Research Article

Discrepancies and Disagreements in Classical Chamber Musicians' Characterisations of a Performance

Neta Spiro[1,2] and Michael F. Schober[3]

Abstract

To what extent do classical chamber musicians converge in their characterisations of what just happened in their live duo performance, and to what extent do audience members agree with the performers' characterisations? In this study a cello-piano duo performed Schumann's Phantasiestücke, Op. 73, no. 1 as part of their conservatory studio class in which members critique performances in development. Immediately after, the listeners and players individually characterised what had most struck them about the performance, first writing comments from memory and then marking scores while listening to a recording on their personal devices. They all then rated (on a 5-point scale) their agreement with comments by two other class members. Findings demonstrate that classical chamber performers can characterise the performance quite differently than their partner does and that they can disagree with a number of their partner's characterisations, corroborating previous findings in case studies of jazz performance. Performers' characterisations can overlap less in which moments strike them as worthy of comment and in their content than their listeners' characterisations do, and they can agree with a non-partner's characterisations more than with their partner's characterisations. At the same time, the data show that listeners who have played the piece before—though not necessarily those who play the same kind of instrument (strings vs. piano)—can be more likely to endorse comments by others who have also played the piece before, even if the comments they make don't overlap with each other more in timing, content or theme.

Keywords

audience, chamber music, interpretations, data visualisation, performers, shared understanding

Submission date: 04 July 2020; Acceptance date: 31 March 2021

Introduction

To what extent do classical chamber performers understand what just happened in their live duo performance in the same way as each other, and to what extent do audience members understand the performance in the same way as the performers? Anecdotally, performers often feel as if they are intersubjectively connected with their performing partners, and they like to assume that their listeners "get" their expressive intentions as performers. Musically experienced audience members like to feel that they have deep understanding of the piece and fair critical perspectives on the performance. The fact that performers follow the same score—and that listeners may have had access to that same score and may have heard previous performances of the same piece—makes it plausible that the experience, reactions and interpretations of co-performers and (at least musically sophisticated) audience members should largely overlap. But how much do they?

Much can and perhaps must be shared among performers for them to be able to play together (Davidson & Good, 2002), from sharing sufficient understanding of the structure of the piece (e.g., Ragert et al., 2013; Williamon

[1] Centre for Performance Science, Royal College of Music, UK
[2] Faculty of Medicine, Imperial College London, UK
[3] The New School for Social Research, New York, NY, USA

Corresponding author:
Neta Spiro, Centre for Performance Science, Royal College of Music, South Kensington, London, United Kingdom;
Faculty of Medicine, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom.
Email: neta.spiro@rcm.ac.uk

& Davidson, 2002), to interpreting each other's visual and auditory cues (e.g., Bishop & Goebl, 2015; Keller, 2014; Williamon & Davidson, 2002), to predicting and, by some theories, simulating the actions of their performing partner and presenting signals well enough to be able to coordinate (e.g., Keller et al., 2007; Novembre et al., 2019). Of course, how conscious performers are, or even can be, of these processes (Schiavio & Høffding, 2015) and exactly what kinds of representations must be involved is less well understood, but the presumption that much is shared is common.

On the other hand, prior evidence suggests that there can be less overlap between performers' impressions than one might expect. In one study (Wöllner, 2013), string quartet members did not necessarily agree with each other's ratings of their own and each other's expressivity in a joint performance they watched on video. Although it is not in a classical genre, in a previous case study on jazz standard improvisation (Schober & Spiro, 2014) the performers (a saxophonist and pianist) did not fully agree with their partner's characterisations of what occurred in the improvisations – music-analytically, collaboratively and evaluatively – and they agreed with more of a commenting listener's characterisations than their partner's. In a subsequent study of a large set of musically experienced listeners, far fewer listeners agreed with the original performers' judgements than with the commenting listener's judgements (Schober & Spiro, 2016). Similarly, in a case study of free jazz improvisers' shared understanding, the performers did not agree with each other's characterisations of the improvisation more than they agreed with other experienced free jazz performers' characterisations (Pras, Schober, et al., 2017).

Our strategy in this case study, as in the previous case studies (Pras, Schober, et al., 2017; Schober & Spiro, 2014, 2016), was to start with what a pair of experienced classical chamber performers and their listeners independently thought worth articulating about their performance immediately afterwards, and then to assess (a) the extent to which they endorsed each other's comments and (b) the extent to which their patterns of judgment across multiple statements agreed with each other. Unlike in the previous case studies, this time the performance context was a conservatory studio class in which advanced students present works in progress for collective critique. The listeners were 12 musically knowledgeable audience members—members of the class, including two members of another duo who were prepared to play the same piece as a work in progress for critique that day. Rather than using predefined questions or statements about the performance for participants to rate (as in, e.g. Juslin et al., 2011; Platz & Kopiez, 2013; Thompson & Williamon, 2003; Wesolowski, 2016, among many others), we prompted the performers and audience members to individually write down the top three things that had struck them about the performance (like Waterman, 1996). We then asked them to listen to a recording of this performance

on their personal devices, and to mark on scores and write about three moments that struck them as worthy of comment. After all had finished, the response sheets were twice redistributed for each participant to rate their agreement (on a 5-point scale) with, and elaborate on if they wished, two other participants' characterisations.

As we have proposed before (Schober & Spiro, 2016), the range of possible levels of shared understanding we could observe here is actually quite large. At one end of the spectrum we could expect *radical idiosyncrasy*: that people's experiences of music may so differ that participants (whether performers or listeners) will never have identical or even strongly overlapping experiences and interpretations. A less extreme possibility—*minimal overlap*—is that the level of shared understanding depends on the type of information or experience that is focused on: interpretation and experience of some aspects (perhaps mis-tunings, or the music's temporal structures and surface features) might be more likely to be shared, while broader interpretations of intention and evaluations of success or expression might be very different. On the other end of the spectrum, *specific content* views suggest that there might be particular aspects of musical performance that participants can identify more or less accurately, for example aspects of players' communication and alignment (e.g., Keller, 2014; King, 2006), specific expressive, affective or structural characteristics of the music (e.g., Canonne & Garnier, 2015; Hargreaves et al., 2005), judgments of performers' fidelity to the score (e.g., Waddell et al., 2018) or interpretations of musical meaning (e.g., Clarke, 2005). One might expect classical chamber duos—particularly if they have performed together, rehearsed together or discussed the music before (Ginsborg & King, 2012)—to overlap substantially in understanding of specific content, particularly given shared background knowledge of the piece and musical style (Pitts, 2013).

As for listeners' and performers' shared understanding, we propose that the same possibilities are plausible for classical chamber performance as in jazz standard performance (Schober & Spiro, 2016): the greater the experience in the genre, the more likely the overlap in understanding (a *more-experienced-listeners-understand-more-like-performers* hypothesis) and, at the same time, participants in an interaction may experience and understand what happened differently from outsiders (a *listeners-as-outsiders* hypothesis).

In looking at performers and audiences together we are interested in gathering characterisations as close to the moment of experience as possible and to study experience of the performance through participants' (rather than researchers') perspectives. We are thus attempting to capture a fleeting experience without interfering with the moment of music making or listening too much; so unlike some other studies (e.g., Canonne & Garnier, 2015; Waterman, 1996) we do not ask participants to make judgements or identify moments *during* the initial experience of the music. Our focus is on a particular occurrence of music making rather

than general opinions about collaborative skills (e.g., Cosano Molleja et al., 2017; Dobson & Gaunt, 2015). Our approach is open ended in allowing performers to focus on what they think is worth mentioning rather than asking them to focus on specific predetermined topics (e.g., music-analytic features, performers' intentions, emotional responses, aesthetic success, or details of their own or other's playing) or prioritizing particular kinds of characterisations (perceptions, thoughts, feelings, judgments, interpretations, etc.).

Research questions

This case study asks the following research questions about classical chamber duo performers:

(1) Are performers more likely to overlap with each other in their characterisations of a just-completed performance than with non-performing audience members?

(2) Are musicians who have experience playing a piece (have studied, rehearsed or performed it) more likely to overlap with each other in their characterisations of a just-completed performance than with musicians who have no experience with the piece?

(3) Are musicians who play the same kind of instrument (piano vs. strings) more likely to overlap with each other in their characterisations of a just-completed performance than those who play different kinds of instruments?

For each question, we make the comparison by asking whether the musicians being compared

(a) select more of the same moments to comment about
(b) comment more on the same topics
(c) endorse one set of musicians' comments more than others.

To address these research questions, we asked a cello-piano duo to perform once for the other 12 members of their conservatory studio class a chamber piece they had been working on. The class included another duo who were prepared to play the same piece that day and others with a range of experience with the piece. Immediately afterwards, the listeners and players individually wrote the top three things that had struck them about the performance. Then, listening to a recording of this performance on personal devices, they marked on scores and wrote about three moments that struck them as most worthy of comment. After all the participants had finished, the response sheets were redistributed for each participant to rate their level of endorsement of another participant's characterisations, and then of one other participant's characterisations. This allowed us to collect in short order a large number of reactions to a performance and the extent to which those reactions were shared collectively.[1]

This method tests the generality of the findings from the jazz case studies in a classical context. It is not a foregone conclusion that what we see in the studies on jazz will extend to a classical context with its score-based performance practice (Seddon & Biasutti, 2009), and in a pair that has rehearsed together extensively.

Method

Participants

The 14 participants were all members of a May 2015 conservatory class called "Duos for Piano & Strings" at Mannes School of Music, College of Performing Arts, The New School. This is a performance-based course in which advanced chamber performers present, workshop, and critique classical chamber duo repertoire, with a focus on "the development of personal musicianship, the discussion of interpretive issues and strategies, and the improvement of technical expertise related to duo repertoire." All participants had spent notable time with each other over the course of the previous 10 weeks (as well as in other settings) and they therefore had extensive experience discussing and critiquing each other's chamber performances. The only inclusion criterion was being a member of this class.

