Seeking Agreement in Remote Psychiatric Consultations: A Conversation Analytic Perspective

Naphat Jittavisutthikul[1], Patrick Healey[2], Sarah Bourdin[1], Philip McNamee[1, 3]

[1]Wolfson Institute of Population Health, Queen Mary University of London
[2]School of Electronic Engineering and Computer Science, Queen Mary University of London
[3]Newham Centre for Mental health, East London NHS Foundation Trust

Abstract

Meaningful interaction is key to effective mental healthcare. Increasingly, healthcare interactions are being delivered remotely. We do not yet understand what impact this may have on the quality of interaction. We use conversation analytic techniques to explore how people build shared understanding in a small corpus of video-based psychiatric consultations using an application designed to support remote conversations about mental health (Remote DIALOG+). We identify four key issues: i) disruptions to the flow of conversation due to the structure imposed by the standardised sequence of questions (protocol) ; ii) altered patterns of non-verbal communication due to divided attention; iii) the effects of exposing actions (mouse movements, typing) that are normally only partially visible; and iv) changes in the dynamics of authority and agency between service users and clinicians. Our findings suggest that remote interactions tend to reduce engagement and limit exploration of service users' concerns, amplifying issues previously found in face-to-face health consultations. The process of reaching shared agreement between service users and clinicians is vital to the therapeutic goal.

1 Introduction

The quality of the interactions between clinicians and service users (SU) is critical in mental healthcare where the effectiveness of treatment is especially dependent on the ability of SUs and therapists to build an effective relationship, or "alliance" (Flückiger et al., 2018). These relationships are built up over multiple therapeutic sessions. Some conditions, such as Severe Mental Illnesses (SMI) including, Schizophrenia, bipolar disorder, and severe depression, require life-long support. People with psychosis need continuous treatment, medication, and support, making it one of the most expensive mental health conditions to manage. Psychotic disorders affect 0.7% of the UK population (NICE, 2013) and in 2012, the NHS spent £11.8 billion on treatment alone (Ride et al., 2020).

Remote mental health consultations are seen as a promising way to help address this need. It reduces waiting time, costs, and travel time and is also convenient for people with restricted mobility (Dixon et al., 2016; Hollander & Carr, 2020). In certain settings, such as the one used in this study, remote consultations provide access for rural populations for whom face-to-face (F2F) appointments are especially challenging. Early interventions substantially reduce societal and financial burdens (Addington et al. 2007), while wider practical benefits include reduced environmental impact (Barnett et al., 2021).

Although mental health consultations are conventionally delivered F2F, the COVID-19 pandemic forced a sudden switch to remote delivery via telephone or video call. This rapid introduction provided a platform for longer-term adoption of remote consultation. In just three years, the use of online (video) GP appointments has increased by almost 12-fold (NHS, 2025a). Over 38,000 mental health therapy sessions now take place online (NHS, 2025b). Although the implementation of remote services has been ad hoc and highly variable (Barnett et al., 2021).

Importantly, we do not yet understand the impact of switching to remote interaction on the quality of the communication. This paper seeks to understand the impact of remote psychotherapy, through detailed analysis of how remote interactions affect the specific conversational processes through which clinicians and SUs build shared understanding.

Proceedings of the 29th Workshop on the Semantics and Pragmatics of Dialogue, September 3–5, 2025, Bielefeld, Germany.

1.1 Therapeutic Conversations

Building shared understanding in healthcare interactions can be challenging. There are differences in people's expectations, knowledge and experience, in their ways of talking about health, in their understanding of the goals of the interaction, and of how treatments work. There may also be more fundamental differences on what the nature of the problem is, especially for SUs with psychosis who may be suffering from delusions and paranoia (Themistocleous et al., 2010; McCabe & Healey, 2018), potentially leading to additional mistrust of both digital technologies and services.

Conversation analysts have studied how participants in healthcare interactions work to build shared understanding. In this context *reformulations* where one speaker recasts or rephrases another speaker's utterance- play an important role in arriving at an agreed description of events or situations. Davis (1986) examined how therapeutic issues are reformulated in psychotherapy settings. Peräkylä (2019), and Weiste and Peräkylä (2013) have highlighted the ways therapists and clients use reformulations to highlight, rephrase, relocate, and exaggerate each other's descriptions of events and experiences.

Conversation analysis (CA) has also been used to understand the effects of digital interventions in this context. For example, Ekberg et al. (2013) analysed how therapists react to their client's communication of emotion in text-based Cognitive Behavioural Therapy (CBT) sessions. Impala et al. (2023) used CA to explore the ways in which the alliance between clients and their therapists' ruptured and were repaired.

Here we focus on sequences at the end of the therapeutic session where future courses of action are discussed i.e. where the SU and clinician try to agree on what they are going to do before the next session. We focus on these moments because a) the requirement to formulate and explicitly document agreed actions at the end of the consultation provides a useful practical criterion of what constitutes shared understanding in this context b) It involves concurrent co-ordination of the conversation and the video and application software c) these moments expose key differences and possible misinterpretations between participants. This is especially useful in revealing the conversational processes that are used to build shared understanding (Schegloff, 1992; Healey et. al, 2018; Law et. al., 2022).

The Remote DIALOG+ Application

The DIALOG+ application is designed to support routine mental health consultations. The software follows a solution-focused protocol developed by Priebe et. al. (2017) constructed to ensure key aspects of an SU's life circumstances are explored. The tool is designed to encourage inputs by both clinicians and SUs -either adjusting numeric ratings using sliders or adding text to action items and notes. The software aims to promote discussion and joint formulation of 'Action Items' to support the SU's management of their mental health condition by making all actions visible to both participants and by enabling input by either of them.

