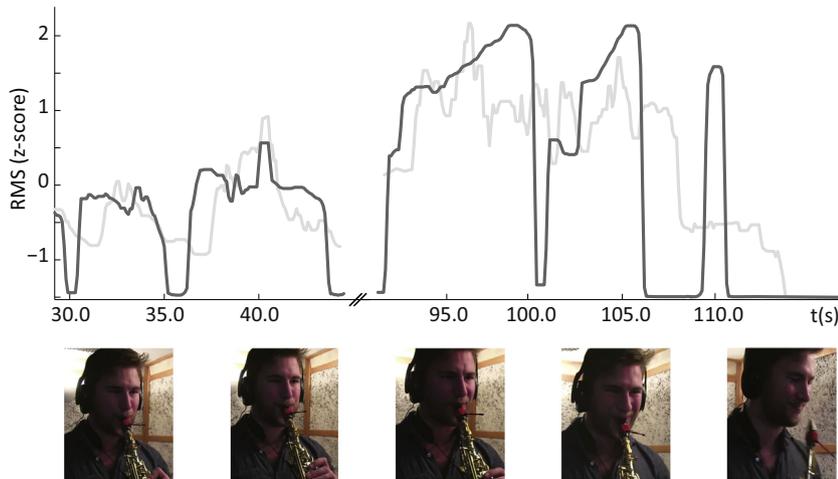
## 5.2. Results

The  $n = 63$  extracts grouped by participants clustered into three types of vertical coordination in spectral/harmonic space. Fig. 4b shows the distribution of the three clusters across the 5 types of social behaviours encoded in the original music. The first group of duets ( $n = 7$ ) was annotated as sounding "identical, fusion-like, unison, similar, homogeneous and mirroring", and was never used in domineering, insolent and disdainful interactions. The second group of duets ( $n = 23$ ) was annotated as "different, independent, contrasted, separated and opposed", and was never used in caring interactions. The third group ( $n = 33$ ) was annotated as "different, but cooperating, complementary, rich and multiple", and was most common in conciliatory interactions. As an illustration, Fig. 4a shows patterns of consonance/dissonance in one such prototypical conciliatory interaction.

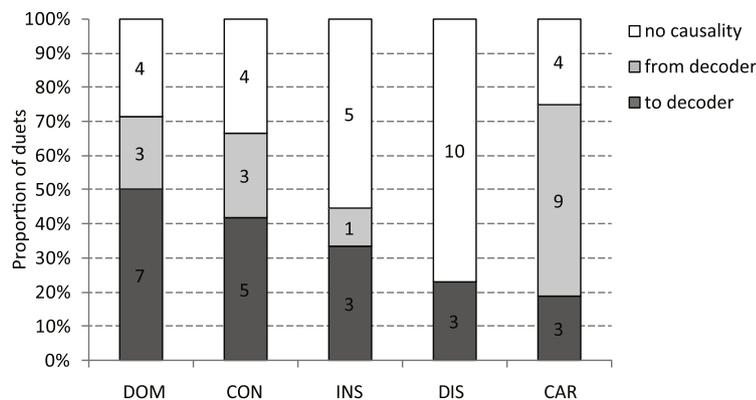
## 5.3. Discussion (Studies 3 & 4)

Together, the computational and psychoacoustical analyses of Study 3 & 4 have shown that the types of relational intentions

(a) Time-series of RMS energy in a prototypical domineering interaction, with time-aligned photographs of the encoder’s facial expression while playing. Dark grey : encoder (soprano saxophone); light grey : decoder (piano).



(b) Statistically-significant conditional Granger causalities between the musician’s time-series of RMS energy, categorized by encoded intent. Dark grey : encoder-to-decoder causalities ; light grey : decoder-to-encoder causalities ; white : non-significant in either direction



**Fig. 3.** (Study 3) Acoustical analysis of Granger causality. In prototypical domineering interactions, sound energy in the encoder’s channel was the Granger-cause for the energy in the decoder’s channel (a). Decoder-to-encoder causalities were maximum in caring interactions, while practically absent from insolent or disdainful interactions (b). Abbreviations: CAR: caring; CON: conciliatory; DIS: disdainful; DOM: domineering; INS: insolent.

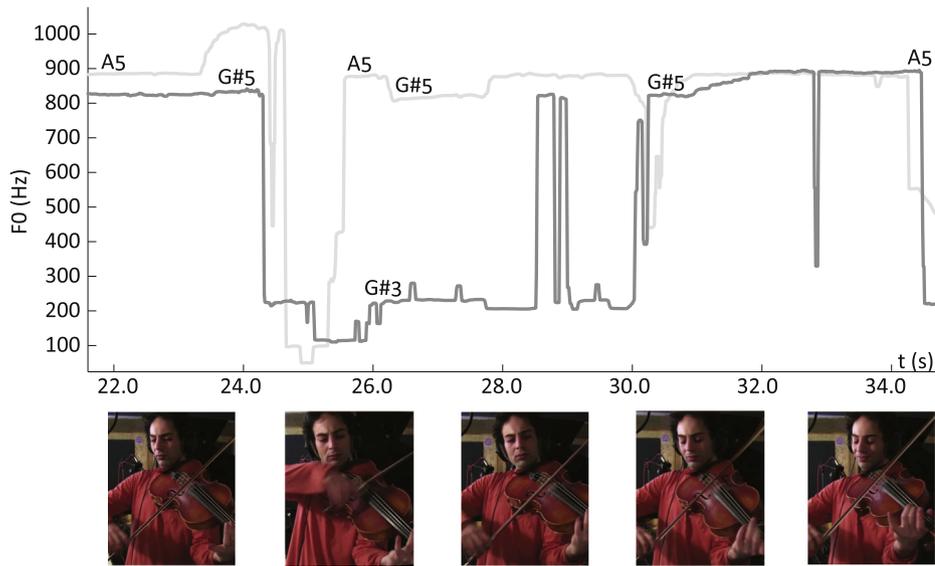
communicated in Study 1 and 2 covaried with the amount of simultaneous playing time, the direction of Granger causality of RMS energy between the two musicians’ channels, and their degree of harmonic/spectral complementarity. For instance, domineering interactions were characterized by long overlaps, encoder-to-decoder causality and no occurrence of harmonic fusion; insolent and disdainful interactions were associated with short overlaps, no decoder-to-encoder causality, and harmonic contrast; caring, with long overlaps, decoder-to-encoder causality, and no occurrences of harmonic contrast.

