BRAIN
A JOURNAL OF NEUROLOGY

Mentalizing under influence: abnormal dependence on prior expectations in patients with schizophrenia

Valerian Chambon,[1] Elisabeth Pacherie,[2,*] Guillaume Barbalat,[3,4,*] Pierre Jacquet,[1]
Nicolas Franck[1,4] and Chlöé Farrer[5,6]

1 Centre de Neuroscience Cognitive, Université de Lyon, CNRS, France
2 Institut Jean Nicod, EHESS, DEC-ENS, CNRS, Paris, France
3 Laboratoire Langage, Cerveau, Cognition, Université de Lyon, CNRS, France
4 Centre Hospitalier le Vinatier, Lyon, France
5 Centre de Recherche Cerveau et Cognition, Université de Toulouse, UPS-CNRS, Toulouse, France
6 Faculté de Médecine de Rangueil, Toulouse, France

*These authors contributed equally to this work.

Correspondence to: Valerian Chambon,
Centre de Neuroscience Cognitive – UMR 5229 CNRS;
67, bd Pinel, 69675 BRON Cedex,
France
E-mail: valerian.chambon@isc.cnrs.fr

An impaired ability to appreciate other people's mental states is a well-established and stable cognitive deficit in schizophrenia, which might explain some aspects of patients' social dysfunction. Yet, despite a wealth of literature on this topic, the basic mechanisms underlying these impairments are still poorly understood, and their links with the clinical dimensions of schizophrenia remain unclear. The present study aimed to investigate the extent to which patients' impaired ability to appreciate other people's intentions (known as mentalizing) may be accounted for by abnormal interaction between the two types of information that contribute to this ability: (i) the sensory evidence conveyed by movement kinematics; and (ii) the observer's prior expectations. We hypothesized that this is not a generalized impairment, but one confined to certain types of intentions. To test this assumption, we designed four tasks in which participants were required to infer either: (i) basic intentions (i.e. the simple goal of a motor act); (ii) superordinate intentions (i.e. the general goal of a sequence of motor acts); (iii) social basic; or (iv) social superordinate intentions (i.e. simple or general goals achieved within the context of a reciprocal interaction). In each of these tasks, both prior expectations and sensory information were manipulated. We found that patients correctly inferred non-social, basic intentions, but experienced difficulties when inferring non-social superordinate intentions and both basic and superordinate social intentions. These poor performances were associated with two abnormal patterns of interaction between prior expectations and sensory evidence. In the non-social superordinate condition, patients relied heavily on their prior expectations, while disregarding sensory evidence. This pattern of interaction predicted the severity of 'positive' symptoms. Social conditions prompted exactly the opposite pattern of interaction: patients exhibited weaker dependence on prior expectations while relying strongly on sensory evidence, and this predicted the severity of 'negative' symptoms. We suggest both these patterns can be accounted for by a disturbance in the Bayesian inferential mechanism that integrates sensory evidence (conveyed by movement kinematics) into prior beliefs (about others' mental states and attitudes) to produce accurate inferences about other people's intentions.

Received May 24, 2011. Revised August 25, 2011. Accepted September 27, 2011. Advance Access publication November 21, 2011

Keywords: schizophrenia; mentalizing; sensory evidence; prior expectations; Bayesian integration
Abbreviation: TFT = tit-for-tat

Introduction

One of the most disabling clinical features of schizophrenia is poor social functioning, reflecting impairments in interpersonal communication and relationships (see Corcoran, 2001, for review). Many authors have proposed that some aspects of patients' social dysfunction are a consequence of a deficit in mentalizing, defined as the cognitive ability to attribute mental states (such as intentions) to others and explain and predict their behaviour on that basis (Frith, 2004; Harrington *et al*., 2005; Sprong *et al*., 2007). Extensive research over the last two decades has provided robust evidence for the presence of a stable mentalizing impairment in schizophrenia (Sprong *et al*., 2007; Bora *et al*., 2009). However, both the nature and the extent of this impairment remain widely debated, owing to its extreme heterogeneity among clinical subgroups of schizophrenia (McCabe *et al*., 2004; Harrington *et al*., 2005; Bara *et al*., 2011).

It has been suggested that inconsistent results in the literature may be the consequence of the great variety of tasks used, both in terms of stimulus type (verbal versus iconographic) and complexity (Walter *et al*., 2009). Crucially, the heterogeneity of the data could also result from a lack of control over the variable under examination. Indeed, 'intention' is a term embracing various subtypes, the content of which can vary along two main dimensions: the scope and the target (Chambon *et al*., 2011). The 'scope dimension' refers to the complexity of the intended goal, and differentiates 'basic intentions' directed at simple motor goals (e.g. grasping an object) from 'superordinate intentions' directed at somewhat more complex goals (e.g. quenching one's thirst), the achievement of which typically involves the completion of a number of subgoals (e.g. grasping a glass, opening a tap, filling the glass, closing the tap, etc.) (Pacherie, 2000, 2008). On the 'target dimension', 'non-social intentions' directed at an object can be distinguished from 'social intentions' directed at a third party (Blakemore and Frith, 2004; Ciaramidaro *et al*., 2007). The ability to appreciate other people's intentions thus refers to separate processes that could be differentially recruited depending on the scope and/or the target of the intention being considered. As such, one cannot preclude the possibility that patients may show impaired understanding of one particular type of intention while the appreciation of other intention types is spared.

So far, few studies have directly tested patients' abilities to appreciate distinct types of intention within the same experimental settings, or using the same material across conditions. One study found that disorganized patients were impaired at evaluating superordinate intentions but not basic intentions (Zalla *et al*., 2006). Another recent study suggested that patients may not be impaired in appreciating actions directed at inanimate objects, but specifically in inferring intentions achieved within the context of social interaction (Walter *et al*., 2009). Disentangling this confusing array of findings requires investigating patients' mentalizing abilities at a more fine-grained level of functioning. That is, not only by assessing patients' raw performances in intention recognition tasks, but also by further exploring how individuals with schizophrenia deal with the information that usually contributes to such recognition.

Attributing intentions to an observed agent can be described as a Bayesian inference drawing upon two distinctive types of information: (i) the 'sensory evidence' available from the action scene (derived from the agent's movement kinematics); and (ii) the observer's 'prior expectations' about which intention is the most likely cause of what is observed, given past experience (Baker *et al*., 2006, 2009; Griffiths *et al*., 2008). It has been shown that intention inference is contingent upon an adaptive interplay between these two sources of information, with participants tending to rely progressively more on their prior expectations as the reliability of sensory evidence decreases, and vice versa. Crucially, this interaction has also been found to vary according to the 'type' of intention to be inferred, with participant's prior experience gaining priority over perceptual evidence when inferring intentions from within a social context rather than in isolation (Chambon *et al*., 2011).

Building on these previous findings, we hypothesized that patients' heterogeneous mentalizing abilities could be accounted for by an abnormal weighting of these two classes of information (prior knowledge and sensory evidence), which in turn might depend on the specific dimensions (i.e. the scope and target) of the intention being considered. This assumption echoes Fletcher and Frith's (2009) suggestion that both the aberrant perceptions (hallucinations) and beliefs (delusions) of schizophrenia might be caused by an abnormality in the brain's inferencing mechanisms, resulting in a diminished ability to integrate new experiences (e.g. sensory evidence) with stored knowledge based on previous experiences (e.g. prior knowledge; Hemsley, 2005). Critically, disturbance of this (Bayesian) inferential mechanism could be a good predictor of the severity of schizophrenia symptoms. For example, the mentalizing profile of patients with positive symptoms might be characterized by a tendency to give excessive credit to endogenous, self-generated information (e.g. prior expectations of how people are supposed to behave under some circumstances), whereas patients with negative symptoms might display a stimuli-induced mentalizing style that may be accounted for by an exaggerated tendency to focus on directly observable, external information, rather than inner experiences (Frith, 1994; Taylor, 1994).

