What Do You Think I Think? Theory of Mind and Schizophrenia

Rosemarie McCabe (r.mccabe@qmul.ac.uk)
Unit for Social & Community Psychiatry, Newham Centre for Mental Health
London, E13 8SP UK

Ivan Leudar (i.leudar@manchester.ac.uk)
Department of Psychology, University of Manchester
Manchester, M13 9PL UK

Patrick G.T. Healey (ph@dcs.qmul.ac.uk)
Department of Computer Science, Queen Mary, University of London
London E1 4NS.

Abstract

Having a ‘theory of mind’ (ToM) means that one appreciates one’s own and others’ mental states. Experimental studies have provided evidence that ToM is impaired in schizophrenia. Conversational interaction depends on the ability to make judgements about other participants’ intentions and beliefs. Consequently, a key prediction that follows from the experimental studies is that people with schizophrenia should display specific communicative problems associated with these judgements. We present a detailed analysis of 35 naturally occurring encounters between health professionals and people with schizophrenia. Using conversation analytic techniques, we provide evidence that patients a) show intact first- and second-order ToM in their discussion of their own and others’ mental states b) are able to use third position repair, which depends on diagnosing another participant’s possible misunderstanding and c) retain the ability to plan complex conversational moves on the basis of what their interlocutor knows.

Introduction

Theory of Mind

‘Theory of mind’ (ToM) has been an influential model of social cognition in psychology. Having a well developed ‘theory of mind’ means that one appreciates one’s own, and others’, thoughts, beliefs and mental states in general. Theory of mind was originally formulated to account for strategic social behaviour in primates. It has since been used to explain how children’s behaviour becomes increasingly socially coordinated (e.g., Wimmer & Perner, 1983) and to account for the problems of children with autism (Baron-Cohen, Leslie & Frith, 1985). More recently, Frith (1992), Frith and Corcoran (1996) and Corcoran (2000, 2001) extended the model to explain the symptoms of schizophrenia.

Theory of mind and Schizophrenia

Frith (1992, 1995) has proposed that a single general mechanism may underlie the heterogeneous symptoms characteristic of schizophrenia: defective metarepresentation (or representing of representations). For example, a primary representation such as ‘John is sad’ can be associated with a metarepresentation that encodes knowledge about mental states such as ‘Mary believes ‘John is sad’. In this framework, all mental states – beliefs, pretence, feelings – require metarepresentation (Frith, 1992). Frith’s main thesis, put simplistically, is that the symptoms of schizophrenia arise because the mechanism that enables metarepresentation fails. Hence, for the patient, the primary representation becomes detached from the patient’s knowledge about it so that, for example, other’s intentions are no longer interpreted in the context of one’s knowledge about them but in isolation from it. Frith (1992) provides an example concerning one’s boss’s intentions: ‘My boss wants of me “you must be on time”. If metarepresentation is impaired, this could become “you must be on time” in the form of an auditory hallucination, i.e., voices talking to or about the patient. The importance of this thesis lies in the proposal of a single underlying cognitive framework, i.e., impaired knowledge about mental states (ToM) that would explain the symptoms seen in schizophrenia.

The experimental approach

Most of the evidence concerning ToM problems in schizophrenia comes from experimental studies. This evidence is, however, inconsistent both with respect to whether there is a ToM impairment and whether this is a trait or state characteristic of schizophrenia. Typically, these studies use a false belief task, which centres on the respondent supplying the correct answer to a puzzle. Bloom and German (2000) have discussed the problems associated with the false belief task as a test of ToM. Firstly, the false belief task is a difficult. Subjects have to follow the actions of two characters in a narrative, remember that one character could not have observed an event and remember where an object used to be and where it is now. Hence, it places heavy task demands with respect to memory and attention and, as such, is not a ‘pure’ test of ToM. Secondly, there is more to ToM than passing the false belief test: false belief is just one aspect of the capacity to reason about the mental states of others. ToM encompasses a range of reasoning abilities not captured by the false belief task.
We present an alternative approach to the study of ToM: identifying how ToM is used *in situ* in naturally occurring social interaction. Studies of naturally occurring interaction are notably absent from ToM research but crucial because (1) they provide ecologically valid evidence on how ToM is done in practice, (2) deficits are identified in such routine interactions and (3) the major problems experienced by people with ToM deficits and those in contact with them arise in these interactions.