It is not the first time that leader/follower patterns are studied in musical joint-action, using methods similar to the Granger causality metric used in Study 3. Mutual adaption between two simultaneous finger-tapping agents was studied using forward (“B adapts to A”) and backward (“A adapts to B”) correlation (Konvalinka et al., 2010) and phase-correction models (Wing et al., 2014). Granger causality between visual markers of interactive musicians was used to quantify information transfer between

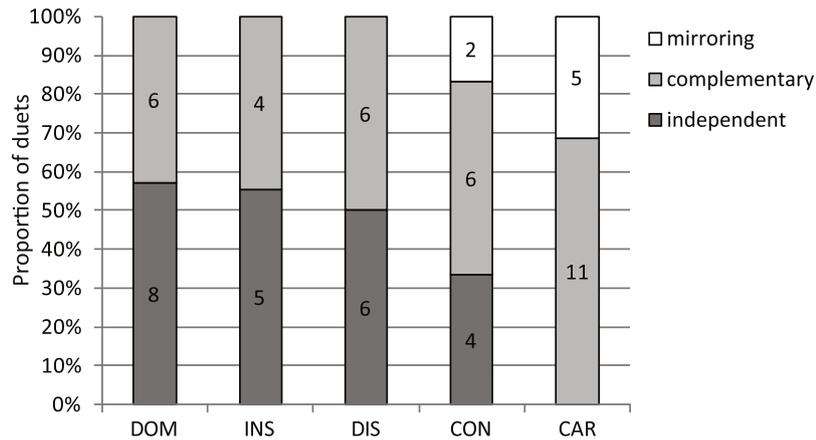
conductor and musicians in an orchestra (D’Ausilio et al., 2012) and between the first violin and the rest of the players in a string quartet (Badino, D’Ausilio, Glowinski, Camurri, & Fadiga, 2014); the same techniques were used on audio data in Papiotis, Marchini, Perez-Carrillo, and Maestre (2014, 2016). Much of this previous work, however, struggled with low sample size: comparing two performances of the same ensemble (D’Ausilio et al., 2012; Papiotis et al., 2014), or two ensembles on the same piece (Badino et al., 2014). Here, based on more than 60 improvised pieces from our corpus, we could use the same analysis to show robust patterns of how these cues are used in various social situations.

In addition, leader/follower patterns in the musical joint action literature have been mostly studied as an index of goal-directed division of labour when task demands are high (Hasson & Frith, 2016). For instance in finger-tapping, leader/follower situations emerge as a stable solution to the collective problem of maintaining both a steady beat and synchrony between agents (Konvalinka et al., 2010). Here we show that these patterns are not only emerg-

(a) Time-series of fundamental frequency in a prototypical CON interaction, with time-aligned photographs of the encoder. Dark grey : encoder (viola); light grey : decoder (violin).



(b) Psychoacoustical categories of harmonic coordination between the encoder and the decoder during sections of simultaneous play, categorized by encoded intent. Dark grey : independent, opposed simultaneous play; light grey : complementary, cooperative play; white : mirroring, unison-like play.



**Fig. 4.** (Study 4) Psychoacoustical analysis of harmonic coordination. In prototypical conciliatory interactions, encoders initiated motions from independent (e.g., maintaining a dissonant minor second G#-A interval) to complementary (e.g., gradually increasing G# to a maximally consonant unison on A) simultaneous play with the decoder (a). Strongly mirroring simultaneous play was never used in domineering, insolent and disdainful interactions (b). Abbreviations: CAR: caring; CON: conciliatory; DIS: disdainful; DOM: domineering; INS: insolent.

ing features of a shared task, but that they are involved in the realization of a communicative intent, and that they constitute social information that is available for the cognition of observers.

That cues of temporal coordination should be involved in social observation of musical interactions is reminiscent of a large body of literature on beat entrainment: the coordination of movements to external rhythmic stimuli is a well-studied biological phenomenon (Patel, Iversen, Bregman, & Schulz, 2009), and has been linked in humans to a facilitation of prosocial behaviours and group cohesion (Cirelli et al., 2014; Wiltermuth & Heath, 2009). The leader/follower patterns of Study 3, however, go beyond simple alignment, reflecting continuous mutual adaptation and complementary behaviours that evolve with time and the social

intentions of the interacting musicians (see Fig. 3a). The present results also go beyond the purely affiliatory behaviours that have been so far discussed in the temporal coordination literature: in our data, the type of temporal causality observed between musicians co-occurred with a variety of attitudes, and rather appeared to be linked to control: causality from musician A for high control interactions (domineering), causality from musician B for low control interactions (caring), and an absence of significant causality in attitudes that were non-affiliatory, but neutral with respect to control (disdainful, insolent).

If temporal coordination is a well-studied aspect of musical joint action, less is known from the literature about the link found in Study 4 between harmonic coordination and social observation.

Group music making and choir singing have been linked to social effects such as increased trustworthiness (Anshel & Kipper, 1988) and empathy (Rabinowitch et al., 2013), but it is unclear whether these effects are mediated by the temporal or harmonic aspects of the interactions (or even the simple fact that these are all group activities). Here, we found that the degree of vertical coordination (or the “harmony” of the musicians’ tones and timbres) covaried with the degree of affiliation of the social relations encoded in the interactions: mirrored for high affiliation attitudes (caring, conciliatory) and opposed for low affiliation (disdainful, insolent). That harmonic complementarity could be a cue for social observation in musical situations is particularly interesting in a context where preference for musical consonance was long believed to be very strongly biologically determined (Schellenberg & Trehub, 1996), until recent findings have shown that it is in fact largely culturally-driven (McDermott, Schultz, Undurraga, & Godoy, 2016).

In sum, Study 2 established that listeners relied on dual-channel information to decode the behaviours, and Study 3 and 4 have identified temporal and harmonic coordination as two dual-channel features that co-varied with them. It remains an open question, however, whether observers actually make use of these two types of cues (i.e. if temporal and harmonic coordination are indeed social *signals* in our context). By manipulating cues directly in the music, Study 5 will attempt to establish causal evidence that it is indeed the case.

## 6. Study 5: effect of audio manipulations on ratings of affiliation and control

In this final study, we used simple audio manipulations on a selection of duets from the corpus of Study 2 to selectively enhance or block dual-channel cues of temporal and harmonic coordination, thus testing for a causal role of these cues in the social cognition of the musical interactions. To manipulate temporal coordination, we time-shifted the two musicians’ channels in opposite directions, in order to inverse leader/follower patterns. To manipulate harmonic coordination, we detuned one musician’s channel from interactions that had a large degree of consonance, in order to make the interactions more dissonant. Because these two manipulations only operated on the relation between channels, they did not affect the single-channel cues of each channel considered in isolation.