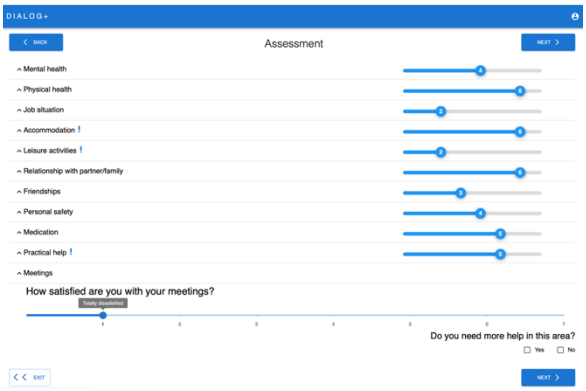

Figure 1: Remote DIALOG+ Software (Rating)

The DIALOG+ protocol (Priebe et al., 2017) consists of 11 items: "Mental Health", "Physical Health", "Job Situation", "Accommodation", "Leisure Activities", "Relationship with Partner/Family", "Friendships", "Personal Safety", ""Medication, "Practical Help" and "Meetings", each with a 7-point Likert scale implemented as a slider so that together they provide a graphical overview of the SU's responses (*Figure 1*). The SU also has the option (a tick box) to ask for more help for each item. The same 11 items are assessed at each session. Current ratings can be compared (graphically) to selected prior sessions to see how an SU's responses may have changed between sessions and over time. Then, up to three items are selected for further discussion using a solution-focussed framing, consisting of questions such as "why this rating and not lower?", "what is the best-case scenario?" and "what can the SU/ clinician/ family do?". The questions are designed to

encourage dialogue between SUs and clinicians. The final step is to agree on 'action(s)' and document them to return to check in on the progress in subsequent meetings. Law et al. (2022) describe the use of DIALOG+ on a shared tablet in face-to-face consultations.

Here we explore how remote delivery impacts patient-clinician interactions, relationship, and outcomes. In particular, how SUs and clinicians manage agreements and disagreements in deciding on action items. The contents of the agreed upon action (i.e. **what** they agree on) are less important than the behavioural 'work' the participants do together (i.e. **how** they both reach agreement) to reach the point of documenting an agreed action using the communicative resources available to them including: utterances, gestures, expressions, posture, the visible actions of typing up text, moving the mouse, moving sliders and any other cues they make use of.

2 Methods

2.1 Participants

Eligible clinicians from the participating Cornwall Partnership NHS Foundation Trust were contacted using purposive sampling on the following criteria: a) They must be an NHS employee regularly delivering psychiatric interventions currently or within last year, b) with at least six months experience in the mental health care field, and c) have access to devices to conduct online consultations sessions.

Participating clinicians were asked to identify SUs under their care who they were meeting regularly with (have an established therapeutic alliance with), who were: a) 18 years or older and had capacity to provide informed consent to participate in the study b) had a clinical diagnosis of psychosis (matching ICD-10 codes F20-F29) and/ or referred to Early Intervention Services (EIS) for psychosis, c) had experience accessing secondary mental health services within the past year, d) able to speak and understand English, and e) have access to at least one device to engage in the online sessions.

The three clinicians who participated in the study were trained by the Remote DIALOG+ research team to use the application, and were asked to provide a recording of their video conference screen (Microsoft Team) while going over the Remote DIALOG+ software contents with their SUs, where consent for video and audio records were provided by the patient (see Ethics section below).

2.2 Conversation Analysis

The videos are analysed using the conversation analytic (CA) approach first developed by Sacks et al. (1974). CA was chosen for its ability to capture subtle social behaviours and its suitability for analysing naturalistic healthcare conversations, in situ (Sidnell, 2012). Jefferson transcription style was used to capture tonal details in addition to timings and overlaps (Fagan, 2017; Park & Hepburn, 2022).

A first-pass transcription of the audio files was generated using WhisperX (Bain et al., 2013), a speech-recognition model which provides a basic transcript of the conversations, timings of utterances, some filler words, and some speaker diarisation. Disfluencies, overlaps, and full diarisation are then checked manually and full Jefferson-style transcriptions (Transcription Key provided in Appendix A) were created for sections where the SU and clinician discussed the action items. Names of people and locations are changed to protect the identity of participants, and any pictures of people included here are blurred.

2.3 Ethics Approval

This paper is a part of a research project investigating the experiences of SUs and clinicians using the Remote DIALOG+ software for remote psychosis interventions. It received approval from the Sheffield Research Ethics Committee, (NHS Health Research Authority) under reference number 22/YH/0194. The video recordings which form the contents of this paper were collected prior to the analysis, for the purpose of evaluating the Remote DIALOG+ software, and both SUs and clinicians were aware of the purpose of the study. The research team members with access to the dataset received Disclosure and Barring Service (DBS) checks to work with confidential and personal data.

3 Results:

Participants consisted of four SUs, three of whom are male. All SUs are in the age range between 18-59, two are from White British ethnic background, while the other two did not answer the demographic questionnaire. A total of four dyads was formed with 3 clinicians. The dataset consists

of 7 screen-recordings of the remote psychiatric consultation, each ranging from 33 to 53 minutes. The sessions took place between April to August 2023.

3.1 Basic Observations

Settings:

The settings of the conversations vary considerably. Most are recorded from a quiet place (SU's house, clinic). In one of the videos, the SU accepted the call inside a parked car. While some SUs used multiple devices to take the video call and see the Remote DIALOG+ software (e.g. phone and laptop/personal computer) some used only their phones. These differences highlight the diversity of situations of use typical of ordinary interactions between SUs and their clinicians in natural settings.

