Psychomusicology: Music, Mind, and Brain
2014, Vol. 24, No. 2, 173–183

0275-3987/14/$12.00 http://dx.doi.org/10.1037/pmu0000047

Perception of Musical Cooperation in Jazz Duets Is Predicted by Social Aptitude

Ana Pesquita, Timothy Corlis, and James T. Enns
University of British Columbia

Skilled jazz musicians are adept at coordinating their musical actions to produce an auditory outcome that is more than the sum of its parts. Whereas previous studies have investigated the cognitive mechanisms supporting ensemble music *production*, the present study focuses on the *perception* of this collaboration. The stimuli in this study were recorded duets of improvised New Orleans jazz standards, varying in the opportunity musicians were given for collaboration, from fully live performances (2-way feedback), to studio dubbed performances (1-way feedback), to studio mixes (no feedback). Participants listened to these duets in a random order and either made an explicit judgment of whether or not they were live recordings (Experiment 1) or rated the recordings on four dimensions of musicality (Experiment 2). Participants in both experiments were also categorized according to their social aptitude (Autism Quotient) and according to their musical training (Musical Expertise Questionnaire). The results showed that many listeners are sensitive to musical collaboration in this setting, and among listeners with the least musical training this sensitivity was linked to their social aptitude. These findings demonstrate that the human ability to assess the quality of a social interaction (Blakemore & Decety, 2001) is present even when the interaction is auditory, nonverbal, and in a medium in which the listeners themselves are not skilled. They also imply an important link between social aptitude and the ability to perceive the quality of a musical interaction (Phillips-Silver & Keller, 2012).

Keywords: music perception, ensemble musical performance, empathy, musical expertise, social aptitude

Supplemental materials: http://dx.doi.org/10.1037/pmu0000047.supp

The origins of music have been linked to the ability to communicate emotions through acoustic signals in early hominids (Mithen, 2006). This social feature of music is still central to the way we produce and experience music today, as seen in the popularity of musical performances that involve some degree of spontaneous collaboration (e.g., jazz, folk, rock).

Previous studies have investigated the cognitive mechanisms supporting ensemble music *production* (Keller, 2008). Yet there is little systematic research focusing on ensemble music *perception*. Anecdotal observation of audiences during live musical performances suggests that listeners greatly appreciate the expressive exchanges that occur between musicians. But to what extent are listeners sensitive to the degree of musical interaction that occurs in these performances?

Ana Pesquita, Department of Psychology, University of British Columbia; Timothy Corlis, Department of Music, University of British Columbia; James T. Enns, Department of Psychology, University of British Columbia.

This work was supported by a PhD scholarship to author A.P. from the Portuguese Fundação para a Ciência e Tecnologia (SFRH/BD/76087/2011) and an NSERC (Canada) Discovery Grant to author J.T.E. The authors thank Sarah MacDonald, Veronica Galvin, Jessica Lueng, and Mia Wilson for their help in data collection. The authors are grateful to the musicians Alan Matheson, Evan Arntzen, and Jon Roper.

Correspondence concerning this article should be addressed to Ana Pesquita, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver BC, BC V6T 1Z4 Canada. E-mail: anapesquita@psych.ubc.ca

Before describing how we approached the study of ensemble music *perception*, we briefly review evidence characterizing the affective, cognitive, and motor processes sustaining ensemble music *production.* This will help set the stage for our approach to studying perception in these settings.

A central component of ensemble music performance is affective synchronization (Phillips-Silver & Keller, 2012). It is commonly believed that musicians engage in affective exchanges that are mediated by expressive instrumental sounds. Musicians also collaborate to construct and achieve a shared musical goal. This implies a high degree of interpersonal coordination. To achieve a unified sound, the musical actions performed by each musician must be guided by the continuous anticipation and monitoring of their own sounds, the sounds produced by other musicians, and the resulting overall sound. Thus, joint musical production can be described as a sensorimotor loop in which individual performers dynamically adapt their musical actions with the aim of creating a joint auditory Gestalt (Keller, 2007, 2008).

Experimental evidence indicates that an effective joint performance is predicated on the musicians' ability to flexibly adapt to each other's sounds. Using transcranial magnetic stimulation methodology, Novembre, Ticini, Schütz-Bosbach, and Keller (2012) showed that musicians represent their own actions as well as the actions of their coperformers. Interestingly, for our interest in *perception*, these authors observed that cortical excitability associated with the representations of coperformers' actions was positively correlated with self-reported empathy. This suggests that the ability to adopt another's perspective may play an important role in joint musical production.

In addition to forming internal representations of individual musical actions (i.e., one's own and the coperformers' actions), musicians also seem to monitor the overall sound outcome. For example, Goebl and Palmer (2009) showed that the temporal dynamics of each musician in a duet is dependent on auditory and visual cues coming from the other musician. Loehr and Palmer (2011) also observed that performers in a duet corepresented their partner's musical actions to modulate the temporal features of their own performance. In a study measuring event-related potentials in musicians playing piano duets, Loehr, Kourtis, Vesper, Sebanz, and Knoblich (2013) showed that the P300 signal in event-related potentials was larger when an experimentally induced pitch alteration disrupted the joint harmonic outcome, in comparison with when it did not. These results imply that musicians attend closely to the overall product of their musical interactions.

This literature on ensemble *production* strongly suggests that ensemble musicians produce a joint auditory Gestalt through continuous cooperation. Here we ask the complementary question of whether listeners are sensitive to the auditory outcomes of this cooperative exercise. To test this question, we developed an experimental design inspired by the notion of *configural perception* in Gestalt theory. In *configural perception*, a composite with an emergent property is formed when the combination of an element and a context results in a stimulus that is more discriminable than the elements alone (Pomerantz, 2003; Pomerantz, Sager, & Stoever, 1977). As shown in Figure 1A, two diagonal lines placed within the context of the same right angle will combine to form a triangle versus a trident. Discriminating the triangle from the trident is done more rapidly and accurately than discriminating the two diagonal lines, even though the context has added no value on its own. A similar result applies to the discrimination of line position, as shown in Figure 1B. Here the two lines create the emergent property of a step versus a wall, when they are combined with the two occluding squares. Importantly, not all elements when combined with a common context become more discriminable. Therefore, when the composite is easier to discriminate than the elements, it can be concluded that the participant is perceptually better tuned to the emergent property than they are to the elements (Pomerantz, 2003).

In the present study, we applied this logic to the perception of musical synergy in jazz duets. We recorded pairs of skilled musicians improvising on New Orleans jazz standards, using two separate audio tracks so we could record each musician separately. We then combined the recorded tracks in various ways to form a continuum of possible musical cooperation. Each pair of musicians played the same small collection of songs on numerous occasions, so that we were able to sample a rich range of improvisational variation. For any given song, the musicians listened to a common click track while playing, to ensure that they played at the same tempo on every occasion that they played that song. This allowed us to combine segments from the various tracks later in the studio.

As illustrated in Figure 1C, we used individual instrument tracks as both *elements* and *contexts*. *Contexts* corresponded to instrument tracks recorded by one musician engaged in a live performance with a coperformer. In the example shown in Figure 1C, this is a recording of a clarinetist on a single occasion, labeled A. Our manipulation consisted of combining this *context* track with one of three *element* tracks. In the *live duet* condition, the *element* is a second instrument track produced by a coperformer engaged in two-way auditory interaction with the musician producing the *context* track (a flugelhorn player on the same occasion, A). In the *dub duet* condition, the *element* corresponds to an instrument track produced by a musician playing to a prerecorded *context* track, without knowledge that the other musician is not live (a flugelhorn player on a previous occasion, B). In the *mix duet* condition, the *element* is a separate live recording that was combined later in the studio with the *context* track to create a studio-mix duet (a flugelhorn player playing live with the same clarinetist but on a different occasion, C).

