

Quirky Conversations: How people with a diagnosis of schizophrenia do dialogue differently

Christine Howes
University of Gothenburg
christine.howes@gu.se
ORCID: 0000-0002-2794-1586

Mary Lavelle
Queens University, Belfast
mary.lavelle@qub.ac.uk
ORCID: 0000-0002-3951-0011

Abstract

People with a diagnosis of schizophrenia (PSz) have difficulty engaging in social interaction, but little research has focused on dialogues involving PSz interacting with partners who are unaware of their diagnosis. Using quantitative and qualitative methods on a unique corpus of triadic dialogues of PSz first social encounters, we show that turn-taking is disrupted in dialogues involving a PSz. Specifically, there are on average longer gaps between turns in groups which contain a PSz compared to those which do not, particularly when the speaker switch occurs from one control participant to the other. Furthermore, the expected link between gesture and repair is not present in dialogues with a PSz, particularly for control participants interacting with a PSz. As well as offering some insights into how the presence of a PSz affects an interaction, our results also demonstrate the flexibility of our mechanisms for interaction.

Keywords Dialogue; Schizophrenia; Turn-taking; Gesture; Repair

1 Introduction

Schizophrenia is diagnosed in approximately 1% of the population and is characterised by social dysfunction (DSM-IV). Difficulty engaging in social interaction is one of the most debilitating aspects of the disorder with significant consequences for the lives of those diagnosed. People with a diagnosis of schizophrenia have low rates of employment (1), smaller social networks, and are one of the most socially excluded groups in society (2). Social functioning difficulties are present prior to the onset of other diagnostic symptoms such as hallucinations or delusional beliefs, are persistent over time and associated with poorer prognosis (3; 4). In line with this, social support has been shown to be a protective factor in this diagnostic group, thus social deficits further compound their prognostic outcomes (5; 6).

Our understanding of the social difficulties in schizophrenia is derived primarily from studies exploring social cognition in people with a diagnosis of schizophrenia (PSz), referring to the mental operations that underpin the process of social interaction such as perception and interpretation of social cues (7). Such studies infer social skill from performance in off-line tasks completed outside the context of social interaction. Examples include discriminating facial expressions in pictures; attributing emotional states to the protagonists in short narratives; and

inferring intentions in abstract problem solving contexts. People with a diagnosis of schizophrenia consistently have poor performance on such tasks (7; 8), yet it is unclear how this translates to the complex and nuanced dynamics required to navigate a real world social interaction.

1.1 Dialogue

In contrast to written language, talk in interaction is characterised by incomplete utterances, non-word fillers (*hesitation particles*: ‘uh’ ‘erm’), pauses, and self-repairs such as repeated words or phrases or reformulations of the utterance in progress (9). Such disfluencies are often taken to be symptomatic of problems with communication, particularly in terms of self-monitoring one’s own speech production (10). However, disfluencies in an individual’s utterances are also affected by the interaction itself (11). Self-repairs may be produced as a direct response to feedback, including non-verbal behaviour in the form of head nods, facial expressions or hand gestures, and are also themselves linked to increased gesture use in the speaker (12; 13; 14; 15). Much work has been done on gesture in terms of types of gesture, timing and function (16; 17; 18; 19, among many others). In this paper we do not address such detailed distinctions of gesture, but take a neutral approach, for which we use automatically derived hand movement measures as a proxy (see section 2.4 for details).

Additionally, dialogue involves multiple individuals who have to organise their turn-taking appropriately so that they are not all talking at once, and nor are there large gaps between turns. The classic Conversational Analysis (CA) account of turn-taking (20), takes speaker change to be licensed at Transition Relevance Places (TRPs), which occur after Turn Constructional Units (TCUs); segments of speech which are in some sense complete. However, speaker change is not obligatory at a TRP since the current speaker may continue with a new TCU. Where turn changes do occur, Sacks et al. (20) distinguish between cases where the current speaker selects the next speaker (for example by directly addressing a question to someone or gazing and/or gesturing towards a specific individual as they reach a TRP) and those in which the next speaker self-selects (e.g. by answering a question that was directed to a group of people). In contrast, cue-based models (21) emphasise the types of embodied behaviours that speakers make use of to manage turn-taking behaviour, including multimodal factors such as gaze and gesture, but does not fully consider the effects of the (potential) next speaker’s behaviours. While there is some debate in the literature, there is good evidence that gaze can facilitate turn yielding in face-to-face dialogue (22), although it is not be a completely reliable signal, especially in multiparty dialogue (23). Additionally, pauses or hesitation particles are also associated with such interaction management (24), and can signal turn yielding or floor-holding.

For people with a diagnosis of schizophrenia, individual tasks suggest that they have difficulty monitoring their own behaviour (25) and mismatches between speech and gesture (26). Role-play studies also show that PSz are less effective at meshing their turns (27) and have atypical patterns of gesture (28). Gesture performance in particular is linked to social functioning prognosis longitudinally (29). Evidence from genuine interactions is limited however, and contradictory, with some studies reporting that PSz use fewer self-repairs than people without a diagnosis (30), for example, and others reporting that they use more (31; 32). These studies may not be comparable due to the contexts of the interactions; often with a therapist or interviewer who is aware of the diagnosis.

Studies using a unique corpus of triadic interactions in which half of the dialogues include a person with a diagnosis of schizophrenia, but their interacting partners are not aware of the diagnosis (see section 2), show that PSz use fewer gestures whilst speaking (33), and have reduced coordination between gesture and speech (34) and between gesture and repair (35). Analysis of

disfluencies on the same data (36) show that in contrast to the evidence from therapist interactions (32), PSz use fewer self-repairs than both their interacting partners and controls in groups without a PSz, suggesting that the context of the interaction is a key factor, rather than issues with self-monitoring per se. PSz and their interacting partners use fewer hesitation particles than controls, which may be due to reduced competition for the floor, an empirical question requiring more in depth analyses of different corpora. For unfilled pauses (defined as gaps of greater than 200 milliseconds between utterances where the same speaker speaks before and after the gap), these are more common within the turns of people interacting with a PSz. We hypothesise that this may be because floor-holding and turn yielding cues are less useful in dialogues with a PSz.

The triadic nature of the interactions in this corpus offers a unique opportunity to investigate the dynamics of turn exchange when there is competition for the position of speaker/addressee. This is not possible in dyadic interactions, and is more complex in larger multiparty interactions.

The findings derived from this corpus to date on disfluencies and gesture (36; 33; 34) are based on data collated at the level of participant, so cannot directly address issues around the timing or dynamics of turn-taking at the level of the intertwining utterances. The current mixed methods analysis builds on the previous analyses, investigating the dynamic nature of disfluencies, gesture and their multimodal relationship in the negotiation of turn exchange in these triadic interactions.

