Play together, think alike: Shared mental models in expert music improvisers

Clément Canonne, Jean-Julien Aucouturier

Play together, think alike: Shared mental models in expert music improvisers

Clément Canonne
Department of Musicology, Université de Bourgogne, Dijon (France)

Jean-Julien Aucouturier
STMS (IRCAM - CNRS UMR 9912 - UPMC), Paris (France)

When musicians improvise together, they tend to agree beforehand on a common structure (e.g., a jazz standard) which helps them coordinate. However, in the particular case of collective free improvisation (CFI), musicians deliberately avoid having such referent. How, then, can they coordinate? We propose that CFI musicians who have experience playing together come to share higher-level knowledge, which is not piece-specific but rather task-specific, i.e., an implicit mental model of what it is to improvise freely. We tested this hypothesis on a group of 19 expert improvisers from the Parisian CFI community, who had various degrees of experience playing with one another. Drawing from the methodology of team cognition, we used a card-sorting procedure on a set of 25 short improvised sound sequences to elicit and represent each participant's mental model of the CFI task. We then evaluated the similarity between the participants' models, and used the measure in a nearest neighbor classification algorithm to retrieve clusters of participants who were known to play together. Conformly to our hypothesis, we found that the degree of similarity in participants' mental models predicted their degree of musical familiarity with better-than-random accuracy: musicians who played together tended to "think" about improvised music in the same way.

Introduction

Improvised music is invented as it is being performed, but it never really is *creatio ex nihilo*. Every form of improvisation is built on pre-existing musical atoms, e.g. the 300-or-so prototypical elements of the Iranian radif repertoire (Nettl, 1987), the II-V-I harmonic progressions favored by the jazz musicians (Berliner, 1994), or the embodied virtuosic gestures of certain musicians (Sudnow, 2001). Besides pre-existing material, most forms of improvisation also rely on a pre-existing structure or "referent": an "underlying formal scheme or guiding image specific to a given piece, used by the improviser to facilitate the generation and editing of improvised behavior" (Pressing, 1984, p. 346). A referent like, e.g. the 32 bars, theme and chord progression of *Autumn Leaves* gives an improviser formal elements (a large-scale structure, a pre-organization of the different musical "roles"), musical content (themes, chord changes), poetic associations and stylistic guidelines to draw from.

Sharing a referent ("calling a tune") fulfills an essential function for musicians who improvise together: it facilitates their coordination by establishing explicit common knowledge (Lewis, 1969). From the moment a band chooses e.g. *My Funny Valentine* as a basis for improvisation, they know (and know that each of them knows) that: the style of music is more likely to be jazz than twelve-tone atonality; the musicians will have to follow a given chord progression and go into cycles around it; the tempo is likely to be slow, the mood relaxed and so on.

However, some forms of collective improvisation exist that deliberately try to avoid using referents. Such Collective Free Improvisation (CFI) is encountered in a wide spectrum of musical styles, e.g. in the early experiments of Free Jazz or contemporary "non-idiomatic improvisation" (Bailey, 1993). The practice of CFI asks tough questions about the cognitive mechanisms that enable musical coordination: how can musicians improvise together without agreeing on a pre-existing structure? How can they produce coherent music together in real-time in the absence of explicit common knowledge? Because it has no referent, CFI is a particularly pure and paradigmatic case of the coordination problems which are present, to some degree, in every form of collective improvisation (or even, arguably, collective behaviour). In particular, studying CFI separates the question of coordination ("how do they improvise together?") from the more often studied question of signal generation ("how does one improvise?"), for which convincing cognitive (Pressing, 1988) and computational (Pachet, 2012) models already exist.

Coordination between musicians can be defined as their converging over successive sets of mutually consistent musical decisions, with an outcome that satisfies their individual preferences (Schelling, 1960; Lewis, 1969). There are many aspects of such convergence in CFI, e.g., temporal (i.e., organizing the interactions between players in time), strategic (i.e., building a satisfying musical form in real-time) and aesthetic (i.e., building congruency out of the possibly divergent preferences of the players). There are also many possible ways in which parts of this problem can be solved. First, it can be argued that, even though they claim not to, CFI improvisers inevitably end up setting up conventions to help them coordinate. Classic jazz jam sessions for instance, despite their image of pure spontaneity, are known to have such hidden rules, e.g., an "etiquette" which requires that everybody played as many choruses as the first musician who soloed

(Becker, 2000). Second, low-level sensorimotor strategies may suffice to solve a large part of the coordination problem. For instance, if asked immediately after they played, CFI musicians were found to largely agree on what specific points in time were the structural milestones/coordination cues in their performance (Canonne & Garnier, 2012), and these points tend to correspond to salient events of the acoustic stream (in e.g. pitch, dynamics, timbre, articulation) (Canonne, 2013; see also Moran, 2010; Keller, in press). But even with such mechanisms at play, a more-encompassing proposal would be to suppose that, in order to successfully coordinate, CFI musicians must come to share some higher-level knowledge, which is not piece-specific like a referent would be, but task-specific; not a structure, but a meta-structure; an implicit mental model of what it *is* to improvise without referent. It is this last hypothesis that this paper sets to test.

The notion of shared mental model (SMMs) has been at the forefront of team cognition research for the past 15 years (Mohammed, Ferzandi, & Hamilton, 2010). Mental models are defined as "organized knowledge structures that allow individuals to recognize and remember relationships among components of their environment and to construct expectations for what is likely to occur next" (Rouse & Morris, 1986). If members in a team share similar mental models, each can draw from their own model to choose behaviours that are consistent and coordinated with teammates (Cannon-Bowers, Salas, & Converse, 1993; Klimoski & Mohammed, 1994). This notion seems particularly natural in the context of collective musical improvisation: when asked to comment retrospectively on their own performance, improvisors often ascribe meaning (is this a transition? Is this the conclusion of a sequence?) and intentions (does he want to take a solo? Do they want to imitate each another?) to musical situations (Canonne & Garnier, 2012). SMM theory predicts that, if musicians engaged in the same CFI performance share a mental model linking, say, a sudden change of register (e.g. going from low to high) to the meaning of "introducing a new idea", then they will be able to coordinate quickly if one player suddenly executes such a move: they will decide e.g., to join or oppose it. It also seems natural that improvisers who have a longer history of performing together should develop a common understanding about certain musical situations.