Procedure:

DIALOG+ is explicitly designed to support real-time editing throughout the discussion of scale items and future actions by both participants (Priebe et al., 2017). However, we observed that SUs rarely took the initiative to fill in the ratings or type up their answers themselves. In the present recordings, all but one session was led by the clinicians (c.f. Law *et al.* 2022).

The general sequence of agreeing on an action is observed to be as follows:

1. Clinician prompts the SU for a response to a question they can both see on the Remote DIALOG+ screen.
2. SU initiates a response.
3. Clinician starts to document the answer by typing it up while the SU is speaking.
4. Clinician occasionally suspends typing to seek clarification or elaboration and/ or highlights the SU's response and/ or reformulates the response.
5. SU responds with either agreement/ disagreement or clarification.
6. Clinician finishes documenting the answers and makes a verbal offer that finalises the action item.

In some instances, the clinician acknowledges the SU's responses but does not document them.

3.2 Reformulation

The process of agreeing on an action sometimes involves the clinician pursuing a sequence of reformulations of the SU's initial description (Weiste & Peräkylä, 2013). For example, in *Excerpt 1* there is a sequence of descriptions of how the SU's family could help them with their mental health.

1	C	.hhh an and- wh-what about=umm (.) other people. (.) the wider circle of people. (.) um yer partner family friends,=w-wh-what could they possibly do to- to help you? .hh ((hand stroking chin) (6.9)
2	SU	yeah **comment**. (.) on how i'm doing,
3	C	so↑ some-some **feedback**? ((starts typing)) yeah↑ (9.4)
4	C	and how does that normally help you (.) when you hear back from-from other people ((drinks water)) (1.4)
5	SU	i think they jusst sometimes have a **clearer >pers<pective** of what's going on than I do
6	C	((starts typing)) so thee th- ob- **objective** (.) **perspective** from other people helps ((typing, 5.4)) .hh and is that because its uhhh **encouraging**? (1.9)
7	SU	just **realistic**.
8	C	°**realistic** okay° ((starts typing)) helps provide .h (.) uh a realistic point of view (.) okay↑
9	C	.hh are you happy with what we've done on [that section so far↑
10	SU	[((nods)) yep
11	C	course this type of thing is something that we can return [to cus we're gonna
12	SU	[((blinks, nods))
11	C	do two sessions of this↑ ((hovers mouse over 'add')) so we just add ((clicks 'add')) that↓ for a minute,

Excerpt 1: Dyad 3 session 1, clinician (C) and service user (SU).

The wording in the sequence (bolded for emphasis) moves from the SU's initial "comment" to the clinician's "feedback", to the SU's "clearer perspective" to the clinician's "objective" to the SU's "realistic". There are several things to notice in this sequence. First, the SU does not directly align on the clinician's reformulations but offers their own instead (e.g. "realistic" in response to "objective" and "encouraging"). Second, there are

marked pauses after the clinician's proposed reformulations and the SU's responses. Third, these pauses are not simply lapses in the conversation; the clinician's reformulations are presented as questions directed to the SU, they are combined with gaze to the SU's video window as a *response mobilising* signal and the clinician does not follow up until the SU responds. In contrast, the SU's responses are not formatted as questions and the clinician responds quickly despite being concurrently engaged in typing. While antipsychotic medications can cause speech disturbances, particularly slower speech and increased length of pauses (de Boer et al. 2020), it is notable that these pauses were absent from other reformulations sequences with this SU, suggesting these extensive pauses were not caused by medication but rather by the SU's misalignment with the clinician's formulations.

Overall, the impression is that the clinician is pursuing a particular type of *reformulation/ reinterpretation* which the SU is reluctant to accept. When the clinician then types out the actions (visible character-by-character to the SU) the documented version foregrounds his own formulation but from the first-person perspective of the SU. For example, "Others to provide **feedback** on how they think I'm doing", "The **objective perspective** from others helps provide a **realistic** point of view".

Writing action items formulated from the first-person SU perspective was observed in all dyads except one (*Exerpt 2*), where the clinician exclusively used the action steps section to write out her own tasks to complete before the next session.

1	C	so-ahv put an agreed, actions=>Carmen to contact the surgery to follow up on Dr Martin's< outpatient le'er. re (.) Zopiclone↓ ((types "Zopiclone")) (0.8)
2	SU	ye:ah
3	C	and=antidepressant (2.4)
4	SU	mm hmm, (2.6)
5	C	you happy wit that↑ (0.8)
6	SU	yeah yeah,
7	C	s'there anythin else you want me to, add. as an. (.) action (1.1)
8	SU	no a=dont think [so (1.8)
9	C	[((shakes head))
10	SU	[we can get those two things [done
11	C	[((moves mouse)) [yep-
12	SU	°mumbles°[
13	C	[((clicks 'add'))

Excerpt 2: Dyad 4, session 1

In this excerpt, as the writer (clinician) is the one formulating her own actions ("Carmen to contact…"). While there is no evidence of disagreement in this dyad, the SU was less active in formulating the action items, as shown by the lack of negotiation. The aim of the DIALOG+ application is to encourage patient-centric and solution-focused action items, and reformulations provide a way for SUs to attempt to reshape those action items.

3.3 Sequencing

A recurring issue with structured protocols is managing the tension between the order of business suggested by a protocol and participants' own understanding of how their circumstances relate to each of the questions (c.f. Law et. al. 2022). This can be seen happening at two levels. First, the encoding of categories of activity in the protocol may not align with the personal circumstances of the SU or with what they consider to be the natural connections or *topic ties* that can be made between them. Second, in the practicalities of managing the selection and rating of items. Both disrupt the flow of the conversation.