2 Methods

2.1 Data

The corpus, described elsewhere (33) consists of 40 triadic interactions. Half of these involve a person with a diagnosis of schizophrenia (6 male and 14 female) and two control participants (21 male and 19 female), and the other half involve three control participants (34 male and 26 female). In each triad, people were unfamiliar to each other. In PSz interactions, control participants were unaware that they were interacting with a person who had a diagnosis of schizophrenia. Interactions were motion captured and audio-visually recorded. All procedures were approved by a NHS Research Ethics Committee (07/H0711/90). All participants provided written informed consent. One control group and one PSz group were excluded from the analysis due to issues with video data. A further PSz group was excluded due to missing motion capture data.

2.2 PSz sample

Exclusion criteria included those presenting with motor side effects from antipsychotic medication (e.g., muscle stiffness and involuntary muscle spasms) and non fluent English speakers. Seventeen of our PSz participants were taking anti-psychotic medication (2 typical; 15 atypical) and three were medication free. All PSz participants were diagnosed as having schizophrenia and were regularly attending psychiatric outpatient appointments. The positive and negative syndrome scale for schizophrenia (PANSS) (37) assessed their positive, negative, and general symptoms. PSz symptoms scores were relatively low (PANSS Positive symptoms $M=15.8$, $sd=6.76$; PANSS Negative symptoms $M=9.95$, $sd=3.36$; PANSS General $M=28.41$, $sd=10.42$). PSz participants were not displaying overt symptoms at the time of the interactions, e.g. verbally responding to auditory hallucinations.

2.3 Task

Participants were instructed to discuss a moral dilemma called the Balloon Task. This task (described in detail elsewhere, see e.g. 38) requires participants to reach agreement on which of four passengers should be thrown out of a hot air balloon that will otherwise crash, killing all the passengers, if one is not sacrificed.

2.4 Analysis

All speech data was transcribed in ELAN (39). The duration between utterances was extracted automatically in milliseconds. Positive durations correspond to gaps and negative durations correspond to overlaps. For the turn-taking analysis presented here we only considered cases where there was a speaker change. We do not consider whether the speaker change occurs at a transition relevance place or not – in some cases it will, in other cases it will not (for example where there is an apparent speaker switch for a mid-turn backchannel from another speaker which will split a single turn from a speaker into two utterances, with an intervening backchannel). Of course this has consequences for the analysis and interpretation, however, the methodology applies equally to both PSz and control groups so differences between them are still relevant and meaningful. See also (40) for discussion of related issues.

Self-repairs were annotated using STIR (STrongly Incremental Repair detection; 41) which automatically detects speech repairs on transcripts. STIR is trained on the Switchboard corpus (42), and has been shown to be applicable to therapeutic dialogue, with high rates of correlation to human coders in terms of self-repair rate (43).

Hand movement was automatically extracted from the raw motion capture data. In order to control for individual variation, for each participant we extracted the movement from each of the three hand/wrist markers, and calculated mean and standard deviation of movement in any direction by frame in mm. For frames with missing markers, if this was fewer than 50 frames (frame rate 60/second), we imputed the missing data using a linear trajectory, otherwise left the data as missing. Following the methodology in (33; 44), to account for individual variation, for each pair of frames we calculated whether the movement between them was greater than the individual's mean movement plus one standard deviation, for any of the three markers, and if so marked this as movement. The use of all three wrist and hand markers helps to mitigate the points where single markers dropped out e.g. due to occlusion. The hand movement data was imported to ELAN. Visual inspection of the data suggested that using an individual's mean + 1 standard deviation is generally a good proxy for hand movement. However, this is not the case where this value was very low (due to minimal or no movement, or extreme cases of marker drop out) in which case the algorithm was oversensitive to minor non-gestural movements caused by e.g. posture shifts. It was also not accurate in cases where the value of the mean plus one standard deviation was very high (individuals who gesture a lot), in which case the algorithm was undersensitive to genuine gestures. For this reason, we introduced a lower and upper threshold for the movement values. These were set at 2mm per frame for the lower bound and 5mm per frame for the upper bound. These refinements to the movement calculation result in a more reliable and sensitive index of hand movement than has been adopted in previous analyses of this corpus (e.g. 33).

It should be borne in mind that although we believe that our automatically derived hand movement measures are a good proxy for gesture, they do not distinguish between gestural hand movement and other hand movement (e.g. scratching, fidgeting). It is also the case that the automatic hand movement annotation captures only the movement phases of a gesture – including preparation and retraction (16), and will not pick up any hold phases of gestures,

which are known to be interactionally relevant (see e.g. 45), particularly in respect to turn-taking.

Analyses were performed in SPSS 28.

3 Results

3.1 Turn transitions

In order to assess differences in turn-taking in the control and PSz dialogues, we compared the duration between each turn by group, participant type and who speaks next.

		mean	s.d.	N
Group	Control Group	-59.66	780.56	3776
	PSz group total	-7.18	723.49	3294
Participant Type	PSz in PSz group	-8.38	720.55	937
	C in PSz group	-6.70	724.81	2357
Next speaker	C to C in PSz group	15.64	724.73	1418
	C to PSz in PSz group	-40.44	724.01	939

Table 1: Turn change duration by group, participant type and speakers

As seen in table 1, both the PSz and control groups have turn changes which are on average below zero (i.e. in overlap). One-way ANOVAs show that there is a significant difference between the groups with turn exchanges in the control groups occurring faster than those in the PSz group ($f_{(1,7068)} = 8.513, p = 0.004$). However, drilling down in the PSz groups suggests that this is not the complete picture. While there is no significant difference in the gap following a turn by the PSz and Cs in the PSz groups, this masks the difference which emerges when we look not only at the identity of the person who speaks the turn *prior* to the turn change but also the turn *following* the turn change, as shown in figure 1 ($f_{(3,7066)} = 3.880, p = 0.009$). Post hoc tests with a Bonferonni correction show no significant differences between groups except for between C to C in the control and PSz groups ($p = 0.008$, 95% CI -137.31 to -13.29). This means that turn exchanges between the two Cs in the PSz group have a longer gap than turn exchanges between Cs in the control group.

These results, particularly in conjunction with the less fine-grained results presented in (36), provide evidence that in the dialogues with a person with a diagnosis of schizophrenia, there are differences in the timing of turn transitions. We hypothesise that this could indicate that the cues for turn ending (or specific next speaker selection) are less clear in the PSz interactions, i.e. they may be missing or ambiguous. It could also be the case that what the PSz has said may be more difficult to formulate an appropriate response to. It is particularly striking that the turn exchanges most affected by the presence of a PSz are those between their C interlocutors. We interpret this as suggesting that there are specific points (such as TRPs; 20) at which the Cs in the PSz groups are expecting (or encouraging) the PSz to take the turn but the PSz is not doing so, with the result that after a pause the other control steps in. However, more detailed analyses of whether the gaps are consistently occurring at TRPs is required to validate this interpretation. This may also suggest that PSz are less responsive to turn taking cues or more reluctant to select as next speaker, but again more targeted analyses are required to ascertain if this is the case.