Empirical research in team cognition abounds with proposals on how to characterize whether, and how much, mental models are shared in a team (DeChurch & Mesmer-Magnus, 2010). Adapting this work to a musical context, we set to test, first, whether CFI musicians can be said to share mental models about the task of free improvisation, and second, to evaluate whether the amount of mental model similarity is linked to team familiarity. In other words, we aim to test whether musicians who play together "think" alike about the task of CFI.

To do so, we study an ecological set of professional musicians from the French Parisian CFI community, and consider their naturally varied degrees of experience playing with one another as our independent variable. Adapting the card-sorting methodology of the SMM literature (Edwards, Day, Arthur, & Bell, 2006) to the context of musical improvisation, we use a 'sound-sorting" procedure, in which we ask these musicians to organize musical excerpts into "classes of actions", i.e., to *"place in the same group the sounds to which [they] would react in a similar way if [they] were to hear them in a CFI setting"*. We then compute similarity indices between the participants' groupings/mental models and test whether such similarity can be used to predict the existing relations of familiarity between participants better than at random.

Methods

Participants

Nineteen young adults (male: N=13; M=26, SD=3.5) participated in the experiment, all career musicians actively involved in collective free improvisation. Participants were recruited via the CFI masterclasses of the [institution name removed]. Thirteen were currently affiliated with [institution name removed]; the others were graduates of the same institution (N=6). All had very substantial musical training (3-5 years of improvisation practice, 2-5 years of CFI practice) and musical aptitude (reported average grade M=92.5/100 at the [institution name removed] entrance aptitude test).

Stimuli

Twenty-five 15-second original musical recordings were used to create the experimental stimuli. We selected these extracts from a larger, unreleased dataset of 400 studio recordings of 20 improvising musicians (unrelated to our subjects: we checked that our subjects where not playing in the same bands than the authors of our experimental stimuli) whom we each asked to 20 perform short (15-20s.) "clippings" of improvised musical material that they would commonly use in a CFI situation. These 400 recordings were then manually edited down to a duration of 15 seconds, and subjected to automatic content analysis in order to select a subset of 25 recordings that acted as a representative sample (in terms of dynamics, timbre, rhythm and tonality) of the larger set (see Figure 1 and Appendix for details).

Additionally, we constructed two control conditions by deriving two degraded variants for each of the 25 selected excerpts using signal processing manipulations aiming to preserve only the excerpt's timbre, or only its temporal morphology. Timbre-only variants were constructed from the original excerpt using splicing (Aucouturier & Defreville, 2009) (see Appendix). They sounded like an abstract, dense and busy texture of the same timbre quality as the original, but with none of its original time structure (Figure 2b). Morphology-only variants were constructed from the original excerpt using a noise-vocoder procedure Gygi, Kidd, and Watson (2004) (see Appendix). They sounded like a single noisy instrument playing phrases of the same temporal morphology as the original, but with unrecognizable timbre/instrument (Figure 2c). All stimuli (25 excerpts, in 3 variants each) were loudness-normalized to 75dBA using the ma_sone Matlab procedure (Pampalk, 2004).

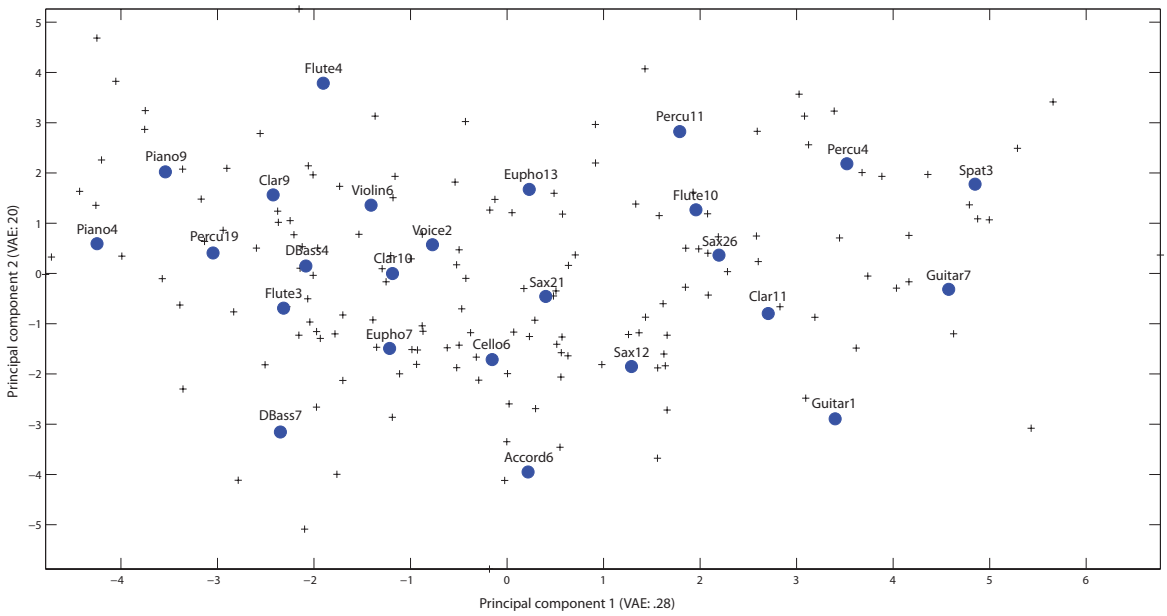
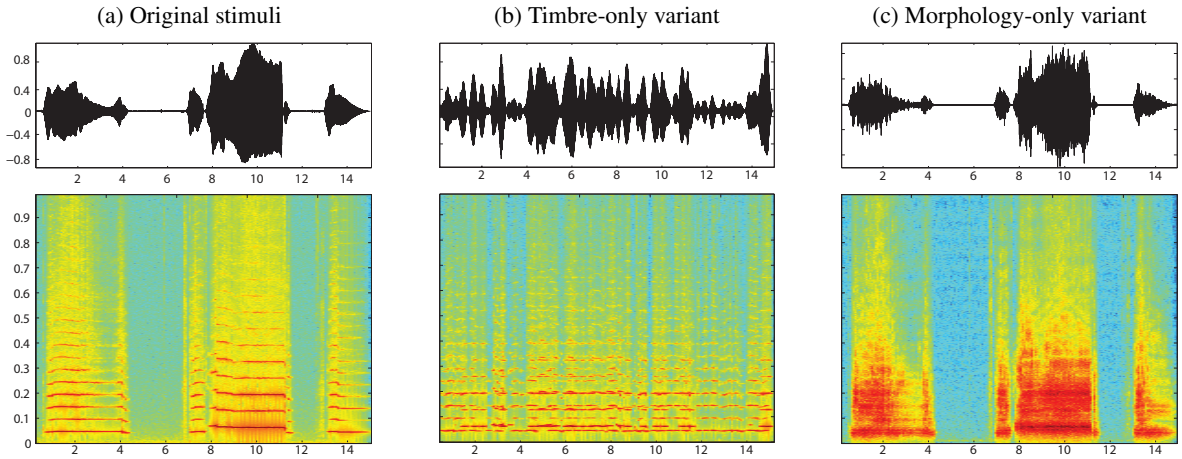