In the present study, we directly tested the above assumption by assessing patients' understanding of the basic or superordinate intentions of an agent performing an action in either isolation, or within the context of social reciprocation. Both sensory and prior information were manipulated by: (i) varying the completeness of action sequences; and (ii) selectively increasing the probability of a particular intention occurring within the sequence, at the expense of competing intention types. We then looked at (i) whether patients' performances on each intention inference task may be accounted for by an abnormal dependence on prior knowledge and/or sensory evidence, and (ii) whether this abnormal dependence—if observed—correlated with the scale for the assessment of positive (Andreasen, 1984), negative (Andreasen, 1983) or disorganization symptoms of schizophrenia.

Materials and methods

Participants

All patients fulfilled DSM-IV criteria of schizophrenia (American Psychiatric Association, 1994) with no other psychiatric diagnosis on DSM-IV Axis I. Exclusion criteria included history of neurological illness or trauma, alcohol or drug dependence according to DSM-IV criteria, analphabetism and being >60 years of age. All patients were receiving antipsychotic medication and were clinically stable at the time of testing. Comparison participants reported no psychiatric problems (Table 1), and were systematically matched with patients for age, handedness (Oldfield, 1971) and years of education (Table 1). All participants reported normal or corrected-to-normal visual acuity. After receiving a complete description of the study, written informed consent was obtained according to the Declaration of Helsinki. This research was approved by the local Ethical Committee (B80631-60) and all participants received 10 euros for taking part.

Four distinct groups of controls ($n = 30$ for each group) and patients ($n = 20$ for each group) performed the four distinct tasks. Individuals with schizophrenia were selected to obtain four groups of patients matched for the severity of negative (scale for the assessment of negative symptoms; Andreasen, 1983), positive (scale for the assessment of positive symptoms; Andreasen, 1984) and disorganization symptoms (Table 1). The disorganization score was computed by summing the following subscores: bizarre behaviour, positive formal thought disorder (from the scale for the assessment of positive symptoms), alogia and inappropriate affect (from the scale for the assessment of negative symptoms). These items have been shown to constitute regular and fundamental components of the disorganization dimension (Hardy-Bayle *et al*., 2003). In the social basic task, one patient was excluded because of poor performance [i.e. >2 standard deviations (SDs)] from the group mean).

Common procedure in the four tasks

In each task, participants were instructed to infer the intention of an actor manipulating non-meaningful objects. The specific contributions of sensory evidence and prior knowledge to the intentional inference were manipulated by varying the amount of visual information (i.e. the completeness of action sequences) and the probability of occurrence associated with each different intention, respectively [see Chambon *et al*. (2011) for detailed descriptions of the video clips used in each task].

Each task consisted of two experimental sessions. First, a baseline session, characterized by a flat (unbiased) probability distribution, in which all intentions had the same probability of occurrence across trials. Secondly, a bias session, in which prior knowledge was manipulated by increasing the probability of one intention (the 'likely' intention, 55% of the trials) to the detriment of the others ('unlikely' intentions, 22% each), resulting in biasing participants towards the likely intention. This bias was randomly assigned so that each intention was equally biased across participants.

Table 1 Clinical and demographic characteristics

Characteristics	Age (years)	Education (years)	Handedness	Duration of illness	SANS score	SAPS score	Disorganization score[a]
Experiment							
Non-social basic							
Healthy ($n = 30$)	35.1 (7.5)	11.9 (2)	0.87 (0.14)				
Patients ($n = 20$)	34 (9.3)	11.1 (1.7)	0.83 (0.16)	10.3 (7.5)	40.5 (15)	31.9 (23.5)	16.1 (12.9)
P-value	0.65	0.12	0.37				
Non-social superordinate							
Healthy ($n = 30$)	36.5 (8.9)	12.1 (1.5)	0.81 (0.17)				
Patients ($n = 20$)	34.6 (8.8)	11.6 (1.8)	0.78 (0.17)	12.3 (8.1)	43.2 (21.6)	29.9 (15.6)	12.9 (5.3)
P-value	0.46	0.26	0.61				
Social basic							
Healthy ($n = 30$)	34.2 (10.5)	11.4 (1.8)	0.82 (0.14)				
Patients ($n = 19$)	35.2 (9)	11.2 (1.7)	0.79 (0.19)	11 (8.4)	44 (24.1)	28.5 (22.3)	14.5 (12.8)
P-value	0.74	0.61	0.57				
Social superordinate							
Healthy ($n = 30$)	35.4 (8.8)	12.3 (1.9)	0.85 (0.13)				
Patients ($n = 20$)	33.8 (10)	11.7	0.81 (0.18)	11.9 (8.6)	44.8 (23.9)	29.4 (15.1)	11.5 (6.9)
P-value	0.56	0.27 (1.4)	0.48				
				all $P > 0.43$	all $P > 0.5$	all $P > 0.64$	all $P > 0.14$

[a] Sum of the scores for bizarre behaviour, positive formal thought disorder from the SAPS, and alogia and inappropriate affect from the SANS.
SANS = Scale for the Assessment for the Negative Symptoms; SAPS = Scale for the Assessment of Positive Symptoms. Data are mean (SD).

The amount of visual information was manipulated by varying the duration of the video clips. Actions were thus either presented with a very high (1880 ms after movement onset), high (1640 ms), moderate (1560 ms), or low (1480 ms) amount of visual information [see Chambon *et al*. (2011) for the selection and control of these amounts].

The baseline and the bias sessions were composed of two types of interleaved blocks: ‘overt’ blocks, in which the actions were shown with a very high amount of visual information (1880 ms) to allow participants to clearly distinguish the different intentions, and ‘covert’ blocks, in which actions were of varying durations (1480, 1560 or 1640 ms) (Fig. 1). The overt blocks were used to bias participants in favour of one particular intention (i.e. the *likely* intention), whereas the covert blocks were used to test the effect of the bias on action sequences shown with varying amounts of visual information.

Each experimental sequence (one overt block followed by one covert block) was repeated nine times over each session. The order of trials was randomized and varied between participants. Furthermore, each clip was presented only once to prevent any influence of memorized kinematic parameters on participants' performances.

All clips were filmed using a digital camera (Sony®- HDR-SR7) and were acquired and tailored using the software Adobe Premiere®. They were presented on a computer monitor (IIYAMA® 19′) at a distance of 60 cm from the participant. Finally, prior to each task, a training session was conducted with distinct clips from those used in the experimental sessions.

Non-social tasks

In both the non-social basic and the non-social superordinate tasks, video clips depicted a single actor manipulating (rotating, lifting or transporting) rectangular cubes. The cubes were of similar size (3 × 6 cm) and orientation, and placed at an equal distance (16.8 cm) from the starting position of the actor's hand (Fig. 2A and B).

Non-social basic task

In the non-social basic task, participants were first required to observe one incomplete manipulation of a single cube (lasting for 1480, 1560 or 1640 ms after movement onset). A response screen representing the first letter of each possible non-social, basic intention (to transport, lift or rotate) then appeared for 2500 ms, during which participants had to press the keyboard button corresponding to the intention inferred (transport, lift or rotate) as quickly and accurately as possible. In the bias session, the non-social basic intention for which the probability of occurrence was increased (i.e. the likely intention) was counterbalanced across participants.