We predicted that listeners would be most likely to perceive musical cooperation when *element* and *context* were created with two-way auditory feedback between the duet musicians (live duet). We considered the dub and mix duets to be two different versions of a duet that should be lacking this critical feature. Note that all three conditions involve musicians playing with the intent to record skilled live music. The difference between the live and dub condition is that in the dub condition there is no opportunity for the previously recorded track to respond to anything the live musician offers by way of improvisation. The difference between the live and mix conditions is that the two-way feedback influencing the improvisation is either from the same recoding session (live) or from a different and independent recording session (mix).

The primary goal of this study was to measure listeners' sensitivity to musical interaction. We began in Experiment 1 by investigating the listeners' ability to distinguish live from not-live duet performances, along with their confidence in this decision. We also asked whether this ability varied with the social aptitude and musical skills of the participants. In Experiment 2 we tested whether live duets lead to different subjective experiences than dubbed and mixed duets. We did this by asking listeners to rate the duets on the dimensions of emotionality, engagement, creativity, and synergy. These scales were selected based on past research showing their relevance to the appreciation of solo music, as we review briefly below.

Research on the perception of solo music has shown that ratings of *emotionality* are influenced by expressive variations in tempo and amplitude, corresponding in intuitive ways to both positive (e.g., louder, faster tempo) and negative (e.g., softer, slower tempo) emotions (Bhatara, Tirovolas, Duan, Levy, & Levitin, 2011). *Engagement* represents the degree to which a particular piece of music grabs the listener's interest. The experience of being engaged in a piece of music is often described as finding the *groove*, and is used equally by ensemble musicians and the audience to describe that pleasant feeling of wanting to move some part of the body in relation to some sound pattern (Madison, 2006; Madison, Gouyon, Ullén, & Hörnström, 2011). Iyer (2002) describes *groove* as "an isochronous pulse that is established collectively by an interlocking composite of rhythmic entities" and an "attentiveness to an additional unifying rhythmic level below the level of the tactus (pp. 406)." Theoretical treatments of *creativity* in musical ensembles propose that it is predicated on three components: improvisation (online adaptation to the actions of the coperformer during the performative moment), collaboration (the joint product of the interactional dynamics between musicians), and emergence (musical patterns that supersede either individual performance; Sawyer, 2006). The fourth dimension of *synergy* was drawn directly from the Gestalt theory notion that there is added value that occurs when musicians play together to produce music

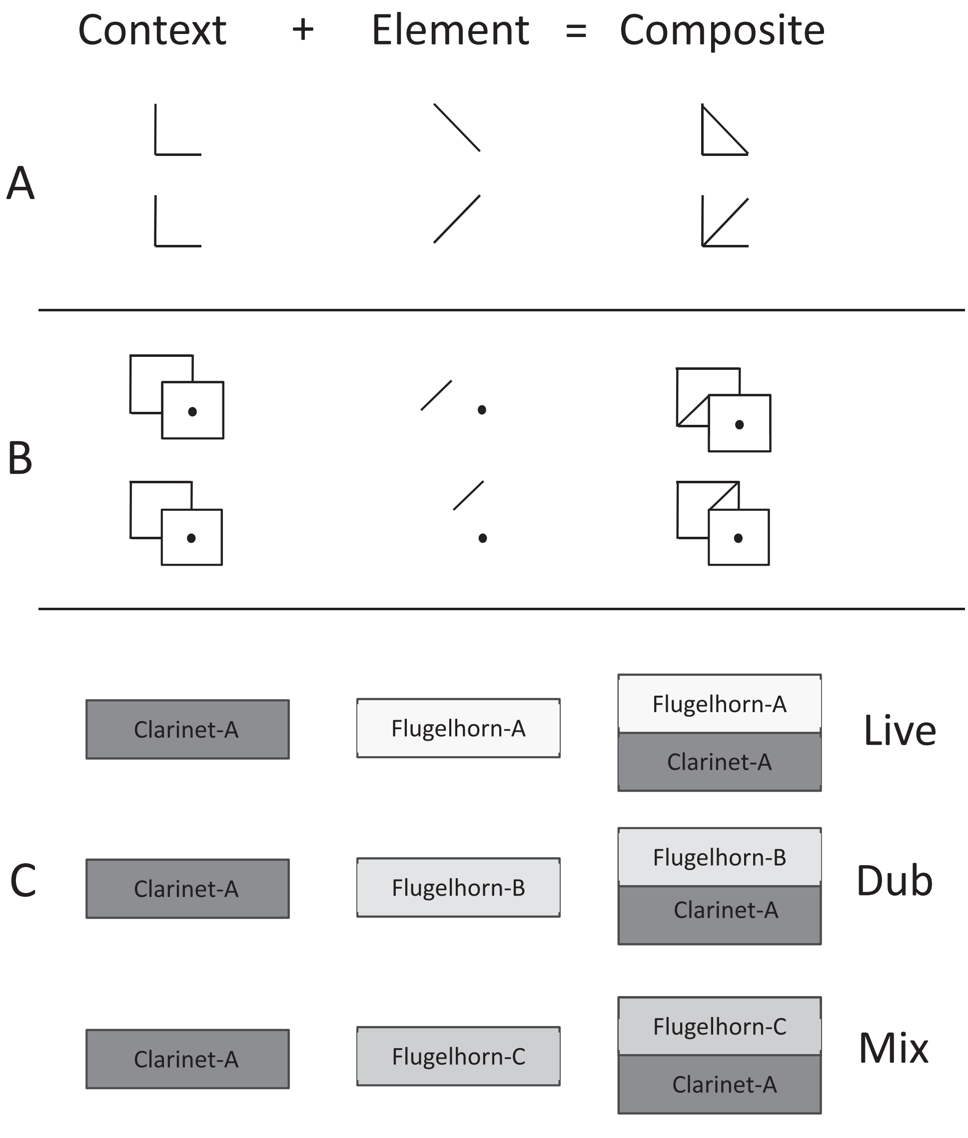

Figure 1. The logic of configural perception. A *composite* with an emergent property is formed when the combination of a *context* and an *element* results in a stimulus that is more discriminable than the *elements* on their own. (A and B) Examples of configural perception in vision. (C) Application of this logic to the construction of jazz duets. The emergent property of social synergy is expected to be greater for live duets than for dub or mix duets.

that cannot be achieved through individual performances, that is, the whole is greater than the sum of the parts.

A secondary goal of our study was to test whether listener differences in social aptitude would modulate sensitivity to the perceptual signals of cooperation. Previous research outside the music realm has shown that humans are extremely sensitive to subjective states in observed social interactions (Blakemore & Decety, 2001). Other research has shown that skills in social aptitude and empathy play a crucial role in the interpretation of social exchanges (Keysers, Kaas, & Gazzola, 2010). In the present study, we had participants fill out the 50-item Autism Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) after completing the listening session, thereby allowing us to test the hypothesis that the auditory nonverbal signals present in recorded jazz duet performances would be more accessible to listeners with relatively higher levels of social aptitude and empathy. Such a correlation would provide evidence that musical cooperation involves mechanisms that are similar to those used in other social domains.