All participants were experienced classical piano and string players (see Table 1 for details). They were all older than 18 years of age, and ranged in age up to their late 20's (see Table 2 for additional self-reported demographic characteristics). In terms of familiarity with the piece performed in this study, 11 of 14 had heard it, 9 had played it before, and 4 were prepared to play it that day.

Procedure

The piece. The piece was selected in consultation with the instructor, based on likelihood that existing duos in the class could be willing to rehearse and be relatively quickly prepared to perform in the class. The piece, Schumann's Phantasiestücke, Op. 73, no. 1 (1849), was originally written for clarinet and piano and subsequently arranged for cello or violin and piano; it has since been arranged for several more instruments (including double bass, horn, bassoon, all with piano accompaniment). As a frequently performed piece that has been called "a veritable chestnut in Schumann's chamber music repertory" (Reissenberger & Hoeprich, 2014, p. 449), it was likely to be known by many of the participants in our study, whether they played it themselves or had heard it performed. The piece is marked "Zart und mit Ausdruck" (tender and with expression), and it is in a ternary form (A (bars 1-21), B (bars 22- 37), A' (bars 37 – 59), Coda (bars 59-69)) that includes exact and varied repetition of musical material.

Before the day of performance. The instructor consulted with the class members to gain consent for their participation in the study, with clarity that participation was optional and could be

Table 1. Self-reported musical experience.

Domain	Number of participants
Primary instrument	
Piano	8
String	6
	(3 cello, 1 double bass and bass guitar, 1 violin, 1 viola)
Piano and string experience	
Piano (no string)	3
String (no piano)	3
Piano and string instrument	8
Performance genres	
Classical only	10
Classical plus jazz	1
Classical plus rock/pop	2
Classical plus rock/pop, folk, and funk	1
Listening genres	
Classical only	6
Classical plus other genres (jazz, folk, rock/pop, funk, "anything but country")	8
Years of formal training in music theory	
2-5 years	2
5-10 years	3
10 or more years	9
Years of formal training on a musical instrument (including voice)	
5-10 years	1
10 or more years	13

Table 2. Self-reported demographic characteristics.

	Number of participants
Gender	
Female	6
Male	8
Race/ethnicity	
Asian	1
Asian/Pacific Islander	5
Asian/Pacific Islander & White	1
Black	1
Israeli	1
White	4
Would rather not say	1
Comprehension of English language (1 = Poor, 5 = Excellent)	
3/3.5	5
4	3
5	6
Ease of responding	
Very easy	4
Somewhat easy	5
Neither easy nor hard	5

suspended without penalty. Two piano-cello duos were asked to prepare the piece for performance on the day of this study. They were told that only one pair would be asked to perform on the day. These were thus four participants who had a different relationship with the piece (level of most immediately recent familiarity and preparedness to perform with each other) than the rest of the participants.

Participants were instructed to bring their own devices on which they could listen to audio and on which they could receive emails, as well as headphones to allow private listening. The instructor collected email addresses from all participants that they could use on their own devices on the day of performance.

Day of performance. All data collection took place on one day in the room in which students usually had their class. The classroom was set up after a sound check by recording engineer Amandine Pras to determine the best placement for performers and microphones for ensuring acoustic homogeneity between the two instruments, with the goal of achieving "natural" representation of their balance and dynamics. A Royer SF12 stereo ribbon microphone, using the Blumlein stereo system of two coincident bi-directional microphones with an angle of 90 degrees, was used with a portable Tascam dr100 stereo recorder. The instructor introduced the two experimenters, who passed around and collected paper consent forms. A coin was then tossed to decide which of the prepared duos would play.

Performance. The players were instructed to play the piece the whole way through without stopping, and to treat this as a normal performance that they would be doing in this class. The other class members were instructed to listen and pay attention to the performance. The performance then commenced. Immediately after the performance (while participants filled out the first questionnaire), the recording engineer saved the audio recording as an MP3 file in a cloud server, with no postproduction editing, mixing or mastering so as to avoid potentially affecting interpretations of the performance. We then sent an email link to the recorded file to all participants at the email addresses they had provided earlier.

Characterizing general aspects of the performance. Immediately after the performance, the experimenters first handed each participant a pre-assigned private code number (on paper) that would be known only to them and the researchers, for keeping track of responses and being able to assign answer sheets appropriately in the next parts of the study. Researchers then handed out paper copies of the first questionnaire to all participants, including the players, and asked them to write their private code on it. This questionnaire (see Figure 1) instructed participants:

> Taking just a few minutes, please list the top three things that struck you about the *performance* while you were listening or playing. This could be about expressive features of the performance, synchronization, particular moments that most worked or didn't work, overall characterizations of the playing and ensemble work—whatever most struck you. What you write can be about specific moments in the piece or more general statements about the performance. We're NOT looking for

Figure 1. Sample of completed questionnaire 1: General characterisations of the performance.

music-analytic descriptions of the piece itself, nor generalizations about the performers; this is focused on the *performance.*

Participants were instructed that the group would continue with the next part when everyone was done, and they were asked not to discuss the performance or even chat with each other during this part.

Characterizing particular moments in the performance. When all participants had finished Questionnaire 1, the experimenters collected it. They then asked all participants to leave the room with their device and go to the hallway where the wireless connection worked. From the email they had just received, participants were to download the audio file to their device, and then return to the classroom.

Each participant was provided with a copy of the musical score for the Phantasiestücke, each copy marked with a letter. Participants were asked to enter their private code on the score. They were then instructed to listen on their own device to the audio recording they had just downloaded. (They were

told that they were welcome to follow along in the score while listening or not as they preferred). After listening, they were provided with Questionnaire 2 (see Figure 2 for an example), which instructed them to mark on the score the three moments (which could be individual notes, phrases, or entire sections, as they preferred) that struck them as most worthy of comment—positive or negative—and then to enter their three comments about those moments on Questionnaire 2. These comments could be about the players' collaboration, the artistic success or non-success of particular moments, or even about technical aspects if those had struck them while listening. They were also told that what they marked here did not have to correspond at all with the general characterisations they had made on Questionnaire 1.

Participants were further instructed that we needed their marking of the score to be very specific about exactly which bars or notes they were commenting on—if possible, even the minutes and seconds in the recording. We also asked them to be as clear as possible about which parts they were referring to: piano, cello, piano left hand,

Questionnaire 2   Your code___87___

Listen on your own device to the audio recording you just downloaded. (You can follow along in the score while listening or not as you prefer).

Taking just a few minutes, your task is to mark ON THE SCORE the *three moments* (could be individual notes, phrases, or entire sections—up to you) that strike you as most worthy of comment—positive or negative—and then write your comment on the sheet below. Please make sure to mark your chosen moments "1," "2" and "3" on the score. Your comments can be about the players' collaboration, the artistic success or non-success of particular moments, or even about technical aspects if those really struck you while you were listening. If the moments that were particularly worthy of comment were for another reason, that's fine—we want to know what you noticed. What you comment on here does not have to correspond at all with the overall/general comments you made beforehand.

We need your marking of the score to be REALLY specific—about exactly which bars or notes you are commenting on. If possible, list the minutes and seconds in the recording you are talking about. We also need your marking to be as clear as possible about which part you mean: piano, cello, piano left hand, piano right hand, or ensemble.

Feel free to listen more than once, or to relisten to part of the recording, as needed in order to make your best judgment.

We'll move on to the next part once everyone has finished.

Moment 1.
The dynamic change from pp to f should be bigger

Moment 2.
I like how this f wasn't too aggressive, but more lyrical, it was given a buoyancy that made it soar.

Moment 3.
Need to linger more on this downbeat to make the surprise harmony more apparent

Figure 2. Sample of completed questionnaire 2: Characterisations of particular moments in the performance.

piano right hand, or the ensemble. Participants were told that they were welcome to listen more than once, or to relisten to part of the recording, as needed in order to make their best characterisation. Participants were told that they could take as much time as they needed and that we would move on to the next part once everyone had finished.

While participants filled out Questionnaire 2, the experimenters checked the comments in Questionnaire 1 to determine whether any revealed the identity of the commenter. The comments by one participant (one of the performers) were recopied onto a new sheet anonymizing the text so that the performer's identity would not be revealed in the next steps. For example, "there were some places the two of

us were not together" was rewritten as "there were some places the performers were not together."

The scores and Questionnaire 2 were collected individually as they were completed, checked in case any needed to be anonymized (none did), and stapled together with the corresponding Questionnaire 1 filled out by the same individual (based on the private code).

Endorsing another participant's characterisations. The experimenters distributed Questionnaire 3 along with a stapled packet (Questionnaires 1 and 2 and the marked score) following a partially pre-specified scheme based on the private codes; the participants could not tell whose packet they were receiving. The two performers received each other's packets, and the two participants who had been prepared to play that day but didn't received each other's packets. The rest were distributed arbitrarily, making sure that no one would receive their own comments and score.

Participants were instructed to rate on Questionnaire 3 the extent to which they endorsed the characterisations in the packet. For each general statement and for each moment, they were to rate their level of agreement on a 5-point scale, from "strongly disagree" to "strongly agree." They could also select "don't understand" if they didn't understand a comment or marking. They were welcome to elaborate on why they disagreed or had a different interpretation in a "comments" section (see Figure 3 for an example). They were welcome to listen again to the recording in order to make their determination. Participants were told that they could take as much time as they needed, and that we would move on to the next part once everyone had finished.

Endorsing a different participant's characterisations. After the experimenters had collected the packets that participants had thus far rated in Questionnaire 3, they now distributed the packets to different listeners. The scheme was that the performers now rated characterisations by the prepared-but-not-performing player of their performing partner's instrument, and that the prepared players now rated characterisations by the performing player of their partner's instrument. All other participants again rated another arbitrarily assigned participant's characterisations, making sure that they did not receive the same packet they had rated before or their own.