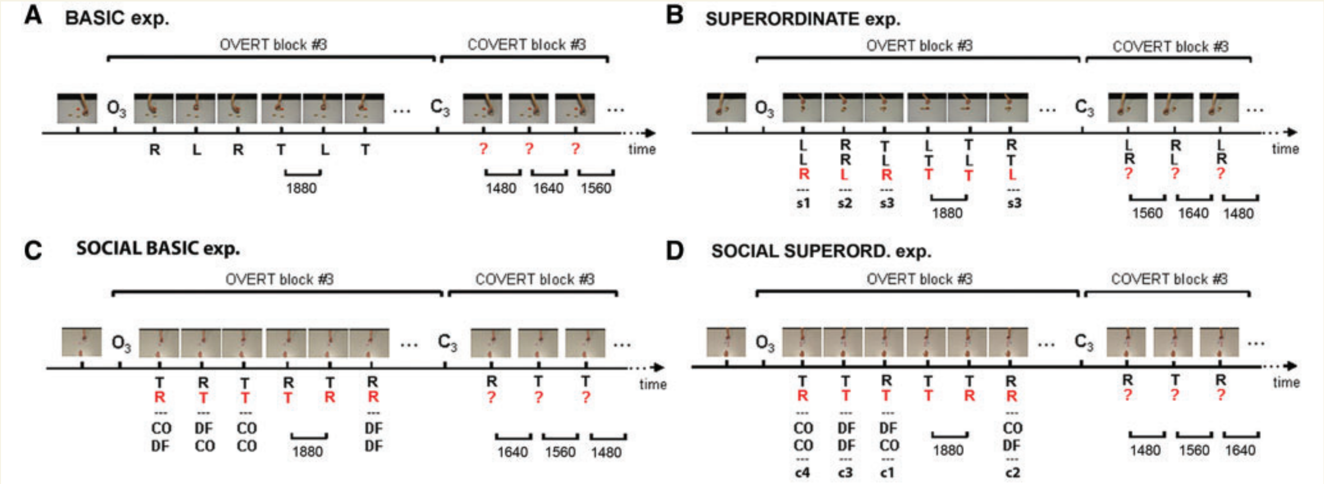


Figure 1 Task design. Examples of the typical experimental sequence (one overt block followed by one covert block) used in both the baseline and the bias sessions. Overt blocks (O): 18 movies with a very high, constant amount of visual information (1880 ms). Covert blocks (C): nine movies with three different amounts of visual information (1480, 1560 and 1640 ms). In the four tasks, the probability of all intentions was held constant across the block, except in the overt blocks of the bias session, where one particular intention had a greater probability of occurring than the others. (**A**) In the basic task, subjects had to identify a single intended action (L = lift; R = rotate; T = transport). (**B**) In the superordinate task, subjects had to identify the final intended action (indicated by a red letter) of an action sequence leading to shapes 1, 2 or 3 (s1 = shape 1, etc.). (**C**) In the social basic task, subjects had to identify the intended action of the second player (red letter). (**D**) In the social superordinate task, subjects had to identify the intended action of the second player (red letter) leading to configurations 1, 2, 3 or 4 (c1 = configuration 1, etc.). In both the social basic and social superordinate tasks, the action or the configuration achieved by each player indicated either a cooperative or a defective strategy (CO = cooperate; DF = defect). In each experiment, a probabilistic bias was assigned to one particular action (basic), shape (superordinate) or strategy (social). The red question mark indicates the action for which the amount of visual information varied (basic: a single action; superordinate: the last action of the sequence).

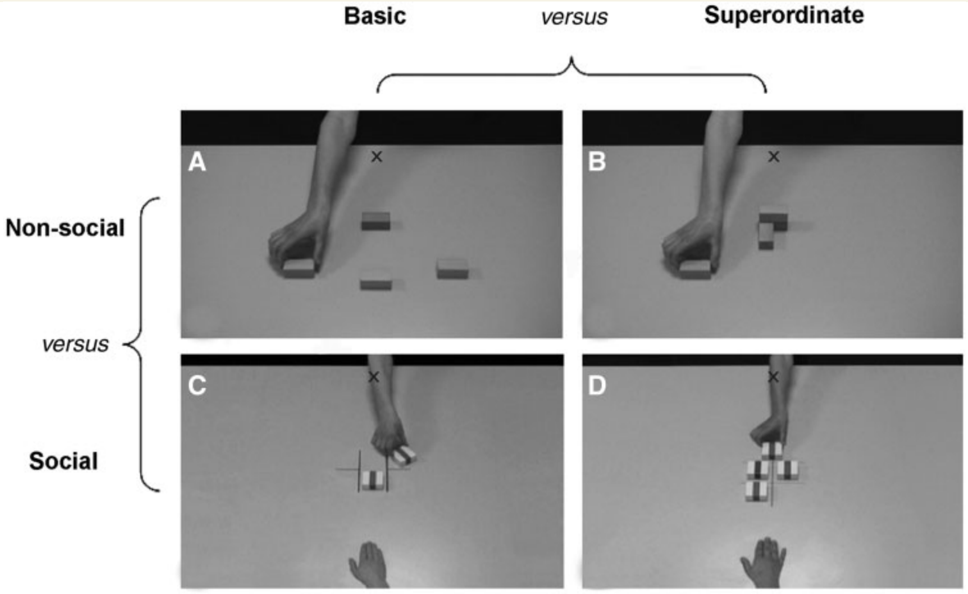


Figure 2 Examples of stimuli for each of the four tasks. (**A**) Non-social basic intention task; (**B**) non-social superordinate intention task; (**C**) social basic intention task; and (**D**) social superordinate intention task. The black cross indicates the starting position of the hand.

Non-social superordinate task

Video clips showed a sequence of three actions (e.g. to transport, to rotate or to lift a cube) leading to the construction of one out of three possible non-meaningful shapes (s1, s2 or s3; Fig. 1B). Each sequence was therefore characterized by a superordinate intention to build one of these three final shapes. The duration of the video sequences was varied (lasting 1480, 1560 or 1640 ms after movement onset) so that the last action was rendered incomplete. Participants were instructed to infer the superordinate intention and to give a response indicating the nature of the last, incomplete action in the sequence by pressing the corresponding keyboard button as quickly and accurately as possible.

Crucially, to ensure that participants were biased towards the superordinate intention itself and not merely towards the final action, commutative (i.e. interchangeable) sequences were used so that each shape could be constructed from multiple, distinct sequences of actions. Sequences shown in the covert blocks were thus distinct from those used in the overt blocks (e.g. the shape s1 could be obtained from the sequence 'lift–lift–rotate' in an overt block, but from the sequence 'lift–rotate–lift' in a covert block). In the bias session, the probability of building one of the three final shapes was increased at the expense of the other two, whilst keeping the probability of each simple action occurring during shape-building equal. The specific shape that was biased was counterbalanced across participants.

Social intention tasks

In the two social tasks (social basic and social superordinate), participants were instructed to infer whether a social intention was of either a cooperative or defective nature. They observed two players engaged in a social game, in which they either cooperated by coordinating their actions in order to achieve a shared goal, or defected by refusing to coordinate their actions. One after the other, the two players either transported the cube closest to them towards the centre of the board, or rotated it so that it remained at the same place (Fig. 2C and D). The first player's action was always shown entirely to the participants, while the second player's action was made incomplete by varying its duration across the trials (1480, 1560 or 1640 ms after its onset). Participants had to infer the nature of the second player's social intention (i.e. cooperative or defective). To do so, they were instructed to give a response about the nature of the incomplete action (i.e. to rotate or to transport) which unambiguously denoted the social intention, by pressing the corresponding button as quickly and accurately as possible (R for rotate, T for transport).

In the video clips, the second player's social intention either differed from that of the first player (i.e. the first player defected and the second cooperated, or the first player cooperated while the second defected) or it mirrored the first player's intention (i.e. both players cooperated or defected). This second type of response strategy is known as a 'tit-for-tat' (TFT) strategy. In situations of iterative cooperation, a TFT strategy is known to frequently be a more intuitive and successful strategy than alternative ones, such as 'always cooperating', 'always defecting' or 'acting randomly' (Axelrod, 1997; André and Day, 2007; Chambon *et al*., 2011). We thus chose to experimentally strengthen this existing *a priori* bias by increasing the probability that the second player adopts a TFT strategy, i.e. uses a strategy that mirrors their opponent's. In the bias session, the probability that the second player responded TFT was therefore increased so that, on average, he was more likely to cooperate (rather than defect) if the first player had previously cooperated, and to defect (rather than cooperate) if the first player had previously defected. In the baseline session, however, the probability of a TFT response was equal to that of responses using alternative strategies (i.e. cooperation in response to defection, or defection in response to cooperation).