We tested the influence of the two manipulations on dimensional ratings of the social behaviour of musician A (the “encoder”) towards musician B (the “decoder”), made along the two dimensions of affiliation (the degree to which one is inclusive or exclusive towards the other) and control (the degree to which one is domineering or submissive towards the other) (Pincus et al., 1998). For temporal coordination, we predicted that  $\mathcal{H}_1$ : low-control interactions (e.g. caring duets) in which channels were time-shifted to increase encoder-to-decoder causality would be evaluated higher in control than the original; conversely, that  $\mathcal{H}_2$ : high-control interactions (e.g. domineering duets) in which channels were time-shifted to increase decoder-to-encoder causality would be evaluated lower in control than the original; and that  $\mathcal{H}_3$ : both manipulations of temporal causality would leave ratings of affiliation unchanged. For harmonic coordination, we predicted that  $\mathcal{H}_4$ : high-affiliation interactions (e.g. caring and conciliatory duets) in which one channel was detuned to increase dissonance would be evaluated lower in affiliation than the original, and that  $\mathcal{H}_5$ : the manipulation of consonance would not affect ratings of control. Note that for practical reasons, we could not test the opposite manipulation (making a dissonant interaction more consonant). In addition, participants were non-musicians, trusting that the discrepancy seen in Study 2 was an effect of the decoding task that would disappear in this new setup.

## 6.1. Methods

**Participants:**  $N = 24$  non-musicians participated in the study (female: 17,  $M = 21.2$ yo,  $SD = 2.7$ ). All were undergraduate students recruited via the INSEAD-Sorbonne Behavioral Lab. Two participants were excluded because they declared not understanding the instructions and were unable to complete the task.

**Stimuli:** We selected 15 representative shorts extracts ( $M = 24$ s.) from the set of correctly decoded recording from Study 1, and organized them in 3 groups: one (CTL+) corresponding to attitudes with originally low levels of control (caring: 5) for which we wanted to test an increase of control with an increase of encoder-to-decoder causality; one (CTL-) corresponding to attitudes with originally high levels of control (domineering: 5) for which we wanted to test a decrease of control with a decrease of encoder-to-decoder causality; and one (AFF-) corresponding to attitudes with originally high levels of affiliation (caring: 3, conciliatory: 2) for which we wanted to test a decrease of affiliation with decreased harmonic coordination.

Stimuli in the CTL+ group were manipulated in order to shift musician B backward in time by a few seconds while preserving the timing of musician A. Because extracts in the CTL+ group corresponded to *caring* attitudes, the original recordings had a primarily decoder-to-encoder temporal causality (Study 3), with musician A reacting to, rather than preceding, musician B. By shifting musician B backward in time (or, equivalently, shifting musician A ahead of musician B), we intended to artificially reverse this causality by creating situations where musician A was perceived to temporally lead, rather than follow, the interaction. Time shifting was applied with the Audacity software, by inserting the corresponding amount of silence at the beginning of the track. Because the amount of time shifting was small ( $M = 2.3$ s; exact delay was determined for each recording based on the speed and phrasing of the music) compared to the typical rate of harmonic changes in a recording, the manipulation only affected the extract’s temporal coordination, but preserved its degree of harmonic coordination as well as the single-channel cues of both channels considered separately.

Stimuli in the CTL- group were manipulated with the inverse transformation, i.e. shifting musician B ahead of musician A by a few seconds. Because extracts in the CTL- group corresponded to *domineering* attitudes, the original recordings had a primarily encoder-to-decoder causality, which this manipulation was intended to artificially reverse and create situations where musician A was perceived to temporally follow, rather than lead, the interaction.

Finally, stimuli in the AFF- group were manipulated in order to detune musician A by one or more semitones while preserving the tuning of musician B, thus creating artificial dissonance (or decreasing harmonic coordination) between the two channels. Pitch-shifting was applied with the Audacity software (Team, 2016), using an algorithm which preserves the signal’s original speed. Therefore, the manipulation only affected the interaction’s vertical coordination, but preserved both cues of simultaneous playing time and temporal causality, as well as the single-channel cues of both channels considered separately. See Table 2 for details of the manipulations in the three groups.

**Procedure:** The  $n = 30$  extracts (15 original and 15 manipulated versions) were presented in two semi-random blocks of  $n = 15$  trials, separated with a 3-min pause, so that the two versions of a same recording were always in different blocks. Stimuli were presented dichotically as in the stereo condition of Study 2 (musician A on the left, musician B on the right), and participants were asked to judge the attitude of musician A with respect to musician B, using two 9-point Likert scales for affiliation (not at all affilia-

**Table 2**  
(Study 5) Stimuli in the high-affiliation (AFF-), low-control (CTL+) and high-control groups (CTL-), with mean ratings of affiliation and control obtained before and after acoustic manipulation (pitch shifting in AFF-, time shifting in CTL+ and CTL-). Abbreviations: CON: conciliatory; CAR: caring; DOM: domineering; s.t.: semitone; s: second.

Extracts	Musician A	Musician B	Attitude	Before		Manipulation		After	
				Affiliation (M)	Control (M)	Start-End (Dur.)	Transf.	Affiliation (M)	Control (M)
<i>AFF- GROUP</i>									
1_1	Piano	Saxophone	CON	6.7	3.7	1:00–1:14 (0:14)	A + 1 s.t.	<b>6.6</b>	4.5
2_9	Piano	Saxophone	CAR	7.0	5.3	2:11–2:53 (0:42)	A + 1 s.t.	<b>6.7</b>	5.6
5_5	Viola	Violin	CAR	5.5	5.3	0:17–0:48 (0:31)	A + 1 s.t.	<b>4.6</b>	4.6
5_9	Viola	Violin	CON	5.0	4.4	0:53–1:23 (0:30)	A + 2 s.t.	<b>4.5</b>	5.3
5_10	Violin	Viola	CAR	5.7	4.5	0:38–0:58 (0:20)	A + 1 s.t.	<b>5.3</b>	3.9
<i>CTL+ GROUP</i>									
3_7	Bassoon	Saxophone	CAR	5.3	3.9	0:11–0:32 (0:21)	B + 2s.	4.7	<b>4.8</b>
7_3	Guitar	Clarinet	CAR	6.8	4.0	0:29–0:46 (0:18)	B + 1.4s.	4.1	<b>4.8</b>
8_2	Saxophone	Dbl bass	CAR	6.0	3.3	1:43–2:25 (0:42)	B + 5s.	5.9	<b>4.0</b>
9_10	Viola	Double bass	CAR	6.0	3.4	2:02–2:35 (0:33)	B + 3s.	6.0	<b>3.8</b>
10_5	Flute	Viola	CAR	5.2	4.2	0:04–0:25 (0:21)	B + 2s.	5.4	<b>4.5</b>
<i>CTL- GROUP</i>									
1_7	Piano	Saxophone	DOM	2.8	6.3	0:35–0:44 (0:09)	A + 1s.	2.5	<b>5.5</b>
1_7	Piano	Saxophone	DOM	2.7	6.0	0:56–1:11 (0:15)	A + 3s.	2.5	6.5
3_6	Saxophone	Bassoon	DOM	3.4	6.0	0:18–0:33 (0:15)	A + 0.8s.	3.3	<b>5.7</b>
3_9	Bassoon	Saxophone	DOM	4.2	6.8	0:44–1:12 (0:28)	A + 3s.	3.2	6.8
4_8	Saxophone	Piano	DOM	5.5	4.3	0:22–0:45 (0:23)	A + 2.5s.	5.5	<b>3.8</b>