Reporting previous experience and demographics. Participants then filled out a final questionnaire that asked them about their familiarity with the piece (had they played it before? Heard it before?), their instrumental experience, and their demographic characteristics. Once all questionnaires had been collected, the experimenter spent the remaining class time in a debriefing discussion with the course instructor and class members about the research project, its context, and the next steps in data analysis.

Data coding and analysis strategy

Commenting on the same moments. To examine the extent to which participants commented on the same moments, we plotted the start and end bars and beats of each moment notated on the score by each participant.

Content overlap. To assess the extent of potential content overlap among different participants' comments our strategy was to err on the side of finding potential overlap rather than focussing on differences. Both authors together grouped comments that included the same or semantically related words. For example we grouped "lazy," "slow," "too slow," "could be faster," and "too settled" in one cluster (Cluster 1, Appendix 1). We then further differentiated clusters that broke into clear subgroups based on the fuller context of the comment. For example, we broke a possible cluster about the performers not being together into two, one about the ensemble as a whole (Cluster 8) and the other about being together at particular moments (Cluster 2).

Topic overlap. To examine topic overlap, we first identified a set of broad topics that we saw the full set of comments as encompassing, both evaluatively positive and negative. The first author first identified a set of topics that seemed to represent the range of commentary and then annotated each comment according to the one or more categories from this set that the comment fitted into. Both authors then checked each annotation and for the low-frequency categories merged those that reasonably grouped together. The final set of topics was: dynamics, balance, tempo, timing/rhythmic motion, synchronization, communication, expressivity, tone quality, and other kinds of comments. Table 3 gives examples of comments that we saw as overlapping on the topic of tempo; see Appendix 2 for the full set of comments with all annotations and Supplemental Table 1 for the details of the coding scheme and procedure.

Based on the codes assigned to each comment, we counted topic overlap as the number of comments (out of the possible total of 6—three general comments and three comments about particular moments in the performance) that overlapped at all in any topic, for each possible pairing of the 14 participants. To do this, we focused on the content of the three comments about moments independent of which beats the moments were about, taking the generous assumption that two statements about tempo or synchronization in some sense overlap in topic even if they are about different moments. We counted two comments within the same topic as overlapping even if the wording was quite different, but we did not count comments as overlapping if they were coded in the same category but with opposite valence (e.g., one comment praised the balance and the

Figure 3. Sample of completed questionnaire 3: Endorsement another participant's characterisations.

other criticized it), and we only counted "other" comments as overlapping if they did indeed overlap in content. For the few comments that we thought might or might not belong within a particular topic category (those assigned question marks in Appendix 2), we counted them as belonging to the category for this analysis. For each pairing, we counted the largest number of comments that could be seen as overlapping; so if 4 comments by one party overlapped in content with one comment by the other, that pairing would be assigned a count of 4 overlaps.

Table 3. Comments identified as including the topic of tempo. Check marks represent the topics assigned; question marks represent a possible classification. Comments judged as evaluatively positive are represented in green, and as evaluatively negative in red. Each comment can address more than one topic.

General comments.

Comment author	Comment	Dynamics	Balance	Tempo	Timing/ Rhythmic motion	Synchronization	Communication	Expressivity	Tone quality	Other
A Performing cellist	The rhythmic pulse was a little bit lazy. I think it could have more flow.			?	✓					
B Performing pianist	Tempo began slow but turned out to be perfect.			✓						
C Prepared cellist	I think that the two performers decided to take a too slow tempo.			✓						
E String listener, played it	But I think the tempo could be faster or more flowing.			✓						
I Pianist listener, played it	First, this performance is well gone through, but regardless of how good it is, it was a bit too settled and less emotional than my interpretation.			?				✓		✓

Comments about specific moments.

Comment author	Comment	Dynamics	Balance	Tempo	Timing/ Rhythmic motion	Synchronization	Communication	Expressivity	Tone quality	Other
B Performing pianist	The tempo felt slower and dragged more than the beginning.			✓						
C Prepared cellist	In this section (middle section) there is something that doesn't match. The duples against the triples are not exactly together. In addition to that the cello line is too "flat"; I think it should be with more variations of sound and tempo.			✓		✓		✓	✓	
E String listener, played it	I think the cellist could play that section a little bit hurriedly, just like the cello wants to say something with the pianist at the same time.			✓						
I Pianist listener, played it	Part of last pages sounds dragging because there are too many *rubatos*. Might be a better idea to wait, not slowing down too soon, until next section.			✓	✓					

Results

Research Question 1: Did the performers overlap with each other in their characterisations of their just-completed performance more than with non-performing audience members?

We address this research question in the following ways.

Commenting on the same moments. Were the performers more similar to each other than listeners in terms of the moments they chose to comment on? Figure 4 plots the time spans for the three moments commented on by each participant.[2] As is visible in Figure 4, the two performers did not overlap at all in the moments they chose to comment on, even under the most generous interpretation of overlapping (any beats of the moments chosen overlap, even if much else doesn't). Each of the three moments that the performers chose was, however, also chosen by at least one listener, under this generous interpretation. All listeners overlapped with another listener at least once in the moments that they chose, averaging 2.42 (of 3 possible) overlapping moments. Based on moments chosen, the performers were not more similar to each other in their judgments than the listeners.

Content overlap. To what extent did performers' comments overlap in content more than with their audience members' comments? Appendix 1 groups all the statements of the 84 that we see as potentially (at least partially) overlapping in content. From this clustering, the performers overlapped in content for two statements, one about a (temporarily) too-slow tempo and another about not being together, although in both cases these concerns were raised by at least one other audience member as well. We also see (at least partial) overlap in the content of 2 or 3 other statements by the performers with other audience members. Based on this view of the content, the performers' comments do not overlap in content more than with other participants'.

Topic overlap. Was there greater overlap in the themes that performers' comments addressed, even if the precise content of the comments differed? Using the method described earlier, we see that the performers overlapped in topic in some way for 5 of their 6 comments. Both performers overlapped in topic more (all 6 comments) with at least one participant other than their co-performer, and they overlapped in topic with at least two other participants as much (5 comments) as with their co-performer. As Appendix 2 shows, the pattern is the same for the general comments and for the comments about particular moments: the performers' comments overlap in topic at least as much with more than one other participant's comments as they do with their performing partner's comments. Of course, 6 characterisations by each participant surely do not exhaust everything that they thought about the performance, but the comments do give some indication of what they thought was most important to articulate. At least by this coding scheme and counting method, we see no evidence that the performers' comments overlap in topic more with each other than with their audience members.

Endorsement of the partner's comments. Did the performing partners endorse their partner's comments more than they endorsed comments by others? Of course, this data set only includes performers' ratings of two others' characterisations—their performing partner's and one prepared non-performer's—so we see this evidence as suggestive rather than definitive. Nevertheless, both performers endorsed characterisations by the person who was not their partner at least as much as their partner's characterisations. As Figure 5 shows, the pianist endorsed the prepared cellist's comments (giving a rating of "agree" or "strongly agree") more (5 of 6 comments, 83%) than their partner's (2 of 6 comments, 33%), and the cellist endorsed no more of their partner's comments (3 of 6, 50%) than the prepared pianist's comments (3 of 6, 50%). The pattern is the same if we adjust the ratings based on the raters' (optional) write-in elaborations about their ratings, which in some cases qualified the rating. For example, the pianist gave an "agree" rating to their cellist partner's comment "The rhythmic pulse was a little bit lazy. I think it could have more flow," but the pianist's comment along with the rating ("I think it was on the slow side, but I wouldn't use the term 'lazy.' I prefer 'relaxed.'") suggests less than complete endorsement.

In sum, we see no evidence that the performers overlap with each other in their characterisations of their just-completed performance more than with non-performing audience members. They did not select more of the same moments to comment about, address more similar topics, or endorse their partner's characterisations more.

Research Question 2: Did musicians with experience playing the piece overlap with each other in their characterisations of the performance more than with musicians who did not?

We address this research question by comparing overlap in characterisations between those who had played the piece before (4 who were prepared to play the piece the day of the study and 5 who had played the piece before but had not prepared to play the piece that day) and those who had never played the piece before (5 participants).

Commenting on the same moments. Did those with more experience with the piece choose more of the same moments to comment on? Every one of the 9 participants who had played the piece (the first nine rows of Figure 4) chose at least one of the same moments, using the generous interpretation of overlapping as including any overlapping beats of the moments chosen. Of the three moments chosen, these nine participants averaged 2.1 moments that
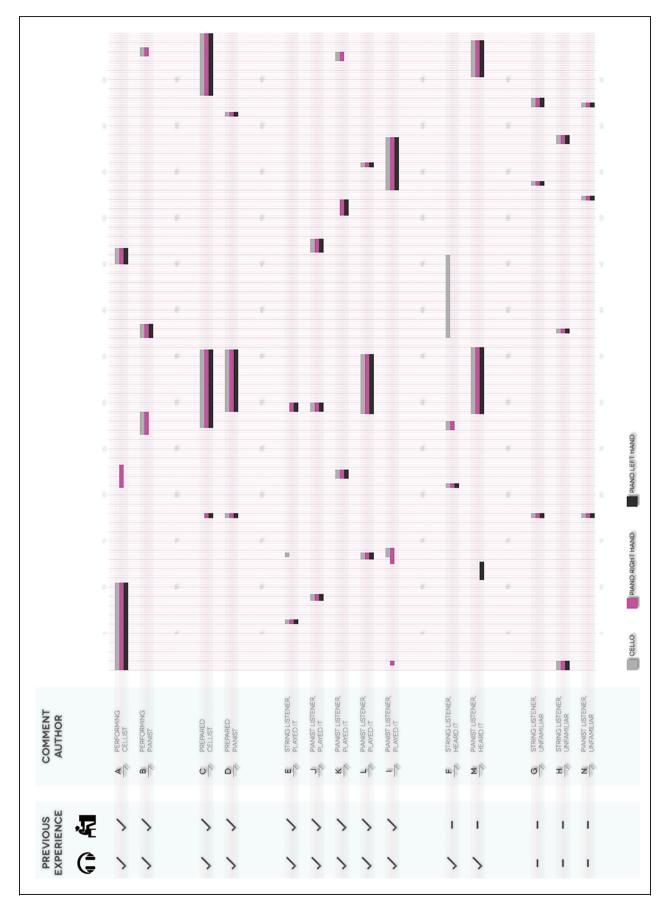
Figure 4. Locations in piece of moments selected for commentary. Vertical lines represent bars (thicker lines) and beats (thinner lines).
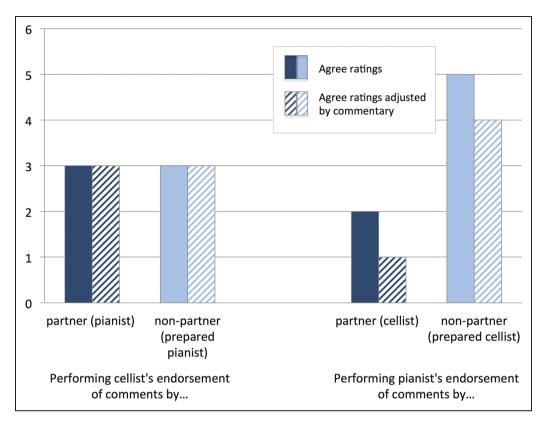
11