For example, (*Excerpt 3*) on completing the rating of a section, an SU talks about his improved mood and initially indicates that his answer is probably complete ("I suppose that's it"). However, he immediately initiates an attempt to tie in another topic ("just") but self-repairs and then announces he is pausing the new topic ("wait"). Although the clinician responds that it is "fine" to depart from the protocol in this way, the clinician also simultaneously proceeds to click the button to add the notes to the system (visible to the SU) and moves on to the next question in the protocol.

1	SU	ermmm ((looks up)) (2.6) uhhh yea. i=spose thats it(hh)
2	C	okay, ((typing))
3	SU	just- ((covers mouth)) (2.3) but yeah i think its=similar to the other thing where (.) now theres more (.) ((gestures off-screen)) oh sorry i'll wait til the actual questions ((rubs nose)) i-[

4	C	[no that's fine, ((rest hands under chin and hovers mouse over 'add' button))
5	SU	forgot theres actual questions ((index finger over philtrum)) (2.5)
	C	((presses 'add')) so what is (.) whats the best case scenario

Excerpt 3: Dyad 2 session 1

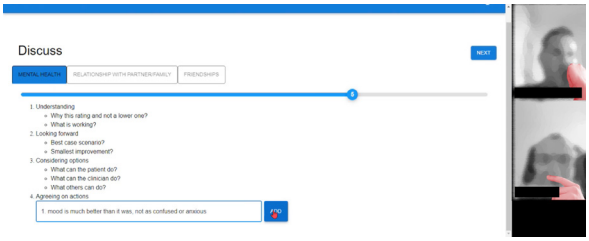

Figure 2: Clinician (top right), SU (bottom right), visible mouse (bottom centre)

This short sequence illustrates several points about the interplay between the video window, application window and the conduct of the interaction. First, there is the interleaving of successive verbal and non-verbal contributions by the clinician and SU. The SU's initial gesture (hand over mouth) and pause appear to orient both to the apparent 'deviation' from the protocol and to co-ordinate with the clinician's ongoing typing. The verbal attempts to tie in the new topic are relatively quiet and are interleaved with gestures. The clinician is attending to the tool and not the video window at this point and does not provide evidence of verbal or non-verbal uptake. The SU explicitly abandons his attempt with "oh sorry" and the announcement that he will wait and, in line 5, the account that he "forgot" the other questions.

In addition to showing their orientation to following the protocol this also illustrates how the clinician's activities take precedence. Although the clinician explicitly verbally sanctions ("that's fine") the SU's attempt to introduce an interjection and suspends typing to rests his jaw on his fist (*Figure 2,* gestures highlighted in red), suggesting he is providing an opportunity for the SU to take up the turn to talk the SU does not speak. The visible movements of the cursor or text input in the application window provide cues to the clinician's current and projected course of activity for the researchers. In normal circumstances (i.e. without screen sharing) SUs are unable to see this in the interface.

It seems likely that one of the issues here is that the clinicians, and to a lesser extent the SU, are contending with multiple tasks –typing, listening and looking- and that this involves trade-offs of attention that may lead to, for example, the failure of uptake seen above. Substantial time is spent by the clinicians during the session typing up the action items. The clinician's actions in the interface are sometimes inconsistent with the verbal trajectory of the conversation (e.g. mouse moving while saying "that's fine") and the clinician's visible actions take priority over the SU. A recurring observation across the recorded sessions is that the non-verbal signals in the video windows are less effective when screens are split between the video windows and the shared tool.

3.4 Disengagement

During some sessions, examples of disengagement by the clinician could be found (*Excerpt 4*).

1	C	wot=ould be the best case scenario w-for work tha-((stutters)) in that term
2	SU	right >okay okay< ((nods)) In that case yeah- um (.) i'd like a- a fulltime (.) ((gestures)) umm (.) [permanent job inn=
3	C	[((starts typing))
4	SU	=(.) a (.) specialism of IT that i'm happy with=so (.) obviously ((looks up, gestures palm in circle)) after 12 months im hopin od- uhh (.) have enough to (.) be applyin for (.) um (.) ((looks at screen)) jobs that I'm really (.) yeah. yeah ((nods, raises hand towards screen)) (.) umm so that. would obviously then mean it would be my sole (.) ((gestures)) popur-pu-purpose of a job to (.) focus on, ((clears throat)) uh=its (.) [perhaps with [*company*] I don't=
5	C	[((glance right, scratch nose))
6	S	=know=umm (.) but (.) yeah (.) [so (.)
7	C	[((mouse moves off-screen))
8	S	ah that im happy with [((looking at screen))
9	C	[((mouse moves back into DIALOG+ window))
10	S	andd (.)hopefully a nice li'ul [<pay rise↓=
11	C	[((starts typing))

12	S	=as well,> hhh [hahhahah
13	C	[hahah ((laughs while looking down typing))
14	S	erm[(.) uh and hopefully yeah. (.)u[mm
15	C	[((smile drops))
	C	[yea h cool okai (.) so (.) let me see so then lets ((starts typing)) put SI for smallest improvement

Excerpt 4: Dyad 1, session 1

The absence of clear verbal and non-verbal feedback from the clinician in this sequence is striking. In addition to the absence of explicit verbal feedback, appropriately timed pragmatic facial expressions e.g. nods and raised eyebrows are also lacking. The only clear evidence of uptake by the clinician here is the delayed reciprocated laughter on line 13 which occurs only after the SU adds "as well", not at the mention of the "pay rise".