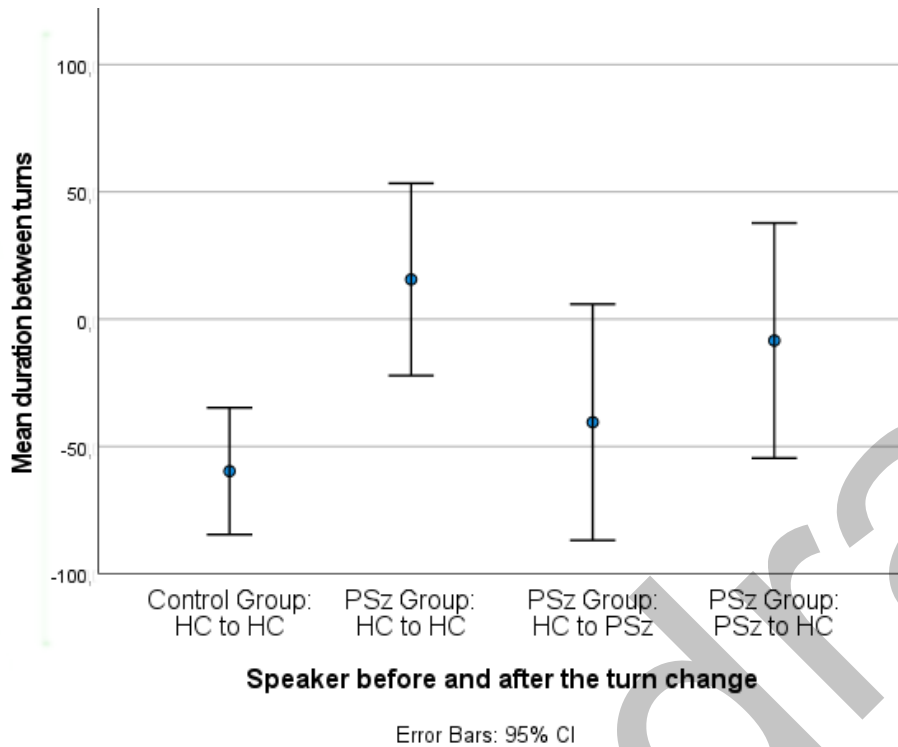


Figure 1: Turn change duration by speakers before and after the turn change

3.2 Qualitative analysis

Alongside previous results on disfluencies, the above quantitative analysis suggests that in PSz interactions there are bigger gaps between turns between the two Cs in interactions including a PSz. But what does this mean in practice? We now turn to some examples from our data. Note that our examples are for illustrative purposes only. We do not claim that our discussion of them is exhaustive and nor do we describe the examples in the level of detail of Conversational Analysis.

3.2.1 Passing up the opportunity to take the floor

The clearest case of the pattern of pauses which we hypothesise to be more common in the PSz interactions, is that where the PSz is passing up an apparent opportunity to take the floor, which is subsequently taken by the other control participant, after a gap. An example of this can be seen in example (2). In this example, C3 produces the first part of a question-answer adjacency pair (46; 47) in line 3, table 2. This creates the expectation for the second pair part – namely an answer. As can be seen in (2a), as C3 is asking the question, both the PSz and C1 are looking at C3 as he speaks. C3's gaze and postural orientation is towards the PSz, which is taken as an indication – at least by C1 who does not immediately proffer an answer to the question – as C3 selecting PSz as the next speaker (23). In line 4 (2b), following a short gap in which none of the participants alters their posture or shifts their gaze, and no answer is forthcoming from PSz, C3 further specifies his initial question, increasing the expectation of a response. After approximately a second of the 2.39 second gap in line 5, rather than providing an answer, which both C3 and C1 seem to expect, PSz leans back slightly and shifts his gaze towards C1, at which

point she takes the floor and provides the second part of the original adjacency pair. The 2.39 second gap in line 5 is over twice as long as the one second “standard maximum” silence in dialogue proposed by Jefferson (48), and such a long silence may go beyond a gap to become a lapse in the conversation (49). In this case, it is only after the standard maximum silence duration that the PSz shifts his posture and gaze and C1 appears to take this as an indication that her expectation that he would answer the question posed by C3 is incorrect, leading to her self-selection as next speaker.¹

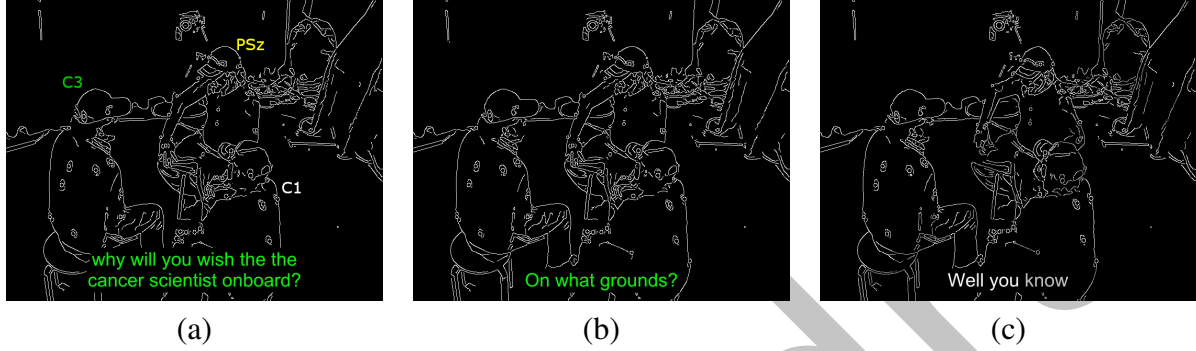


Figure 2: PSz non response to a question

- | | | |
|---|-----|---|
| 1 | C3 | yeah but listen the thing is that erm erm (0.19) |
| 2 | | for arguments sake (0.18) |
| 3 | | if you're going with the kid (1.19) |
| 4 | | #why will you wish the the cancer scientist onboard? |
| | fig | #2a |
| 5 | | (0.31) on# what grounds? |
| | fig | #2b |
| 6 | | (2.39) |
| 7 | C1 | Well # (0.44) you know (0.21) what's the kid going to |
| | fig | #2c |
| 8 | | contribute to society apart from music? |

Table 2: PSz passing up the opportunity to take the floor by not responding to a direct question

3.2.2 Explicit turn exchange cues

The example sequence in figure 3 and table 3, shows one of the strategies adopted by the controls in dialogue with a PSz. At the beginning of this example, in (3a) the speaker, C2, is actively orienting her body, gaze, head and gestures towards the PSz, with C3, also orienting towards the PSz. The PSz is unresponsive and as C2 begins to articulate the final word of the utterance in line 1-2 (“thing”) she turns her gaze towards C3, who simultaneously turns her gaze towards C2 and provides an acknowledgement in overlap (3b), line 3, as well as a non-verbal nodding response. This creates an environment in which the agreement or disagreement of the PSz becomes relevant — even without C2’s bodily orientation towards her, which the PSz does not take up in the following 1.81 second long gap. Interestingly, even before C3’s agreement, C2’s utterance in line 1-2 (which is oriented towards the PSz) seems to be looking to elicit

¹We thank our anonymous reviewer for this point.

agreement from the PSz, as it builds on a recent previous utterance from her; less than 10 seconds previously the PSz stated “I don’t reckon it’s that hard to fly a balloon”.