Figure 1. : Two-dimensional projection of the 400-excerpt dataset (crosses), and the 25 excerpts selected therefrom (circles) for the experiment. Dimensions correspond to the two principal components analysed from a set of 28 acoustic features computed using the MIRToolbox. See Appendix for details and legend of excerpt labels

Figure 2. : Waveform (top) and spectrogram (bottom) representations of one of the experimental stimuli (Flute10) in its original version (2a), its timbre-only variant obtained by splicing (2b) and its morphology-only variant obtained by noise vocoding (2c). Note the degradation of the temporal information in the waveform of 2b, and the degradation of timbre information in the spectrogram of 2c.



Procedure

Participants were subjected to three successive blocks of the same free sorting task, each for a set of 25 sounds. In each block, sounds 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 all sound stimuli were recordings of musical phrases that had been used by a musician in the context of a collective improvisation. They were instructed to organize these sounds into groups so as to reflect the sounds' "pragmatic similarity" if they were to hear them in the context of an ongoing CFI performance, i.e., to *"place in the same group the sounds to which [they] would*

react in a similar way if [they] were to hear them in a CFI setting. For each sample, to ask [themselves]: what would I do/play/respond to this, if someone played it to me?". 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 took approximately 20 minutes to evaluate one block.

All participants had to do the task for three consecutive blocks: the first two blocks consisted of either the 25 spliced sounds of the timbre-only condition or 25 noise-vocoded sounds of the morphology-only condition (counterbalanced over all participants), and the last block always consisted of the 25 original excerpts. All three blocks were evaluated independently from one another, and within each block, the participants were instructed to only focus on the current set of sounds. The block consisting of the original excerpts was always presented last, in order to minimize the possibility of identifying the degraded variants in the other two blocks.

In addition to the free sorting task, each participant was presented a list including the names of all the other participants in the experiment, and to rate musical familiarity with each of them using a 3-level scale[1]: (0) *I may or may not know this person, but I have never played with her in the context of a collective free improvisation*; (1) *I know this person, and I have played occasionally with her in the context of CFI, for instance during a masterclass*; (2) *I know this person, and I play regularly with her in the context of CFI. For instance, we have a band together*. All participants (N=19) gave ratings for all other participants (N=18), yielding a complete matrix (19x19) of participant-to-participant familiarity.

Apparatus

The sounds were played by a Macintosh Mac Pro (Mac OS X v10.4 Tiger) workstation with a Motu Firewire 828 sound card (Motu audio, Massachusetts, USA). The stimuli were presented dichotically with Sennheiser HD280 headphones (Sennheiser electronic GmbH, Wedemark, Germany). Participants were seated in a double-walled sound isolation booth (IAC, New York, USA). Levels were calibrated using a Bruël & Kjær 2238 Mediator sound-level meter (Bruël & Kjæ r Sound & Vibration, Nærum, Germany). The software used to run the experiment was TclLabX 0.3.10 (Gaillard, 2010).

Data analysis

We analysed the participants' grouping data in order to test the hypothesis that the degree of similarity between the groupings made by two participants can predict their musical familiarity. First, we derived a distance measure comparing groupings, i.e., participant mental models, to one another. Second, we used the participants' ratings of familiarity to form clusters of participants that regularly play together. Finally, we tested whether the former measure can be used to classify participants into the latter clusters with significantly better precision than random (see e.g., Glady, Thibaut, & French, 2013). All three experimental conditions were analysed separately.

In each condition, groupings made by each participant were compared to one another using the Jaccard Index (JI) of partition similarity (Saporta & Youness, 2002)[2]. The JI between a grouping P and a grouping Q counts the proportion of pairs of stimuli that are simultaneously joined (i.e. included in the same group) by both P and Q. For two groupings P and Q, we compute JI as:

$$JI(P,Q) = \frac{a}{a+b+c} \tag{1}$$

where a is the number of pairs of stimuli belonging to the same group in P *and* to the same group in Q, b is the number of pairs of stimuli belonging to different groups in P but to the same group in Q, and c is the number of pairs of stimuli belonging to the same group in P but to different groups in Q. The JI gives values between 0 (if none of the pairs joined in P are found joined in Q) and 1 (if the two partitions P and Q perfectly overlap).