Moreover, the clinician produces a number of non-verbal signals that are usually avoided during F2F conversations including scratching, yawning, nose picking, and looking around. These signals create an impression of disengagement and the SU's responses are consistent with this; characterised by frequent gestures, filled pauses, repetitions, cutoffs, restarts and self-repairs (Bavelas et al., 2000). Ruusuvuori (2001) demonstrated similar behaviour from SUs who would become more disfluent (such as pausing) when clinicians read from or write on patient records during a session. Heath (1984) argues these behaviours can also be a method to re-capture attention.

The clinician's visible mouse moments and typing also provide no clear evidence that he is responding to the SU's explanation. While some cases of SU disengagement occurred throughout some dyads, it seemed less consequential for the conversation because clinicians were the ones leading the conversations in most sessions.

3.5 Authority, Exposure, and Agency

As one SU puts it, "It's so different when you're put in the hot seat" (*Excerpt 5*). In one session where the clinician could not get the DIALOG+ software working, the SU agreed to take on the role of facilitating the screen sharing and recording (i.e. typing) the agreed actions. However, this led to additional 'work' (typing, checking screens) on behalf of the SU that led to the clinician taking the supporting role in this conversation. The clinician shows some marked verbal uptake in response to the SU noting down actions. The clinician seems to be more animated in terms of facial expressions, quicker to respond, and displaying 'appropriate' laughter, smiles, and responses, suggesting possible formulations when the SU indicates problems with talking or typing.

1	SU	ahh °t-th-tuu° andd what shall i put- i d- [((laughs)) [um
2	C	[glance up [smirks
3	SU	its so different when you're put in the hot seat and you've gotta[
4	C	[((laughs))
5	SU	eh-cuz you gotta type things up its like wow↑ um (.) because it's the way you want to put it ((finger gestures list)) versus the way you're gonna ((points to screen)) write it because obviously it's just a note[
6	C	[yea, ((nod))
7	SU	umm↑ (.) umm,
8	C	i su'pose you could also put[(.) =
9	SU	[((move hands to hover over keyboard))
10	C	=with regards (.) umm sort of the work side of things,[
11	SU	[((starts typing "with regards work"))

Excerpt 5: Dyad 1 Session 2

In this example, the SU adopts the wording of the clinician's suggestion without any negotiation (lines 8-11) and starts typing before the clinician has finished speaking. The video recording also shows the SU turning their body to hover over the keyboard as the clinician is still preparing to speak signalling their readiness to type as soon as the clinician speaks (*Figure 3*).

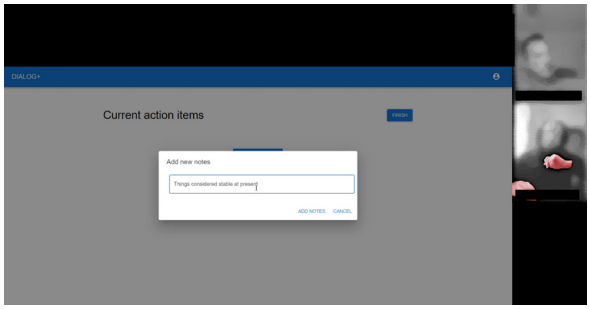

Figure 3: SU (bottom right) moving hands from resting to hover over keyboard.

This highlights how, even though the person typing has practical control of the tool, they don't necessarily have the authority to decide what is being typed. Typing 'exposes' the typer's interpretation and (re)formulation of the discussion to the other participant, in some participants this led to awkwardness. This may partially explain SUs' reluctance to take control in co-writing the action items, in addition to being unfamiliar to the software.

This example illustrates how, within the dynamics of SU-clinician relationships, clinicians have more authority to decide how actions are formulated. In the case of *Excerpt 2* and *5*, the writers themselves are the ones who are going to accomplish the tasks they set out (e.g. SU writes, SU acts upon). However, in all other cases, the SUs are the ones who are committing to carrying out an action formulated and written by the clinician, leading to negotiations and reformulations as seen in *Excerpt 1*, exposing a sense of implicit resistance. It is also reflected in examples of explicit downgrading of agency, on part of the SU e.g., *Excerpt 3*.

This differences in authority and agency afforded to each participant plays a part in the (im)balance of the interaction. Although this tension also exists outside of remote health consultations (Labrie & Schulz, 2015), the effects appear to be intensified in this context because when certain cues are not recognised due to the shared attention between windows, this leads to the disruption of the ability to identify and appropriately respond to the other speaker, hampering the achievement of the conversational goal, in this case – agreeing upon action items together.

Discussion

Meaningful interactions are critical to the success of the relationship between clinician and SU. It has marked effects on outcome, adherence, and compliance, especially in psychotherapy. These excerpts from a small corpus of mental healthcare consultations illustrate some specific challenges for meaningful interaction during remote communication.

The evidence presented here from the details of the verbal and non-verbal communication between clinicians and SUs, visibility of mouse movements and records and timings of action item notes suggests that remote interaction substantially alter the processes through which people build shared understanding. While there isn't a clear definition of what a "good" therapeutic session is, if we step back to ask the question – what does a meaningful interaction look like? Our observations suggest that the process of reaching agreement forms a part of that interaction, as it is a product of the shared understanding and suitable interpretation/ uptake of cues. These findings add to the wider discussion of how agents use cues to build towards shared understanding and reach an agreement within the study of institutional talk in healthcare settings (Heritage & Clayman, 2010).