Towards the end of the long gap, C2 turns her gaze back towards the PSz, followed by asking her a direct question (3c), line 4. As in the previous example, this first pair part sets up the expectation for the second part of the adjacency pair. This means that the PSz no longer simply has the opportunity to take a turn, but there is a stronger – socially normative – obligation for her to do so. At the start of the question in line 4, C3 is still gazing at C2, but she turns her gaze towards the PSz over the (one second) duration of the question, showing that she also orients to the expectation of a response from the PSz (50).

Following a 540 millisecond gap where the expected response is not forthcoming, C2 provides, in line 5, a reason for the question, which demonstrates that she expected the PSz to take a turn. The PSz once again avoids the now more explicit turn-taking cue by actively looking away, precisely as C2 begins the utterance in line 5 “you haven’t...” (3d), (3e) turning her gaze back to C2 just after C2 utters “older” in line 6 before finally taking the floor in (3f), line 8. As can be seen in this example, the Cs are doing a lot of interactive work to include the PSz in the interaction.

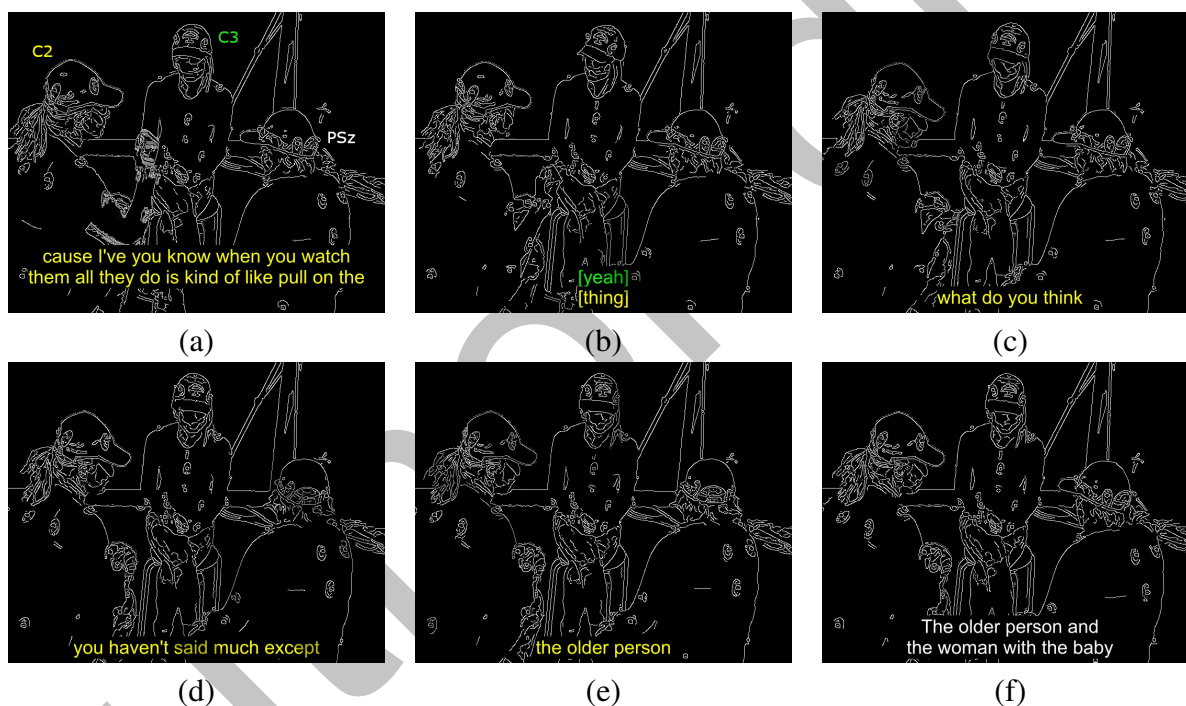


Figure 3: Increasingly explicit turn cues

Taken together, the quantitative results and these examples suggest that there are cases in the PSz interactions where there are differences in turn-taking patterns. This may be because the PSz lacks awareness of the normative turn-taking expectations, or because they are actively avoiding cues directed towards them.

We now turn to the relationship between gesture and repair in dialogues with a PSz.

3.3 Hand movement and repair

As discussed, previous work shows that PSz have decreased gesture use whilst speaking (33) and use fewer self-repairs than controls (36), even when normalised for the fact that PSz speak less than their interlocutors. One possible explanation for the reduction in gesture in PSz is

1	C2	cause I've you know when you watch them all they		
2		do is kind of like# pull on the (0.89) #[thing]		
	fig		#3a	#3a
2	C3			#[yeah]
3		(1.81)		
4	C2	what# do you think (0.54)		
	fig		#3c	
5		you haven't said much# except (0.27)		
	fig		#3d	
6		the (0.36) #older person		
	fig		#3e	
7		(0.02)		
8	PSz1	#The older person and the woman with the baby		
	fig		#3f	

Table 3: Increasingly explicit turn exchange cues from controls towards PSz

precisely the link with repair, since people typically increase their gesture use in problematic turns, as indexed by rates of repair (51).

Firstly we should note that, in contrast to previously reported results based on participant level values and a less sensitive measure of gesture (35; 33), we did not find that PSz gestured less as speakers. When we look at hand movement on a frame by frame level (see figure 4), we find that PSz gesture *more* whilst not speaking, whilst the controls in the PSz group gesture proportionally *less* while speaking, as shown in figure 5. A GLMM (with a binomial probability distribution and logit link function) by frame with presence or absence of hand movement as the dependent variable, presence or absence of speech and group/condition as fixed factors, and triad as random variable² showed a significant main effect of speaking ($f_{1,1062168} = 44607.14, p > 0.001$), such that hand movement is much more likely when someone is speaking, a significant main effect of group/condition ($f_{2,1062168} = 66.02, p > 0.001$), such that PSz produce more gesture than their C interlocutors, and a significant interaction effect of group and condition by speaking ($f_{2,1062168} = 106.86, p > 0.001$), with significant differences in the non-speaking condition between the PSz in the PSz group and both Cs in the PSz group ($p < 0.01$) and Cs in the control group ($p = 0.016$) and in the speaking condition between the PSz and Cs in the PSz group ($p < 0.001$). This means that PSz produce more gestures than both C groups when not speaking, and their interacting partner Cs produce fewer gestures than the PSz when speaking.

Turning to the relationship between gesture and repair (looking only at the hand movements which co-occur with speech), we can see that the presence of repair in a turn is a good predictor of amount of hand movement, but only in the control groups, as shown in figure 6. A GLMM (using a normal distribution and identity link function) with number of speaking frames containing hand movement as the dependent variable, group type and presence or absence of repair and number of words as independent variables and triad, age and gender as random effects showed a main effect of group type (such that control groups contain more hand movements: $f_{1,7103} = 5.23, p = 0.022$), a main effect of number of words (such that longer

²Models including frame number as an additional effect did not alter the main pattern of results of account for any more of the variance in the data so we report models without these here.