The participants' ratings of musical familiarity were made symmetric and converted to dissimilarity scores $d(a,b)$ using:

$$d(p,q) = 2 - max(f(p,q), f(p,q)) \tag{2}$$

where p and q denote participants, $f(p,q)$ is the rating of familiarity declared by participant p for participant q. The 19x19 dissimilarity matrix was then subjected to agglomerative hierarchical clustering (using average linkage), of which we retained three clusters of roughly equivalent cardinality (Figure 3).

[institution name removed]

Cluster1 (N=6) correspond to musicians who perform together as part of the [name removed] collective. Cluster 2 (N=5) and Cluster 3 (N=8) regroup graduates of the two [institution name removed] improvisation cursus and who maintained a joint and regular CFI practice.

We tested the ability to recover these familiarity relationships given the distance measures between the participants' groupings using a k-nearest neighbor classification procedure. In each condition, we computed the $k = 3$ nearest neighbors of each participant based on the inter-participant JI distances, and predicted her familiarity cluster (1,2 or 3) based on the cluster most represented in the neighborhood. For each participant, we then compared the predicted cluster to the participant's actual cluster, giving a total of n_{test} correct predictions. If the participants' JI distances are only randomly related to their familiarity, then n_{test} would not be significatively different from a random $k = 3$ majority voting

[1] Each participant explicitly agreed, by participating in this study, to reveal his identity to the other participants.

[2] Note that several alternative measures to JI have been proposed to compare card sorts P and Q of the same set, such as the Rand index (Saporta & Youness, 2002) or the edit/transfer distance (Deibel, Anderson, & Anderson, 2005), which counts the minimum number of moves to transform P into Q. We chose JI on the basis of Denœud and Guénoche (2006). In addition, as a posthoc verification, we replicated the analysis using the transfer distance, and results were identical.

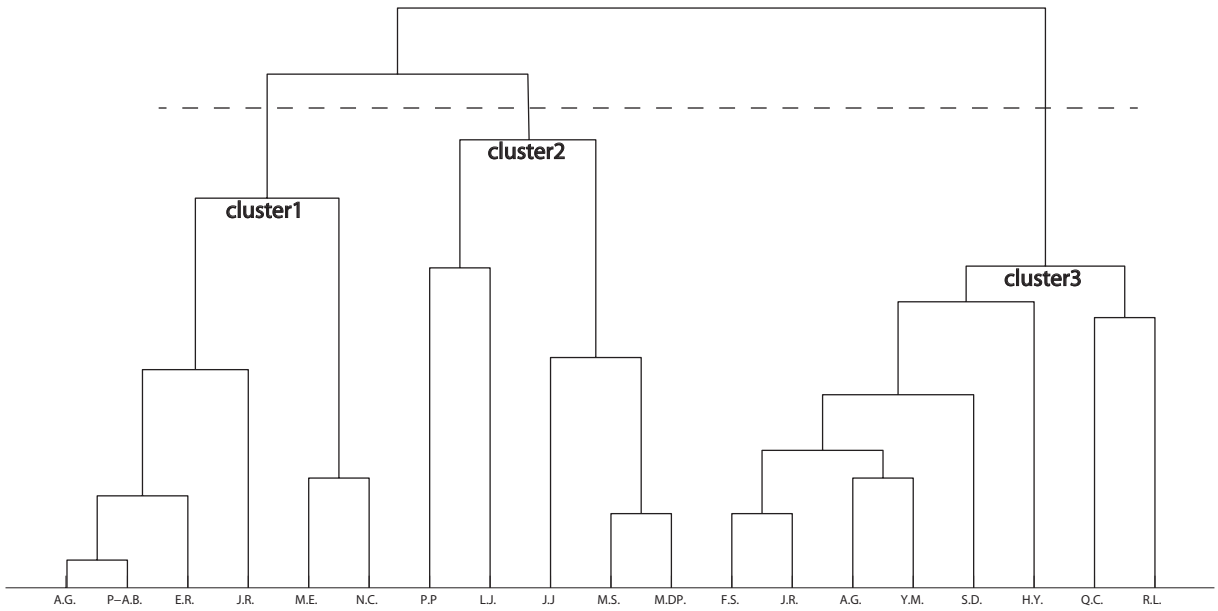


Figure 3. : Hierarchical clustering dendrogram showing the three clusters of participants derived from the participants' ratings of musical familiarity with one another. Cluster1 (A.G., P-A.B., E.R., J.R., M.E., N.C.) correspond to musicians who perform together as part of the [name removed] collective. Cluster 2 (P.P., L.J., J.J., M.S., M.DP.) and Cluster 3 (F.S., J.R., A.G., Y.M., S.D., H.Y., Q.C., R.L) regroups graduates of the two [institution name removed] improvisation cursus and who maintained a joint and regular CFI practice.

procedure, for which the number of correct cases n_{random} can be enumerated as the number of times a participant's cluster is found for either 3 or 2 of her $k = 3$ neighbors[3]:

$$n_{random} = \sum_{i=1}^{N} \frac{\binom{C_i-1}{k} + (N - C_i - 2)\binom{C_i-1}{k-1}}{\binom{N-1}{k}} \quad (3)$$

where i indexes participants, $N = 19$, $k = 3$ is the number of neighbors in the k-NN procedure, C_i is the size of the familiarity cluster of participant i, and $\binom{n}{k}$ is the "n choose k" binomial coefficient (i.e. the number of distinct k-element subsets that can be formed from a set of n elements). Using the C_i's corresponding to the three familiarity clusters 1,2,3, Equ. 3 yields n_{random}=4.17, corresponding to 4.17 correct predictions and 14.83 incorrect predictions. We computed the n_{test} number of correct predictions (and $N - n_{test}$ incorrect predictions) from the JI distances in all three conditions, and tested whether these were statistically significantly more precise than n_{random} (and $N - n_{random}$) using a one-tail Fisher's exact test (preferred over χ^2 because $n_{random} < 5$).