**Intentionality in interaction**

Arguably, recognition of interlocutor’s intentions is an essential pre-condition for all communication (Grice, 1989; Clark, 1996). However, people also address themselves directly to other people’s understanding during the course of normal social interaction. For example, they may be explicitly engaged with the problem of discovering or changing the intentions and beliefs of others. Individuals who have problems representing their own and others’ intentionality should display specific problems with these aspects of conversation. This paper addresses the following question: Do people with schizophrenia explicitly recognise, monitor and infer other peoples’ mental states in routine conversational interaction?

**Method**

**Data**

Two sets of clinical interactions were analysed: routine psychiatrist-patient consultations and cognitive behaviour therapy sessions (CBT), both in outpatient settings. The corpus consisted of 32 audio-visually recorded psychiatrist-patient consultations (Co) and 3 complete audio-recorded courses of CBT (the number of sessions ranged between seven and fifty two). Altogether, the database comprised over 80 hours of recordings.

**Participants**

All 35 patients met DSM-IV criteria for a diagnosis of schizophrenia. Fifty-seven percent of the sample was male, the age range was 28-66 and they had a mean length of illness of 13.9 years (SD 9.8). There were 10 clinicians: 7 psychiatrists (6 were consultants) and 3 clinical psychologists. For the sake of brevity, we use materials from the first 3 sessions of each of the CBT cases and 7 of the psychiatric consultations.

**Analysis**

Conversation analysis was used to analyse the transcripts. This method has previously been applied to study everyday conversation, children’s acquisition of ToM, talk in aphasia, autism, medical assessment and therapeutic interactions (e.g. McCabe et al., 2002; Peräkylä, 1997). Conversation analysis aims to characterise what people do, it avoids abstract category systems, relying instead on a detailed consideration of particular interactional episodes. This enables us to examine in detail where exactly individuals with schizophrenia encounter their own and other peoples’ intentionality in conversation and how they deal with it. Analysis is always accompanied by the data and can be checked publicly.

Talk was transcribed according to conversation analytic conventions (Sacks, Schegloff & Jefferson, 1974) retaining the characteristics of speech delivery (e.g., pauses, overlap, stress, intonation). For the audiovisual recordings, we also transcribed visual and tactile features of the participants’ conduct in relation to the talk. Simplified transcripts are presented here.

**Findings**

**Conversational sequences displaying intact ToM**

Three kinds of evidence in the recordings and transcripts were identified that would clearly indicate ToM skills:

1. How patients, spontaneously or in response to a query, describe their own and others’ mental states (i.e., beliefs, intentions and feelings) and engage in metacognitive reflections.
2. What patients say to their interlocutor on the basis of what the patient must appreciate about their interlocutor’s state of mind.
3. What patients say to their interlocutor on the basis of the patient having an adequate representation of their interlocutor’s state of mind and how they use this information to plan an elaborate series of conversational turns.

1. Appreciating others’ states of mind

In extract 1, the client reports to the assessor the abuse he suffers from people on account of his mental illness (lines 1-3). The ensuing interaction lays bare the ‘theory of mind’ as it is done in situ. The reported abuse is in itself an intrinsically intentional behaviour and the interviewer indeed takes the report to entail a tacit attribution of knowledge by the client to the abusers (this knowledge being that the client is mentally ill).

(1) CBT ST&IE:1:407

1. C people shout out windows. ah you nutter.
2. Int DO they.
3. C yeh.
4. (.7)
5. Int what do you make of that.
6. (2.4)
7. C =s: get on with it, what's the point? ()
8. Int how do all these people know that you've had er ( ) sort of mental health problem or ( ) whatever.
9. (1.2)
10. C how do they ( .7) how do they know
Detailed conversation analysis demonstrates clear facility with ToM. He represents his own and other people’s beliefs, compares them, grounds them in his experience and warrants them, and appreciates practical consequences of intentional states. Could he nevertheless be an exception? In fact, all of the interactions we examined included sequences where the patient represented coherently and effectively mental states of others. In the rest of this section we provide representative examples of these.