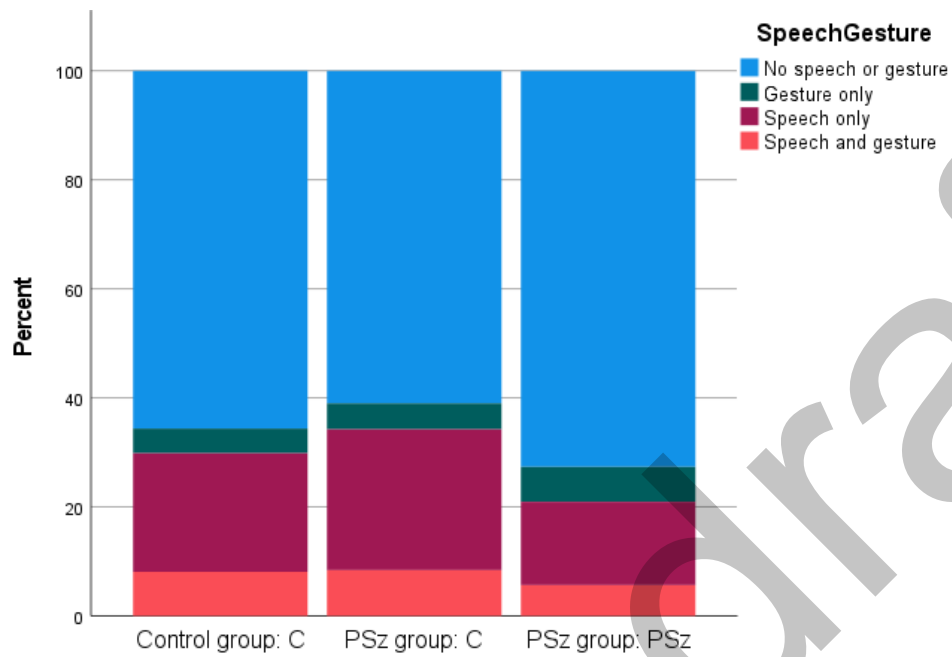


Figure 4: Gesture and/or speech by number of frames

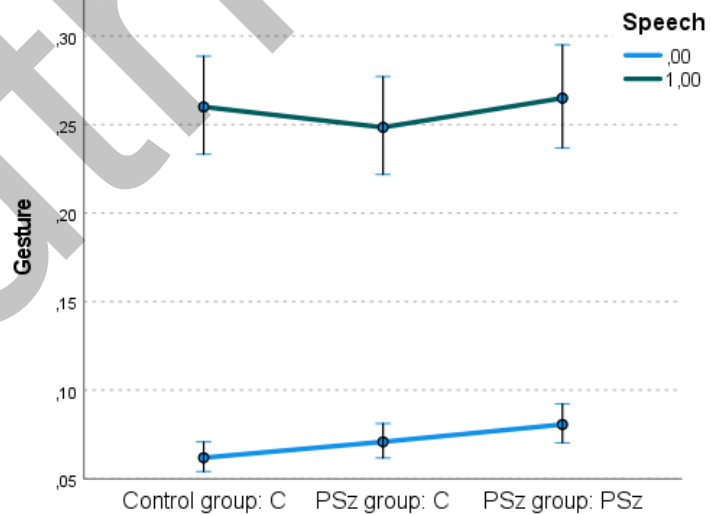


Figure 5: Gesture by speech and group/condition

turns are more likely to have more hand movements: $f_{1,7103} = 4992.50, p < 0.001$) and an interaction between group type and repair ($f_{1,7103} = 28.40, p < 0.001$, such that there was a difference between the groups in the repair condition, but no significant difference in turns without repair: $t = 3.68, p < 0.001$).³ This means that there is less hand movement in turns with repair in the PSz group than in the control group, and demonstrates that the main effect of group is driven by the differences in turns which contain repair. There was no main effect of repair ($f_{1,7103} = 2.08, p = 0.150$).

Interestingly, when we drill down by participant type, we see that the effect is more pronounced in the Cs in the PSz group (6b), with amount of hand movement of a PSz unaffected by the presence or absence of repair. A GLMM with the same settings and factors except with group/condition instead of group showed the same pattern of main and interaction effects.⁴ Post hoc tests show that in the turns with repair, Cs in the control group use more gesture than Cs in the PSz group ($t = 3.95, p < 0.001$), and that Cs in the control group use more gesture in turns with repair ($t = 4.756, p < 0.001$) than in turns without repair, in line with previous findings. In contrast, Cs in the PSz group show the opposite pattern, producing less gesture in turns containing repair ($t = -2.66, p = 0.008$). We will return to the possible reasons for this in the discussion.

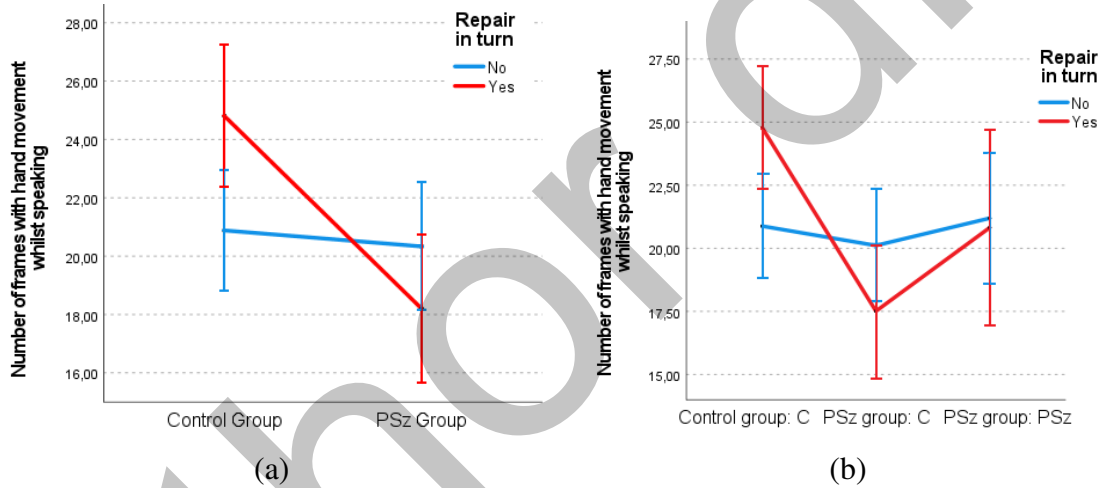


Figure 6: Gesture by repair and group/condition and next speaker

Further, this difference is sensitive to the turn-taking dynamics discussed in section 3.1. As shown in figure 7, when we also consider who is the next speaker,⁵ we see that the unexpected effect of Cs producing less gesture in turns containing a repair only holds when the PSz is the next speaker ($t = -3.289, p = 0.001$).