Figure 5. Performers' endorsement of comments by their partner and prepared non-partner.

overlapped with those chosen by another participant who had played the piece. Those who were prepared to play the piece that day chose an average of 2.5 overlapping moments, and those who had played the piece before averaged 1.8 overlapping moments. While this is consistent with the possibility that greater experience with the piece leads to greater overlap in chosen moments, it is not notably different than the 2.0 moments that participants who had not played the piece overlapped in moments chosen with those who had played the piece. It is also not notably different than the 1.9 moments chosen by participants who had played the piece that overlapped with participants who had never played the piece, nor the 2.5 overlapping moments that those prepared to play the piece that day overlapped with participants who had never played the piece. Based on moments chosen, we do not see strong evidence that experience playing the piece led to greater overlap.

Content overlap. To what extent did comments by participants with greater experience with the piece overlap in content any more with each other than with comments by those with less experience? Of the 17 potential content clusters in Appendix 1, only 2 of them include only comments by participants who were prepared to play that day. An additional 6 include only comments by participants who had played the piece before. The remaining 9 clusters also include comments by participants who had never played

the piece before. While one might argue that it is striking that 16 of the 17 content clusters include a comment made by someone prepared to play that day (and thus perhaps these participants' comments reflect particularly notable aspects of the music-making), it is also striking that most of the clusters also include a comment made by someone who was not. So this pattern does not suggest overwhelming evidence supporting the idea that more experienced players will produce more comments in common.

Topic overlap. Did comments by participants who had played the piece before overlap more in the themes they addressed, as coded in Appendix 2? Using the same counting method described earlier, we see that the four participants who were prepared to play that day overlapped in topic for an average of 4.67 (of 6 possible) comments, ranging from 3 to 6. This seems notably higher than the average of 3.0 (of 6 possible) comments (ranging from 1 to 6) that overlapped in topic for participants who had played the piece before but had not prepared to perform the piece that day, but not notably higher than the average of 4.3 comments (ranging from 2 to 6) that overlapped in topic among participants who had never played the piece. The average overlap in topic for comments by participants who had ever played the piece was 3.63 (ranging from 1 to 6). This pattern of findings does not suggest that comments by participants with greater experience with the piece overlap
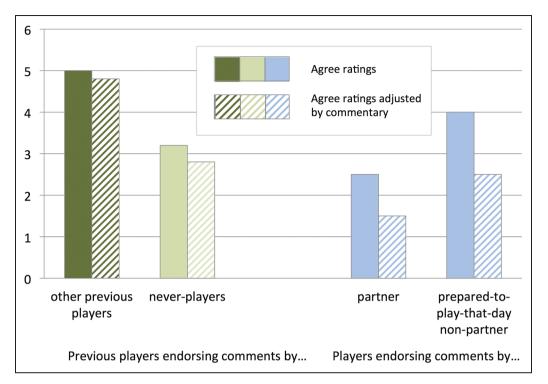
Figure 6. Endorsement of comments by participants who had played the piece before (but were not prepared to play that day) of comments made by others who had played the piece before and comments by participants who had never played the piece. *For comparison, the right half of the figure re-plots average endorsement of comments by players from Figure 5.*

in topic more than comments by participants without experience with the piece.

Endorsing other participants' comments. Did participants with greater experience with the piece endorse comments by other participants with greater experience any more than they endorsed comments by others? Although our assignment of raters who were not asked to prepare to play that day to rating sheets was arbitrary, fortunately, the five participants who had played the piece before each rated their agreement with one other participant who had played the piece before and one other who had never played the piece. These participants endorsed (giving a 4 or 5 on the 5-point scale) an average of 5.0 of the 6 statements authored by others who had played the piece before (83%), and only 3.2 of the 6 statements authored by others who had never played the piece before (53%). If we augment this subset of participants with ratings by the two non-performing participants (who were prepared to play the piece that day) of comments by their partner (who was also prepared to play that day), the average level of endorsement is still higher, 4.2 of 6 (70%), than endorsement of statements authored by others who had never played the piece before.

As Figure 6 shows, this high endorsement by participants who had played the piece before (but were not prepared to play that day) of comments made by others who had played the piece before is strikingly higher than the endorsement level of the participants who were prepared to play that day, which (averaging from Figure 1) was 3.6 of 6

(60%)—an average of 2.5 of 6 (42%) for players' endorsement of their performing partner's comments, and 4 of 6 (67%) for players' endorsement of comments by another who had been prepared to play that day. The overall pattern of findings is consistent with the proposal that players more experienced with the piece are more likely to endorse comments by others who have also played the piece than comments by those who have never played the piece.

Research Question 3: Did musicians who play the same kind of instrument (piano vs. strings) overlap in their characterisations of the performance more than those who play different kinds of instruments?

We address this question by comparing overlap in characterisations between the 8 participants whose primary instrument was piano with the 6 participants whose primary instrument was a string instrument.

Commenting on the same moments. Did those who played the same kind of instrument choose more of the same moments to comment on? Counting all participants (including the performers), and using the generous interpretation of overlapping as including any overlapping beats of the moments chosen, pianists averaged 2.0 moments that overlapped with those chosen by another pianist, and string players averaged 1.7 moments that overlapped with those chosen by another string player. In both cases, they did not overlap
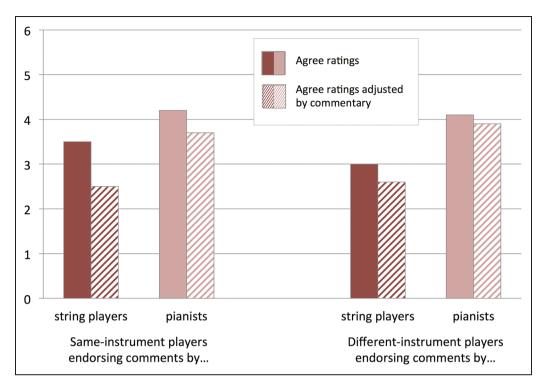
Figure 7. Endorsement of comments by players of same vs. different kinds of instruments as the comment authors (incorporates data from one participant who did not rate the score-based comments for either author by doubling their score to scale it as out of 6 rather than 3).

more with players of the same kind of instrument than with players of the other kind of instrument (string players averaged 2.2 moments that overlapped with pianists' chosen moments, and pianists averaged 2.0 moments that overlapped with string players' chosen moments). This pattern of no greater overlap in moments chosen between players of the same kind of instrument than between players of different instruments is the same even if we exclude the performers from the count (in case performers' unique perspective might skew the results)—same-instrument-kind commenters overlapping on an average of 1.85 moments vs. different-instrument-kind commenters overlapping on 2.0 moments. It is also the same if we also exclude the comments by the two non-performing participants who were prepared to play the piece that day (1.83 vs. 2.0).

This evidence is not consistent with a hypothesis that experience playing the same instrument leads to greater overlap in the moments participants choose to comment on.

Content overlap. Did comments by participants who played the same kind of instrument overlap in content any more with each other than with comments by those who played the other kind? Of the 17 potential content clusters in Appendix 1, 6 of them include only comments by participants who played the same kind of instrument (5 by pianists and 1 by string players). Another 6 include only comments by participants who played different kinds of instruments, and another 5 include comments both by

participants who played same and different kinds of instruments. To look at the data a different way, 11 of the 17 clusters included comments by participants who played different instruments, and 11 included comments by participants who played the same instruments. This pattern does not support the hypothesis that players of the same kind of instrument will produce comments that overlap in content more.

Topic overlap. Did comments by participants who played the same kind of instrument overlap more in the themes they addressed, as coded in Appendix 2? Using the same counting method described earlier, we see that the string players overlapped in general topic for an average of 3.7 (of 6 possible) comments, and the pianists overlapped in topic for an average of 4.0 (of 6 possible) comments. This is higher than the average of 3.2 comments that overlapped in topic for pianists with string players, but not higher than the 4.4 comments that overlapped in topic for string players with pianists. At a more general level, participants who played the same kinds of instruments averaged 3.9 comments overlapping in topic, which is not more than the 4.0 comments overlapping in topic among participants who played different kinds of instruments. This pattern of findings is not consistent with a hypothesis that comments by participants who play the same kinds of instruments overlap in topic more than comments by participants who play different kinds of instruments.

Endorsing other participants' comments. Did participants who played the same kinds of instruments endorse each other's comments more than they endorsed comments by participants who played different kinds of instruments? Given our arbitrary assignment of raters to rating sheets, there was not an even split of raters in each instrument pairing (6 pianist ratings of pianists' comments, 10 pianist ratings of string players' comments, 2 string player ratings of string players' comments, and 10 string player ratings of pianists' comments), and so we interpret these results with caution. The evidence is if anything only weakly consistent with the hypothesis: participants endorsed (giving a 4 or 5 on the 5-point scale) an average of 4.0 of the 6 statements authored by others who played the same kind of instrument (67%), and 3.6 of the 6 statements authored by others who played a different kind of instrument (59%); the difference essentially disappears if we look at the qualified ratings, averaging 3.4 (56%) vs. 3.3 (54%). As Figure 7 shows, the pattern looks highly similar for same- vs. different-instrument ratings, even if ratings of comments by pianists seem to be endorsed at a slightly higher rate.