Biasing the second player's strategy in this way ensured that participants paid attention to the whole action sequence, since to successfully predict the intentions of a player using a TFT strategy it is essential to take into account what the first player has done. Furthermore, using a TFT bias also prevented participants from giving stereotyped responses (e.g. always responding 'cooperate' or 'defect').

Social basic task

In the social basic task, participants were required to infer a social (defective or cooperative) intention that was denoted by the second player's action. This action consisted of either transporting a cube object (printed with a red or a blue line) towards the middle of a grid (termed 'bank') or rotating it so that it remained in its original location.

Social superordinate task

In the social superordinate task, the social intention inferred was achieved by the sequence of both players' actions and therefore corresponded to a final configuration of cubes (Fig. 1D). Players acted in turn with the goal to vertically align three cubes among the four available ones (one cube was printed with a blue line; the other three with a red line). The individual goals of the first and the second players were to align the three cubes printed with a red line (irrespective of the orientation of the lines) or to align three cubes with the same line orientation (irrespective of line colour), respectively. Combining both possible strategies for each player resulted in four possible final configurations of the cubes: both players defected, each preventing the other from achieving his goal (Configuration 4: no alignment); both players cooperated, in order to achieve both of their goals (Configuration 3: cubes were aligned according to both their colour and line orientation); the first player defected, preventing the second from achieving his goal, whilst the second cooperated, helping the first achieve his goal (Configuration 1: cubes were aligned according to their colour); or finally, the first player cooperated, helping the second achieve his goal, whilst the second defected preventing the first from achieving his goal (Configuration 2: cubes were aligned according to line orientation) (Fig. 1D).

As in the non-social superordinate task, commutative sequences were used so that each configuration could be obtained from distinct sequences of actions, ensuring that the second player's intention (e.g. playing TFT) could not be predicted from his single action (e.g. to rotate, or to transport) but only from the entire sequence of actions. Furthermore, the overall probabilities of each strategy (cooperative or defective) and of each single action (to rotate or to transport) were kept equal across the blocks.

Data analyses

Hits and reaction times

In the overt blocks of the bias and baseline sessions, patients' and controls' percentage of correct responses (hits) were compared using two-sample *t*-tests. These analyses were performed to ensure that both groups were equally successful in integrating the flat (baseline session) and biased (bias session) probability distributions associated with each intention.

Note that intentions were equally probable in the baseline session. We therefore referred to as 'future' likely (f-likely) and 'future' unlikely (f-unlikely) those intentions whose probability was increased (likely intention), or decreased (unlikely intention), in the subsequent bias session.

In the covert blocks of the baseline and bias sessions, hits and reaction times were analysed independently using 2 × 2 × 3 mixed-model, repeated-measures ANOVAs with group (controls versus patients) as a between-subjects factor, and intention (f-likely versus f-unlikely intentions) or bias (likely versus unlikely intentions), and amounts of visual information (low, moderate and high) as within-subjects factors. *Post hoc* Fisher tests were then performed to identify differences between conditions.

Whenever the variance structure did not conform to the requirements for parametric analyses, logarithmic transformations were used to obtain the required conformity. Analyses were performed using the statistical software Statistica 7 (www.statsoft .com).

Bias effect

To assess whether the assignment of a bias differently affected the performance of patients compared with controls across the four types of intention, a score reflecting the 'bias effect' was calculated for each subject, in each task. This score was obtained by subtracting the number of correct responses for the likely intention from those of the unlikely ones, in the covert blocks of the bias session. We then performed a 3 × 4 × 2 repeated-measures ANOVA with amount of visual information (low, moderate, and high) as a within-subjects factor, and type of intention (non-social basic; non-social superordinate; social basic; social superordinate) and group (controls versus patients) as between-subjects factors.

Effect of the amount of visual information

We also calculated a score reflecting the influence of the variation in amount of visual information on each participant's performance. This score was obtained, for each participant, in each task, by subtracting the proportion of correct responses obtained in the high visual information condition from that obtained in the low visual information condition. This score was then entered in a 2 × 4 × 2 repeated-measures ANOVA with bias (likely versus unlikely intentions) as a within-subjects factor, and type of intention and group as between-subjects factors.

Relationship to clinical symptoms

Finally, regression analyses were conducted to evaluate the influence of patients' cognitive performance on their clinical symptoms. In particular, we assessed whether an abnormal dependence on the bias and/or on visual information was predictive of the symptom severity on the different dimensions of schizophrenia measured (scale for the assessment of negative symptoms, scale for the assessment of positive symptoms and disorganization scores). For each clinical score, we conducted regression analyses using either the 'bias effect' score or the 'visual information effect' score as predictor variables. We used both these raw scores (simple linear regressions), or their transformed values (simple non-linear regressions with logarithmic, polynomial or exponential transformations). Models with the highest adjusted R^2 and a $P < 0.05$ are reported.

Results

For each session, two-tailed *t*-tests were performed between the two unlikely (or future unlikely) intentions on both reaction times and hits. As no significant differences appeared (all four tasks: all $P > 0.05$; see Supplementary Figs 1, 2A and B), performances for these two unlikely intentions were pooled for subsequent analyses.

Hits and reaction times

Baseline session

Overt blocks (containing very high and constant amount of visual information): in all four tasks, patients performed as successfully as controls when the amount of visual information was very high (mean correct responses > 94.5%, SD < 3.9; between-group comparisons: all $P > 0.05$), revealing that patients and controls were equally successful in integrating the probability distributions associated with each intention.

Covert blocks (containing varying amounts of visual information): the 2 (group) × 2 (intention: f-likely versus f-unlikely) × 3 (visual information) ANOVAs performed on both non-social (Basic and Superordinate) tasks revealed that patients performed the task as successfully as control participants [main effect of group, all F's (1,48) < 1.25, all $P > 0.26$]. Furthermore, there were no significant differences in hits and reaction times between the 'future' likely intention (i.e. the one that participants will be biased towards in the subsequent bias session) and the 'future' unlikely intention, indicating that prior to biasing, there was no *a priori* bias towards one intention over another [main effect of intention (f-likely versus f-unlikely): all F's (1,48) < 0.03, all $P > 0.84$; group × intention interaction effect, all F's(1,48) < 0.12, all $P > 0.72$].

As the amount of visual information increased, intentions were discriminated both faster and more successfully [main effect of visual information: all F's (2,96) > 251.1, all $P < 0.001$]. This improvement did not differ between patients and comparison participants [group × visual information interaction effect, all F's (2,96) < 0.56, all $P > 0.57$]. The group × intention × visual information interaction was not significant, indicating that increasing the amount of visual information improved both groups' performance equally, and independently of the type ('future' likely versus 'future' unlikely) of intention [all F's(2,96) < 0.61, all $P > 0.54$] (Supplementary Fig. 1A and B).

The 2 (group) × 2 (intention: f-likely versus f-unlikely) × 3 (visual information) ANOVAs performed on both social (Basic and Superordinate) tasks revealed that patients tended to be less successful than controls at recognizing intentions [main effect of group, all F's (1,47-48) > 2.93, all $P < 0.084$)]. More specifically, in the social superordinate task, we found a significant interaction effect for hits between group and intention ('future' likely versus 'future' unlikely) factors, indicating that, prior to being biased, control participants displayed an early preference towards inferring a TFT compared with other strategies, which was not found for the patient group [group × intention interaction, $F(1,48) = 6.3$, $P = 0.014$; *post hoc* test comparing TFT versus other strategies in control participants, $P < 0.001$; *post hoc* test comparing controls versus patients on responding TFT, $P = 0.034$] (Supplementary Fig. 3). In the social basic task, controls also inferred a TFT response more frequently than schizophrenic patients (two-sample *t*-test, $t = -2.38$, $P = 0.028$), but the group × intention interaction effect did not reach significance [$F(1,47) = 2.7$, $P = 0.11$].