Mean affiliation and control ratings which changed in the predicted direction after manipulation are indicated in bold.

tory...very affiliatory) and control (not at all controlling...very controlling).

To help participants keep track of the relation between the simultaneous music channels, each trial had visual information which reinforced the auditory stimuli: image and name of instrument A in the left-hand side, image and name of instrument B in the right-hand side, and an arrow going from left to right illustrating the direction of the relation that had to be judged.

To help participants judge the interactions using the two scales of affiliation and control, participants were given detailed explanations about the meaning of each dimension (see [SI Text 2](#)). In addition, at the beginning of the experiment, they were shown an example video in which a male (standing on the left) and a female comedian (standing on the right) act up an argument to the ongoing sound of Symphony No. 5 in C minor of Ludwig van Beethoven (see [SI Video 2](#)). Snapshots of the video were also used as trailing examples to illustrate the instructions ([SI Fig. 2](#)), and participants were encouraged to think of the extracts “as the soundtrack of a video similar to the one they were just shown”, judging the attitude of an hypothetical left-standing male comedian with respect to a right-standing female comedian.

**Statistical analysis:** Mean ratings for affiliation and control were obtained for each participant over the 5 non-manipulated and the 5 manipulated trials in each stimulus group, and compared separately within each group using a paired *t*-test (repeated factor: before/after manipulation).

## 6.2. Results

Affiliation and control ratings, before and after manipulation, are given for all three groups in [Table 2](#). In the AFF- group, A-musicians (“encoders”) involved in interactions with decreased harmonic coordination (detuned extracts) were judged significantly less affiliative to their partner ( $M = 5.5$ ) than in the original, non-manipulated interactions ( $M = 6.0$ ,  $t(21) = -2.50$ ,  $p = .02$ ), while the manipulation did not affect their perceived level of control ( $t(21) = 0.72$ ,  $p = .48$ ; see [Fig. 5](#)). In the CTL+ group, A-musicians in interactions with increased encoder-to-decoder causality (i.e. shifted ahead of their partner) were judged significantly more controlling to their partner ( $M = 4.4$ ) than in the original interactions ( $M = 3.7$ ,  $t(21) = 2.50$ ,  $p = .02$ ), while the manipulation did not

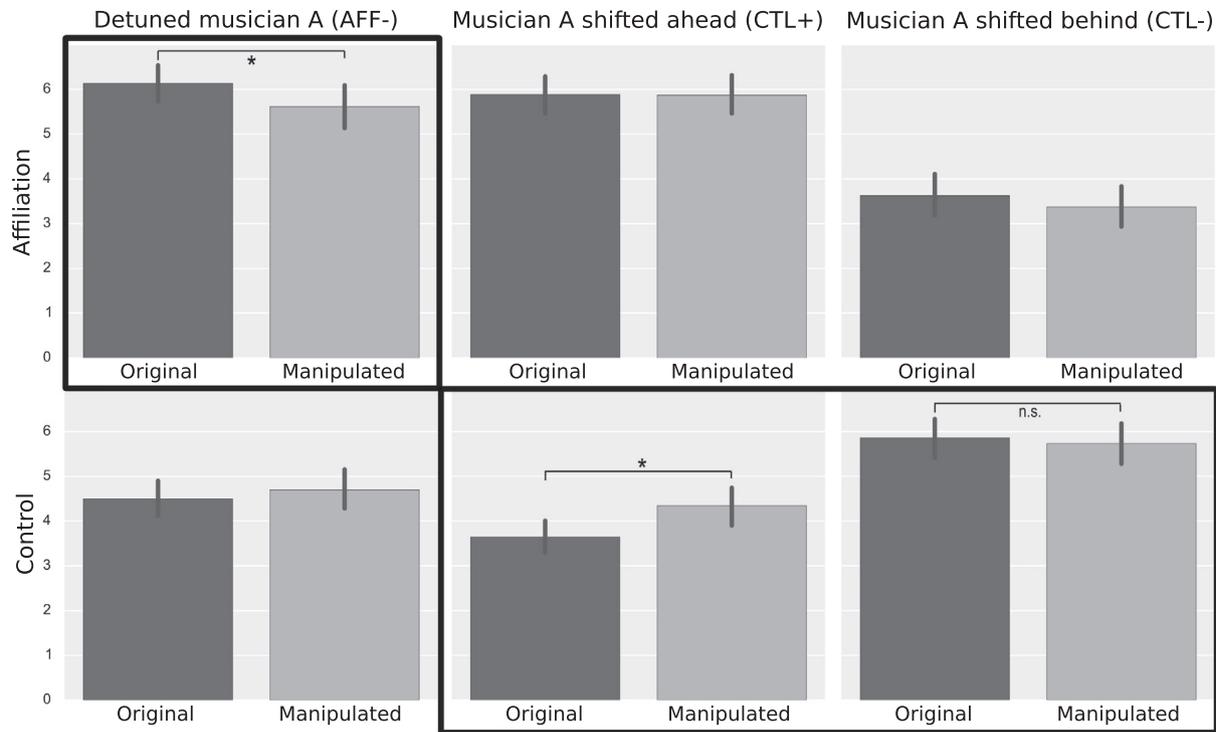
affect their perceived level of affiliation ( $t(21) = -0.09$ ,  $p = .93$ ). In the CTL- group, the effect of the manipulation, while also in the predicted direction, was neither significant for control ( $t(21) = -0.45$ ,  $p = .65$ ) nor for affiliation ( $t(21) = -1.15$ ,  $p = .26$ ).