Exploratory analyses of collective response to performance

Beyond our research questions, the nature of this data set allows a new way of looking at the perspectives of multiple listeners at the same time through different lenses. Supplementary Tables 2–5 list all 84 comments grouped by which were fully endorsed by both raters, less than fully endorsed by both, rejected by both raters and the few that were only rated by one participant. Each comment includes any additional elaboration, which in some cases further explain the rater's level of endorsement but in other cases qualify their rating.

These tables show the ways in which raters disagreed with each other on virtually every topic covered by the comments. At least anecdotally, some of the perspectives in the comments seem predictable based on the commenters' and raters' instruments and prior experience; for example, two pianist listener raters disagreed with the string listener who thought that the pianist's playing was too loud relative to the cello, in one case thinking the opposite and in the other thinking the balance was just right.

Seen collectively, raters' elaborations also highlight some divides in the group about general principles of chamber music interpretation, or at least contrasting ideas about how this particular piece should be performed. Clearly some participants believe that recapitulations should be played differently than expositions, in principle, while others believe they should be played identically (see L Pianist Listener's and K Pianist Listener's opposite reactions to H String Listener's comment in Supplemental Table 3: "In my opinion, one should not play the recap the same way, so I agree" vs. "If it's called 'recap' why would you want to hear something different?"). Some participants believed

that when two parts repeatedly have different rhythmic notations against each other (in this piece, cello duples against piano triplets), they should be played as matching, at least in Schumann's chamber works, but others believed that the contrast should be highlighted (for example, F String Listener's elaboration on disagreeing with I Pianist listener's comment in Supplemental Table 3). And there were clear differences in judgment of the appropriateness of and success of *rubatos* and *rallentandos*.

See https://3milychu.github.io/notesonperspective_jour nal/ (Chu, 2020) presenting the subset of score-based comments about specific moments in the piece and Supplemental Video 1 for a non-interactive version of the visualization. This kind of visualization, linked to an audio-recording of a performance and a score, can allow a user to select alternate views of a dataset as it unfolds over time— e.g., viewing how comments that were agreed upon vs. disagreed upon cluster, or focusing only on comments by string players or those who had played the piece before or comments about rhythm.

Exploration of this visualization allows simultaneous experience of multiple aspects of the data—the timing of comments, their authorship, their content, and their endorsement by different parties—whose confluence is hard to see at the same time using static representations like Supplemental Tables 2–5 (see Pras, Spiro, et al., 2017, for a different approach to a time-based visualization). Figure 4 already showed which moments in the performance attracted more attention than others. The visualization allows exploration of additional aspects (within the restrictions of the data set, with each participant having marked three spots on the score and provided endorsement ratings of two others' comments): which kinds of attention by which kinds of authors at which points.

For an example of the kinds of insights that this sort of visualization can afford, consider Supplemental Video 2, which focuses on the moment the transition begins at 1:24. The visualization concretely shows that this moment generated sudden substantial attention with six comments (see Figure 8 for a still image of this moment). The comments highlight the radical difference in interpretation that the same performance can generate among different audience members. Three comments included positive evaluation (coded gold), four included negative evaluation (coded violet), and one was mixed; four were endorsed by both raters and two by only one. While some commentary praises the pianist's tone quality and expressivity at exactly this moment, other comments critique the pianist's rhythmic flow and the balance in the pianist's left and right hand parts; some comments praised the shifting prominence of cello and piano while others wanted more variation in sound and tempo and critiqued the balance between the two instruments. The prepared cellist and pianist visibly disagree in their interpretations here. Supplemental Video 3 shows the same moment, this time focusing only on comments by participants who have played the piece before,

C Prepared cellist 1 👍 1 👎

'In this section (middle section) there is something that doesn't match. The duples against the triples are not exactly together. In addition to that the cello line is too "flat". I think it should be with more variations of sound and tempo.'

tempo synchronization expressivity
tone quality

M Pianist listener, heard it 2 👍 0 👎

'In bars 29-35, I want greater difference in dynamic level between the three musical parts: the melody in the cello, the countermelody in the right hand and the murmuring triplets in the left hand of the piano.'

dynamics balance

L Pianist listener, played it 2 👍 0 👎

'This whole transition part could be somewhat more moving and flowing in both parts. The pianist could use the triplets to push and rhythmic flow like an ongoing current and the 8th notes in both parts can be played with less emphasis on each note so that it can sound like a big gesture.'

timing/rhythmic motion expressivity

J Pianist listener, played it 2 👍 0 👎

'Piano, beautifully accompanied; overall there are places where piano overplays/sings the cello.'

balance tone quality

D Prepared pianist 1 👍 1 👎

'At this moment the performers successfully passed the thematic material back and forth. The attention was pulled from cello to piano and back to the cello before going towards the final *forte* on page 2.'

communication

E String listener, played it 2 👍 0 👎

'I really like that the pianist changed the character when the music changed to another color.'

expressivity tone quality

Filter ⊙

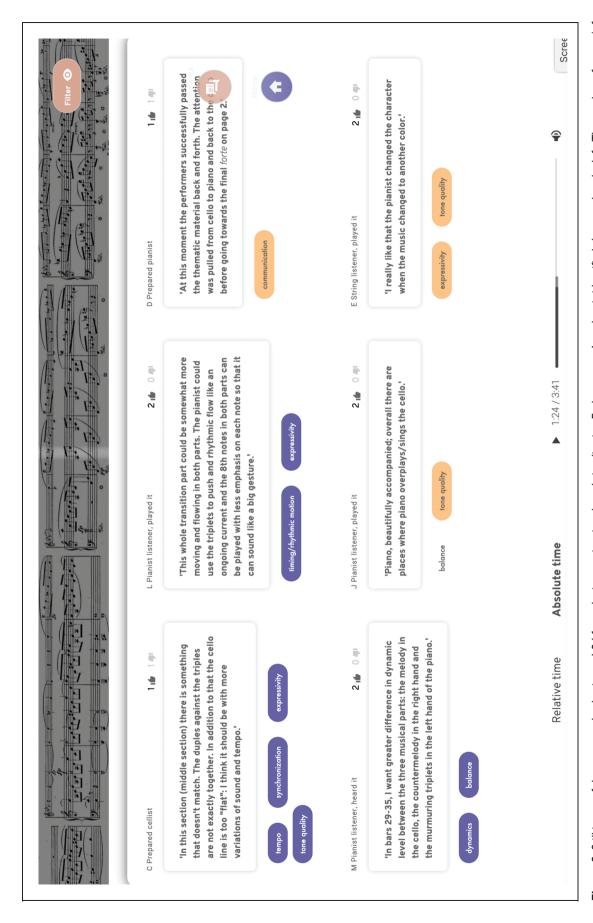Relative time   Absolute time   ▲ 1:24 / 3:41

Scree

Figure 8. Still image of the moment that begins at 1:24 from the interactive time-based visualization Each comment's author is identified above and to the left. The number of raters (of two) who endorsed each comment is given above and to the right. The topics coded appear below each comment in gold (positive comments) or violet (negative comments).

which highlight that these participants found more to critique in this moment than participants who hadn't played the piece before.

Viewing the data set through different filters demonstrates further the range of interpretation across the class, with moments of consensus for different kinds of listeners. We don't want to make too much of the moments of consensus and dissensus that emerge in this relatively small sample of participants, but we believe a visualization tool of this sort could be useful for seeing larger collective patterns in, for example, longitudinal qualitative data sets (Vogl et al., 2018).

Discussion

This case study demonstrates that classical chamber performers can characterise a performance quite differently than their partner does and that they can disagree with most of their partner's characterisations. They can overlap less in their characterisations of a performance than their listeners do, and they can agree with a non-partner's characterisations more than with their partner's characterisations. At the same time, the data show that listeners who have played the piece before—though not necessarily those who play the same kind of instrument as each other (strings vs. piano)—can be more likely to endorse comments by others who have also played the piece before, even if the comments they make don't overlap more in timing, content or theme.

This pattern of findings in a classroom-based classical chamber duo performance echoes those in our prior case studies in other genres, where we saw no evidence that professional musicians had privileged understanding relative to nonperforming listeners in improvising jazz standards (Schober & Spiro, 2014) or free jazz (Pras, Spiro, et al., 2017)—and if anything, potential evidence for greater shared interpretation by nonperforming listeners. The evidence that prior listeners with greater experience with a piece may have greater overlap in their endorsement of each other's characterisations also resonates with the results from Schober and Spiro's (2016) study that included a much larger set of 239 online listeners. In that study, the evidence showed that listeners with genre experience (playing jazz, having experience improvising) were more likely to overlap in patterns of ratings with the performers than those without, which is similar to the current study's finding about experience with this particular piece. In that study, as in the current study, there was also no evidence that listeners who played specific instruments (saxophone or piano) were more likely to endorse statements by the player of their own instrument than the player of the other instrument.[3]

We see this case study of a classical chamber duo's agreement with each other's characterisations as showing what *can* happen rather than what *always* happens. Patterns of overlapping characterisations among performers relative to listeners might look quite different for performers and listeners with different levels of prior experience with a piece, genre, instrument, or each other. Nonetheless, we find it striking that we observe a similar pattern of performers agreeing with each other's comments less than with non-performing listeners in another genre, in a performance with quite different constraints. Unlike in the other case studies, the players here knew each other, had played together before and had experience discussing and critiquing each other's chamber performances; they were performing in a classroom setting at a high-level performance training institution, with an audience of 12 classmates, rather than having been recruited as post-training professional musicians and playing in a concert hall with only the researchers as the audience. Also unlike in the previous case studies, which involved intensive individual post-performance think-aloud interviews and a long delay before participants carried out ratings of a large set of anonymized comments, here the participants provided a smaller set of written comments and marked the score (which was of course not possible in the improvised genres), and ratings of others' comments were immediate. So even with this more efficient data collection method involving more participants but fewer data points per person, the pattern of evidence is consistent.