Finally, we tested whether listener differences in musical expertise would have an influence on their sensitivity to collaboration. The effects of musical expertise on auditory perception and cognition have been extensively documented, and several studies have shown that musicians have an advantage over nonmusicians in several music-related perceptual tasks (Philip, Moore, & Brian, 1993; Pitt, 1994; Rammsayer, Buttkus, & Altenmueller, 2012). In line with this research, we predicted that musical training would increase sensitivity to the auditory features of cooperation. Supporting this possibility is a study by Aglioti, Cesari, Romani, and Urgesi (2008) showing that actions that are more similar to the observer's action repertoire allow for more accurate perception of action outcomes.

Experiment 1

In Experiment 1 we tested whether listeners could detect the presence of real-time cooperation in improvised jazz duets with an explicit report. To assess listeners' ability to discriminate among live, dub, and mix duets, we simply asked listeners to report whether each duet was a live performance (i.e., both musicians could hear and respond to one another in real time) or not (i.e., one or both of the musicians were participating by recording from a previous occasion). Following each forced-choice decision, we also asked participants to give us a confidence rating on a 6-point scale.

Method

Stimuli. Duets were recorded using soundproof recording studios in the Department of Music at the University of British Columbia. Three professional jazz musicians were paid standard studio rates for participating in the recording session. This resulted in three different instrument pairings (flugelhorn-guitar, flugelhorn-clarinet, flugelhorn-saxophone) and six different New Orleans jazz standards (Take The A Train, Beautiful Love, Canal Street Blues, Have You Seen Miss Jones, Mr. PC, and Ornithology).

During the recording sessions, musicians played in separate soundproof rooms that were connected via headphones to a central mixer. For each song, the tempo was controlled by a click track appropriate to the song, so that there was no variation in overall tempo between different takes of the same song. On the first recording session for a given musician–instrument pairing, which lasted ~1 hr, musicians played three different songs at least three different times. Following a short break for coffee, the musicians returned and played these three songs again. However, in this session they were told that on some takes, without further warning, their coperformer would not be live, but consist of the coperformer's track from a take on the previous recording session (studio dub). Whether the coperformer was live or dub was chosen randomly. Our instructions to the musicians, however, remained the same, which were to play to the best of their ability with the intent of making the most interesting improvisations possible on each occasion. A total of about 6 hr of recordings were made in this way, forming the raw material from which we constructed the three different types of duet in this study.

We began by selecting >150 clips of individual instrument tracks from these recordings that were between 40 and 60 s in length. These were taken from the main portion of the song, where the musicians were improvising. In addition, we selected three to four shorter clips of both instruments from the beginning of each song that were 15–30 s in duration, to be used as the introduction or the "head" for each listening epoch. This served to familiarize listeners with a given jazz standard in advance of being asked to judge or rate it. CuBase software was used to clip and recombine the musical clips.

From the full set of clips of improvised music, we eventually selected a set of 25 clips in each of the live, dub, and duet conditions. The constraints on this selection process were that for each of the 25 clips there had to be a base *context* clip for which we had the corresponding live clip, a dub clip, and a mix clip, as illustrated in Figure 1C. In addition, we tried to sample as many of the six songs and three instrument pairings as possible, within these constraints. Examples of these stimulus materials are given in Supplementary Material 1.

The listening experiments were run using Matlab 2010a software on an iMac computer. Participants listened to the jazz duets through Sennheiser HD 202 headphones. Sound volume was set to medium in the iMac soundcard, and the level was kept constant for all participants. Listener's responses were collected via an HP Calcpad 100 keypad. Following the listening session, which took about one half hour, participants completed the AQ and the Musical Expertise Questionnaire (MEQ), which took another 15–30 min.

Participants. Fifty-five participants (33 female), between the ages of 18 and 30 (mean age 20.3, $SD = 2.3$), were recruited from the University of British Columbia Human Subject Pool. Participants received partial course credit in exchange for 1 hr of their time. All participants reported normal hearing. The UBC Behavioral Research Ethics Board approved student participation for credit in this study.

The participant sample was divided into four groups, to test for individual differences based on median splits for social aptitude scores (AQ) and years of formal musical training. The mean AQ score was 17.15 ($SD = 6.19$), and a median split meant that participants scoring between 7 and 18 were considered higher in social aptitude than participants scoring 19–29. The average years of formal music training was 3.00 ($SD = 4.38$), and a median split meant that participants with no formal training formed the novice group and those with 1–14 years formed the musically trained group. These divisions resulted in 16 participants in the novice-lower social group, 17 in the novice-higher social group, 12 in the trained-lower social group, and 9 in the trained-higher social group. The data from one participant were incomplete because he/she failed to fill out the AQ, leaving us with 54 participants in total for the analyses of individual differences.

Procedure. Participants sat in a comfortable armchair viewing a computer screen on table. The experimenter informed the participants that the recordings they were going to hear had been generated as described above, namely, that one third of them consisted of live musicians improvised on standard tunes, one third of them were studio dubs, and one third were studio mixes. Participants listened to a total of 18 different song clips, sampling from all 75 possible clips, but with an equal number of live, dub, and mixed duets. These clips were presented in a different random order to each participant, and each successive participant's sampling of the 75 possible clips was done with replacement of the previous participant's clips.

Each listening trial began with the visual presentation of the word "introduction" on the screen, followed by the audio presentation of the head for that song. All heads corresponded to an excerpt near the beginning of a live interaction for a given song. Following a pause of 2–3 s, the word "song" appeared in its place, followed by the audio presentation of the test clip. On conclusion of the test clip, the computer screen displayed "Was this song recorded in the live condition?" and participants were instructed to respond by pressing one of two specially marked keys indicating "yes" and "no." The response was followed by a visual line on the screen, demarcated with numbers 1–6 and the phrases "not confident at all" and "extremely confident" at each end. Participants used a numbered keypad to indicate their confidence. This response initiated the onset of the next trial.

Before beginning the 18 test trials, participants undertook a practice trial to demonstrate that they were familiar with all aspects of the task. This practice trial was repeated if necessary. Participants did not receive any feedback about the accuracy of their answers in either practice or testing. On completion of the listening session, participants completed the MEQ and the AQ online.

The AQ (Baron-Cohen et al., 2001) consists of 50 items that have been grouped into five subscales: social skill (e.g., I prefer to do things with others rather than on my own.); attention switching (e.g., I prefer to do things the same way over and over again.); attention to detail (e.g., I often notice small sounds when others do not.); communication (e.g., Other people frequently tell me that what I've said is impolite, even though I think it is polite); imagination (e.g., If I try to imagine something, I find it very easy to create a picture in my mind). AQ scores measure the degree to which an individual of normal intelligence shows autistic traits. AQ scores range from 0 to 50 points, with higher points corresponding to a larger number of autistic traits. An AQ score of ≥32 points is suggested by Baron-Cohen et al. (2001) to be a useful cutoff for distinguishing individuals with clinical levels of autistic traits. It is important to caution that this scale is not used alone in clinical diagnosis (see Supplementary Material 2).