## 6.3. Discussion

Study 5 manipulated dual-channel cues of temporal and harmonic coordination in a selection of duets, and tested the impact of the manipulations on non-musicians’ judgements of affiliation and control. For harmonic coordination, we predicted that increased dissonance would decrease affiliation ratings ( $\mathcal{H}_4$ ) while not affecting control ( $\mathcal{H}_5$ ); both predictions were verified. For temporal coordination, we predicted that higher encoder-to-decoder causality would increase control ratings ( $\mathcal{H}_1$ ) while not affecting affiliation ( $\mathcal{H}_3$ ), and that lower encoder-to-decoder causality would decrease control ( $\mathcal{H}_2$ ); both  $\mathcal{H}_1$  and  $\mathcal{H}_3$  were verified, but  $\mathcal{H}_2$  wasn’t.

By manipulating these cues and verifying the above predictions, Study 5 therefore establishes a direct causal relationship between both temporal and spectral/harmonic coordination in music and non-musicians’ judgements of the level of affiliation and control in the underlying social relations. It demonstrates that these two types of cues constitute social signals for our observers, and suggests that they explain, at least in part, the gain in decoding performance observed in Study 2 when judging stereo compared to mono extracts. In addition, these results provide novel mechanistic insights into how human listeners process social relations in music: somehow, even naive, non-musically trained listeners are able to co-represent simultaneous parts of the signal, and extract social-specific information from their fluctuating temporal causality on the one hand, and their degree of harmonic complementarity on the other hand. The former type of cue is used to compute judgements of social control, differentiating e.g. between domineering and caring interactions; the latter, to compute judgements of social affiliation, discriminating e.g. caring from disdainful.

It is important to stress that dual-channel coordination cues are obviously not the only source of information for the listeners’ mind-reading of the interacting agents. Rather, they are processed in complementarity to the single-channel, “prosodic” cues contained in the expression of each individual musicians (e.g. loud,



**Fig. 5.** (Study 5) Affiliation and control judgements of the attitude of musician B, before (“original”) and after (“manipulated”) three types of acoustic manipulations: decrease harmonic coordination (detuned musician A, AFF- condition), increased encoder-to-decoder causality (musician A shifted ahead, CTL+ condition), decreased encoder-to-decoder causality (musician A shifted behind, CTL- condition). Conditions enframed in dark lines correspond to predicted differences (e.g., AFF- on affiliation), and others to predicted nulls (e.g., AFF- on control). Hypothesis tests of paired difference with uncorrected *t*-tests, \* indicating significance at the  $p < .05$  level, error bars represent 95% CI on the mean.

fast music by musician A to convey domination or insolence). These single-channel cues notably allowed both musicians and non-musicians to show better-than-chance performance even in the mono condition of Study 2. In our view, they also provide a possible explanation why manipulating temporal coordination in the high-control group did not affect control ratings in Study 5 ( $\mathcal{H}_2$ ). When comparing single-channel cues of RMS energy in the signal from musician A across our corpus, domineering and insolent interactions are found associated with higher signal energy (Root-mean-square (RMS);  $F(4,59) = 12.58$ ,  $p = .000$ ) and higher RMS variations ( $F(4,59) = 12.3$ ,  $p = .000$ ) than the other attitudes (see *SI Text 3* for details). Because the temporal manipulation of Study 5 did not affect these single-channel cues, it is possible that their saliency in the CTL- group prevented participants from processing more subtle dual-channel cues in order to modulate their judgements of control. In other words, while it was possible for our manipulated musical interactions to lead with a whisper ( $\mathcal{H}_1$ ), it was harder for them to appear submissive while shouting ( $\mathcal{H}_2$ ).

In Study 2, non-musicians had seemed unable to use dual-channel cues to recognize attitudes in stereo extracts. We hypothesized that this was not a consequence of different cognitive strategies, but rather of the experimental characteristics of the task. Study 5 confirmed this interpretation by showing that non-musicians were indeed sensitive to these cues when judging affiliation and control. Several experimental factors may explain this discrepancy. First, it is possible that the task of recognizing categorical attitudes in music (Study 2) carried little meaning for non-musicians, while musicians were able to map these to stereotypical performing situations (caring: accompaniment, domineering: taking a solo) for which they had already formed social-perceptual categories. Second, participant training in Study 5 may have been more appropriate, especially with the addition of

framing the evaluation task with an initial video example. Finally, the combination of rating scales and comparisons between two versions of each duet in Study 5 may have simply decreased experimental noise compared to the hit ratedata of Study 2.

Finally, by collecting dimensional ratings rather than category responses, Study 5 provides more general conclusions than the previous studies, allowing to make predictions for social effects of music that go beyond the five attitudes tested in Study 1–4. For instance, if singing in harmony in a choir improves trustworthiness (an affiliative attitude) or willingness to cooperate (a non-controlling attitude) in a group (Anshel & Kipper, 1988), one may predict that degrading the consonance of the musical signal or enforcing a different temporal causality between participants (e.g. using manipulated audio feedback) will reduce these effects. Similarly, if certain social situations create a preference bias for affiliative or controlling relations (e.g. social exclusion and the subsequent need to belong Loersch & Arbuckle, 2013), then music with certain characteristics (e.g. more vertical coordination) should be preferred by listeners placed in these situations. Finally, the link between musical consonance and social affiliation bears special significance in the context of the psychopathology and neurochemistry of affiliative behaviour, and suggests e.g. mechanisms linking the cognition of musical interactions with neuropeptides like vasopressin and oxytocin Chanda and Levitin (2013).

## 7. General discussion

In this work, we created a novel experimental situation in which expert improvisers communicated relational intentions to one another solely through their musical interaction. Study 1 established that participating musicians were able to decode the social attitude of their partner at the end of their interaction, with

performances similar to those observed in traditional speech and music emotion decoding tasks. Study 2 took Study 1 offline and confirmed that both musically- and non musically-trained third-party listeners could recognize the same attitudes in the recordings of the interactions. In addition, by comparing stereo and mono conditions, Study 2 showed that this ability relied, to a sizeable extent (18.1% recognition accuracy) and at least for trained listeners, on dual-channel cues which were not present in any one of the participants' channels considered in isolation. Using a combination of computational and psychoacoustical analyses, Study 3 & 4 showed that the types of relational intentions communicated in Study 1 and 2 covaried with acoustic cues related to the temporal and harmonic coordination between the musicians' channels. Finally, by manipulating these cues in the recordings of the interactions and testing their impact on the social judgements of non-musician listeners, Study 5 established a direct causal relationship between both types of coordination in the music and the listeners' social cognition of the level of affiliation and control in the underlying social relations: namely, that harmony is for affiliation, and time for control.