In social basic and social superordinate tasks, the performance of both groups increased with the amount of visual information [main effect of visual information: all F's(2,94–96) > 198.64, all $P < 0.001$], but that increase was larger for patients than for controls [group × visual information, all F's(2,94–96) > 3.17, all $P < 0.05$]. This increase was due to patients inferring TFT less frequently in the condition of a low amount of visual information, whilst inferring TFT as often as controls for medium and high amounts (*post hoc* test comparing percentage of hits between controls versus patients for low amount of visual information, $P < 0.05$; no significant differences found for the other amounts). The group × intention × visual information interaction effect, however, was not significant [all F's (2,94–96) < 0.17, all $P > 0.21$] (Supplementary Fig. 2A and B).

In summary, in all four tasks intentions were recognized both faster and more successfully as the amount of visual information increased. In the non-social tasks, hits and reaction times for 'future' likely and unlikely intentions did not differ between groups, whereas in the social tasks, control participants exhibited an early preference for TFT strategies, prior to assignment of any probabilistic bias. This preference for inferring TFT over alternative strategies was not found in patients, which may account for their tendency to perform less successfully than control participants in social tasks, even when probabilities were not manipulated. Importantly, control participants tended to make more TFT responses as the amount of visual information decreased. This resulted in 'mechanically' reducing differences in the rate of likely responses between all three (low, medium and high) amounts of visual information. This effect was not observed in patients, due to their initial lack of preference for TFT.

Bias session

Overt blocks (very high and constant amount of visual information): in all four tasks both controls and patients performed the task successfully when the amount of visual information was very high (mean correct responses > 95%, SD < 3.1; between-group comparisons: all $P > 0.05$), indicating that patients and controls were equally successful in integrating the (biased) probability distribution associated with each intention.

Covert blocks (varying amounts of visual information): four distinct 2 (group) × 2 (bias: likely versus unlikely) × 3 (visual information) ANOVAs have been performed on each task (non-social basic, non-social superordinate, social basic and social superordinate). In all four tasks, participants were both more accurate and faster in recognizing the likely intention [i.e. the intention whose probability of occurrence was increased at the expense of the other competing ones; main effect of bias: all F's(1,47–48) > 33.41, all $P < 0.001$]. Similarly, performance increased with the amount of visual information in all tasks [main effect of visual information, all F's(2,94–96) > 181.8, all $P < 0.001$]. This effect was significantly modulated by the bias factor [visual information × bias interaction effect, all F's (2,94–96) > 15.62, all $P < 0.001$], with participants responding more frequently toward the likely intention as the amount of visual information progressively decreased, a finding consistent with previous results (Chambon *et al*., 2011).

In the non-social basic task, patients performed the task as successfully as control participants [main effect of group $F(1,48) = 0.3$, $P = 0.85$], while in the non-social superordinate condition they exhibited significantly poorer performances than controls [main effect of group $F(1,48) = 9.17$, $P = 0.003$]. In the non-social superordinate task, the group × bias interaction was significant [$F(1,48) = 5.47$, $P = 0.023$]. Interestingly, decomposing this effect using *post hoc* Fisher tests revealed that patients chose the unlikely intention less frequently than controls ($P < 0.001$), but chose the likely intention as frequently as controls (Supplementary Fig. 4A). Furthermore, in this task, increasing the amount of visual information resulted in a larger increase of the rate of 'likely' responses for controls than for patients [group × visual information, $F(2,96) = 3.25$, $P = 0.04$; *post hoc* tests comparing per cent of hits between controls versus patients for medium and high amounts of information, all $P < 0.05$; no significant difference was found for the low amount]. No significant group × bias or group × visual information interactions were found in the non-social basic task.

In both social tasks, patients tended to recognize intentions less successfully than control participants [main effect of group all F's (1,47–48) > 2.8, $P < 0.09$]. The group × bias interaction effect was significant in the social superordinate task only [$F(1,48) = 6.37$, $P = 0.014$]: in this condition, patients were less likely to choose a TFT intention (least significant difference test, $P = 0.02$), while choosing the other, unlikely strategies as often as controls (Supplementary Fig. 4B). Crucially, we found that patients' performance increased to a larger extent than controls' as the amount of visual information increased in the two social tasks [group × visual information, all F's(2,94–96) > 3.11, all $P < 0.05$]. As in the baseline session, that increase was due to patients inferring TFT less frequently in the condition of a low amount of visual information (*post hoc* tests comparing per cent of hits between controls versus patients for the low amount of information, all $P < 0.05$; no significant differences were found for the other amounts). It is noteworthy that this pattern of performance (i.e. fewer responses toward TFT intentions and a greater effect of amount of visual information) was exactly the opposite of that observed in the non-social superordinate task.

Finally, we did not find any significant group × bias × visual information interaction effect in any of the four tasks [all F's (2,94–96) < 2.32, all $P > 0.1$], indicating that the controls' preference for TFT strategies was not modulated by the amount of visual information available (Supplementary Figs 1C and D, and 2C and D).

In summary, in all four tasks, both groups were more accurate and faster when responding toward the likely (i.e. biased) intention, and showed increased preference to this intention as the amount of visual information decreased. It is of note that this finding is consistent with predictions made by a Bayesian estimation scheme: in situations of sparse or incomplete data, participants tend to compensate for visual uncertainty by appealing to their prior knowledge (Chambon *et al*., 2011).

In the non-social superordinate task, we found patients had difficulties in disengaging from their prior expectations (i.e. the likely intention) to select a response congruent with the unlikely intention, while relying less on visual information to make their decision. Such difficulties were associated with poor performances in this condition.

While patients also tended to perform less successfully than controls in the two social tasks, their performances however exhibited an opposite pattern of interaction between sensory and prior information: they were less sensitive than controls to the TFT bias, which resulted in their performance increasing to a larger extent than controls as a function of the amount of visual information.

Effect of the bias across different types of intentions

The 3 (visual information) × 4 (type of intention) × 2 (group) ANOVA first revealed a significant effect of the type of intention [main effect of type of intention, $F(3,191) = 11.7$, $P < 0.001$], with participants relying more on the bias to infer both social and non-social superordinate intentions, compared to basic ones (*post hoc* tests comparing superordinate and basic tasks, all $P < .001$). This difference interacted with group [group × intention interaction effect, $F(3,191) = 11.06$, $P < 0.001$]. Indeed, the bias exerted a greater influence on patients' response than controls' in the non-social superordinate condition ($P = 0.003$) while it exerted a smaller influence on their response compared to controls in both social conditions (all $P < 0.008$) (Fig. 3).

This difference could reflect the preference for TFT that controls already exhibited in the baseline session (see above), rather than reflecting a pure probabilistic bias effect. Therefore, to assess whether responses toward TFT strategies increased to the same extent with biasing across groups, we performed an additional 2 (group: controls, patients) × 2 (session: TFT-baseline, TFT-bias) repeated-measures ANOVA. No significant difference was observed [group × session interaction effect: both social conditions, all F's (1,47–48) < 0.74, all $P > 0.39$], indicating that the group difference for responding TFT in the bias session was due

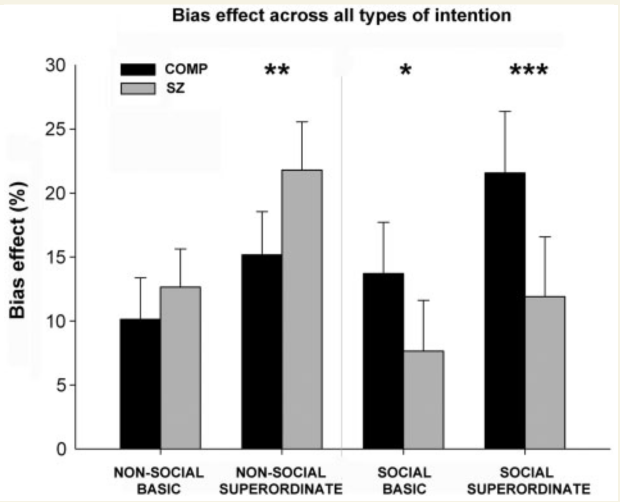


Figure 3 Bias effect (%) for all types of intention tasks. The greater the bias effect, the more participants respond toward the likely (i.e. biased) intention. SZ = patients with schizophrenia; COMP = comparison participants. *P < 0.05; **P < 0.005; ***P < 0.001.