The MEQ consists of 15 items. Items focused on both musical training (e.g., "How many years of formal musical training have you completed?"; "How often do you improvise when you play?"; "How often do you play in a group or with other musicians?") and on musical preferences ("Which genre of music do you listen to the most?"; "How often do you attend live music concerts?"). The MEQ was designed to measure a participant's experience with music, with expertise being reflected in higher scores on questions relating to formal training and musical practice (see Supplementary Material 3).

Results

The chance of correctly identifying a live musical interaction was 0.33, as one third of the trials were live. The mean proportion of hits (correct detection) was 0.530 for live duets and 0.436 and 0.470 for the dub and mix duets, respectively. The d-prime (d') measure was used to measure sensitivity to live musical interactions, because it measures accuracy unaffected by any bias to respond "live" or "nonlive." This analysis indicated that the mean d' of 0.265 was significantly greater than 0.0, $t(54) = 2.09$, $p < .04$, indicating that the participants as a group could discriminate the difference in musical collaboration between live and not-live duets at a level significantly above chance.

Because we asked for a confidence rating along with each detection response, we were also able to combine participant's detection rate (i.e., *live* or *not live*) with their confidence level (i.e., 1 meaning very little confidence and 6 meaning maximum confidence) to yield a more sensitive combined score that ranged from 1–12. For example, judging a duet as being *not live* with the highest degree of confidence resulted in a score of 1; at the opposite end of the spectrum judging a song as being *live* with the highest degree of confidence resulted in a score of 12. The correlation between d' and the levels of confidence was positive at a group level, $r(4) = .56$, meaning that higher levels of confidence generally corresponded to higher levels of d'. This moderate correlation suggests that there is additional information in the confidence scores that is not captured by the d' measure alone. As such, we used the increased resolution we could get by combining the detection response and the reported confidence into a single 12-point scale.

Figure 2 shows the mean combined scores attributed to performances in the live, dub, and mix conditions. We computed a mixed analysis of variance with repeated measures factors of condition (live, dub, mix), and between groups' factors of musical training (novice, expert) and social aptitude (lower, higher). This analysis revealed a significant main effect of condition, $F(2, 100) = 3.91$, $p < .02$. Follow-up tests using the Fisher's LSD procedure indicated that scores in the live condition were significantly higher than scores in either the dub condition, $F(1, 100) = 7.11$, $p < .01$, or in the mix condition, $F(1, 100) = 4.25$, $p < .04$. The difference between dub and mix conditions was not significant, $F < 1.0$. These significant differences provided our justification for comparing the difference between the live and the not-live condition scores as a composite score, for examining potential influences of social aptitude and musical training.

Figure 3 shows the mean difference in the combined score between live and not-live conditions for each of the four groups: musical novices with lower and higher social aptitude, and musical experts with lower and higher social aptitude. An analysis of variance examining the difference between live and not-live scores

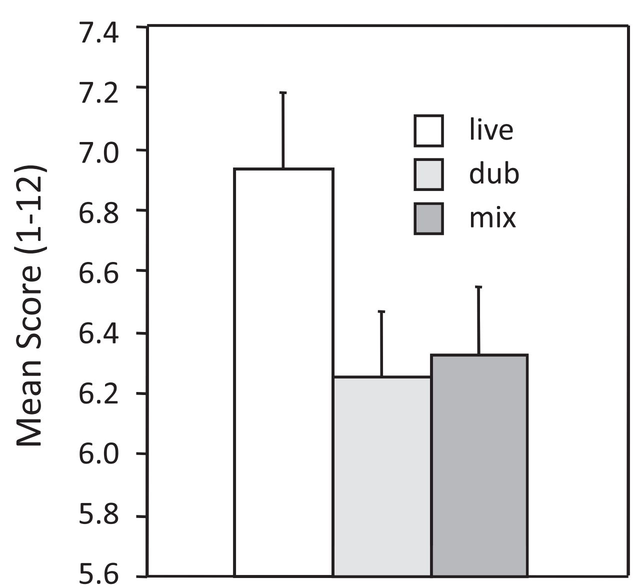

Figure 2. Results of the detection task in Experiment 1. Mean combined detection and confidence scores attributed to jazz duets in the live, dub, and mix conditions. Error bars represent one standard error of the mean.

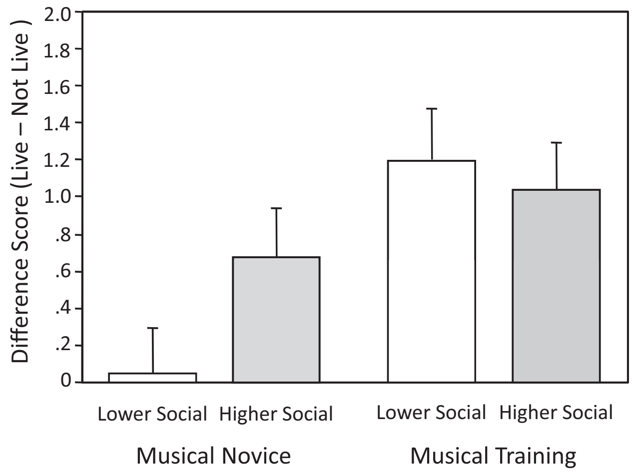

Figure 3. Mean difference in the combined score between live and not-live conditions in Experiment 1, separated for two levels of musical training and social aptitude. Error bars represent one standard error of the mean.

for the four participant groups did not reveal any significant differences between groups: social aptitude, $F(1, 50) < 1.0$, musical training, $F(1, 50) = 1.99$, and the interaction, $F(1, 50) < 1.0$, but it did indicate that the mean difference score was significantly greater than zero, $F(1, 50) = 7.87$, $p < .01$. Fisher's LSD comparisons of these mean difference score from zero within each group indicated that the musical novice-lower social group did not show significant sensitivity: $t(50) = 0.14$), but that the other three groups did (musical novice-higher social: $t(50) = 2.42$, $p < .05$, musical experts-lower social: $t(50) = 4.34$, $p < .05$, and musical experts-higher social: $t(50) = 3.76$, $p < .05$).

These trends suggest that low social aptitude and no musical training may be associated with little sensitivity to collaboration, whereas either social aptitude alone, or musical training alone, may be sufficient to enable participants to detect the difference between live collaboration versus its absence. We pursued this question in the next experiment where we investigated listener's subjective experience of these musical interactions, using four different scales relevant to musical appreciation. We also tried to increase the statistical power in two ways, first by testing a larger sample size, and second, by recruiting a sample of participants that had an even greater range of formal musical training than in Experiment 1.

Experiment 2

In Experiment 2 we investigated whether the cooperation between musicians would influence listeners' subjective experience of the jazz duets. To do so we asked listeners to indicate their appreciation of the duets in terms of the subjective dimensions of *emotion*, *engagement*, *synergy*, and *creativity*. We also tested for the influences of *social aptitude* and *musical expertise*, by seeing how ratings on the four subjective dimensions varied with the individual differences among our participants on the AQ and the MEQ. In an effort to recruit a greater number of participants with musical expertise, we also recruited students pursuing a postsecondary degree in musical performance.

Method

Participants. We recruited listening participants from both psychology classes and from a school of music. Seventy psychology undergraduate students (36 females; mean age = 24.3 years, $SD = 4.3$) and 41 students in an undergraduate music performance program (25 females; mean age = 23.2 years, $SD = 4.1$ year) were recruited at the University of British Columbia. Students participated in a 1-hr session. Psychology students received one course credit, and music students received CAD $20. All participants reported normal hearing. Student participation was approved by the UBC Behavioral Research Ethics. The data from two of the music students were incomplete in either the AQ or the MEQ, leaving 39 music students in the final analysis.