Results

JI distances between participants' groupings in the original condition (unmanipulated stimuli) allowed to correctly predict the participant's familiarity cluster for $n_{test} = 9$ (47%) of the $N = 19$ participants, a 25% absolute improvement over random which reached statistical significance (1-tail FET: p=0.033*).

As a control, neither the JI distances between grouping of the timbre-only stimuli ($n_{test} = 7$, FET p=0.11) nor of the morphology-only stimuli ($n_{test} = 4$, FET p=0.31) allowed significantly better than random prediction of the participants' familiarity. Table 1 gives the contingency tables for all 3 conditions.

Discussion

"After 600 shows, the Wayne Shorter Quartet functions almost telepathically", once said a music critic of the long-standing band of the jazz saxophonist[4]. Our results seem to corroborate this intuition: the manner in which CFI musicians organized musical excerpts in our experiment revealed mental models which similarity significantly concorded with the familiarity they have with one another. By only looking at mental model similarity, a k-nearest neighbor algorithm was able to predict the correct familiarity cluster of a musician 47% of the time, a 25% improvement over random. Additionally, this familiarity effect is not trivial, as it is not

[3] To simplify the combinatorics of eq. 3, we automatically counted a situation where 3 clusters were represented in the $k = 3$ neighborhood as a fail. This situation occurred only once for the N=19 predictions. Eq. 3 is only valid for the case $k = 3$, which we chose to match the size of the familiarity clusters.

[4] Thomas, G. (2012, April 26). Saxophonist-composer Wayne Shorter takes the longer view after his many decades in jazz. The New York Daily News. Retrieved from http://www.nydailynews.com.

Table 1
: Contingency tables comparing the observed numbers of correctly classified cases n_{test} to the expected numbers n_{random} if classification was random, in the timbre-only (1a), morphology-only (1b) and original conditions (1c). Observed proportions marked with * was significantly different from random at the p=.05 level (FET)

(a) Timbre-only condition

	Number of cases		
Criteria	Correct	Incorrect	Total
Groupings	7	12	19
Random	4.17	14.83	19
Total	12.17	26.83	38

(b) Morphology-only condition

	Number of cases		
Criteria	Correct	Incorrect	Total
Groupings	4	15	19
Random	4.17	14.83	19
Total	8.17	29.83	38

(c) Original condition

	Number of cases		
Criteria	Correct	Incorrect	Total
Groupings	9*	10*	19
Random	4.17	14.83	19
Total by case	13.17	24.83	38

observed in two control conditions with musical stimuli that share some of the acoustical properties of the original but have less natural musical relevance.

In the team cognition literature, the empirical evidence for a link between shared mental models and team familiarity is remarkably contrasted. Familiar teams sometimes perform better (Harrison, Mohammed, McGrath, Florey, & Vanderstoep, 2003), but sometimes don't (Cooke, Gorman, Duran, & Taylor, 2007). Interventional studies that form new teams and train them for an experimental task over hours (Mathieu, Heffner, Goodwin, Salas, & Cannon-Bower, 2000) or even weeks (Edwards et al., 2006) have often failed to find any relation between time and model similarity. They sometimes even demonstrated increased divergence over months (Levesque, Wilson, & Wholey, 2001). It is possible that team familiarity fosters shared mental models in certain situations, e.g., in military teams actively engaging in performance feedback (Mathieu et al., 2000) or, here, in collective musical improvisation, but not in others, e.g., in tasks which rather benefit from skill specialization, such as software development (Levesque et al., 2001). It is also possible that team familiarity cannot be robustly elicited in interventional studies that form ad-hoc experimental teams, but rather needs to be assessed in real, existing teams with a richer performance history, as we did here (see also Smith-Jentsch, Kraiger, Cannon-Bowers, & Salas, 2009 for air traffic control teams).

Because of the observational design of the experiment, we could not control all aspects of the sample and it should be noted that the variable of familiarity may be partially confounded by age and/or expertise. There was a statistically significant difference of participant age ($F(2,16)=7.87$, $p=0.004$), with participants in cluster 1 older (M=29.6) than in cluster 2 (M=23.8) and 3 (M=25). Additionally, we asked participants to classify their expertise with both improvised music in general, and CFI in particular. All members in cluster1 declared more than 10 years of improvisation, and more than 5 years of CFI practice, while the most reported category was between 5-10 years of improvisation, and between 2-5 years of CFI practice in both clusters 2 and 3. However, these differences were not statistically significant to a χ^2-test (Yates' corrected), neither for general improvisation practice ($\chi^2=8.41$, $p=0.20$) nor CFI ($\chi^2=5.15$, $p=0.74$). While an effect of expertise cannot be ruled out (i.e., more experienced musicians may also think in a more similar way), this effect is likely to be at the ceiling level given the high overall expertise of the sample. Moreover, such an effect would not explain that the mental models notably contributed to discriminate

between clusters 2 and 3.

The way we elicit mental models in this experiment informs us on the nature of knowledge shared by familiar musicians. First, our procedure was designed to elicit knowledge that was related to the task of improvisation rather than the team. The stimuli were not associated with musicians known to any of the participants, and therefore were unlikely to prompt any questions related to team relations, such as "who's the leader", "who likes to do what", etc.What was shared by familiar musicians was therefore a task mental model (i.e., a common ways to frame the CFI situation) rather than a team mental model (i.e., common knowledge about teammates and their interactions - see Rentsch & Hall, 1994).