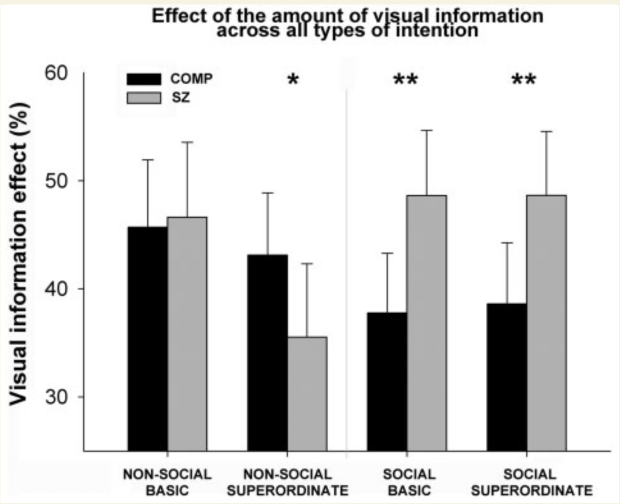


Figure 4 Effect of the amount of visual information (%) for all types of intention considered. The greater this score, the more participants' performance improved as the amount of visual information increased. SZ = patients with schizophrenia; COMP = comparison participants. *P < 0.05; **P < 0.005.

to controls' initial preference for responding TFT in the baseline session.

Effect of the amount of visual information across different types of intentions

The 2 (bias: likely versus unlikely) × 4 (type of intention) × 2 (group) ANOVA revealed a significant main effect of group [$F(1,191) = 4.12$, $P = 0.04$] showing that overall, patients' performances improved to a greater extent than controls' when increasing the amount of visual information. However, the significant interaction between group and type of intention [$F(3,191) = 6.12$, $P < 0.001$] further revealed that, while increasing the amount of visual information improved patients' performance more than controls' in both social basic and social superordinate conditions (*post hoc* tests, all $P < 0.005$), patients' performance improved to a lesser extent as this amount increased in the non-social superordinate condition (*post hoc* test, $P = 0.03$) (Fig. 4).

Clinical symptoms: regression analyses

Bias effect

In the non-social superordinate task, the bias effect significantly and positively predicted both scale for the assessment of positive symptoms ($R^2 = 0.39$, $P = 0.003$) and disorganization ($R^2 = 0.21$, $P = 0.04$) scores (Fig. 5A and B). The higher the effect of the bias on patients' performances (i.e. the more they relied on their prior knowledge to make their decision), the more likely patients were to be disorganized and exhibit positive symptoms. In both social tasks, the bias effect was found to significantly and negatively predict the scale for the assessment of negative symptoms score ($R^2 = 0.22$, $P = 0.03$ for the social basic task, and $R^2 = 0.32$, $P = 0.008$ for the social superordinate task). Therefore, the smaller the effect of the bias on patients' performances (i.e. the less they relied on their priors for inferring a social intention), the more severe the negative symptoms (Fig. 5C and D).

Effect of visual information

In the social superordinate intention task, the effect of visual information significantly predicted the scale for the assessment of negative symptoms score ($R^2 = 0.44$, $P = 0.001$) and—but to a lesser extent—the disorganization score ($R^2 = 0.28$, $P = 0.01$). The higher the effect of visual information on patients' responses (i.e. the more they relied on the visual information to make their decision), the more severe their negative and disorganization symptoms (Fig. 6B). In the social basic task, this effect tended to predict the scale for the assessment of negative symptoms score but the regression coefficient did not reach significance ($R^2 = 0.18$, $P = 0.066$) (Fig. 6A).

Discussion

The present study aimed to investigate whether the impaired ability of schizophrenic patients to appreciate other people's intentions is confined to a particular type of intention, as opposed to being generalized. To test this hypothesis, we designed a series of tasks that required the identification of different types of intentions, varying on the dimensions of scope (basic, superordinate) or target (non-social, social). We further hypothesized that, if present, a localized deficit would be accounted for by abnormalities in the interplay between prior knowledge and sensory evidence, which normally underlies the ability to infer others' intentions (Chambon *et al.*, 2011).

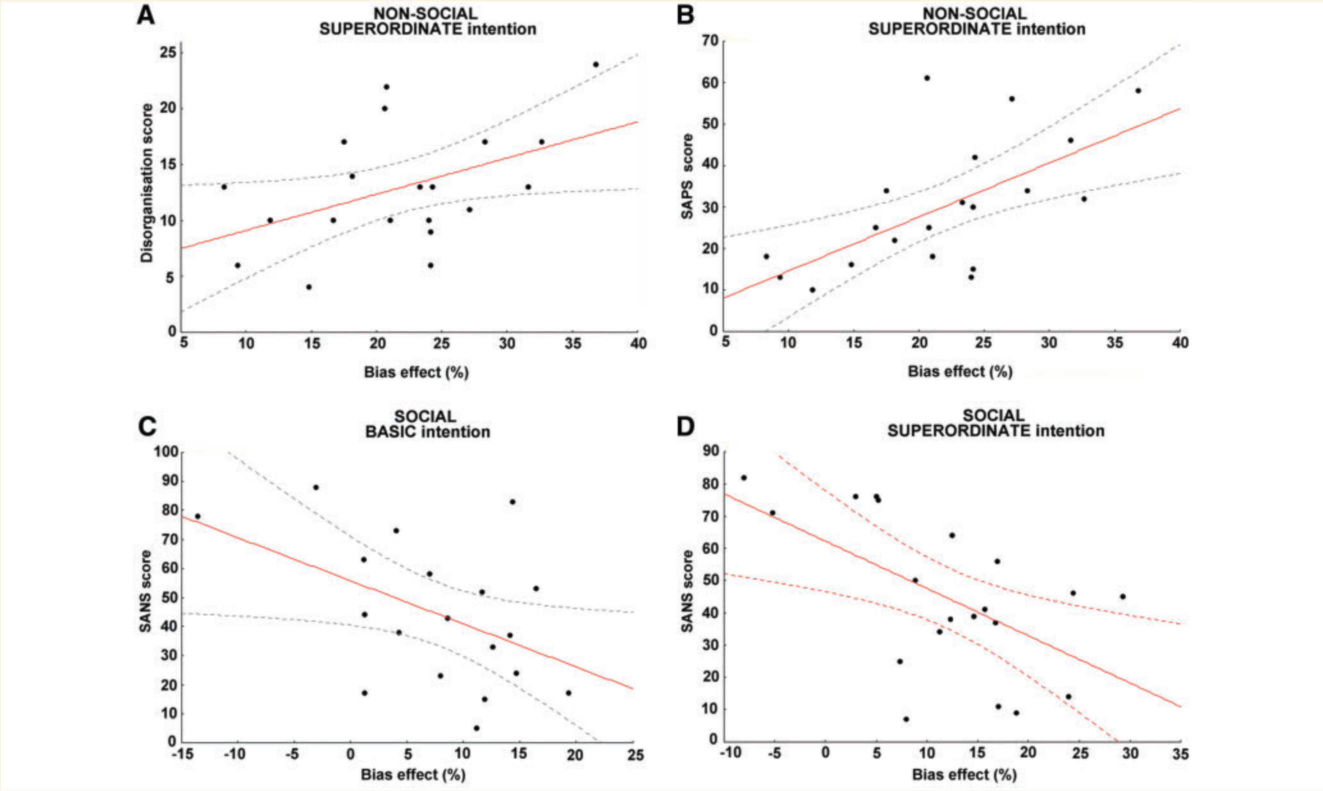


Figure 5 The magnitude of the bias effect predicts the severity of clinical symptoms. The linear regression lines derived from the linear regression analyses between the 'bias effect' (explanatory factor) and patients' disorganization (**A**), positive (**B**) and negative symptoms (**C** and **D**) are shown in red. Note that the correlation is positive for non-social intentions, and negative for social intentions. The 95% confidence intervals for the regression lines are shown in grey. SANS = Scale for the Assessment for the Negative Symptoms; SAPS = Scale for the Assessment of Positive Symptoms.