The participant sample was divided into four groups, to test for individual differences based on years of formal musical training and median splits for social aptitude scores based on the AQ. With regard to music, the entire sample reported a mean of 7.06 years of formal musical training ($SD = 6.06$ years), and a median split meant that the novice group had 0–1 year of formal musical training whereas the trained group reported 3–20 years. Most of the formal training was in classical music, with only four students indicating jazz as their main musical genre. Students with formal musical training reported high levels of ensemble experience, with 65.9% reporting they played with other musicians on a daily basis, 19.5% reported playing in a group at least once a week, and the remaining 14.6% played in group once a month or less. About one half of the music students (48.8%) reported that they rarely if ever engaged in improvised playing, 12.2% of the music students reported practicing improvised playing daily, 17.1% improvised at least once a week, and 21.9% improvised only once a month or less.

The mean AQ score was 19.06 ($SD = 5.90$) and a median split meant that participants scoring between 9 and 17 were considered higher in social aptitude than participants scoring 18–32. The four groups that resulted from these divisions meant that 31 participants were considered novice-lower social, 39 were novice-higher social, 16 were musically trained-lower social, and 23 were musically trained-higher social.

Stimuli. This experiment used the same pool of 75 clips from the three different types of jazz duets (25 live, 25 dubbed, and 25 mixed clips). As in Experiment 1, each listener randomly sampled a set of 18 clips (6 in each condition) from this larger set.

Procedure. The procedure was the same as Experiment 1, with the following exceptions. After each listening trial, four phrases appeared on screen (e.g., "working together") in succession. For each phrase, listeners were told to indicate the extent to which that phrase matched the duet they had just heard. The phrases were selected randomly from those shown in Table 1. Eight phrases were created for each dimension (i.e., emotion, engagement, creativity, and synergy), with four phrases representing a positive formulation and four representing a negative formulation. Scores from negative phrases were reversed-coded. On each trial, phrases were selected randomly from this set, with the constraints that one phrase had to be drawn from each of the dimensions and that an equal number of phrases were selected from the positive and negative columns. Listeners indicated their ratings using a scale from 1 ("strongly disagree") to 6 ("strongly agree"). The purpose behind making the rating experience this

Table 1
Phrases Used to Rate the Jazz Duets in Experiment 2

Dimension	Phrases: Positive	Phrases: Negative
Emotion	Good vibe	No real vibe
	Expressive	Dull
	Emotional	Not emotional
	Energetic	Not energetic
Engagement	Groovy	Cannot feel the groove
	Toe-tapping	Lame
	Grabs my attention	Fails to hold my interest
	Fascinating	Boring
Synergy	In sync	Out of sync
	Working together	Ignoring each other
	Has synergy	Does not have synergy
	Having a conversation	Playing to themselves
Creativity	Innovative	Lacking in innovation
	Creative	Formulaic
	Sophisticated	Unsophisticated
	Masterful	Amateur

diverse (i.e., four phrases selected randomly from four dimensions and two levels of valence) was to prevent participants from responding in stereotyped ways, independent of the question being asked.

Results

Figure 4 shows the mean ratings for each of the four musical dimensions. These results show that ratings across all of the dimensions were generally higher for live duets than for the dub and mix duets, indicating that collaboration in the improvised duets influenced participants' subjective enjoyment of the music. A second finding was that some dimensions were generally rated more highly across all the duets than other dimensions (i.e., mean ratings were highest for emotion, followed by engagement, with synergy and creativity receiving the lowest ratings). These observations were supported by the following statistical analyses.

A repeated-measures analysis of variance examining condition (*live*, *dub*, or *mix*) and dimension (*emotion*, *engagement*, *synergy*, and *creativity*) indicated significant main effects for both factors, condition: $F(2, 210) = 5.29$, $p < .01$, $MSE = 0.487$; dimension: $F(3, 315) = 9.17$, $p < .001$, $MSE = 0.397$; and no significant interaction between these factors, $F(6, 630) < 1.0$. The main effects were examined in greater detail with Fisher's LSD tests, which showed that duets in the live condition were rated significantly higher than in the dub condition, $F(1, 210) = 8.72$, $p < .01$, and in the mix conditions, $F(1, 210) = 7.07$, $p < 0.05$, but that the dub and mix conditions did not differ significantly, $F(1, 210) < 1.0$. A similar exploration of the dimension effect indicated that the rank ordering of emotion ratings, followed by engagement, synergy, and creativity ratings, were all significantly different from one another ($p < .05$) except for engagement versus synergy, and synergy versus creativity ($p > .05$). No other effects were statistically significant.

As in Experiment 1, we next examined these data as a function of individual differences in social aptitude and musical training. Figure 5 shows the average ratings across all four dimensions between live, dub, and mix conditions for these four groups. These data increase support for the trend suggested in Experiment 1, that among participants with relatively little musical training (the upper panel of novices in Figure 5), sensitivity to musical interaction is greatest among those with relatively higher levels of social aptitude. However, among participants with formal musical training, there are two interesting deviations from this pattern (the lower

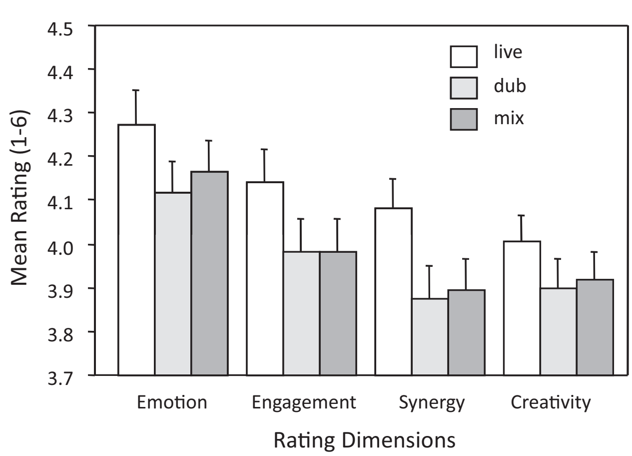

Figure 4. Mean ratings for four musical dimensions in Experiment 2. Error bars represent one standard error of the mean.

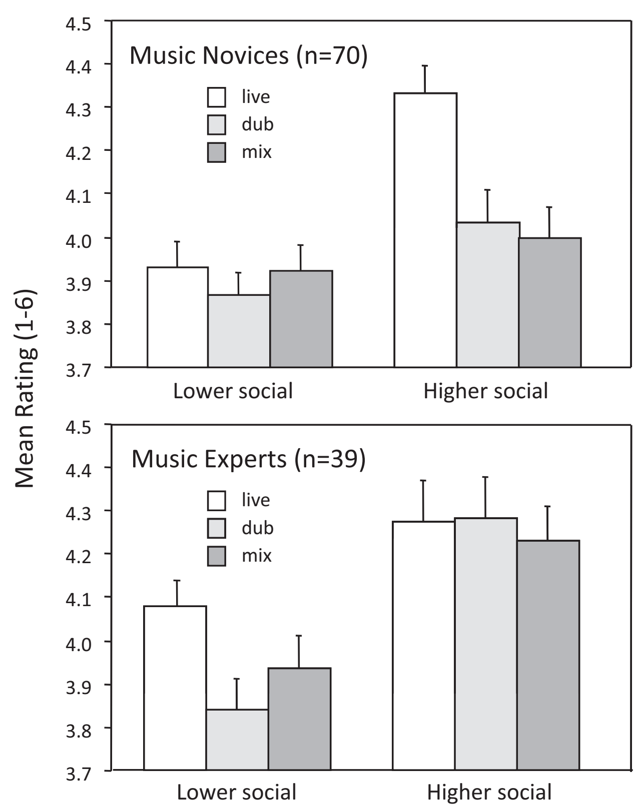

Figure 5. Mean ratings over all four dimensions in Experiment 2, separated for two levels of musical training (upper and lower) and social aptitude (left and right). Error bars represent one standard error of the mean.

panel of music experts in Figure 5). First, listeners with *lower* levels of social aptitude as measured by their AQ score showed greater sensitivity to the differences between live and not-live duets. Second, expert musicians with relatively higher levels of social aptitude tended to give high ratings to all of the duet conditions. These observations were supported by the following statistical analyses.