Second, because our participants were players of a large variety of instruments of different families (keyboards, strings, woodwinds, etc.), similarities observed in the participants' groupings are unlikely to be based on their specific instrumental responses (a motor action, such as pressing a key), but more likely on higher-level, non instrument-specific musical responses. We argue that the observed similarities therefore reflect partially shared classes of musical actions: participants ascribed a similar functional role to musical excerpts that form a given class.

Finally, because the effect of familiarity was only found for non-degraded stimuli, and neither for the timbre-only nor morphology-only versions, the shared knowledge appears to hold on high-level, emergent musical constructs that cannot be degraded to simpler perceptual constructs like musical instruments or prototypical temporal sequences. It is not just because a given group of participants have been exposed to e.g., certain instruments more than others that their mental models coincide. It is the way they frame the usage of such instruments, in time and in interaction, that shows similarity.

This demonstration of knowledge shared by familiar musicians is reminiscent of a large literature on implicit learning in music cognition. Because of cultural familiarity with a musical style, even non-musicians were shown to acquire implicit knowledge of e.g., the rules of western harmony (Bigand & Poulin-Charronnat, 2006), the metrical structure of non-western rhythms (Drake & Ben El Heni, 2003) or cues used for emotional communication with music (Balkwill & Thompson, 1999). However, it is unclear with the present procedure whether the representations shared by the participants are based on implicit or explicit knowledge. If anything, the procedure did not force participants to use explicit knowledge representations[5]: we did not ascribe meanings to the sounds ("how would you react to someone doing a solo") or to the groups ("group the sounds to which you would react by a solo"), nor did ask participants to verbalize the meaning of the groups they formed. However, the participants are professional musicians with very refined explicit musical representations, and it remains a possibility that their mental models are based on explicit, rather than implicit knowledge.

It is also an open question whether the knowledge shared by participants can be described as conventions. For a given fact $\mathcal{A}$ to be a convention by Lewis, 1969, participants should both know $\mathcal{A}$ and know that other participants know $\mathcal{A}$. Although our results indicate that this is, at most, only partially the case, we did not test this "expectation of knowledge" explicitely. Conventions also have normative value, i.e., in a given group everyone expects that others will follow it (e.g., a given musical situation should inevitably entail a given musical action). We find it unlikely that the mental models shared by participants have such value: they are both less determined and less constraining.

In the specific context of CFI, the demonstration of shared mental models in familiar teams shows, first, that it is important to take into account top-down processes when analyzing CFI situations. In fact, CFI situations are probably best seen as a complex mix of top-down and bottom-up processes: even in the absence of referent, musicians' choices are driven both by adhoc cues and by a larger context which dictates the range of possible responses to these cues, based on some shared understanding of the CFI situation - the result of the improvisers' familiarity with one another.

Second, the existence of SMMs gives a precise content to the common idea that some impro-oriented bands - like the Wayne Shorter Quartet - develop over time a "telepathy"-like sense of interaction, that its members are "on the same page". Conversely, the absence of SMMs may explain some of the most patent cases of "dis-coordination" in CFI, which occur when the musicians have contradictory representations of a given situation. Canonne and Garnier (2012) discuss the case of an energetic intervention designed by its author as a way to reinvigorate the collective production, but understood by the other members as a clear statement for the beginning of a solo, thus leading to a long and awkward silence. Shared mental models could then be seen as a key ingredient of CFI's success, by allowing more confident "mind-reading" of the intentions of fellow improvisers, more frequent cognitive consensus in the course of a performance (i.e., the improvisers having the same representation of a given situation in the course of the performance) and swifter repair of "communication errors" when there is cognitive divergence among the improvisers (see also Fuller and Magerko (2011) for a qualitative study of these phenomena in the domain of improvisational theater).

As a final note, one must also consider that SMMs may not always be desired by CFI musicians. Bailey (1993), for instance, found the idea of long-lasting CFI bands paradoxical, fearing that the implicit conventions, the emerging stylistic identity and the shared framing of the CFI situation would prevent the spontaneity and unpredictability that is sought after by CFI practitioners. In fact, it was even proposed that short-lived bands, with musicians who have no history of playing together, was the ideal situation for CFI (Watson, 2004). Contrary to the more-studied contexts of industrial work or military activity, performance and even coordination metrics are not easily defined in hedonic tasks such as mu-

[5] It should also be noted that explicit methodologies such as interviews may not be applicable to elicit similar shared classes of musical actions, given the musicians' tendancy to undermine, for ideological reasons, the presence of "conventionalized" features in the context of free improvisation (Léandre, 2008).

sic improvisation. In CFI, perfect or complete coordination, made possible by strongly overlapping mental models, may not be what is sought (Sawyer, 2003); it is perhaps in the dramatic compromise between the impossibility of coordinating without pre-existing structure and the unstoppable normalizing force of familiarity that the creativity of CFI musicians is best expressed. Sympathy, rather than telepathy, is the true epitome of CFI: the experience of a situation that we know could break down, and that just won't.

Acknowledgement

This research was supported by European Research Council project CREAM (335536) and a FABER grant (9201AAO047S01200) provided by the Conseil Régional de Bourgogne.

References

Aucouturier, J.-J., & Defreville, B. (2009). Judging the similarity of soundscapes does not require categorization: Evidence from spliced stimuli. *Journal of the Acoustical Society of America*, *125(4)*, 2155-61.

Bailey, D. (1993). *Improvisation: Its nature and practice in music*. New York: Da Capo Press.

Balkwill, L., & Thompson, W. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception*, *17*.

Becker, H. (2000). The etiquette of improvisation. *Mind, Culture, and Activity*, *7(3)*, 171-176.