4 Discussion

This mixed methods analysis provides an in-depth investigation of turn exchange behaviour in triadic interactions involving a person with a diagnosis of schizophrenia. The findings demon-

³Post hoc tests were carried out using pairwise comparison and the Bonferroni correction.

⁴Main effect of number of words: $f_{1,7101} = 4991.74, p < 0.001$; group/condition $f_{2,7101} = 4.39, p = 0.012$ and interaction effect group/condition by repair $f_{1,7101} = 14.48, p < 0.001$.

⁵Once again, the pattern of significant effects is the same: number of words: $f_{1,7061} = 5085.76, p < 0.001$; group/condition/next speaker $f_{3,7061} = 4.13, p = 0.006$ and interaction effect group/condition/next speaker by repair $f_{3,7061} = 11.52, p < 0.001$.

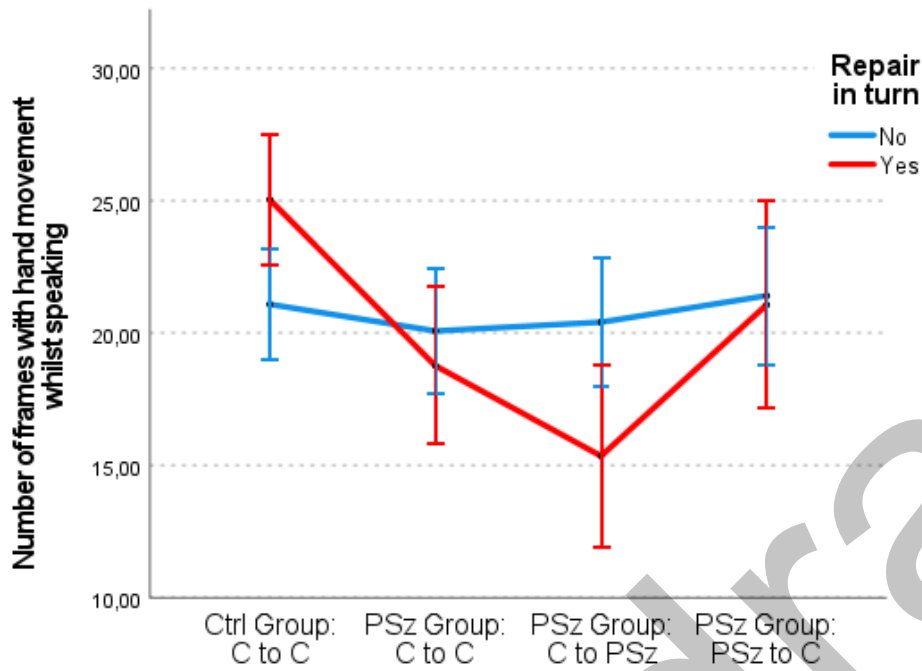


Figure 7: Gesture by repair and group/condition

strate that patients' turn taking behaviour deviates from expectations, in often subtle ways, resulting in adaptation by others. Interactions involving PSz showed reduced competition for the floor, and a lack of clarity about who should take the floor. PSz' use of floor change cues deviates from control participants' expectations; missing or avoiding subtle turn exchange cues and failing to provide such cues during their own turns. This difference in response time is also consistent with work which suggests that competition for the floor in multiparty interactions usually reduces turn exchange times (52) – as is the case in the control groups, but not in the PSz groups, where competition seems to be reduced.

We hypothesise that the differences seen in the turn-taking behaviours may be because the PSz lacks awareness of the normative turn-taking expectations, or because they are actively avoiding cues directed towards them: Our analysis here, based on an automatic detection of silences between turns, does not distinguish between gaps which occur at a TRP and those that do not, which may be a factor here (though note that our broad observations about differences between the groups hold despite this lack of nuance). Future work will address these questions using both human annotation and recently developed automatic methods to detect TRPs (53), and will also consider the content and type of turns (38; 54), since our qualitative analysis suggests that adjacency pairs may be used in the PSz dialogues to make turn-taking cues more explicit.

Furthermore, the multimodal coordination between gesture use and self-repair that persists in control interactions, is not present in interactions involving a person with schizophrenia. PSz' use of gesture was not related to their use of self-repair, while controls interacting with a PSz displayed *fewer* gestures in utterances that included self-repair. Overall, the presence of a person with a diagnosis of schizophrenia in a social interaction changes the behaviour of those they are interacting with, despite their diagnosis being undisclosed.

Our findings align with those of previous studies of communication in PSz derived from

role-play or task based methodologies, specifically, PSz have difficulty meshing their turns (27), have deficits in their use of gesture (28) and a mismatch between gesture and speech (26). This study demonstrates that these difficulties persist in naturalistic social interaction and have a significant influence on others' communication. Furthermore, research from the field of social cognition suggests that PSz have difficulty interpreting social cues in pen and paper assessments (7). Although our analysis does not investigate PSz's ability to interpret social cues, it does suggest that they may fail to utilise such cues when offered in conversation.

Our previous analysis of this corpus was at the level of participant and the index of gesture used speed of movement (33). When manually inspecting the gesture categorisation alongside the video footage this method was found to under and/or over categorise hand movements as 'gesture'. As such, in the current analysis we employed a more sensitive measure of hand gesture based on the raw motion capture data of three motion capture markers, rather than one, alongside manual inspection of the data. This analytic approach revealed that, compared to controls in either group PSz displayed more hand movements categorised as 'gesture' when they were not speaking. Observational annotation of these movements would be required to identify their nature, however one working hypothesis is that these movements are indicative of displacement behaviours, which are self-directed behaviours, found to be correlated with states of heightened arousal, or anxiety (e.g. 55). This raises potential questions about the validity of our methodology, when interpreted as 'gesture', as discussed in section 2. However, we believe that our methods bring advantages in terms of the scale of data that can be analysed without time-consuming human annotation effort. We also believe that the insights into the shifting dynamics through an interaction brought out by our automatic hand-movement detection can be complementary to more traditional gesture annotation and qualitative methods.

In the current analyses, controls interacting with a PSz produced *fewer* hand gestures when speaking. Taken alongside the lack of competition for the floor, this pattern may suggest a reduction in the need to employ floor holding techniques such as hand gesture. Although this may be the case, drilling further into the turn exchange dynamics we also identified a disruption in the relationship between gesture and speech in both PSz and their control partners, compared to the control group, where we saw the expected relationship between repair and increased gesture use (in line with 14; 51).

The most prominent finding is seen in turn exchanges where controls pass their turn to the PSz; here Cs employed significantly fewer gestures when they have verbal difficulty, as indexed by use of self repair. In interactions with a PSz, the relationship between speech repair and gesture is disrupted, not only for PSz but for the interaction as a whole. This suggests that when passing a turn to a PSz, controls prioritise use of one modality rather than coordinating modalities.