A mixed analysis of variance examining the repeated measures factors of condition and dimension, along with the between-group factors of musical training (novice, expert) and social aptitude (lower, higher), indicated a significant main effect of social aptitude, $F(1, 105) = 5.95$, $p < .02$, along with a significant three-way interaction of condition × musical training × social aptitude, $F(2, 210) = 3.42$, $p < .04$. The main effect of social aptitude reflected that participants with higher social aptitude tended to give higher ratings to the duets in general than participants with a lower social aptitude (see right-hand panels in Figure 5).

The significant three-way interaction was explored further by examining the difference between live and not-live (dub, mix) ratings for the four groups of participants. The results can be seen in Figure 6. This interaction, when expressed as a difference score that indexes participants' sensitivity to musical collaboration, was significant, $F(1, 105) = 6.11$, $p < .02$. Fisher's LSD tests indicated that among the musical novice group, participants with higher social aptitude were more sensitive to musical interaction than those with lower social aptitude, $F(1, 105) = 6.54$, $p < .02$. However, among the musically trained participants, social aptitude was not a significant influence, $F(1, 105) = 1.40$, $p < .24$, though the trend was for those with lower social aptitude to show somewhat greater sensitivity to musical interaction. No other effects were statistically significant.

These analyses suggest that for participants with little musical expertise there is a positive correlation between sensitivity to musical collaboration and social aptitude, whereas for participants with greater musical expertise there is a negative correlation. This hypothesis was examined in greater detail by computing the correlations between the five subscales that make up the AQ (social skill, attention switching, attention to detail, communication, imagination), and each participant's sensitivity to collaboration scores. An overall score was obtained for each participant by calculating the difference between live and not-live conditions across all four dimensions (emotion, engagement, synergy, creativity). Each of these correlations was conducted separately for the group of 70 novice musicians and the group of 39 expert musicians. We note that these correlations should be interpreted cautiously, as many of them were not significantly different from zero on their own, or from one another. Nevertheless, their global pattern can be used instructively to guide future theorizing and research.

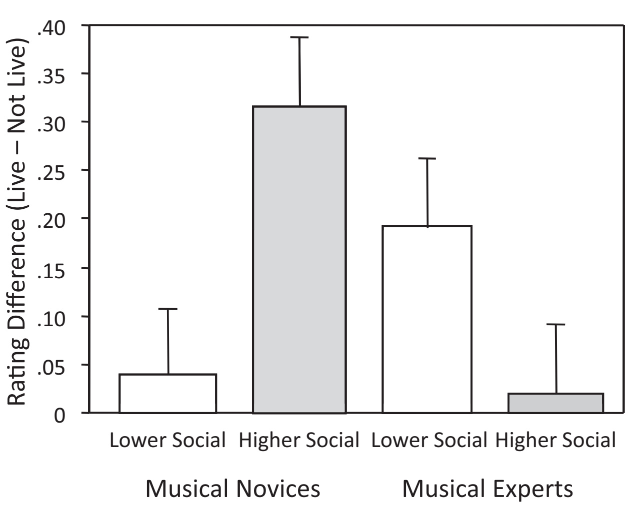

Figure 6. Mean difference in ratings for live and not-live ratings in Experiment 2, separated for two levels of social aptitude and music training. Error bars represent one standard error of the mean.

The correlations between sensitivity differences scores and AQ scores were uniformly negative for the novice group (*r*'s ranged from −.021 to −.215) and uniformly positive for the expert group (*r*'s ranged from .202 to .346). Most importantly, the strongest correlation in the novice group was between the sensitivity difference score and the total AQ ($r(68) = -.215$), indicating that the global AQ score was the best predictor of sensitivity to musical collaboration in this group. Thus, in the novice group a relatively higher global AQ correlated with a decrease in sensitivity to musical collaboration. However, among the expert group, the strongest correlation was a positive one with the subscale of attention to detail ($r(37) = .346$), suggesting that higher scores in AQ were associated with greater sensitivity to differences between duet conditions. This might reflect that listeners in this group tended to focus strongly on particular details, thereby approaching the task analytically rather than using their intuition about social interactions.

General Discussion

Musicians cooperate with one another during ensemble performances to create an auditory outcome with affective, cognitive, and motor consequences (Keller, 2007, 2008; Phillips-Silver & Keller, 2012). In this study, we measured the sensitivity of *listeners* to these signals. The results from two experiments indicated that many listeners, even those untrained in music, are sensitive to the outcomes of musical cooperation in improvised jazz duets. This sensitivity to musical cooperation was evident in both explicit judgments of whether the recorded duets involved two cooperating musicians (Experiment 1) and in ratings on four dimensions of subjective musical appreciation (Experiment 2).

This first and most general finding of our study is consistent with previous literature indicating that humans are sensitive to the quality of an observed social interaction (Decety & Jackson, 2006). This background literature indicates that the capacity is evident in more than one sensory modality. For example, using only vision, participants were able to identify the person occupying a leadership position at levels greater than chance, after studying static pictures depicting pairs of people working together (Mast & Hall, 2004). In the auditory domain, Brennan and Williams (1995) showed that participants listening to a question–answer dialogue were able to use the paralinguistic features of the utterances in the dialogue (i.e., pauses, intonation, and interjections) as predictive cues of the speakers' relative state of knowledge. Using recordings that contained both video and auditory signals, Cappella (1997) showed that a naïve observer could predict with above chance accuracy the future success of a marriage after watching video snippets of a couples' conversation.

These examples all support the idea that humans have the remarkable capacity to gauge the quality of a social interaction as third-person observers. The present data add to this picture by showing that this is possible for improvised jazz. Notably, the listeners in our study only *heard* the musical performance (they did not see the musicians play together and so had no visual cues to judge the interaction), the performance involved strictly *instrumental* music (so there were no verbal cues as in previous studies in the auditory domain), and many of the participants showing this sensitivity had very *limited knowledge* or experience with this musical genre (unlike participants in previous studies who have lifelong experience evaluating everyday social exchanges at home and at work). Nonetheless, these listeners were able to distinguish whether they were listening to a live recording or not (Experiment 1) and reported greater appreciation for duets when they offered greater opportunity for social collaboration compared with when they had been created without mutual collaboration (Experiment 2).