The method developed for this study represents a new attempt at tapping into the unfolding collective cognition of musical performers and their audience. While in one sense the data set is far sparser in content than the richer commentary we elicited in our prior studies of this kind (Pras, Schober, et al., 2017; Schober & Spiro, 2014), the method efficiently elicits specific reactions that more general-purpose jury-style ratings would not, and then allows quick assessment of collective agreement. As the time-based visualization and the various lenses through which one can see it demonstrate, even this relatively small data set is rich and complex, and its blend of qualitative and quantitative views is new.

As Radbourne et al. (2013) argue (see also Egermann et al., 2019), new methods for gathering new kinds of data on audience experience are needed given the proliferation of contexts in which listeners experience collaborative musical performance—far beyond standard concert halls or home listening to recordings. With digitally streamed video of ensemble performance—whether the performers are live in the same physical space, in different spaces, or having recorded multiple tracks separately that later are merged—that listeners can experience alone on their mobile device or in a theatre with other audience members whose responses may feed into their own, new questions about how audience members' thoughts and feelings overlap with performers' arise. Variants of the method used here—perhaps with all materials online on mobile devices—could allow new time-based accounts of collective experience that are both more fine-grained and larger-scale than have been available before, even with audiences that do not have score-notating experience.

Future studies could test the generality of this case study's findings across different performance settings, audiences with different degrees and types of musical experience, and performers, and to test variations of the method–for example, asking for more characterisations from each participant or asking participants to rate characterisations by more people. Future studies can also ask new research questions that clearly arise from this set of findings. For example, would the pattern of results be different for performers who have played together for many years, presumably having worked out differences in their aesthetic sensibilities, and having had more chances to discuss their performances and reach a shared vocabulary about their ensemble work?[4] Would long-term fans of a particular ensemble be more likely to overlap in characterisations with the performers and with each other? Do individual differences in performers' or audience member's coordination and empathic perspective taking abilities (e.g., Novembre et al., 2019) predict degrees of overlap? Would a larger sample show more nuanced evidence for the *more-experienced-listeners-understand-more-like-performers* hypothesis, and further evidence on the *listeners-as-outsiders* hypothesis?

One additional pedagogical note: although the data collection procedure developed for this study was not intended as a classroom intervention, we found it intriguing that it allowed elicitation of thoughtful commentary from all class members simultaneously and privately in a way that isn't possible in a typical classroom discussion, and it allowed class members to consider and respond to others' characterisations simultaneously and privately. We speculate that creating the opportunity for this kind of anonymized interchange prior to public discussion may embolden students who don't always speak up and reveal thinking that might otherwise never be uncovered for all sorts of reasons—from worries about offending or looking foolish, to discomfort with the pressures of public speaking, to needing more time to formulate an opinion, to fitting the social requirement of responding to prior public comments rather than responding directly and individually to the music. We also speculate that immediate close re-listening to what was just heard provides additional opportunities for greater thoughtfulness that may well be useful in the kind of classroom setting studied here.

Despite the many remaining questions about these findings, the data presented here corroborate what we have seen in our case studies of jazz standard and free jazz improvisation: performers' understanding of what happened is not necessarily privileged relative to outsiders' understanding. Performer and performer-listener differences in interpretation may not be unique to the jazz world, but may extend across different genres of music making.

Action editor

Isabel Martínez, Universidad Nacional de La Plata, Facultad de Bellas Artes.

ORCID iDs

Neta Spiro https://orcid.org/0000-0002-5437-2227
Michael F. Schober https://orcid.org/0000-0003-1531-6326

Notes

1. The protocol was deemed exempt from Institutional Review Board by The New School Human Research Protection Program, #2015-1036, May 2015.
2. The figure plots the time spans for the first score marking, which in most cases was the only one. Four of the 14 participants (neither of the performers) noted that at least one of their comments could apply elsewhere in the score when the same issue repeated; this happened for a total of 7 of the 42 comments about moments, or 8 if one counts a comment whose content suggests repetition even though the score wasn't so marked.
3. This data set did not include judgments by listeners who played instruments other than strings or piano. Therefore, of course,

the current data set does not allow us to test the generality of the Schober and Spiro (2016) finding that those who had played the same instruments (saxophone or piano) as the players were more likely to endorse comments by the performers than listeners who did not play those instruments.

4. As discussed by Schiavio and Høffding (2015), even members of a string quartet who have played together for many years may not notice the same things about each other's physical or emotional states and can be surprised by the direction a well-rehearsed and often-performed piece takes in a particular performance.

Supplemental material

Supplemental material for this article is available online.

References

Bishop, L., & Goebl, W. (2015). When they listen and when they watch: Pianists' use of nonverbal audio and visual cues during duet performance. *Musicae Scientiae*, *19*(1), 84–110. https://doi.org/10.1177/1029864915570355

Canonne, C., & Garnier, N. (2015). Individual decisions and perceived form in collective free improvisation. *Journal of New Music Research*, *44*(2), 145–167. https://doi.org/10.1080/09298215.2015.1061564

Chu, E. (2020). *Collective Understanding Visualisation*. https://3milychu.github.io/notesonperspective_journal/

Clarke, E. F. (2005). *Ways of listening: An ecological approach to the perception of musical meaning*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195151947.001.0001

Cosano Molleja, A., Zych, I., & Ortega Ruiz, R. (2017). Attention, communication and development of interpretation skills in chamber music pianists' education. *Psychology, Society and Education*, *9*(2), 283–292. https://doi.org/10.25115/psye.v9i2.715

Davidson, J. W., & Good, J. M. M. (2002). Social and musical co-ordination between members of a string quartet: An exploratory study. *Psychology of Music*, *30*(2), 186–201. https://doi.org/10.1177/0305735602302005

Dobson, M. C., & Gaunt, H. F. (2015). Musical and social communication in expert orchestral performance. *Psychology of Music*, *43*(1), 24–42. https://doi.org/10.1177/0305735613491998

Egermann, H. W., Toelle, J., & Weining, C. (2019, February). Measuring the audience experience - How can we capture audience experience with quantitative and qualitative methods? *Audience Research in the Arts Conference*. http://www.sparc.dept.shef.ac.uk/audience-research-in-the-arts-conference-3-5-july-2019/

Ginsborg, J., & King, E. (2012). Rehearsal talk: Familiarity and expertise in singer-pianist duos. *Musicae Scientiae*, *16*(2), 148–167. https://doi.org/10.1177/1029864911435733

Hargreaves, D. J., MacDonald, R., & Miell, D. (2005). How do people communicate using music? In D. Miell, R. MacDonald, & D. J. Hargreaves (Eds.), *Musical communication* (pp. 1–25). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198529361.003.0001

Juslin, P. N., Liljeström, S., Laukka, P., Västfjäll, D., & Lundqvist, L.-O. (2011). Emotional reactions to music in a nationally representative sample of Swedish adults: Prevalence and causal influences. *Musicae Scientiae*, *15*(2), 174–207. https://doi.org/10.1177/1029864911401169

Keller, P. E. (2014). Ensemble performance: Interpersonal alignment of musical expression. In D. Fabian, R. Timmers, & E. Schubert (Eds.), *Expressiveness in music performance: Empirical approaches across styles and cultures* (pp. 260–282). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199659647.001.0001

Keller, P. E., Knoblich, G., & Repp, B. H. (2007). Pianists duet better when they play with themselves: On the possible role of action simulation in synchronization. *Consciousness and Cognition*, *16*(1), 102–111. https://doi.org/10.1016/j.concog.2005.12.004

King, E. C. (2006). The roles of student musicians in quartet rehearsals. *Psychology of Music*, *34*(2), 262–282. https://doi.org/10.1177/0305735606061855

Novembre, G., Mitsopoulos, Z., & Keller, P. E. (2019). Empathic perspective taking promotes interpersonal coordination through music. *Scientific Reports*, *9*(1), 12255. https://doi.org/10.1038/s41598-019-48556-9

Pitts, S. E. (2013). Amateurs as audiences: Reciprocal relationships between playing and listening to music. In J. Radbourne, H. Glow, & K. Johanson (Eds.), *The audience experience: A critical analysis of audiences in the performing arts* (pp. 83–93). Intellect Ltd.

Platz, F., & Kopiez, R. (2013). When the first impression counts: Music performers, audience and the evaluation of stage entrance behaviour. *Musicae Scientiae*, *17*(2), 167–197. https://doi.org/10.1177/1029864913486369

Pras, A., Schober, M. F., & Spiro, N. (2017). What about their performance do free jazz improvisers agree upon? A case study. *Frontiers in Psychology*, *8*, 966. https://doi.org/10.3389/fpsyg.2017.00966

Pras, A., Spiro, N., & Schober, M. F. (2017). Audience-based visualization of free jazz improvised performance. In A. Williamon & P. Jónasson (Eds.), *Abstracts of the International Symposium on Performance Science 2017* (p. 169). Iceland Academy of the Arts. http://performancescience.org/wp-content/uploads/2018/08/isps-2017-program.pdf

Radbourne, J., Glow, H., & Johanson, K. (2013). Introduction. In J. Radbourne, H. Glow, & K. Johanson (Eds.), *The audience experience: A critical analysis of audiences in the performing arts* (pp. xvi–xviii). Intellect Ltd.