Extract 3 is from a routine psychiatrist-patient consultation. The patient asks the psychiatrist what he thinks about the cause of her symptoms (lines 1-11). The psychiatrist responds by asking the patient “what do you think I think?” (line 12). In asking this particular question, he treats the patient as capable of considering his beliefs about her beliefs. Later in the sequence, the psychiatrist says that he thinks the patient has an illness, and that is what is troubling her (lines 15-17). In line 18, the patient recognises that the doctor knows that she knows that the psychiatrist holds different beliefs about the origin of her symptoms. Even when there is a conflict between the contrasting beliefs related to her psychotic symptoms (i.e., people are after her), she is aware that the psychiatrist has beliefs different from her own.

1. P SO DO Do you think what I'm telling you even
2. when I was working in (place) I asked my
3. supervisor (0.8) b'cos she was dealing with the
4. psychiatry people an (1.0) do th do they exist that
5. there are people that are causing this (0.2) eh
6. sickness (0.6) b'cos I'm fully confide fully satisfied
7. now it's not the medication that makes me with all
8. the symptoms (0.4) it's the( . ) those people that
9. I'm (0.3) that ( . ) ar (after me) that I
10. Dr yeah mhhm
11. P feel sick an everything(.) I blame them (.)
12. Dr yeae (1.2) well what do you think I think?
13. ((smiling))(0.2)
14. P mhm? (0.8)
15. Dr well I th I think you have an illness that's fairly
16. well under control at the moment (0.1) but th eh
17. an that's what's troubling you (0.2) buht (0.8)
18. P it's not the people you think? Ha-ha-ha-ha (1.2)
19. Dr that's not mhy opinion
(P = patient; Dr. = psychiatrist)

In extract 4 (a pre-CBT assessment) the patient is telling the interviewer about his weekly income and accounting for his expenditure. (He gives half of his income to his mother and with the other half he buys alcohol and ‘chocolate’, i.e., cannabis). This is straightforward informing: telling somebody something that s/he does not know. Telling the psychologist that he buys cannabis is, however, less straightforward. He acknowledges that it might be a

---

What interests the interviewer is how the abusers know that he, the client, has mental health problems. This means that the interviewer presupposes that the client is able to represent knowledge that others have of him - that is why he can ask how come 'these people know'? (lines 8-10). This question reflects the fact that in general, it is not enough to make attributions of mental states, one may have to justify them (e.g. Williams, 2001). And of course the assessment of the attributions in a mental state interview as a justified belief and not a delusion will depend on providing a believable account. And C indeed does provide one (lines 14-16). The knowledge of his psychiatric problems was probably spread by the builders who had access to his home when they were working there. Note that providing this account involves C being reflexive about his own knowledge - he believes that they know for particular reasons. So, the client in (1) is able to represent the knowledge others have of him and can justify these attributions. The interviewer presumes that this is so and draws on this ability of his client.

An important aspect of the ToM account is that the individual ought to be able to de-couple representations from what is represented (Leslie, 1987). For example, the individual with a fully developed ToM will realise that different people may have different beliefs about the same object. In extract 2, client C clearly has this facility when he compares different people's beliefs about himself.

---

C trusts his parents, but not his younger sister (line 1). The reason he gives for this compares their respective assessments of his problem (lines 6-7). The parents 'know' that his problems are not his fault, his sister thinks 'he is putting it on'. C is then sensitive to social distribution of knowledge, and his position in it - he clearly knows what his parents know, and rejects what his sister thinks as false. In terms of ToM, in providing this comparative account, C distinguishes between knowing and thinking (line 6), as well as between the mental states of different people with regard to the same object.
problematic thing to reveal to her (line 8). In saying "you're not supposed to know about that", he takes account of the interviewer's position in relation to this statement and distinguishes what others should and should not know.

(4) CBT ST&IE:1:752
1. C I usually get me benefit (0.5) give me mum hundred quid
2. Int mhm (.4)
3. C fifty for the support- fifty quid a week, () and the other fifty quid I usually (0.4) get alcohol with (.) a'n I 'ave got me (.)
4. Int mhm
5. C chocolate (.) yeR not supposed to know about.

The extracts so far demonstrate that these individuals can spontaneously and successfully express beliefs about present and absent others' states of mind, as well as about their own. Additional, more complex, evidence of ToM can also be identified in the form of sophisticated conversational moves which involve:

(1) planning and developing a conversational sequence, which can only succeed if one appropriately anticipates the other participant's mental state and likely response to a particular utterance, i.e., appreciation of sequential expectancy.