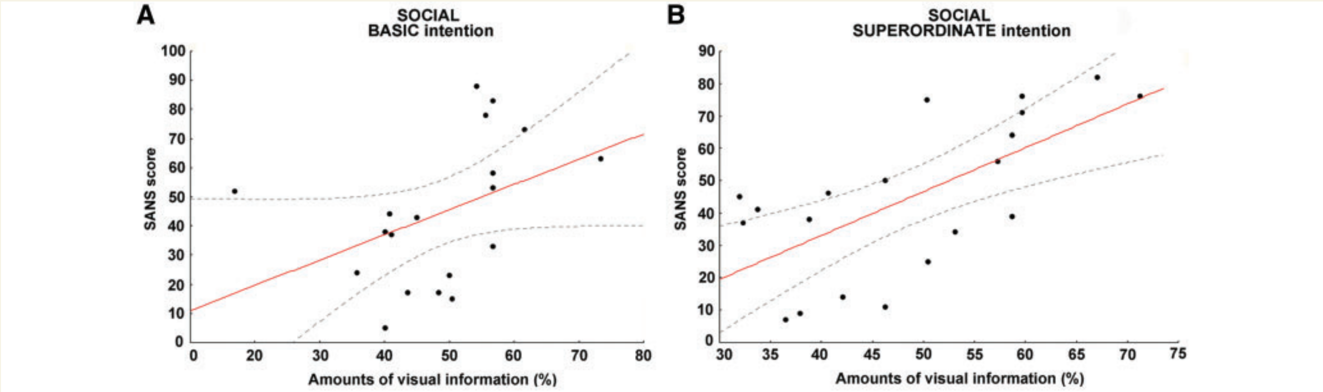


Figure 6 In social basic (**A**) and superordinate (**B**) tasks, abnormal dependence on visual information predicted scale for the assessment of negative symptoms. The linear regression lines derived from the linear regression analyses between the effect of the amount of visual information (explanatory factor) and patients' negative symptoms are shown in red. The 95% confidence intervals for the regression lines are shown in grey. SANS = Scale for the Assessment for the Negative Symptoms.

We first showed that controls and patients are sensitive to both types of information, and their interaction. First, all participants were more successful in recognizing underlying intentions when the visual information conveyed by the action scene was increased. Secondly, they showed more accuracy and were faster when recognizing likely compared to unlikely intentions. Finally, both groups' performances exhibited a strong bias effect, which progressively increased as the amount of visual information decreased, and vice versa. Crucially, we also observed specific differences between the two groups. Depending on the scope or target of the presented intention, patients with schizophrenia showed an abnormal weighting of either prior knowledge or sensory information, which was further associated with the severity of positive or negative symptoms of the condition.

Inferring non-social intentions

Whilst performing as successfully as controls when inferring intentions involving a single action (basic intentions), patients exhibited poorer performance when recognizing intentions involving a 'sequence' of basic motor acts (superordinate intentions). These poorer performances cannot be due to an increased attentional load, resulting from paying attention to a sequence of three actions as opposed to a single act, since patients were equally successful to controls in the baseline session of the superordinate condition. Indeed, differences between the patient and control groups were only found in the bias session, where the task required them to properly sample the probability distribution associated with each type of (likely or unlikely) intention.

Specifically, poorer performance in the bias session was characterized by a decreased number of responses toward unlikely intentions, with patients having difficulty disengaging from their biased expectations to select less likely alternatives. Crucially, this abnormal dependence on biased expectations may not be primarily due to a faulty weighting of probabilities—patients responded towards likely intentions as often as controls—but to an inability to revise prior expectations in light of new evidence. This assumption is supported by the fact that increasing the amount of visual information was of less benefit to patients' performances than controls', indicating that patients relied on visual information to a lesser extent than comparison participants to make their decision.

This inability to disengage from prior, self-generated expectations, together with a tendency to disregard external/sensory evidence, echoes specific biases observed in schizophrenia across a wide range of studies, such as the so-called 'bias against disconfirmatory evidence' (Woodward *et al*., 2008), or a tendency to make hasty decisions ('jumping to conclusion': Garety *et al*., 1991). Indeed, in tasks of probabilistic reasoning, individuals with schizophrenia tend to make judgements based upon less evidence than comparison participants and/or to hold prior beliefs despite little evidential support (Brankovic and Paunovic, 1999; Jones *et al*., 1999). Our results in the superordinate condition reveal that this bias may not be specific to the domain of reasoning, but may also extend to mentalizing, potentially underlying such abnormalities in schizophrenia. When having to make decisions about other people's (non-social) intentions, patients preferentially relied on previously formed expectations, i.e. on beliefs about how the observed agent is most likely to behave, while neglecting potentially disconfirmatory visual information.

It is noteworthy that this pattern of performance was only observed in the superordinate, but not in the basic condition. We believe that this finding may be accounted for by the specific property of the intention manipulated. Indeed, a superordinate intention is achieved by a sequence of interchangeable basic actions, and, therefore, cannot be directly deduced from the current action *per se* (Pacherie, 2000; Jacob and Jeannerod, 2005). Inferring superordinate intentions thus requires the participant to refer to a distal representation of the goal, which is not directly available from observation and consequently tends to be less challenged by visual evidence (Chambon *et al*., 2011). Intentions which are not predictable from merely observing the current action, i.e. superordinate intentions, may thus aggravate patients' tendency not to revise their beliefs in the face of progressively disconfirmatory evidence from the action scene.

An impaired revision process could signal a disturbance in the (Bayesian) inferential mechanism, which compares new sensory evidence with stored knowledge of the world, or prior beliefs (e.g. beliefs about what is the most likely cause of an observed behaviour, out of the possible alternatives). Under normal circumstances, a difference between expected and observed information gives rise to a prediction error that can be used to update one's model of the world (Kilner *et al*., 2007*a*, *b*; Fletcher and Frith, 2009). A disturbance in this error-dependent updating mechanism, possibly caused by alterations in the dopaminergic circuitry (Gradin *et al*., 2011), may result in patients having an abnormal degree of certainty in their beliefs about other people's intentions. If these beliefs are not challenged by external evidence and, if necessary, replaced with contextually appropriate beliefs, patients' inferences about others' mental states would be based on an outdated model of the current situation. This could result in patients exhibiting maladaptive or bizarre behaviours (Chambon *et al*., 2008; Barbalat *et al*., 2009) and/or holding incorrect (i.e. deluded) beliefs about the real causes driving other's behaviour (Fletcher and Frith, 2009). Precisely in line with this assumption, we found that patients' abnormal dependence on prior expectations predicted the severity of positive symptoms of schizophrenia: the more patients relied on their priors to make their decision, the more severe these symptoms were.

This observation sheds new light on previous evidence of patients with passivity symptoms misattributing actions to non-agents, or over-attributing intentionality where there is none (Abu-Akel and Bailey, 2000; Bentall *et al*., 2001; Blakemore *et al*., 2003), especially in situations that require continuous monitoring of visual signals arising from the action scene (Franck *et al*., 2001). This 'hyper-intentionality' may indeed result from the quantitative over-generation of hypotheses and an inability to revise them in light of disconfirmatory evidence, potentially resulting in paranoid (over-)interpretation of other people's goals (Abu-Akel and Bailey, 2000; Bara *et al*., 2011). Ultimately, over-reliance on unchallenged, internal expectations, whilst dismissing external evidence, could lead patients to make abnormal distinctions between the real causes driving other people's behaviour, and their subjective beliefs about what these causes should, or might be. Such a confusion between external and internal states of affairs would undermine a patient's ability to separate their own intentions from those of others (Frith and Corcoran, 1996), or to disentangle external experiences from inner experiences (Walter *et al*., 2009), a feature that frequently accompanies paranoid and/or passivity symptoms, such as the well-documented 'delusion of control' (Brüne *et al*., 2008).