A key design goal of remote DIALOG+ is to facilitate more equal interaction between SUs and clinicians by supporting synchronous 'negotiation' of actions by both participants. However, in the data collected here, there is no evidence that SUs made use of this feature, only one participant was the active user of the system during a session and they, in effect, acted as note taker rather than note formulator. We observe that agreement alone does not translate to active patient participation in the process of negotiating action steps, such as what can be seen the case of *Excerpt 4* where the SU lets go of his attempt at initiating a continuation of the discussion and *Excerpt 3* where no reformulations were made. The dynamics between the SU and clinician ultimately shapes the goal of interaction and how they reach an agreement during the psychiatric consultations, and the clinician's opinions are sometimes accepted without negotiation (*Excerpt 5*).

Some issues, such as the way a therapeutic protocol interferes with the natural flow of conversation, are also encountered in face-to-face conversations (e.g. Law et. al. 2022; NHS Improvement, 2018). However, the influence of the protocol's structure appears to be amplified in the remote environment, as the shared screen displays the sequence and adds auxiliary cues (including people's concurrent actions in the interface) which reinforce the procedural relevance of the protocol and project for future actions. In face-to-face sessions interface actions are only partially visible to the SU e.g., they can see the clinician is doing something but not what it specifically relates to (Law et. al. 2022). In one sense the process in the remote interaction is more

transparent to SUs and clinicians, but also exposes people's actions and projected actions in ways that can cause difficulties.

In several examples, participant's default focus of visual attention is not on each other but on the current item and their progress through the sequence. Our evidence suggests that this is not only because of the general attenuation of non-verbal signals commonly seen in video conferencing (c.f. Heath and Luff, 1992) but also because the set-up of the remote consultations divides people's attention between windows, sometimes across different devices. There are also distractions from the concurrent use of non-shared applications e.g. email or interactions with other objects or people in the room.

Another recurring feature of these remote consultations are the reduced effectiveness of gestures and facial expressions. Although the inset video windows provide a channel for the use of non-verbal communication this appears much less effective than it would be F2F. Participants still produce gestures and facial expressions, especially when evidence of uptake is weak or absent, but these signals appear much less effective than in F2F interactions. It seems that with the additional task of typing and working with the software, the clinician is sometimes unable to recognise, respond, and engage with the SU effectively.

The strength of conversation analysis is its ability to identify the specific interactional processes that are used to achieve agreement. Automated analysis tools, such as the transcriptions provided by LLMs are not currently capable of detecting these signals. Computer assisted qualitative data analysis software (CAQDAS) have been used since the 1990s (Jenkins et al., 2023). WhisperX, used here, generates a useful first-pass transcription but disfluencies, overlaps, and diarisation are not fully captured. Deep learning programmes such as MediaPipe Holistic can be used to aid analysis of non-verbal signals e.g., to capture the head tilt and eyebrow movements (Kuznetsova & Kimmelman, 2024).

However, the state-of-the-art is still unable to capture the interactional significance of each laugh, eyebrow furrow, and glance. Their interpretation depends on the ability to analyse the interaction turn-by-turn including concurrent non-verbal feedback and sequential context of response and acknowledgement of all participants.

This leads to broader questions surrounding the use of AI and LLMs in healthcare consultations. In our case study of remote therapy sessions with SUs with psychosis, the process of negotiation and forming an agreement of action outcomes forms a part of the therapy itself – in that the SU is able to share their opinions, listen to an alternative viewpoint, and work to form an action plan to look forward in the future. Here, the way clinicians shapes the SUs' process in framing their issues as a problem that could be acted upon and resolved are key to its success. DIALOG+ is designed to encourage this through the process of typing up the actions and the text being visible to both participants, as they manage this tension and reach a common ground.

Say that the remote session takes place as usual in a remote video conference platform, then they use the automatically generated summary of the session to keep a record of their discussion- What would it generate? For example, not every part of the discussion ends up being typed up. The current state-of-the-art in generative AI can neither participate in nor summarise the details of these exchanges. It can, however, produce an overall summary of the finally agreed actions. However, if the purpose is to build a meaningful interaction with another person across the screen, then the process of reaching agreement itself is what forms the goal of the therapy session.

As with other forms of remote psychiatric consultations, Remote DIALOG+, is not intended to replace in-person care. However, it can provide an alternative where traditional modes of service are inaccessible. The excerpts presented in this paper show that workarounds are needed in order to overcome the limitations of remote communication. This has implications for the acceptability and adoption of remote psychiatric services. Future designs of such software should consider the issue of cognitive load (typing and co-ordinating between screens), auxiliary clues, and relationship dynamics (authority and agency) which influences the conversation between participants; perhaps by having a built-in video feed into the protocol screen, or by having the SUs themselves leading the protocol to diminish differences in authority between the participants.

Limitations

This study uses excerpts from real-life conversations between English-speaking SUs with

psychosis and their clinicians during remote consultations using the Remote DIALOG+ software, based on a small sample of clinicians and their SUs in rural Cornwall. Therefore, generalising this evidence in other settings (i.e. across other populations, other health conditions, or other teleconsultation softwares) should be made with caution. Further examples in broader settings or in larger samples could help explore the ways in which people (in general) reach shared agreement in remote psychiatric consultations. No claims were made about how the participants' sex, age, geographic locations, and ethnic background could influence their interactions with their clinicians, nor on achieving desirable treatment outcomes.

While clinicians were trained to use the software by the DIALOG+ research team, the SUs were not. Moreover, some SUs have more experience with video conference systems than others due to a variety of factors (such as age, occupation, and personal experience). This may have limited SUs' confidence in using the software during the recorded conversations, which could reduce their agency and intention to lead conversations and note-taking within the protocol, or aggravate the asymmetrical authority shared between the SU and their clinicians that is noted in this study.

Although the conversations are ostensibly between two people, participants were aware that the recordings would be analysed by the research team which could introduce participation bias, and it is not possible to definitely rule out the presence of overhearers, such as family or partners during the calls which could alter the SUs' behaviours within the sessions.