(2) diagnosing another participants misunderstanding of of the patient's own talk and explicitly addressing this, i.e., third position repair and

2. Conversational devices requiring appreciation of sequential expectancy established by an interlocutor

Further evidence would be conversational moves which display not only an understanding of what the interlocutor has just said (which could consist in routine, literal understanding of language) but moves which display planning of the sort which it would be implausible to suppose could be matters of rote learning. A useful account is provided by Drew (1995) for what he, following Goody (1995), calls 'anticipatory interational planning'. At least some conversational sequences visibly require (and display) knowledge of others' states of mind: those are the sequences in which participants set up (in CA terms, project), a certain, normative, bi-lateral development of the interaction. In other words, sequences which start when the action of A makes sense only as a precursor to a future action, the delivery of which is contingent on an expectation of B's contribution. A good example of such a sequence, following Drew, is the use of a preliminary utterance that 'sets the scene' for a subsequent action. For example, I might ask 'do you know Jane?' above. In extract 6, the patient uses a question format as the first move in a sequence, which is later revealed as setting the scene for an account about what happened with another doctor.

(6) Co27:21
1. P .hhh so I thought it might get better =
2. Dr <yeah> () it often does =
3. P you know () and it <didn't> () and I went to see Dr (Name) <I don't know if you got a letter from him>
4. Dr no
5. P I went to see him () he then explained the problems of feeling terrible it <affected my chest as well as my heart kept speeding up

In lines 4, 5 the patient asks 'I don't know if you got a letter from him?'. This is in question format, but it also has the conversational effect of introducing the business that the patient transacted with Dr Smith. Thus, when the interlocutor says 'no' in line 6, he confirms that he did not receive the letter, and licenses the patient to explain what his business with Dr Smith was ('I explained the problems of feeling terrible' and so on). In other words, the patient has successfully initiated a sequence of moves that allows him successfully to recruit his interlocutor into hearing a 'news report' about what had happened with a third party. To do so, the patient had to be able to project not only his own future move (making the report) and to launch it by 'setting the scene' in that oblique manner, but also had to gauge his interlocutor's likely response so that the project would come off.

In extract 7, it is the interviewer who initiates a trajectory, and it is the patient who has to 'decode' the first move and come up with an appropriate response. Just before the start of this extract, the psychologist asks the patient how he coped with visions. The patient replies that he would drink 'enough' cans of beer to make them disappear.
1. Int and how many would have to be enough
2. (1.2)
3. C three cans of special brew?
4. Int right that's quite a lot isn't it really cos it's
5. strong
6. C yeh
7. Int strong stuff
8. (1.0)
9. C I tried it with other beers but 't special brew or
10. the stronger ones seemed to get rid of it
11. quicker.
12. Int were you drinking before you got the visions

At line 3 the patient could reasonably think that he has given a
full answer to the question 'three cans of special brew'. He
might expect the psychologist to acknowledge the answer and
move on. But the psychologist calls attention to the strength
of the drink. It is at this point that we see the patient display
a fully competent appreciation of a planned interactional
sequence.

To explain the sequence it might be helpful to use an
illustration from a different mundane interaction. Consider
this example from Schegloff (1988, transcription simplified).
Carol has returned from the shops:
1. S Hi Carol. =
2. C [i::
3. R [CA:ROI, HI::
4. S You didn't get an ice-cream sandwich,
5. C I knew, hh I decided that my body
6. didn't need it,
7. S Yes but ours di:d=

At line 4, S 'notices' that Carol has not brought back
something. That is not understood as a mere factual statement.
On the contrary, Carol takes it as a call to account for a failure
on her part to bring back an 'ice-cream sandwich'. In other
words, Carol is 'seeing beyond' the literal meaning of S's
statement.

Returning to extract 7, S's noticing is like the psychologist's
noticing, at line 4, that special brew is 'strong'. The patient is
put in Carol's position. He has to understand the
psychologist's 'observation' as a call to account for drinking
that much strong beer. At first (line 6) he merely says 'yeh'.
The psychologist then says 'strong stuff' (line 7), hearable as a
reiteration of the implied request for an account.