Berliner, P. F. (1994). *Thinking in jazz : the infinite art of improvisation*. Chicago: The University of Chicago Press.

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.

Cannon-Bowers, J. A., Salas, E., & Converse, S. A. (1993). Shared mental models in expert team decision making. In J. N. J. Castellan (Ed.), *Current issues in individual and group decision making* (p. pp. 221-246). Hillsdale, NJ: Erlbaum.

Canonne, C. (2013). Focal points in collective free improvisation. *Perspectives of New Music*, *51(1)*.

Canonne, C., & Garnier, N. (2012). Cognition and segmentation in collective free improvisation. In *Proceedings of the 12th international conference on music perception and cognition* (pp. 197–204).

Cooke, N. J., Gorman, J. C., Duran, J. L., & Taylor, A. R. (2007). Team cognition in experienced command-and-control teams. *Journal of Experimental Psychology: Applied*, *13*, 146–157.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). Measuring shared team mental models: A meta-analysis. *Group Dynamics: Theory, Research, and Practice*, *14(1)*, 1–14.

Deibel, K., Anderson, R., & Anderson, R. (2005). Using edit distance to analyze card sorts. *Expert Systems*, *22(3)*.

Denœud, L., & Guénoche, A. (2006). Comparison of distance indices between partitions. In *Studies in classification, data analysis, and knowledge organization* (p. 21-28).

Drake, C., & Ben El Heni, J. (2003). Synchronizing with music: intercultural differences. In *Annals of the new york academy of sciences, volume 999, the neurosciences and music* (p. 429–437).

Edwards, B. D., Day, E. A., Arthur, W., Jr., & Bell, S. T. (2006). Relationships among team ability composition, team mental models, and team performance. *Journal of Applied Psychology*, *91*, 727–736.

Eerola, T. (2011). Are the emotions expressed in music genre-specific? an audio-based evaluation of datasets spanning classical, film, pop and mixed genres. *Journal of New Music Research*, *40(4)*, 349–366.

Fuller, D., & Magerko, B. (2011). Shared mental models in improvisational theatre. In *Proceedings of the 8th acm conference on creativity and cognition, new-york* (p. 269-278). ACM.

Gaillard, P. (2010). Recent development of a software for free sorting procedures and data analysis. In *3rd lam/incas3 workshop, free sorting tasks and measures of similarities on some structural and formal properties of human categories, paris, france.*

Glady, Y., Thibaut, J.-P., & French, R. (2013). Using a novel method of eye-tracking scanpath analysis to distinguish children's and adults' analogy strategies. In *Proceedings of the annual meeting of the cognitive science society.*

Gygi, B., Kidd, G., & Watson, C. (2004). Spectral-temporal factors in the identification of environmental sounds. *J. Acoust. Soc. Am.*, *115(3)*, 1252-1265.

Harrison, D. A., Mohammed, S., McGrath, J. E., Florey, A. T., & Vanderstoep, S. W. (2003). Time matters in team performance: Effects of member familiarity, entrainment, and task discontinuity on speed and quality. *Personnel Psychology*, *56*, 633–699.

Keller, P. E. (in press). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures.* Oxford: Oxford University Press.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, *20*, 403-437.

Lartillot, O., Toiviainen, P., & Eerola, T. (2008). A matlab toolbox for music information retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data analysis, machine learning and applications* (p. 261-–268). Berlin, Heidelberg: Springer.

Léandre, J. (2008). *A voix basse* (E. MF, Ed.).

Levesque, L. L., Wilson, J. M., & Wholey, D. R. (2001). Cognitive divergence and shared mental models in software development project teams. *Journal of Organizational Behavior*, *22*, 135-144.

Lewis, D. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.

Mathieu, J., Heffner, T., Goodwin, G., Salas, E., & Cannon-Bower, J. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology*, *85(2)*, 273–283.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). *Journal of Management*, *36(4)*, 876-910.

Moran, N. (2010). Improvising musicians' looking behaviours: Duration constants in the attention patterns of duo performers. In *Proceedings of the 11th international conference on music perception and cognition (icmpc11), seattle, usa.*

Musil, J., El-Nusairi, B., & Müllensiefen, D. (2012). Perceptual dimensions of short audio clips and corresponding timbre features. In *Proceedings of the 9th international symposium on computer music modelling and retrieval (cmmr 2012), london, uk* (p. pp. 311-318).

Nettl, B. (1987). *The radif of persian music: Studies of structure and cultural context in the classical music of iran*. Champaign: Elephant and Cat.

Pachet, F. (2012). Musical virtuosity and creativity. In J. Mc-

Cormack & M. D'Inverno (Eds.), *Computers and creativity.* Springer.

Pampalk, E. (2004). A matlab toolbox to compute similarity from audio. In *Proceedings of the ismir international conference on music information retrieval (ismir'04), barcelona, spain.*

Pressing, J. (1984). Cognitive processes in improvisation. In W. R. CROZIER & A. CHAPMAN (Eds.), *Cognitive processes in the perception of art* (p. 345-363). Amsterdam: Elsevier.

Pressing, J. (1988). Improvisation: Methods and models. In J. Sloboda (Ed.), *Generative processes in music* (pp. 129–178). Oxford: Clarendon.

Rentsch, J. R., & Hall, R. J. (1994). Members of great teams think alike: A model of team effectiveness and schema similarity among team members. *Advances in Interdisciplinary Studies of Work Teams*, *1*, 223-261.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin*, *100*, 349-363.

Saporta, G., & Youness, G. (2002). Comparing two partitions: Some proposals and experiments. In *Proceedings of the 2002 computational statistics conference* (p. 243-248).

Sawyer, R. K. (2003). *Group creativity : Music, theater, collaboration*. Mahwah, NJ: Erlbaum.