The fact that social behaviours, such as those of dominating, supporting or scorning a conversation partner, can be conveyed and recognized in purely musical interactions provides vivid support to the recently emerging view that music is a paradigmatically social activity which, as such, involves not only the outward expression of individual mental states, but also direct communication acts through the intentional use of musical sounds (Cook, 2013; Cross, 2014; Turino, 2008). That music should have such “social power” is an important component in many theoretical arguments in favour of music's possible adaptive biological function (Fitch, 2006). However empirical evidence for the possibility to induce or regulate social behaviours with music (e.g., links between group music-making and empathy or prosocial behaviour) so far had been mostly indirect. Here we showed that music does not only mediate social behaviour, but can directly communicate (i.e. encode and be decoded as) social relational intentions. This finding, complete with the computational and mechanistic insights reported here, holds potential to bring new understanding to a variety of musical cognitive phenomena that common approaches, based e.g. on syntactic, emotional or sensorimotor mechanisms, have fallen short of explaining so far. These range from e.g. the link between beat entrainment and prosocial behaviour (Cirelli et al., 2014; Wiltermuth & Heath, 2009) or strong musical experiences and empathy (Egermann & McAdams, 2013; Eerola et al., 2016), to the musical induction of narrative visual imagery (Gabrielsson & Bradbury, 2011; Osborne, 1989) and musical cognitive impairments in populations with autism (Bhatara et al., 2010) or schizophrenia (Wu et al., 2013). Perhaps most importantly, these results open avenues for vastly more diversified views on musical expressiveness than the garden-variety basic emotions (joy, anger, sadness, etc.) that have dominated the recent literature (Eerola & Vuoskoski, 2013; Juslin & Västfjäll, 2008). In short, musical cognition is not only intra-personal, but also inter-personal (Keller, 2014; Keller, Novembre, & Hove, 2014).

Beyond music, it is interesting to compare the coordination cues identified in this work and how they link to social attitudes, to the social processes at play in linguistic conversation pragmatics: in speech, one is expected to give floor when agreeing and take the floor when contradicting (Mast, 2002), a behaviour that is possibly common with other systems of animal signalling (Naguib & Mennill, 2010). Here, on the contrary, sustained playing together with the interlocutor was a typical feature of affiliative behaviours, and well-segregated turn-taking was associated with disdain (cf. Fig. 2a). In addition, interacting musicians systematically manipulated the complementary or contrasting character of their synchronous signalling to suggest e.g. an initial conflict being

resolved in a conciliatory manner (cf. Fig. 4a). On the contrary, in speech, one does not signal affiliation by “talking” simultaneously over one's conversation partner, a major third apart. In sum, the production and perception of these cues likely constituted a different cognitive domain than the interpretation of similar cues in speech.

A closer cognitive parent to the attribution of social meaning to temporal and harmonic coordination in musical interactions may rather be found in the mechanisms involved in non-verbal backchanneling. First, back-channels too can be both simultaneous and affiliative: vocal back-channels for instance (e.g. continuers like “mmhm” or “uhhuh”) are short but very frequent (involved in 30–40% of turn transfers Levinson & Torreira, 2015) and overwhelmingly affiliative (73% of agreements in Levinson & Torreira (2015)). Second, like musical interactions, their temporal dynamics are carefully coordinated with the main channel: gestural feedback expressions, like nods or head shakes, are precisely calibrated with the structure of the interlocutor's speech and can convey a range of affiliative or non-affiliative attitudes (Stivers, 2008). Finally, the co-represented characteristics of the back- and main-channels are informative about the social relational characteristics of the dyad: observing the body language of two conversing agents is sufficient to judge their familiarity or the authenticity of their interaction (McDonnell et al., 2009; Miles et al., 2009; Moran et al., 2015). By providing a precise computational and psychoacoustic characterization of the coordination cues involved in the social information processing of music, and by showing that this capacity is found in both musicians and non-musicians - and therefore quite general - the present work provides mechanistic insights not only into the cognition of musical interactions, but into the social cognition of non-musical auditory interactions in general (Bryant et al., 2016), or even non-verbal interactions as a whole (Trevarthen, 2000). We are especially curious of applications of our musical paradigm to contribute to the debate about online and offline social cognition (Gallotti & Frith, 2013; Schilbach, 2014), as well as to inter-brain neuroimaging techniques such as EEG hyperscanning (Lindenberger, Li, Gruber, & Müller, 2009).

Beyond the cognitive sciences, finally, these results provide empirical grounding to the philosophical debate on musical formalism and expressiveness (Young, 2014), notably giving support to persona-based positions - i.e., that expressive music prompts us to hear it as animated by agency of a certain sort, a more or less abstract musical persona (Levinson, 1996), rather than for resemblance-based positions, which postulate that hearing music as expressive is simply to register the resemblance between the music's contours and the bodily or vocal expressions of such or such mental state (Davies, 1994). These results also support a growing body of critical music theory arguing for a social agentive view of written and improvised musical interactions - that we, in fact, can listen to music as if it embodied social dynamics between real or fictive agents (Maus, 1988), but also that music making itself can be seen as a situation which allows and enables the exploration of a large array of interpersonal relations through sonic means (Higgins, 2012; Monson, 1996; Warren, 2014).

At the root of our hearing music lies our disposition to interpret musical sounds as expressive or communicative actions. Like in a cathartic playground, musical sounds lead us to experience not only the full spectrum of mundane emotions, but also the multiple ways of relating to the other(s) - from alliances to conflicts, from mutual assistance to indifference, from friendship to enmity.

## Acknowledgments

The authors wish to extend their most sincere thanks to Uta and Chris Frith for their support and feedback at various stages of writing this manuscript. We acknowledge Emmanuel Chemla for

kindly suggesting the idea behind Study 5, and thank Hugo Trad who helped with data collection for Study 2. Finally, the authors thank all the improvisers who participated in Study 1 for their inspiring music-making. All data reported in the paper are available on request. The work described in this paper was funded by European Research Council Grant StG-335536 CREAM (to JJA) and Conseil Régional de Bourgogne (to CC).

## Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.cognition.2017.01.019>.