The findings were based on analysing only what can be 'observed' through these set of remote psychiatric sessions, hence, inferring about external factors were avoided, such as the participants' backgrounds, symptoms, past experiences with remote telepsychiatry, and societal or institutional influences. Future research could utilise other non-observational or quantitative methods to study other aspects of this topic not covered by conversation analysis (or indeed a mix-methods study), such as content or text analysis to reveal larger-scale patterns of agreements/ disagreements across consultation sessions (Angus et al., 2012; Birkett et al., 2017; Labrie & Schulz, 2015).

Acknowledgements

This work was supported by NIHR through the project "Remote delivery of an app-based intervention (DIALOG+) in community mental health care development" NIHR 201680 E72268. NJ is funded by the Wellcome Trust PhD programme - health data in practice: human-centred science (218584/Z/19/Z).

References

Addington, D., Addington, J., & Patten, S. (2007). Relapse rates in an early psychosis treatment service. Acta Psychiatrica Scandinavica, 115(2), 126–131. https://doi.org/10.1111/j.1600-0447.2006.00879.x

Angus, D., Watson, B., Smith, A., Gallois, C., & Wiles, J. (2012). Visualising Conversation Structure across Time: Insights into Effective Doctor-Patient Consultations. PLOS ONE, 7(6), e38014. https://doi.org/10.1371/journal.pone.0038014

Bain, M., Huh, J., Han, T., & Zisserman, A. (2023). WhisperX: Time-Accurate Speech Transcription of Long-Form Audio (No. arXiv:2303.00747). arXiv. https://doi.org/10.48550/arXiv.2303.00747

Barnett, P., Goulding, L., Casetta, C., Jordan, H., Sheridan-Rains, L., Steare, T., Williams, J., Wood, L., Gaughran, F., & Johnson, S. (2021). Implementation of Telemental Health Services Before COVID-19: Rapid Umbrella Review of Systematic Reviews. Journal of Medical Internet Research, 23(7), e26492. https://doi.org/10.2196/26492

Bavelas, J. B., Coates, L., & Johnson, T. (2000). Listeners as co-narrators. Journal of Personality and Social Psychology, 79(6), 941–952. https://doi.org/10.1037//0022-3514.79.6.941

Birkett, C., Arandjelović, O., & Humphris, G. (2017). Towards objective and reproducible study of patient-doctor interaction: Automatic text analysis based VR-CoDES annotation of consultation transcripts. 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2638–2641. https://doi.org/10.1109/EMBC.2017.8037399

Davis, K. (1986). The process of problem (re)formulation in psychotherapy. Sociology of Health & Illness, 8(1), 44–74. https://doi.org/10.1111/1467-9566.ep11346469

de Boer, J. N., Voppel, A. E., Brederoo, S. G., Wijnen, F. N. K., & Sommer, I. E. C. (2020). Language disturbances in schizophrenia: The relation with

antipsychotic medication. Npj Schizophrenia, 6(1), 1–9. https://doi.org/10.1038/s41537-020-00114-3

Dixon, L. B., Holoshitz, Y., & Nossel, I. (2016). Treatment engagement of individuals experiencing mental illness: Review and update. World Psychiatry, 15(1), 13–20. https://doi.org/10.1002/wps.20306

Ekberg, S. J., Barnes, R. K., Kessler, D. S., Malpass, A., & Shaw, A. R. G. (2013). Managing the therapeutic relationship in online cognitive behavioural therapy for depression: Therapists' treatment of clients' contributions. Language@Internet, 10(4).

Fagan, J. (2017, February 13). Jefferson Transcription System – A guide to the symbols. University Transcription Services. https://universitytranscriptions.co.uk/jefferson-transcription-system-a-guide-to-the-symbols/

Flückiger, C., Del Re, A. C., Wampold, B. E., & Horvath, A. O. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. Psychotherapy, 55(4), 316–340. https://doi.org/10.1037/pst0000172

Healey, P. G. T., Mills, G. J., Eshghi, A., & Howes, C. (2018). Running Repairs: Coordinating Meaning in Dialogue. Topics in Cognitive Science, 10(2), 367–388. https://doi.org/10.1111/tops.12336

Heath, C., 1984. Participation in the medical consultation: the co-ordination of verbal and nonverbal behaviour between the doctor and patient. Sociol. Health Illn. 6, 311–388. https://doi.org/10.1111/1467-9566.ep10491964

Heath, C., & and Luff, P. (1992). Media Space and Communicative Asymmetries: Preliminary Observations of Video-Mediated Interaction. Human–Computer Interaction, 7(3), 315–346. https://doi.org/10.1207/s15327051hci0703_3

Heritage, J., Clayman, S., 2010. Dimensions of Institutional Talk, in: Talk in Action. John Wiley & Sons, Ltd, pp. 34–50. https://doi.org/10.1002/9781444318135.ch4

Hollander, J. E., & Carr, B. G. (2020). Virtually Perfect? Telemedicine for Covid-19. New England Journal of Medicine, 382(18), 1679–1681. https://doi.org/10.1056/NEJMp2003539

Impala, T., Okamoto, A., & Kazantzis, N. (2023). Alliance rupture and repair in cognitive behavior therapy. In Rupture and repair in psychotherapy: A critical process for change (pp. 119–139). American Psychological Association. https://doi.org/10.1037/0000306-006

Jenkins, N., Monaghan, K., Smith, M., 2023. Did they really say that? An agential realist approach to using computer assisted transcription software in qualitative data analysis. Int. J. Soc. Res. Methodol. 26, 97–109. https://doi.org/10.1080/13645579.2021.1965037