Ragert, M., Schroeder, T., & Keller, P. E. (2013). Knowing too little or too much: The effects of familiarity with a co-performer's part on interpersonal coordination in musical ensembles. *Frontiers in Psychology*, *4*, 368. https://doi.org/10.3389/fpsyg.2013.00368

Reissenberger, B., & Hoeprich, E. (2014). Deconstructing Robert Schumann's Fantasiestücke, op.73. *Early Music*, *42*(3), 449–459. https://doi.org/10.1093/em/cau068

Schiavio, A., & Høffding, S. (2015). Playing together without communicating? A pre-reflective and enactive account of joint musical performance. *Musicae Scientiae*, *19*(4), 366–388. https://doi.org/10.1177/1029864915593333

Schober, M. F., & Spiro, N. (2014). Jazz improvisers' shared understanding: A case study. *Frontiers in Psychology*, *5*, 808. https://doi.org/10.3389/fpsyg.2014.00808

Schober, M. F., & Spiro, N. (2016). Listeners' and performers' shared understanding of jazz improvisations. *Frontiers in Psychology*, *7*, 1629. https://doi.org/10.3389/fpsyg.2016.01629

Seddon, F., & Biasutti, M. (2009). A comparison of modes of communication between members of a string quartet and a jazz sextet. *Psychology of Music*, *37*(4), 395–415. https://doi.org/10.1177/0305735608100375

Thompson, S., & Williamon, A. (2003). Evaluating evaluation: Musical performance assessment as a research tool. *Music Perception: An Interdisciplinary Journal*, *21*(1), 21–41. http://www.jstor.org/stable/10.1525/mp.2003.21.1.21

Vogl, S., Zartler, U., Schmidt, E. M., & Rieder, I. (2018). Developing an analytical framework for multiple perspective, qualitative longitudinal interviews (MPQLI). *International Journal of Social Research Methodology*, *21*(2), 177–190. https://doi.org/10.1080/13645579.2017.1345149

Waddell, G., Perkins, R., & Williamon, A. (2018). Making an impression: Error location and repertoire features affect performance quality rating processes. *Music Perception: An Interdisciplinary Journal*, *36*(1), 60–76. https://doi.org/10.1525/mp.2018.36.1.60

Waterman, M. (1996). Emotional responses to music: Implicit and explicit effects in listeners and performers. *Psychology of Music*, *24*(1), 53–67. https://doi.org/10.1177/0305735696241006

Wesolowski, B. C. (2016). Assessing jazz big band performance: The development, validation, and application of a facet-factorial rating scale. *Psychology of Music*, *44*(3), 324–339. https://doi.org/10.1177/0305735614567700

Williamon, A., & Davidson, J. W. (2002). Exploring co-performer communication. *Musicae Scientiae*, *VI*(1), 53–72. https://doi.org/10.1177/102986490200600103

Wöllner, C. (2013). Speaking with one voice? String quartet members' perception of each others' performances. In A. Williamon & W. Goebl (Eds.), *Proceedings of the International Symposium on Performance Science* (pp. 211–216). Association Européenne des Conservatoires. http://performancescience.org/wp-content/uploads/2018/08/isps2013_proceedings.pdf

Appendix 1. Comments clustered by potential overlap in content. Terms judged as overlapping are in bold.

Cluster	Comment	Comment author	Heard?	Played?	Instrument kind
	The rhythmic pulse was a little bit **lazy**. I think it could have more flow.	A Performing cellist	yes	yes	string
	Tempo began **slow** but turned out to be perfect.	B Performing pianist	yes	yes	piano
1	I think that the two performers decided to take a **too slow** tempo.	C Prepared cellist	yes	yes	string
	But I think the tempo **could be faster** or more flowing.	E String listener, played it	yes	yes	string
	First, this performance is well gone through, but regardless of how good it is, it was a bit **too settled** and less emotional than my interpretation.	I Pianist listener, played it	yes	yes	piano
	There are some places where we [the performers] were **not together**. But I like the flexibility we [they] had in our [their] performance. I think that kind of ensemble work - responding to each other - should come out nicely to the audience.	A Performing cellist	yes	yes	string
2	Ending cut-off was **not quite together** but it was very natural.	B Performing pianist	yes	yes	piano
	Some of the endings of the phrases were **not together**.	K Pianist listener, played it	yes	yes	piano
	I think that the balance was OK, although **I [the cellist] couldn't really hear the piano** while he was playing.	A Performing cellist	yes	yes	string
	The performers seemed to have **different approaches to the sound of the music;** the piano was generally more open versus a softer cello.	D Prepared pianist	yes	yes	piano
3	**Piano is too loud and covering the cello.**	H String listener, unfamiliar	no	no	string
	Two players' sync was good but somewhat I want to hear **more careful and delicate gesture in piano part** (I would rather say balance between cello and piano than this).	I Pianist listener, played it	yes	yes	piano
	I think that the **balance was OK**, although I [the cellist] couldn't really hear the piano while he was playing.	A Performing cellist	yes	yes	string
4	The ensemble was **good** both in terms of synchronization and **balance**.	M Pianist listener, heard it	yes	no	piano
	The blend of the textures between the cello and piano was very **well balanced**.	N Pianist listener, unfamiliar	no	no	piano
5	The rhythmic pulse was a little bit lazy. I think it **could have more flow.**	A Performing cellist	yes	yes	string
	The overall musical shapes **could be more flowing.**	L Pianist listener, played it	yes	yes	piano
	The **relationship between the cello eighth notes and the melodic right hand of the piano could have been cleaner rhythmically speaking.**	B Performing pianist	yes	yes	piano
6	In this section (middle section) there is something that doesn't match. The **duples against the triples are not exactly together.** In addition to that the cello line is too 'flat': I think it should be with more variations of sound and tempo.	C Prepared cellist	yes	yes	string
	The tempo felt **slower and dragged more** than the beginning.	B Performing pianist	yes	yes	piano
7	Recap of the melodic theme. **Sounds a little bit dull.** I wish I could hear a different statement about it.	H String listener, unfamiliar	no	no	string

Appendix I. (continued)

Cluster	Comment	Comment author	Heard?	Played?	Instrument kind
8	Both [of us] performers had the same idea of a heavier *rubato* on each of the 4 eighth notes, but we **executed them differently**. The first and second notes of the last set of **4 were not together** in comparison to the 3rd and 4th.	B Performing pianist	yes	yes	piano
	Ensemble **not together**.	K Pianist listener, played it	yes	yes	piano
9	Sometimes I think the two performers **don't have the same music idea**.	C Prepared cellist	yes	yes	string
	The performers seemed to have **different approaches** to the sound of the music; the piano was generally more open versus a softer cello.	D Prepared pianist	yes	yes	
10	The feeling of the performance is calm and **intimate**.	C Prepared cellist	yes	yes	string
	I thought the performance was **intimate**.	G String listener, unfamiliar	no	no	string
	Here there is the first real *FORTE* but the piano is missing this dynamic a bit; especially the left hand with the chord **should be louder**.	C Prepared cellist	yes	yes	string
11	At this moment the cello *crescendoed* to a climactic moment while the **piano became softer**. It seems that the sound of the **piano did not successfully support the cello** during one of the most expressive moments. The same occurrence happened later as well (see moment 1a)*.	D Prepared pianist	yes	yes	piano
	*[The sound of the piano did not match the expressive intent of the cello. It seems the cello craved a true arrival while the piano wished to back away to the chord.] That's the first real *f* moment in this piece. But in the recording, **I couldn't really hear big differences**.	G String listener, unfamiliar	no	no	string
12	Both performers had **similar approaches to time** and *rubato* in the performance.	D Prepared pianist	yes	yes	piano
	Could tell they **discussed a lot** about piece (**where to take time** etc.)	J Pianist listener, played it	yes	yes	piano
13	Two players' **sync was good** but somewhat I want to hear more careful and delicate gesture in piano part (I would rather say balance between cello and piano than this).	I Pianist listener, played it	yes	yes	piano
	The ensemble was **good** both in terms of **synchronization** and balance.	M Pianist listener, heard it	yes	no	piano
14	The performers were very **aware of melodic phrases** and **successfully passed them back and forth.**	D Prepared pianist	yes	yes	piano
	Fluid communication between players on phrasings and last minute ideas	F String listener, heard it	yes	no	string
15	They were both **really expressive**.	J Pianist listener, played it	yes	yes	piano
	The performance was **very expressive**.	K Pianist listener, played it	yes	yes	piano
16	There **could be a lot more different characters**.	L Pianist listener, played it	yes	yes	piano
	I felt the interpretation **could have been more imaginative** in terms of *rubato*, colors, textures, and vocal inflection.	M Pianist listener, heard it	yes	no	piano
17	Ending: the cellist is playing beautifully, but **I wish the pianist would have played softer, really *pp*** so it would feel like dying (or falling asleep).	H String listener, unfamiliar	no	no	string
	The dynamic **change from *pp* to *f* should be bigger**.	N Pianist listener, unfamiliar	no	no	piano

Appendix 2. Topics identified as broad content areas across comments. Check marks represent the topics assigned; question marks represent a possible classification. Comments judged as evaluatively positive are represented in green, and as evaluatively negative in red. Each comment can address more than one topic.

General comments.