## References

- Alluri, V., Toiviainen, P., Jääskeläinen, I. P., Glerean, E., Sams, M., & Brattico, E. (2012). Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. *Neuroimage*, *59*(4), 3677–3689.
- Anshel, A., & Kipper, D. A. (1988). The influence of group singing on trust and cooperation. *Journal of Music Therapy*, *25*(3), 145–155.
- Badino, L., D'Ausilio, A., Glowinski, D., Camurri, A., & Fadiga, L. (2014). Sensorimotor communication in professional quartets. *Neuropsychologia*, *55*, 98–104.
- Bailey, D. (1992). *Improvisation. Its nature and practice in music*. New York: Da Capo Press.
- Bente, G., Leuschner, H., Al Issa, A., & Blascovich, J. J. (2010). The others: Universals and cultural specificities in the perception of status and dominance from nonverbal behavior. *Consciousness and Cognition*, *19*(3), 762–777.
- Bhatara, A., Quintin, E.-M., Levy, B., Bellugi, U., Fombonne, E., & Levitin, D. J. (2010). Perception of emotion in musical performance in adolescents with autism spectrum disorders. *Autism Research*, *3*(5), 214–225.
- Bigand, E., & Poulin-Charronnat, B. (2006). Are we experienced listeners? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*(1), 100–130.
- Blacking, J. (1973). *How musical is man?* University of Washington Press.
- Bryant, G. A., Fessler, D. M., Fusaroli, R., Clint, E., Aarøe, L., Apicella, C. L., ... Chavez, B. (2016). Detecting affiliation in laughter across 24 societies. *Proceedings of the National Academy of Sciences*, *113*(17), 4682–4687.
- Chanda, M. L., & Levitin, D. J. (2013). The neurochemistry of music. *Trends in Cognitive Sciences*, *17*(4), 179–193.
- Chen, J. L., Penhune, V. B., & Zatorre, R. J. (2008). Moving on time: Brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. *Journal of Cognitive Neuroscience*, *20*(2), 226–239.
- Cirelli, L. K., Wan, S. J., & Trainor, L. J. (2014). Fourteen-month-old infants use interpersonal synchrony as a cue to direct helpfulness. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *369*(1658), 20130400.
- Clarke, E. F. (2014). Lost and found in music: Music, consciousness and subjectivity. *Musicae Scientiae*, *18*(3), 354–368.
- Cook, N. (2013). *Beyond the score: Music as performance*. Oxford University Press.
- Cross, I. (2014). Music and communication in music psychology. *Psychology of Music*, *42*(6), 809–819.
- Dalla Bella, S., Peretz, I., Rousseau, L., & Gosselin, N. (2001). A developmental study of the affective value of tempo and mode in music. *Cognition*, *80*(3), B1–B10.
- D'Ausilio, A., Badino, L., Li, Y., Tokay, S., Craighero, L., Canto, R., ... Fadiga, L. (2012). Leadership in orchestra emerges from the causal relationships of movement kinematics. *PLoS One*, *7*(5), e35757.
- Davies, S. (1994). *Musical meaning and expression*. Cornell University Press.
- De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Does social interaction constitute social cognition? *Trends in Cognitive Sciences*, *14*(10), 441–447.
- Edelman, L. L., & Harring, K. E. (2015). Music and social bonding: The role of non-diegetic music and synchrony on perceptions of videotaped walkers. *Current Psychology*, *34*(4), 613–620.
- Eerola, T., & Vuoskoski, J. K. (2013). A review of music and emotion studies: Approaches, emotion models, and stimuli. *Music Perception: An Interdisciplinary Journal*, *30*(3), 307–340.
- Eerola, T., Vuoskoski, J. K., & Kautiainen, H. (2016). Being moved by unfamiliar sad music is associated with high empathy. *Frontiers in Psychology*, *7*, 1176.
- Egermann, H., & McAdams, S. (2013). Empathy and emotional contagion as a link between recognized and felt emotions in music listening. *Music Perception: An Interdisciplinary Journal*, *31*(2), 139–156.
- Elvers, P. (2016). Songs for the ego: Theorizing musical self-enhancement. *Frontiers in Psychology*, *7*.
- Fitch, W. T. (2006). The biology and evolution of music: A comparative perspective. *Cognition*, *100*(1), 173–215.
- Gabrielsson, A., & Bradbury, R. (2011). *Strong experiences with music: Music is much more than just music*. Oxford University Press.
- Gaillard, P. (2010). Recent development of a software for free sorting procedures and data analysis. In *Third Lam/Incas3 workshop on free sorting tasks and measures of similarities on some structural and formal properties of human categories*, Paris, France. .
- Gallotti, M., & Frith, C. D. (2013). Social cognition in the we-mode. *Trends in Cognitive Sciences*, *17*(4), 160–165.
- Hagen, E. H., & Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Human Nature*, *14*(1), 21–51.
- Hasson, U., & Frith, C. D. (2016). Mirroring and beyond: Coupled dynamics as a generalized framework for modelling social interactions. *Philosophical Transactions of the Royal Society B*, *371*(1693), 20150366.
- Healey, G. (2004). Messiaen and the concept of 'personnages'. *Tempo*, *58*(230), 10–19.
- Hevner, K. (1935). The affective character of the major and minor modes in music. *The American Journal of Psychology*, 103–118.
- Higgins, K. M. (2012). *The music between us: Is music a universal language?* University of Chicago Press.
- Johnson-Laird, P. N., & Oatley, K. (1989). The language of emotions: An analysis of a semantic field. *Cognition and Emotion*, *3*(2), 81–123.
- Justin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, *129*(5), 770.
- Justin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences*, *31*(05), 559–575.
- Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian & R. Timmers (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 260–282). Oxford: Oxford University Press.
- Keller, P., Novembre, G., & Hove, M. (2014). Rhythm in joint action: Psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philosophical Transactions of the Royal Society B*, *369*(20130394).
- Klorman, E. (2016). *Mozart's music of friends: Social interplay in the chamber works*. Cambridge University Press.
- Knoblich, G., Butterfill, S., & Sebanz, N. (2011). Psychological research on joint action: Theory and data. *Psychology of Learning and Motivation—Advances in Research and Theory*, *54*, 59.
- Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. (2010). Follow you, follow me: Continuous mutual prediction and adaptation in joint tapping. *The Quarterly Journal of Experimental Psychology*, *63*(11), 2220–2230.
- Korman, J., Voiklis, J., & Malle, B. F. (2015). The social life of cognition. *Cognition*, *135*, 30–35.
- Levinson, J. (1996). *Musical expressiveness, the pleasures of aesthetics*. Cornell University Press.
- Levinson, J. (2004). Music as narrative and music as drama. *Mind & Language*, *19*(4), 428–441.
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, *6*, 731.
- Lindenberger, U., Li, S.-C., Gruber, W., & Müller, V. (2009). Brains swinging in concert: Cortical phase synchronization while playing guitar. *BMC Neuroscience*, *10*(1), 22.
- Loersch, C., & Ar buckle, N. L. (2013). Unraveling the mystery of music: Music as an evolved group process. *Journal of Personality and Social Psychology*, *105*(5), 777.
- MacDonald, R., & Wilson, G. (2014). Musical improvisation and health: A review. *Psychology of Well-being*, *4*, 20.
- Mast, M. S. (2002). Dominance as expressed and inferred through speaking time. *Human Communication Research*, *28*(3), 420–450.
- Maus, F. E. (1988). Music as drama. *Music Theory Spectrum*, 56–73.
- McDermott, J. H., Schultz, A. F., Undurraga, E. A., & Godoy, R. A. (2016). Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature*, *535*(7613), 547–550.
- McDonnell, R., Ennis, C., Dobbyn, S., & O'Sullivan, C. (2009). Talking bodies: Sensitivity to desynchronization of conversations. *ACM Transactions on Applied Perception (TAP)*, *6*(4), 22.
- Miles, L., Nind, L., & Macrae, C. N. (2009). The rhythm of rapport: Interpersonal synchrony and social perception. *Journal of Experimental Social Psychology*, *45*(3), 585–589.
- Monson, I. (1996). *Saying something: Jazz Improvisation and interaction*. Chicago: University of Chicago Press.
- Moran, N. (2016). 'Agency' in embodied musical interaction. *Routledge companion to embodied musical interaction*. Routledge.
- Moran, N., Hadley, L. V., Bader, M., & Keller, P. E. (2015). Perception of 'back-channeling' nonverbal feedback in musical duo improvisation. *PLoS One*, *10*(6), e0130070.
- Naguib, M., & Mennill, D. J. (2010). The signal value of birdsong: Empirical evidence suggests song overlapping is a signal. *Animal Behaviour*, *80*(3), e11–e15.
- Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, *9*(9), 1186–1192.
- Osborne, J. W. (1989). A phenomenological investigation of the musical representation of extra-musical ideas. *Journal of Phenomenological Psychology*, *20*(2), 151.
- Pachet, F., Roy, P., & Foulon, R. (2016). Do jazz improvisers really interact? The score effect in collective jazz improvisation. *Companion to embodied music interaction*. Routledge.
- Papiotis, P., Marchini, M., Perez-Carrillo, A., & Maestre, E. (2014). Measuring ensemble interdependence in a string quartet through analysis of multidimensional performance data. *Frontiers in Psychology*, *5*.
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681.