This is echoed in the exchange between Carol and her friends.
Carol's account that her 'body didn't need it' is met with 'Yes
but ours did' (line 7). Like the psychologist's 'strong stuff',
this keeps up the pressure for an acceptable account. At this
point, the patient says "I tried it with other beers but 't special
brew or the stronger ones seemed to get rid of it quicker".
This makes sense as an account which meets the
psychologist's repeated, but implicit, queries. Like Carol, the
patient has to 'go beyond' the literal meaning of his
interlocutor's remarks to understand what kind of response
s/he expects. In other words, the patient has successfully
'decoded' his interlocutor's unstated intentions. He has
understood that the psychologist was using the format of a
simple 'observation' or a 'noticing' in a way planned to elicit
an account for a questionable answer.

3. Third position repair

Third position repair is used by speakers to address
misunderstanding of a previous utterance. For example
speaker A speaks (1st position), speaker B responds (2nd
position), and speaker A recognises through B's utterance that
B misunderstood A's original utterance. Hence A speaks
again (in 3rd position) to repair or correct B's
misunderstanding. Instances of third position repair in our
data would be strong evidence that a patient is not relying on
routinised interactional moves but can understand that their
interlocutor has misunderstood a particular utterance and
respond instantaneously to correct that misunderstanding.
Moreover, it shows an ability to diagnose the nature of the
misunderstanding. The following extract is an example of
such an instance.

In extract 5, the patient and psychiatrist have been discussing
the patient's financial situation and the patient has informed
the psychiatrist and he and his wife get a particular allowance
which decreases with age and is disallowed once one reaches
the age of sixty-five:

(5) Co33:101
1. P we get attendance allowance (Name) and I and
2. then we lose quite a lot of money
3. Dr why?
4. P because there is a rule that the older you get, at
5. sixty-five you lose
6. Dr when are you sixty-five?
7. P oh, it's not me, I'm thinking of (Name) she will be
8. in a few years time

In line 6, the psychiatrist asks the patient when he is sixty-
five. However, the patient's concern is in relation to his wife
who will be sixty-five before him. In line 7, the patient
displays that he appreciates the psychiatrist's mental state: the
psychiatrist has identified the wrong referent, the patient
notices this and puts it right at the first interactionally
appropriate opportunity.

Discussion

Thirty-five clinical interactions involving patients with
schizophrenia were analysed to identify whether patients were
impaired in their ability to appreciate that other people have
minds, and what the contents of those minds might be.
Detailed conversation analysis of clinical interactions
indicated that they were able to, spontaneously or in reply to a
question, reflect on their own beliefs, contrast beliefs of different people (including their own versus others), position themselves in socially distributed cognitions, and justify attributions about their own and others’ mental states. They could do third position repair and bring off their part in complex conversational sequences that hinged on an appreciation of their interlocutor’s beliefs. Moreover, clinicians clearly treated the patients as intentional and able to reflect on complex representations of mental states.

However, their routine interaction in clinical settings, even in interviews with psychiatrists, who in reviewing a patient’s mental state actively attempt to elicit unusual ways of thinking, manifest abundantly that patients can indeed appropriately and plausibly appreciate another’s state of mind with varying degrees of complexity.

It is the case that there were many examples of bizarre beliefs and anomalous experiences. The patients claimed that they were God, that they were combating the devil, that rats and spiders were visible in their house, and so on. Elsewhere, we have also examined conversational sequences in which patients are actively psychotic for evidence of ToM problems (McCabe, Leudar & Antaki, 2004). The sequences consisted of ‘symptomatic’ talk about patients’ delusional beliefs and were marked by disagreement between the patient and psychiatrist about these beliefs. However, they did not stem from theory of mind deficits. When talking about these delusional beliefs, patients used ToM to highlight their own beliefs as different from others and seek information about why this was so in an attempt to reconcile others’ disbelief. Although others disagreed with patients about their beliefs, this did not prompt patients to modify their claims. Even in this context, patients used theory of mind abundantly and appropriately.

This is the first study to analyse ToM in real interactions and so overcomes the limitations of using false belief tasks as a test of theory of mind. It makes clear that a simple ToM deficit account of schizophrenia is inadequate. Actively psychotic patients retain subtle interactional abilities that depend on their ability to reason about their own and other’s mental states.

Acknowledgements

Our thanks to Tony Morrison for allowing us access to recordings of CBT sessions.

References


Notes

1 The first part of annotation identifies the type of interaction; Co = consultation, CBT = cognitive behaviour therapy; the digits specify the session number and the number of the 1st line in the original.