This is puzzling. We hypothesise that the use of gesture in turns which contain self-repair is a normally productive strategy, signalling either the presence of a potential problem or the attempt to resolve it. In interactions with a PSz, where turn exchange may have greater ambiguity, this strategy does not seem to be employed. This might suggest that its usefulness is overridden by other considerations in the interaction. Although the current analyses does not identify the reason for this, we present a number of possible explanations. For example, if the potential misunderstanding involves the PSz, which may be more likely in cases where they take the next turn, their diminished responsiveness, potentially even a lack of shared gaze (see the example in figure 3), may mean gestures are less useful in this context, with more explicit verbal requests being favoured as an alternative to engage the PSz. One way to potentially unpick this in PSz interactions where turn exchange is ambiguous would be to analyse turns, not by next speaker, but by the identity of the direct addressee. This would enable us to see if the reduction in

coordination between repair and gesture remains when the turn exchange is more predictable and the addressee is another control, rather than the PSz. Other possible factors which could be investigated in future work are whether the types of repairs are the same (for example, there may be less need for gestural support for articulation repairs) or whether the Cs in interaction with a PSz change their behaviour during the course of dialogue. Another possible direction for future analysis which can also be conducted on the motion capture data, is identifying the types of movements. In (15), they found that maximum hand heights for speakers were higher during disfluencies compared to other moments in interaction, suggesting that there are particular types of gestures associated with repair. If these gestures are interactive in nature (analogously to the verbal hesitation particles which act as a floor holding device), then this might also be a factor in their reduced use when passing the floor to the PSz, due to the already discussed lack of competition for the floor.

The current analysis did not investigate the reasons for the behaviours observed, in either turn-taking or gesture and repair, nor do we suggest that any behavioural pattern is superior to another. This study merely presents an account of how communication in interactions involving PSz differs from those involving controls. We do not know if the nature of the task, or the fact that it took place in a lab environment contributed to a level of anxiety in patients or how this may have contributed to patients' behaviour. Future comparative studies in different contexts, using different conversational topics could explore these questions.

5 Conclusions

These investigations demonstrate the complex interconnectedness of participants in an interaction and the necessity to analyse them as a dynamic unit (56; 57). As well as offering some insights into how the presence of a PSz affects an interaction as a whole, they also demonstrate the flexibility of our mechanisms for interaction. As the controls conversing with a person diagnosed with schizophrenia show, strategies can be –and are– adjusted on the fly to account for deviations in the expected interactive behaviours from one's interlocutors. As usual, there is much work still to be done.

Acknowledgements

This work would not have been possible without our many colleagues and friends, including (but not limited to) Ellen Breitholtz, Robin Cooper, Pat Healey, Julian Hough, Rose McCabe and Matt Purver. The views expressed in this paper – as well as any errors that remain – are our own.

Funding

This work was supported by Riksbankens Jubileumsfond (RJ: P16-0805:1) and the Swedish Research Council (VR 2014-39).

References

- [1] Evensen S, Wisløff T, Lystad JU, Bull H, Ueland T, Falkum E. Prevalence, employment rate, and cost of schizophrenia in a high-income welfare society: a population-

- based study using comprehensive health and welfare registers. *Schizophrenia bulletin*. 2016;42(2):476-83.
- [2] Macdonald E, Hayes R, Baglioni Jr A. The quantity and quality of the social networks of young people with early psychosis compared with closely matched controls. *Schizophrenia research*. 2000;46(1):25-30.
 - [3] Niendam TA, Jalbrzikowski M, Bearden CE. Exploring predictors of outcome in the psychosis prodrome: implications for early identification and intervention. *Neuropsychology review*. 2009;19(3):280-93.
 - [4] Fusar-Poli P, Byrne M, Valmaggia L, Day F, Tabraham P, Johns L, et al. Social dysfunction predicts two years clinical outcome in people at ultra high risk for psychosis. *Journal of psychiatric research*. 2010;44(5):294-301.
 - [5] Norman RM, Malla AK, Manchanda R, Harricharan R, Takhar J, Northcott S. Social support and three-year symptom and admission outcomes for first episode psychosis. *Schizophrenia Research*. 2005;80(2):227-34.
 - [6] Vázquez Morejón AJ, León Rubio JM, Vázquez-Morejón R. Social support and clinical and functional outcome in people with schizophrenia. *International Journal of Social Psychiatry*. 2018;64(5):488-96.
 - [7] Green MF, Horan WP, Lee J. Social cognition in schizophrenia. *Nature Reviews Neuroscience*. 2015;16(10):620.
 - [8] Couture SM, Penn DL, Roberts DL. The functional significance of social cognition in schizophrenia: a review. *Schizophrenia bulletin*. 2006;32(suppl_1):S44-63.
 - [9] Schegloff EA, Jefferson G, Sacks H. The preference for self-correction in the organization of repair in conversation. *Language*. 1977;53(2):361-82.
 - [10] Levelt WJM. Monitoring and self-repair in speech. *Cognition*. 1983;14(1):41-104.
 - [11] Goodwin C. The interactive construction of a sentence in natural conversation. In: Psathas G, editor. *Everyday Language: Studies in Ethnomethodology*. New York: Irvington Publishers; 1979. p. 97-121.
 - [12] Holler J, Beattie G. Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener? *Gesture*. 2003;3(2):127-54.
 - [13] Healey PGT, Lavelle M, Howes C, Battersby S, McCabe R. How listeners respond to speaker's troubles. In: *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Berlin; 2013. .
 - [14] Seyfeddinipur M, Kita S. Gestures and self-monitoring in speech production. In: *Annual Meeting of the Berkeley Linguistics Society*. vol. 27; 2014. p. 457-64.
 - [15] Özkan EE, Gurion T, Hough J, Healey PG, Jamone L. Specific hand motion patterns correlate to miscommunications during dyadic conversations. In: *2021 IEEE International Conference on Development and Learning (ICDL)*. IEEE; 2021. p. 1-6.