- Patel, A. D. (2010). Music, biological evolution, and the brain. In C. Levander & C. Henry (Eds.), *Emerging disciplines* (pp. 91–144). Houston, TX: Rice University Press.
- Patel, A., Iversen, J., Bregman, M., & Schulz, I. (2009). Experimental evidence for synchronization to a musical beat in a nonhuman animal. *Current Biology*, *19*, 827–830.
- Pecenka, N., & Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental Brain Research*, *211* (3–4), 505–515.
- Pincus, A. L., Gurtman, M. B., & Ruiz, M. A. (1998). Structural analysis of social behavior (SASB): Circumplex analyses and structural relations with the interpersonal circle and the five-factor model of personality. *Journal of Personality and Social Psychology*, *74*(6), 1629.
- Rabinowitch, T.-C., Cross, I., & Burnard, P. (2013). Long-term musical group interaction has a positive influence on empathy in children. *Psychology of Music*, *41*(4), 484–498.
- Rosenthal, R., & Rubin, D. B. (1989). Effect size estimation for one-sample multiple-choice-type data: Design, analysis, and meta-analysis. *Psychological Bulletin*, *106* (2), 332.
- Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255.
- Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: Evidence from infant listeners. *Psychological Science*, *7*(5), 272–277.
- Schilbach, L. (2014). On the relationship of online and offline social cognition. *Frontiers in Human Neuroscience*, *8*, 278.
- Seth, A. (2010). A matlab toolbox for granger causal connectivity analysis. *Journal of Neuroscience Methods*, *186*, 262–273.
- Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*(3), 114–123.
- Small, C. (1998). *Musicking: The meanings of performing and listening*. Wesleyan University Press.
- Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, *41* (1), 31–57.
- Team, A. (2016). Audacity(r): Free Audio Editor and Recorder [computer program]."  
Version 2.1.0. <<http://audacity.sourceforge.net/>> Retrieved April 20th 2015.
- Trevarthen, C. (2000). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae*, *3*(1 suppl), 155–215.
- Turino, T. (2008). *Music as social life: The politics of participation*. University of Chicago Press.
- Wagner, H. L. (1993). On measuring performance in category judgment studies of nonverbal behavior. *Journal of Nonverbal Behavior*, *17*(1), 3–28.
- Wan, C. Y., & Schlaug, G. (2010). Music making as a tool for promoting brain plasticity across the life span. *The Neuroscientist*, *16*(5), 566–577.
- Warren, J. R. (2014). *Music and ethical responsibility*. Cambridge University Press.
- Wichmann, A. (2000). The attitudinal effects of prosody, and how they relate to emotion. In *ITRW on speech and emotion*, Newcastle, Northern Ireland, UK, September. .
- Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science*, *20*(1), 1–5.
- Wing, A. M., Endo, S., Bradbury, A., & Vorberg, D. (2014). Optimal feedback correction in string quartet synchronization. *Journal of The Royal Society Interface*, *11*(93), 20131125.
- Wu, K.-Y., Chao, C.-W., Hung, C.-I., Chen, W.-H., Chen, Y.-T., & Liang, S.-F. (2013). Functional abnormalities in the cortical processing of sound complexity and musical consonance in schizophrenia: Evidence from an evoked potential study. *BMC Psychiatry*, *13*(1), 1.
- Young, J. O. (2014). *Critique of pure music*. Oxford University Press.
- Zatorre, R. J. (2013). Predispositions and plasticity in music and speech learning: Neural correlates and implications. *Science*, *342*(6158), 585–589.
- Zentner, M. R., & Kagan, J. (1996). Perception of music by infants. *Nature*, *383*(6595), 29–29.