The results for individual differences in the explicit detection task in Experiment 1, involving a majority of listeners with relatively little musical training, pointed to the possibility that either a relatively higher level of social aptitude, or a relatively higher level of musical training, might each be sufficient to allow listeners to detect the musical cooperation present in the live duets. This raises the interesting theoretical possibility that participants without formal musical training, but with a higher social aptitude, may be accomplishing this feat through their sensitivity to the mutually expressed affective signals in the live duets. It also raises the question of whether participants with greater musical proficiency, but relatively lower social aptitude, may be accomplishing the same feat through a different means, perhaps through a more sophisticated appreciation of the musical structure and its technical realization by the ensemble. The plausibility of this hypothesis is consistent with the view that styles of music appreciation vary considerably between individuals (Chamorro-Premuzic & Furnham, 2007). Qualitative research suggests that for some individuals music listening is primarily linked to the perception of structural features; for other individuals the focus is more on the emotional content of the musical piece (Gabrielsson & Wik, 2003).

In Experiment 2 we tested this hypothesis more directly by making two changes to our procedure. First, we recruited listeners with a much larger range of musical training. Second, we asked listeners to respond with their subjective impressions in response to adjectives on four different dimensions of music appreciation. The first finding of Experiment 2 was that the social synergy signal in the duets had an influence on the overall subjective ratings of musicality. This is important because it demonstrates that the signal is not merely apparent to cognition (e.g., a listener can detect that two musicians are having a live exchange), but that it also has an influence on the listener's affective appreciation (subjective evaluation) of the duets.

The results of Experiment 2 indicated that live duets were experienced as more emotionally expressive than nonlive duets, suggesting that the dynamic cooperation between musicians facilitated a more consistent and coherent emotional experience. This finding in the context of duets implies that there is more to emotion perception in music than is communicated solely by variations in tempo and amplitude (Bhatara et al., 2011). This is because both live and not-live duets in the present study contained similar variation in these features at the level of individual instruments. The way that live duets differed from not-live duets in the present study was in their opportunity for dynamic interaction between the two performers. The finding that live duets were rated as more engaging than not-live duets suggests that listeners were better able to find the pulse underlying the performances of the two musicians. Future research should follow up on this possibility by testing whether live duets create more coherent and unified rhythmic textures (Keller, 2007, 2008), when expressed directly through the body movements of the listeners (e.g., tapping). The finding that live duets were generally judged to be more creative than not-live duets favors the theoretical proposal that *creativity* in musical ensembles is predicated on online adaptation to the actions of the coperformer during the performative moment (Sawyer, 2006).

The present finding that live duets were generally judged to have more synergy suggests that the value added when musicians play together is an accessible aspect of the listener's experience. This supports the use of the theoretical framework of configural perception as a viable way to investigate ensemble music perception (Pomerantz, 2003; Pomerantz et al., 1977). We did this by combining individual instrument tracks into duet *composites*, where the individual tracks function as either *elements* or *contexts*. *Contexts* were instrument tracks recorded by one musician engaged in a live performance with a coperformer. We created *composite* duets by combining the *context* track with an *element* track and then made comparisons among *element* tracks with same *context* track. The *element* tracks were combined with the *context* track in three ways: as a live instrumental duet (allowing for two-way auditory interaction), as an overdubbed recording (allowing only one-way interaction from the live musician), and as mixed tracks (allowing for no interaction because each track was from a different recording session of the same song). The finding that participants were sensitive to these differences is thus an objective measurement of the stronger perceptual configuration that is created when musicians play live music.

We note with interest that the findings of Experiment 2 did not provide any evidence that musical training interacted with listeners' responses to questions about the four dimensions (i.e., emotion, engagement, synergy, and creativity). Instead, all listeners generally rated the live duets as more positive than the dub and mix duets on all four dimensions. This is contrary to the hypothesis that musical novices might use different dimensions to assess the duets than more highly trained musical experts. However, we caution that this null result not be interpreted too strongly. For example, it is possible that musical experts perceive and communicate their experience of joint musical performances using different or more specialized categories than were tested here. Drawing on Sawyer's (2006) theoretical analysis, such dimensions might more directly address what Sawyer considers to be the pillars of music synergy: *improvisation* (online adaptation to the actions of the coperformer during the performative moment), *collaboration* (the joint product of the interactional dynamics between musicians), and *emergence* (musical patterns that supersede either individual performance). Testing these more technical dimensions in the future may allow expert musicians to better communicate their appreciation for the structural and technical features of musical interaction.

A second finding of Experiment 2 concerned the contrasting relations we observed for the role of social aptitude in the ratings of music novices and experts. Among novices it appears that the perception of musical cooperation in the duets was primarily apparent to those listeners with higher levels of social aptitude (upper panel in Figure 5). This finding is consistent with previous evidence that empathy and social aptitude positively influence a listener's ability to perceive musical emotion and expressiveness. For example, Wöllner (2012) asked observers to make continuous ratings of the expressiveness of individual music performances. Participants were tested in visual only, auditory only, and multimodal stimuli conditions. The author reported that higher levels of empathy in listeners facilitated the perception of expressive musical intentions, but only in the multimodal condition. The author's interpretation was that these participants were sensitive to the correlations between the musician's visible body movements and the corresponding sounds, which the author referred to as perception–action circuits that the participants could experience through automatic processes of embodied empathy. Adding to this evidence, a recent study by Quintin, Bhatara, Poissant, Fombonne, & Levitin (2011) showed that adolescents with autism spectrum disorders are impaired relative to intelligence-matched controls and individuals with Williams syndrome at judging the difference in emotional expressiveness in solo piano pieces. Because autism is characterized by impairments in social aptitude and empathizing abilities, this finding emphasizes the importance of social skills for perceiving the expressive and emotional qualities of music. The present finding that social aptitude correlated positively with sensitivity to musical cooperation in jazz duets adds to this literature by showing that listeners are sensitive to correlations between performing musicians even when there are no visual cues, there are no verbal cues, and the listeners have limited knowledge or experience with the musical genre.

In contrast with the pattern found for novice musicians, there was a trend for musical collaboration to be better detected among expert musicians with *lower* levels of social aptitude. When we examined the correlations between this ability and the individual subscales of the Autism Quotient, the overall pattern suggested the interpretation that the link between Autism Quotient and sensitivity to social synergy might be different for novices and experts. In particular, it was the *attention to detail* subscale that seemed to dominate the data for music experts, whereas it was the overall Autism Quotient total that was the best predictor for music novices.

High *attention to detail* is a trait characteristic of the more general tendency for systemization in the autism spectrum (Baron-Cohen, 2010). But high systematizing tendencies are also common outside the disorder spectrum, as for example, among physics students (Baron-Cohen et al., 2001). Kreutz, Schubert, and Mitchell (2013) measured *empathizing* and *systemizing* as specific cognitive traits in professional musicians, amateur musicians, and nonmusicians. The results suggested *empathizing* and *systemizing* were independent contributors to effective musical processing, as the *systemizing* traits increased with musical training, while *empathizing* traits did not. The authors proposed that musical training results in a specific focus on the perception and interpretation of formal musical patterns and structures. The present finding, that music experts with higher levels of *attention to detail* showed an increased sensitivity to musical cooperation, is thus consistent with these and related findings linking systemizing traits to musical training (Bhatara et al., 2011; Kreutz, Schubert, and Mitchell (2013).