- [16] Kendon A. Gesticulation and speech: Two aspects of the process of utterance. In: Key MR, editor. Nonverbal communication and language. The Hague: Mouton; 1980. p. 207-27.
- [17] Goldin-Meadow S. The role of gesture in communication and thinking. Trends in cognitive sciences. 1999;3(11):419-29.
- [18] McNeill D. Gesture and thought. In: Gesture and Thought. University of Chicago press; 2008. .
- [19] Wagner P, Malisz Z, Kopp S. Gesture and speech in interaction: An overview. Speech Communication. 2014;57:209-32.
- [20] Sacks H, Schegloff EA, Jefferson G. A Simplest Systematics for the Organization of Turn-Taking for Conversation. Language. 1974;50(4):696-735.
- [21] Duncan S. Some signals and rules for taking speaking turns in conversations. Journal of Personality and Social Psychology. 1972;23(2):283-292.
- [22] Degutyte Z, Astell A. The role of eye gaze in regulating turn taking in conversations: a systematized review of methods and findings. Frontiers in Psychology. 2021;12:616471.
- [23] Lerner GH. Selecting next speaker: The context-sensitive operation of a context-free organization. Language in society. 2003;32(2):177-201.
- [24] Clark HH, Fox Tree JE. Using *uh* and *um* in spontaneous speaking. Cognition. 2002;84(1):73-111.
- [25] Johns LC, Rossell S, Frith C, Ahmad F, Hemsley D, Kuipers E, et al. Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. Psychological medicine. 2001;31(04):705-15.
- [26] Goss J. Speech-gesture mismatches and dynamic coordination in schizophrenia. The University of Chicago; 2011.
- [27] Mueser KT, Bellack AS, Douglas MS, Morrison RL. Prevalence and stability of social skill deficits in schizophrenia. Schizophrenia research. 1991;5(2):167-76.
- [28] Walther S, Vanbellingen T, Müri R, Strik W, Bohlhalter S. Impaired gesture performance in schizophrenia: particular vulnerability of meaningless pantomimes. Neuropsychologia. 2013;51(13):2674-8.
- [29] Walther S, Eisenhardt S, Bohlhalter S, Vanbellingen T, Müri R, Strik W, et al. Gesture performance in schizophrenia predicts functional outcome after 6 months. Schizophrenia bulletin. 2016;42(6):1326-33.
- [30] Caplan R, Guthrie D, Komo S. Conversational Repair in Schizophrenic and Normal Children. Journal of the American Academy of Child & Adolescent Psychiatry. 1996;35(7):950-958. Available from: <http://www.sciencedirect.com/science/article/pii/S089085670962476X>.
- [31] McCabe R, Healey PGT, Priebe S, Lavelle M, Dodwell D, Laugharne R, et al. Shared understanding in psychiatrist-patient communication: Association with treatment adherence in schizophrenia. Patient Education and Counselling. 2013.

- [32] Howes C, Purver M, McCabe R, Healey PGT, Lavelle M. Helping the Medicine Go Down: Repair and Adherence in Patient-Clinician Dialogues. In: Proceedings of the 16th workshop on the semantics and pragmatics of dialogue (SeineDial). Paris; 2012. p. 155-6.
- [33] Lavelle M, Healey PG, McCabe R. Is nonverbal communication disrupted in interactions involving patients with schizophrenia? *Schizophrenia bulletin*. 2013;39(5):1150-8.
- [34] Lavelle M, Howes C, Healey PGT, McCabe R. Speech and hand movement coordination in schizophrenia. In: Proceedings of the TiGeR Tilberg gesture research meeting. Tilburg; 2013. .
- [35] Howes C, Lavelle M, Healey PGT, Hough J, McCabe R. Helping hands? Gesture and self-repair in schizophrenia. In: Resources and processing of linguistic and extra-linguistic data from people with various forms of cognitive/psychiatric impairments (RaPID) at LREC; 2016. .
- [36] Howes C, Lavelle M, Healey PGT, Hough J, McCabe R. Disfluencies in dialogues with patients with schizophrenia. In: Proceedings of the 39th annual meeting of the Cognitive Science Society. London, UK; 2017. Available from: http://www.christinehowes.com/papers/Howes.etal_CogSci2017.pdf.
- [37] Kay SR, Flszbein A, Opfer LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia bulletin*. 1987;13(2):261.
- [38] Howes C, Breitholtz E, Lavelle M, Cooper R. Justifiable reasons for everyone: Dialogical reasoning in patients with schizophrenia. In: Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue. SEMDIAL; 2021. .
- [39] Wittenburg P, Brugman H, Russel A, Klassmann A, Sloetjes H. ELAN: a Professional Framework for Multimodality Research. In: Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation; 2006. .
- [40] Purver M, Howes C, Gregoromichelaki E, Healey PGT. Split Utterances in Dialogue: A Corpus Study. In: Proceedings of the 10th Annual SIGDIAL Meeting on Discourse and Dialogue. London, UK; 2009. p. 262-71.
- [41] Hough J, Purver M. Strongly Incremental Repair Detection. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Doha, Qatar: Association for Computational Linguistics; 2014. .
- [42] Godfrey JJ, Holliman E, McDaniel J. SWITCHBOARD: Telephone Speech Corpus for Research and Development. In: Proceedings of IEEE ICASSP-92. San Francisco, CA; 1992. p. 517-20.
- [43] Howes C, Hough J, Purver M, McCabe R. Helping, I Mean Assessing Psychiatric Communication: An Application of Incremental Self-Repair Detection. In: Proceedings of the 18th workshop on the semantics and pragmatics of dialogue (DialWatt). Edinburgh; 2014. p. 80-9.
- [44] Battersby S, Healey P. Head and hand movements in the orchestration of dialogue. In: Proceedings of the Annual Meeting of the Cognitive Science Society. vol. 32; 2010. .

- [45] Sikveland RO, Ogden R. Holding gestures across turns: Moments to generate shared understanding. *Gesture*. 2012;12(2):166-99.
- [46] Schegloff EA. Sequence organization in interaction: A primer in conversation analysis. vol. 1. Cambridge/New York: Cambridge University Press; 2007.
- [47] Kendrick KH, Brown P, Dingemanse M, Floyd S, Gipper S, Hayano K, et al. Sequence organization: A universal infrastructure for social action. *Journal of Pragmatics*. 2020;168:119-38.
- [48] Jefferson G. Notes on a possible metric which provides for a standard maximum silence of approximately one second in conversation. *Tilburg papers in Language and Literature*. 1983.
- [49] Hoey EM. Lapses: How people arrive at, and deal with, discontinuities in talk. *Research on Language and Social Interaction*. 2015;48(4):430-53.
- [50] Holler J, Kendrick KH. Unaddressed participants' gaze in multi-person interaction: optimizing reciprocity. *Frontiers in psychology*. 2015;6(98):1-14.
- [51] Healey PGT, Plant N, Howes C, Lavelle M. When words fail: Collaborative gestures during clarification dialogues. In: 2015 AAAI Spring Symposium Series: Turn-Taking and Coordination in Human-Machine Interaction; 2015. .
- [52] Holler J, Alday PM, Decuyper C, Geiger M, Kendrick KH, Meyer AS. Competition reduces response times in multiparty conversation. *Frontiers in Psychology*. 2021;12.
- [53] Ekstedt E, Skantze G. Projection of Turn Completion in Incremental Spoken Dialogue Systems. In: Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue; 2021. p. 431-7.
- [54] Breitholtz E, Cooper R, Howes C, Lavelle M. Reasoning in multiparty dialogue involving patients with schizophrenia. In: Amblard M, editor. (In)Coherence of Discourse. Language, Cognition and Mind. Springer; 2021. .
- [55] Troisi A. Ethological research in clinical psychiatry: the study of nonverbal behavior during interviews. *Neuroscience & Biobehavioral Reviews*. 1999;23(7):905-13.
- [56] Fusaroli R, Raczaszek-Leonardi J, Tylén K. Dialog as interpersonal synergy. *New Ideas in Psychology*. 2013;30:1-11.
- [57] Eshghi A, Gregoromichelaki E, Howes C. Action Coordination and Learning in Dialogue. In: Probabilistic Approaches to Linguistic Theory. CSLI; 2022. To appear.