Kuznetsova, A., Kimmelman, V., 2024. Testing MediaPipe Holistic for Linguistic Analysis of Nonmanual Markers in Sign Languages. https://doi.org/10.48550/arXiv.2403.10367

Labrie, N., & Schulz, P. J. (2015). Quantifying Doctors' Argumentation in General Practice Consultation Through Content Analysis: Measurement Development and Preliminary Results. Argumentation, 29(1), 33–55. https://doi.org/10.1007/s10503-014-9331-5

Law, J. H., Healey, P., & Galindo Esparza, R. (2022). Embodied Interaction in Mental Health Consultations: Some Observations on Grounding and Repair. In S. Dobnik, J. Grove, & A. Sayeed (Eds.), Proceedings of the 2022 CLASP Conference on (Dis)embodiment (pp. 51–61). Association for Computational Linguistics. https://aclanthology.org/2022.clasp-1.6/

McCabe, R., & Healey, P. G. T. (2018). Miscommunication in Doctor-Patient Communication. Topics in Cognitive Science, 10(2), 409–424. https://doi.org/10.1111/tops.12337

NHS. (2025a, March 27). Appointments in General Practice, February 2025. NHS England Digital. https://digital.nhs.uk/data-and-information/publications/statistical/appointments-in-general-practice/february-2025

NHS. (2025b, April 10). NHS Talking Therapies Monthly Statistics Including Employment Advisors, Performance February 2025. NHS England Digital. https://digital.nhs.uk/data-and-information/publications/statistical/nhs-talking-therapies-monthly-statistics-including-employment-advisors/performance-february-2025

NHS Improvement. (2018). Spoken communication and patient safety in the NHS. NHS. https://www.england.nhs.uk/wp-content/uploads/2022/03/spoken-communication-and-patient-safety-in-the-nhs-summary-v2.pdf

NICE. (2013). Psychosis and schizophrenia in children and young people: Recognition and management (Clinical Guideline No. CG155). National Institute of Care Excellence. https://www.nice.org.uk/guidance/cg155/resources/

psychosis-and-schizophrenia-in-children-and-young-people-final-scope2

Park, S. H., & Hepburn, A. (2022). The Benefits of a Jeffersonian Transcript. Frontiers in Communication, 7. https://doi.org/10.3389/fcomm.2022.779434

Peräkylä, A. (2019). Conversation Analysis and Psychotherapy: Identifying Transformative Sequences. Research on Language and Social Interaction, 52(3), 257–280. https://doi.org/10.1080/08351813.2019.1631044

Priebe, S., Golden, E., Kingdon, D., Omer, S., Walsh, S., Katevas, K., McCrone, P., Eldridge, S., & McCabe, R. (2017). Developing the DIALOG+ intervention. In Effective patient–clinician interaction to improve treatment outcomes for patients with psychosis: A mixed-methods design. NIHR Journals Library. https://www.ncbi.nlm.nih.gov/books/NBK424443/

Ride, J., Kasteridis, P., Gutacker, N., Aragon Aragon, M. J., & Jacobs, R. (2020). Healthcare Costs for People with Serious Mental Illness in England: An Analysis of Costs Across Primary Care, Hospital Care, and Specialist Mental Healthcare. Applied Health Economics and Health Policy, 18(2), 177–188. https://doi.org/10.1007/s40258-019-00530-2

Ruusuvuori, J., 2001. Looking means listening: coordinating displays of engagement in doctor-patient interaction. Soc. Sci. Med. 1982 52, 1093–1108. https://doi.org/10.1016/s0277-9536(00)00227-6

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. Language, 50(4), 696–735.

Schegloff, E. (1992). Repair After Next Turn: The Last Structurally Provided Defense of Intersubjectivity in Conversation. American Journal of Sociology - AMER J SOCIOL, 97. https://doi.org/10.1086/229903

Seedhouse, P. (2005). Conversation Analysis as Research Methodology. In K. Richards (Ed.), Applying Conversation Analysis (pp. 251–266). Palgrave Macmillan UK. https://doi.org/10.1057/9780230287853_15

Sidnell, J. (2012). Basic Conversation Analytic Methods. In The Handbook of Conversation Analysis (pp. 77–99). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118325001.ch5

Stivers, T., & and Rossano, F. (2010). Mobilizing Response. Research on Language and Social Interaction, 43(1), 3–31. https://doi.org/10.1080/08351810903471258

Themistocleous, M., McCabe, R., Rees, N., Hassan, I., Healey, P. G. T., & Priebe, S. (2010). Establishing mutual understanding in interaction: An analysis of conversational repair in psychiatric consultations. Communication & Medicine, 6(2), 165–176. https://doi.org/10.1558/cam.v6i2.165

Weiste, E., & and Peräkylä, A. (2013). A Comparative Conversation Analytic Study of Formulations in Psychoanalysis and Cognitive Psychotherapy. Research on Language and Social Interaction, 46(4), 299–321. https://doi.org/10.1080/08351813.2013.839093

A Appendices

Symbol	Meaning
[	Overlap between speakers
(.)	Short pauses (<0.5 seconds)
(1.1)	Pauses in seconds (>0.5 seconds)
((word))	Description of actions
word	Stress emphasised on a word/part of a word
word-	Cut-off word/ speech
word=word	No gap between words
word= =word	Continuation of a turn represented in a new line
wo:rd	Lengthened vowel
word.	Slight falling intonation
word,	Slight rising intonation
↓	Distinct falling intonation
↑	Distinct rising intonation
°word°	Mumbling
>word<	Increased speed
<word>	Decreased speed
hhh	Exhale
.hhh	Inhale

Appendix A: Transcription Key