Comment author	Comment	Dynamics	Balance	Tempo	Timing/ Rhythmic motion	Synchronization	Communication	Expressivity	Tone quality	Other
A Performing cellist	The rhythmic pulse was a little bit lazy. I think it could have more flow.			?	✓					
	I think that the balance was OK, although I [the cellist] couldn't really hear the piano while he was playing.		✓✓							
	There are some places where we [the performers] were not together. But I like the flexibility we [they] had in our [their] performance. I think that kind of ensemble work - responding to each other - should come out nicely to the audience.						✓	✓		
B Performing pianist	Tempo began slow but turned out to be perfect.			✓✓						
	All within pp - mf range	✓								
	Ending cut-off was not quite together but it was very natural.					✓	✓			
	I think that the two performers decided to take a too slow tempo.			✓						
C Prepared cellist	The feeling of the performance is calm and intimate.							✓		
	Sometimes I think the two performers don't have the same music idea.						✓			
	The performers seemed to have different approaches to the sound of the music; the piano was generally more open versus a softer cello.	✓	✓				✓			
D Prepared pianist	Both performers had similar approaches to time and rubato in the performance.				✓	✓				
	The performers were very aware of melodic phrases and successfully passed them back and forth.						✓			
	I like the character of their playing.							✓		
E String listener, played it	But I think the tempo could be faster or more flowing.			✓						
	The cellist's vibrato is too wide and could be slower and have more variety.								✓	

Appendix 2. (continued)

Comment author	Comment	Dynamics	Balance	Tempo	Timing/ Rhythmic motion	Synchronization	Communication	Expressivity	Tone quality	Other
F String listener, heard it	Depth of character transfer between players and phrases, while maintaining individuality						✓			
	Fluid communication between players on phrasings and last minute ideas						✓			
	Wide use of color from the cello which was attentively communicated to the piano and vice versa						✓		✓	
G String listener, unfamiliar	I thought the performance was intimate.							✓		
	I was distracted by the space of the room. It was hard to focus on just their playing.									✓
	I wish the space was darker.									✓
	Both players are very good but I'm not sure I like the piece…									✓✓
H String listener, unfamiliar	I felt tired while listening. Felt like going to sleep.									✓
	Piano is too loud and covering the cello.	✓	✓							
	First, this performance is well gone through, but regardless of how good it is, it was a bit too settled and less emotional than my interpretation.			?				✓		✓
I Pianist listener, played it	The piano's entrance top two notes were a bit striking, but the harmonization and balance at the piano itself were good.	✓							✓	
	Two players' sync was good but somewhat I want to hear more careful and delicate gesture in piano part (I would rather say balance between cello and piano than this).		✓			✓		?		
	Could tell they discussed a lot about piece (where to take time etc.)				✓					
	They were both really expressive.							✓		✓
J Pianist listener, played it	They could communicate and 'feel' each other more, through the piece (related to comment I*). *[Could tell they discussed a lot about piece (where to take time etc.)]						✓			

(continued)

Appendix 2. (continued)

Comment author	Comment	Dynamics	Balance	Tempo	Timing/ Rhythmic motion	Synchronization	Communication	Expressivity	Tone quality	Other
K Pianist listener, played it	The performance was very expressive.							✓		
	Some of the endings of the phrases were not together.					✓				
	The dynamics were excellent.	✓								
L Pianist listener, played it	The overall musical shapes could be more flowing.				?			?		
	There could be a lot more different characters.							✓		
	The ending was pleasingly done with convincing atmosphere.							✓		
	The ensemble was good both in terms of synchronization and balance.		✓			✓				
M Pianist listener, heard it	I felt the interpretation could have been more imaginative in terms of *rubato*, colors, textures, and vocal inflection.				✓			✓	✓	
	Although the tone quality was beautiful in both musicians' playing, I wanted more clearly defined dramatic moments and a point of climax.							✓	✓	
	The blend of the textures between the cello and piano was very well balanced.		✓							
N Pianist listener, unfamiliar	Wistful quality to the performance—sad but not melodramatic.							✓		
	Rubato was enough to be expressive without detracting from the direction of longer phrases.				✓			✓		

Comments about specific moments. Order of comments reflects order in which author listed them, and not necessarily score chronology (e.g., first comment could be about the ending).

Comment author	Comment	Dynamics	Balance	Tempo	Timing/ Rhythmic motion	Synchronization	Communication	Expressivity	Tone quality	Other
A Performing cellist	The opening section (first page) was a little bit stuck. There are a few phrases that didn't have clear direction, and I couldn't really follow the line when I was listening.				✓					
	The triplet in the piano part is too much in tempo, especially at the 2nd line of the 2nd page. I prefer having more flexibility rather than playing like a metronome.				✓					
	I really like the little climax in the measure. Both ensemble working and individual play are satisfying to me when I listened to it, especially the little stretch we [the performers] did in the middle of the bar.	?			✓		✓	?		
	Both [of us] performers had the same idea of a heavier *rubato* on each of the 4 eighth notes, but we executed them differently. The first and second notes of the last set of 4 were not together in comparison to the 3rd and 4th.				✓					
B Performing pianist	The tempo felt slower and dragged more than the beginning.			✓						
	The relationship between the cello eighth notes and the melodic right hand of the piano could have been cleaner rhythmically speaking.					✓				
	In this section (middle section) there is something that doesn't match. The duples against the triples are not exactly together. In addition to that the cello line is too 'flat'. I think it should be with more variations of sound and tempo.			✓		✓		✓	✓	
C Prepared cellist	Here there is the first real *FORTE* but the piano is missing this dynamic a bit; especially the left hand with the chord should be louder.	✓								
	Too much *RALLENTANDO!* They don't cut the sound together in the last note. The cello finishes before the piano.				✓	✓				

26

Comment author	Comment	Dynamics	Balance	Tempo	Timing/Rhythmic motion	Synchronization	Communication	Expressivity	Tone quality	Other
	At this moment the cello crescendoed to a climactic moment while the piano became softer. It seems that the sound of the piano did not successfully support the cello during one of the most expressive moments. The same occurrence happened later as well (see moment 1a)*. *[The sound of the piano did not match the expressive intent of the cello. It seems the cello craved a true arrival while the piano wished to back away to the chord.]	✓	✓					✓	✓	
D Prepared pianist	At this moment the performers successfully passed the thematic material back and forth. The attention was pulled from cello to piano and back to the cello before going towards the final *forte* on page 2.						✓			
	Similar to moment one*, the sound of the piano did not match the expressive intent of the cello. It seems the cello craved a true arrival while the piano wished to back away to the chord. *[At this moment the cello crescendoed to a climactic moment while the piano became softer. It seems that the sound of the piano did not successfully support the cello during one of the most expressive moments. The same occurrence happened later as well (see moment 1a)]]				✓		✓	✓	✓	
	I feel like they ended the phrase with a little *rit*. I really like that disposition.				✓					
E String listener, played it	I really like that the pianist changed the character when the music changed to another color.							✓	✓	
	I think the cellist could play that section a little bit hurriedly, just like the cello wants to say something with the pianist at the same time.			✓						
	The cello leaves us unresolved to be perfectly resolved by the piano's left hand. Impeccable timing.				✓		?			
F String listener, heard it	The cello's descending thirds (duples) are counterbalanced beautifully with the right hand piano (triplets) answers lyrically yet rhythmically.		✓		✓		✓	✓		
	The recap in the cello is more subtle yet more confident and projected, leaving the listener feeling like there is hope and not all fate.							✓		

(continued)

Comment author	Comment	Dynamics	Balance	Tempo	Timing/ Rhythmic motion	Synchronization	Communication	Expressivity	Tone quality	Other
G String listener, unfamiliar	That's the first real *f* moment in this piece. But in the recording, I couldn't really hear big differences.	✓								
	That's the repetition of moment I*, but it sounded more open and intense. But I'm not sure if I feel more expressivity in that spot or he played differently than moment I. *[That's the first real *f* moment in this piece. But in the recording, I couldn't really hear big differences.]							✓		
	For me, I think this spot should be most expressive moment of the piece, but it wasn't special enough.							✓		
	Beginning – very beautiful. Setting the correct mood for the piece.							✓		
H String listener, unfamiliar	Recap of the melodic theme. Sounds a little bit dull. I wish I could hear a different statement about it.							✓		
	Ending: the cellist is playing beautifully, but I wish the pianist would have played softer, really *pp* so it would feel like dying (or falling asleep).		✓	✓				✓	✓	
	The *fp*'s in the right hand in the piano are a bit striking; might be better to do *fp* in the context of moving toward.	✓			?					
I Pianist listener, played it	Piano right hand part could be interpreted as duple eighth notes, not triple eighth notes, in Schumann's works. And cello sighing sound is beautiful.				✓			✓	✓	
	Part of last pages sounds dragging because there are too many *rubatos*. Might be a better idea to wait, not slowing down too soon, until next section.			✓	✓					
	Tension between two is different (and direction).						✓			
J Pianist listener, played it	Piano, beautifully accompanied; overall there are places where piano overplays/sings the cello.		✓?						?	
	Piano, singing but more balance with the cello especially when it's octaves.		✓					✓	✓	
	Cellist and pianist are not together.					✓				
K Pianist listener, played it	The inner voice is too loud in the piano part.	✓								
	Ensemble not together.					✓				

Comment author	Comment	Dynamics	Balance	Tempo	Timing/ Rhythmic motion	Synchronization	Communication	Expressivity	Tone quality	Other
L Pianist listener, played it	Every time this rhythmic pattern of 2 against 3 appears, it lacks a feeling of rhythmic urgency. The duple rhythm could be played more as if it interrupts and with a little more *tenuto* (expressive emphasis).				✓			✓		
	This whole transition part could be somewhat more moving and flowing in both parts. The pianist could use the triplets to push and rhythmic flow like an ongoing current and the 8th notes in both parts can be played with less emphasis on each note so that it can sound like a big gesture.				✓			✓		
	The third beat where it's marked *fp* lacks this painful emotion. It sounded like taking it for granted. One could perhaps use a little more time, while still listening to the note already played, and use more sound to bring out more emotion that can resolve later.				✓			✓		
	I feel the left hand *fp* chords, which are first introduced in bar 10, need to be just as committed to emotionally and brought out as much as the right hand when it has this marking.	✓						✓		
M Pianist listener, heard it	In bars 29-35, I want greater difference in dynamic level between the three musical parts: the melody the cello, the countermelody in the right hand and the murmuring triplets in left hand of the piano.	✓	✓							
	I want more inflectional variety between the repetitive figures in bars 65-69. I sense a 'floating away' quality in dynamics here, but I want the pulse to stretch even more with each additional statement of the melodic fragments.	✓			✓			✓		
	The dynamic change from *pp* to *f* should be bigger.	✓								
N Pianist listener, unfamiliar	I like how this *f* wasn't too aggressive, but more lyrical. It was given a buoyancy that made it soar.							✓		
	Need to linger more on this downbeat to make the surprise harmony more apparent.				✓			✓		