A final result to consider is that expert musicians with higher levels of social aptitude tended to give uniformly positive ratings to all of the duet conditions. This finding is counterintuitive, at least at first glance, because it is a common expectation that expertise is associated with outstanding performance in the relevant domain (Ericsson, 2005). A possible interpretation for this surprising result is that listeners who are both highly social and musically trained are motivated to actively create social interpretations for the duet music. This possibility is consistent with other evidence that people register ambiguous information in ways that align with their desires, sometimes called *wishful seeing* (Dunning & Balcetis, 2013). This interpretation is also consistent with our anecdotal experience of playing music from our three conditions to the original performers some months after the original recordings were made. These musicians who created the stimuli in the study were uniformly unable to distinguish the live music from the other two conditions, often giving especially strong attributions of creativity and interest to the studio mixes they had not even made (though they did not know that).

Another interpretation of this puzzling result is that these listeners have a tendency to positively appreciate both the expected products of cooperation and the unexpected products of miss-cooperation. Supporting this possibility is the conceptualization that expert musicians use moment-to-moment modulations of *expectation* and *surprise* to construct expressive performances (Huron, 2006). Therefore, it is possible that expert musicians value the miss-alignments between instruments in noncooperative music as an unexpected expressive choice, corresponding to *surprise* in Huron's (2006) terms.

In conclusion, this study presents novel evidence on human sensitivity to the quality of musical interaction in improvised jazz duets. The secondary analyses we have conducted of individual differences in this ability point to the possibility that both social aptitude (i.e., human empathy and a theory of mind) and musical expertise are separate contributing influences to this ability. Future studies will need to focus more directly on this hypothesis, along with considering such questions as (a) What are the psychoacoustic characteristics of musical cooperation in improvised jazz duets? and (b) How widely do these findings generalize across musical genres, listener cultures, types of musical expertise, and modes of response?

References

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience, 11*, 1109–1116.

Baron-Cohen, S. (2010). Empathizing, systemizing, and the extreme male brain theory of autism. In I. Savic (Ed.), *Sex differences in the human brain, their underpinnings and implications* (Vol. 186, pp. 167–175). Amsterdam: Elsevier.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism-Spectrum Quotient (AQ): Evidence from asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31,* 5–17.

Bhatara, A., Tirovolas, A. K., Duan, L. M., Levy, B., & Levitin, D. J. (2011). Perception of emotional expression in musical performance. *Journal of Experimental Psychology Human Perception and Performance, 37,* 921–934.

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience, 2,* 561–567.

Brennan, S., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34,* 383–398.

Cappella, J. N. (1997). Behavioral and judged coordination in adult informal social interactions: Vocal and kinesic indicators. *Journal of Personality and Social Psychology, 72,* 119–131.

Chamorro-Premuzic, T., & Furnham, A. (2007). Personality and music: Can traits explain how people use music in everyday life? *British Journal of Psychology, 98,* 175–185.

Decety, J., & Jackson, P. (2006). A social-neuroscience perspective on empathy. *Current Directions in Psychological Science, 15,* 54–58.

Dunning, D., & Balcetis, E. (2013). Wishful seeing: How preferences shape visual perception. *Current Directions in Psychological Science, 22,* 33–37.

Ericsson, K. A. (2005). Recent advances in expertise research: A commentary on the contributions to the special issue. *Applied Cognitive Psychology, 19,* 233–241.

Gabrielsson, A., & Wik, S. L. (2003). Strong experiences related to music: A descriptive system. *Musicae Scientiae, 7,* 157–217.

Goebl, W., & Palmer, C. (2009). Synchronization of timing and motion among performing musicians. *Music Perception, 26,* 427–438.

Huron, D. B. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge: MIT Press.

Iyer, V. (2002). Embodied mind, situated cognition, and expressive microtiming in african-american music. *Music Perception: An Interdisciplinary Journal, 19,* 387–414.

Keller, P. E. (2007). Musical ensemble synchronisation. In C. Stevens, E. Schubert, B. Kruithof, K. Buckley, & S. Fazio (Eds.), *Proceedings of the inaugural international conference on music communication science* (pp. 80–83). Sydney, Australia: Marcs Institute

Keller, P. (2008). Joint action in music performance. In F. Morganti, A. Carassa, & G. Riva (Eds.), *Enacting intersubjectivity: A cognitive and social perspective on the study of interactions* (pp. 205–221). Amsterdam: IOS Press.

Keysers, C., Kaas, J. H., & Gazzola, V. (2010). Somatosensation in social perception. *Nature Reviews Neuroscience, 11,* 417–428.

Kreutz, G., Schubert, E., & Mitchell, L. A. (2013). Cognitive styles of music listening. *Music Perception: An Interdisciplinary Journal, 26,* 57–73.

Loehr, J. D., Kourtis, D., Vesper, C., Sebanz, N., & Knoblich, G. (2013). Monitoring individual and joint action outcomes in duet music performance. *Journal of Cognitive Neuroscience, 25,* 1049–1061.

Loehr, J. D., & Palmer, C. (2011). Temporal coordination between performing musicians. *Quarterly Journal of Experimental Psychology, 64,* 2153–2167.

Madison, G. (2006). Experiencing groove induced by music: Consistency and phenomenology. *Music Perception: An Interdisciplinary Journal, 24,* 201–208.

Madison, G., Gouyon, F., Ullén, F., & Hörnström, K. (2011). Modeling the tendency for music to induce movement in humans: First correlations with low-level audio descriptors across music genres. *Journal of Experimental Psychology Human Perception and Performance, 37,* 1578–1594.

Mast, M. S., & Hall, J. A. (2004). Who is the boss and who is not? Accuracy of judging status. *Journal of Nonverbal Behavior, 28,* 145–165.

Mithen, S. (2006). *The singing neanderthals: The origins of music, language, mind, and body*. Cambridge, MA: Harvard Univ Press.

Novembre, G., Ticini, L. F., Schütz-Bosbach, S., & Keller, P. E. (2012). Distinguishing self and other in joint action. Evidence from a musical paradigm. *Cerebral Cortex, 22,* 2894–2903.

Philip, A., Moore, F., & Brian, C. J. (1993). Frequency analysis and musical ability. *Music Perception: An Interdisciplinary Journal, 11,* 39–53.

Phillips-Silver, J., & Keller, P. E. (2012). Searching for roots of entrainment and joint action in early musical interactions. *Frontiers in Human Neuroscience, 6,* 26.

Pitt, M. A. (1994). Perception of pitch and timbre by musically trained and untrained listeners. *Journal of Experimental Psychology Human Perception and Performance, 20,* 976–986.

Pomerantz, J. (2003). Wholes, holes, and basic features in vision. *Trends in Cognitive Sciences, 7,* 471–473.

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 422–435.

Quintin, E.-M., Bhatara, A., Poissant, H., Fombonne, E., & Levitin, D. J. (2011). Emotion perception in music in high-functioning adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 41,* 1240–1255.

Rammsayer, T., Buttkus, F., & Altenmueller, E. (2012). Musicians do better than nonmusicians in both auditory and visual timing tasks. *Music Perception: An Interdisciplinary Journal, 30,* 85–96.

Sawyer, R. K. (2006). Group creativity: Musical performance and collaboration. *Psychology of Music, 34,* 148–165.

Wöllner, C. (2012). Is empathy related to the perception of emotional expression in music? A multimodal time-series analysis. *Psychology of Aesthetics, Creativity, and the Arts, 6,* 214–223.

Received September 9, 2013
Revision received March 29, 2014
Accepted May 12, 2014 ■