Schelling. (1960). *The strategy of conflict*. Cambridge,MA: Harvard University Press.

Shalizi, C. (2013). *Advanced data analysis from an elementary point of view*. Cambridge University Press.

Shofner, W., & Campbell, J. (2012). Pitch strength of noise-vocoded harmonic tone complexes in normal-hearing listeners. *J. Acoust. Soc. Am.*, *132(5)*, EL398-EL404.

Smith-Jentsch, K. A., Kraiger, K., Cannon-Bowers, J. A., & Salas, E. (2009). Do familiar teammates request and accept more backup? transactive memory in air traffic control. *Human Factors*, *51(2)*, 181-182.

Sudnow. (2001). *Ways of the hand : A rewritten account*. Cambridge, MA: MIT Press.

Watson, B. (2004). *Derek bailey and the story of free improvisation*. London: Verso Books.

Appendix: Signal manipulations

Timbre-only manipulation: timbre-only variants were constructed from the original excerpt using splicing (Aucouturier & Defreville, 2009): the audio signal, sampled at 44100Hz, was framed into 16384-point windows with 4096-point overlap; all windows were randomly permutated in time; the resulting "shuffled" version was then reconstructed by overlap-and-add. The large windows used for splicing preserved much of the note-level cues important of timbre perception (Musil, El-Nusairi, & Müllensiefen, 2012), but largely degraded their temporal structure. However, their random permutation in time largely degraded the temporal structure of the original version: in practice, all timbre-only variants sounded like an abstract, dense and busy texture of the same timbre quality as the original (Figure 2b).

Morphology-only manipulation: morphology-only variants were constructed from the original excerpt using the noise-vocoder procedure described in Gygi et al. (2004): the signal was first decomposed with a bank of 60 ERB filters; in each ERB band, the temporal envelope was extracted by half-rectifying the waveform and passing it through a 4-pole Butterworth LP filter (fc=20Hz); each of the 60 envelopes was then multiplied by a bandpass noise of the same bandwidth; the resulting noise-vocoded version was then reconstructed by adding the 60 waveforms back together. The large number of ERB bands used in the manipulation preserved much of the temporal structure of the original version, including some basic perception of pitch and phrasing (Shofner & Campbell, 2012). However, the replacement of all harmonic structure within bands by noise largely degraded the possibility of identifying instrumental timbres: in practice, all morphology-only variants sounded like an unique noisy instrument playing phrases of the same temporal morphology as the original (Figure 2c).

Appendix: Stimuli selection

Stimuli selection: in order to select 25 experimental stimuli from the dataset of 400 excerpts, we computed a set of 28 acoustic features for each excerpt using the MIRToolbox (Lartillot, Toiviainen, & Eerola, 2008), providing an algorithmic evaluation of various acoustical and perceptual characteristics of the excerpt: dynamics (algorithm used: RMS energy), timbre (algorithms used: spectral flux, spectral flatness, spectral entropy, spectral centroid, spectral spread, and spectral roughness), rhythm (algorithms used: pulse clarity, event density, tempo, frequency and magnitude of fluctuation peak) and tonality (algorithms used: major/minor mode and key clarity). All algorithm parameters were kept at default value, following the methodology described e.g. in Eerola (2011). We then subjected this 28-dimension dataset to principal component analysis (Shalizi, 2013), from which we kept the first two principal components (VAE: 0.48). We used that reduced description space to cluster the 400 excerpts into 25 clusters, using the k-Means algorithm (Shalizi, 2013) with euclidean distance, and finally selected the most central excerpt to each cluster, resulting in 25 excerpts which agnostically summarized the acoustical and (evaluated) perceptual diversity of the original dataset (Figure 1).

Accord6 (accordion) A series of extremely fast and very gestural short phrases in the medium register

Cello6 A series of short rhythmic motives in pizzicati in the low register, build around the chord C-G-Ab-Eb

Clar9 (clarinet) A fast and virtuosic phrase in double-articulation, mainly noisy but letting appear a few pitches

Clar10 (clarinet) A series of noisy sounds, almost like a electronically generated pink sounds

Clar11 (clarinet) A series of high-pitched bisbigliandos

Dbass4 (double bass) A continuous bowing with a lot of pressure, creating many harmonic sounds

Dbass7 (double bass) An energetic phrase, mainly noisy, with a lot of percussive effects

Eupho7 (euphonium) A slow elegiac phrase, almost blues-inflected, with the musician singing in his instrument while playing

Eupho13 (euphonium) A phrase exploring the entire register of the instrument, very soloist-like, made of short notes and with a quasi-pulse effect

Flute3 A delicate and virtuosic phrase, with a lot of trills

Flute4 A few notes with a lot of air in the sound

Flute10 A slow, tone-centered and modal-like phrase in the low register of the instrument

Guitar1 A few guitar notes, essentially build around the lower string, and with a lot of electronic effects

Guitar7 An iterative phrase, mainly made of overlapping Gs with delay

Percu4 (percussion) A phrase made with a flexatone, alternatively struck and scraped, playing with the resonances

Percu9 (percussion) A continuous phrase around the same few notes of Bulbul Tarang

Percu11 (percussion) A very bright chain sound

Piano4 A series of short rhythmic motives with muted sounds in the low register

Piano9 A slow phrase of prepared piano, with a resonant bass

Sax12 (saxophone)A series of soft and hectic slaps

Sax21 (saxophone) A very assertive phrase of soprano saxophone, with a call-to-attention effect, which sounds like some sardinian melody

Sax26 (saxophone) A phrase made of two successive multiphonics

Spat3 (spat) A swarming and high-pitched phrase, mainly made of bird call sounds

Violin6 A phrase build around a single double-stop, starting with an unison before exploring micro-tonal deviations

Voice2 A tone-centered phrase in the medium register, with a few ornamental figures and extreme vibrato