Inferring social intentions

Surprisingly, the patients' pattern of performance on social tasks was the exact opposite of the pattern observed on non-social tasks. On social tasks, patients relied on the available visual information to a greater extent than controls, whilst showing a decreased sensitivity to the bias (i.e. towards TFT intentions). Furthermore, this pattern of performance predicted the severity of negative symptoms: patients with more severe negative symptoms were less sensitive to the bias (i.e. they were less likely to select TFT) and conversely, were more reliant on visual information to make their decision.

At first sight, a lower sensitivity to the bias in the social conditions may seem at odds with patients' excessive reliance on prior knowledge in the previous (non-social) conditions. However, performance in the bias session indicated that patients did normally integrate the (biased) probability distribution of the session, with the number of responses toward TFT increasing with its probability of occurrence, as found in controls. Rather, we found that unlike control participants, patients did not exhibit any early preference for TFT in the baseline session—that is, prior to being biased toward this particular mode of interaction. It is noteworthy that while increasing the probability of TFT was of benefit to both groups, this increase was not enough to compensate for patients' initial deficit.

This absence of an inherent preference for TFT suggests that for patients, social situations may not prompt the same expectations as those typically observed in healthy participants. Indeed, situations identified as involving social interactions are prone to trigger domain-specific expectations concerning the way agents are likely to behave in such situations (Castelli *et al*., 2000; Scholl and Tremoulet, 2000; Kourtis *et al*., 2010). Under normal circumstances, these modular, high-level expectations may contribute to priority being given to some intentional causes at the expense of other, competing causes (e.g. cooperation in response to previous cooperation, defection in response to defection; Chambon *et al*., 2011). Reliance on these domain-specific, prior expectations, which can be induced even by basic movements (such as the relative movements of geometrical figures; Heider and Simmel, 1944), may prove crucial in situations of sparse data, or when sensory evidence is too noisy to guarantee accurate inference-making (Baker *et al*., 2006). Poor performance in social conditions suggests that patients lack the prior expectations which usually bias social inferences, consistent with a previous suggestion that impoverished social knowledge, from which these expectations may be derived, constitutes an intrinsic feature of schizophrenia (Cutting and Murphy, 1990).

Concomitantly, impoverished expectations within the social domain may account for why patients were excessively over-reliant on visual information. As previously suggested, inferring another agent's intention requires the adaptive integration of new external evidence into prior beliefs about the agent's goals and attitudes (Baker *et al*., 2006; Fletcher and Frith, 2009), which is contingent upon the relative reliability of these two sources of information (Chambon *et al*., 2011). Results in the social conditions suggest a disturbance in this integrative mechanism that is exactly the opposite of what was observed in non-social conditions: impoverished expectations within the social domain, resulting in a reduced ability to draw reliable internal predictions, prompted patients to over-weight external evidence.

Crucially, this abnormal (over-)weighting of visual information correlated with the negative symptoms of schizophrenia. Previous observations have similarly shown that patients with negative symptoms, such as anhedonia or alexithymia, tend to excessively focus on directly observable, external information, rather than inner experience (Taylor, 1994). Our results further suggest that these incapacitating features may be accounted for by an impaired ability to make reliable predictions about other's behaviour, rendering patients slaves 'to every (external) influence' (Frith, 1994). It is noteworthy that over-reliance on external evidence can be particularly harmful in social situations, in which many possible intentions are potentially congruent with what is observed, so that it is impossible to infer the agent's intention from environmental cues only. In such situations, it has been shown that participants tend to compensate for sensory uncertainty by appealing to prior knowledge (Csibra and Gergely, 2007). Impoverished prior knowledge in the social domain may therefore result in an incapability to reduce the intrinsic uncertainty of social world. This may have important consequences on how patients suffering from negative symptoms perceive other people's social attitudes and behaviours. Indeed, a pervasive and constant uncertainty may render any observed, or experienced, social interactions fruitless, and could ultimately be responsible for social and motivational disorders that are characteristic of negative symptoms of schizophrenia (Fletcher and Frith, 2009).

It is noteworthy that a lack of preference for TFT in schizophrenia is consistent with previous data. In tasks simulating human cooperation in group interactions, patients do not exhibit any pattern of 'equivalent retaliation', or 'altruistic punishment', e.g. they do not defect when the game partner has previously defected, or they accept unfair offers at a significantly higher rate than did healthy controls (Chung *et al*., 2011; Csukly *et al*., 2011). Interestingly, individuals with schizotypal traits exhibit the same pattern of performance as patients (van't Wout and Sanfey, 2011), suggesting that poor expectations in the social domain may represent a marker of vulnerability to schizophrenia. Moreover, such abnormal expectations may serve as early clinical intervention targets. Indeed, there is growing evidence that cognitive therapies targeting social skills improves long-term prognosis and significantly benefits the patient's everyday life (Horan *et al*., 2011; Ventura *et al*., 2011). We believe accordingly that early detection of an abnormal use of social-specific knowledge may have a positive impact on both patients' social functioning and evolution of the condition. Critically, a lack of preference for the 'TFT' mode of interaction predicted the severity of negative but not positive symptoms. This further suggests that poor expectations in the social domain may also be relevant to the formation of symptom profile, together with being a useful indicator for identification and management of vulnerable individuals.

Neural underpinnings

Together, these results suggest that different neural dysfunctions may underline patients' abnormal performance, depending on their symptom profile and the type of intention considered. According to a recent model, action understanding is achieved through interactions between a ventral pathway where intention priors are formed, and a dorsal network hierarchically organized according to the level at which the observed action is represented (kinematic, motor command, or goal level) (Kilner, 2011). Intention priors in the ventral pathway are used to predict in dorsal areas the most likely action required to achieve the most likely intention, given what is observed. An error signal is generated when the prediction is not accurate. We suggest that undue weight given to prior expectations in patients with positive symptoms may be caused by abnormal encoding of prediction error signals in dopamine-rich brain areas (Gradin *et al*., 2011). This would result in the inability to update intention priors in brain areas of the ventral pathway. On the other hand, lack of preference for TFT in patients with negative symptoms suggests abnormal biasing influences from brain regions that encode social-specific knowledge, such as the medial prefrontal cortex (Overwalle, 2009). Thus, the medial prefrontal cortex might insufficiently bias action prediction in brain areas within the dorsal pathway, resulting in an equal weighting of all possible action alternatives (e.g. cooperation if previous defection, cooperation if previous cooperation, etc.). Such weakening of social-specific influences is likely to reduce the accuracy of prediction error-dependent mechanisms, leading patients to rely on sensory evidence by default. In future work, the use of neuroimaging techniques should allow us to test these assumptions directly.

Conclusion

We identified specific mentalizing impairments in participants with schizophrenia. Rather than being generalized, these impairments were contingent upon the scope (basic versus superordinate) or the target (non-social versus social) of the intention to be inferred, and were further accounted for by abnormal integration of visual information and prior knowledge.

In non-social tasks, patients showed specific difficulties in inferring intentions achieved by a sequence of basic motor acts (superordinate intentions). We found that this poor performance was due to patients over-relying on prior expectations and disconfirming visual evidence. This abnormal pattern of interaction predicted the severity of positive symptoms. We suggested that this faulty interaction may signal a disturbance in the inferential mechanism driving the integration of sensory evidence into prior beliefs, to produce accurate inference about other people's intentions. Such a disturbance could favour a paranoid (over-) interpretation of other people's goals, by hindering the revision of one's prior beliefs, and may ultimately lead patients to distinguish abnormally between their own and others' intentions—a confusion frequently experienced by individuals with passivity symptoms.

Patients also showed difficulties in inferring social intentions. However, their pattern of performance was the exact opposite to that observed in non-social conditions. While they exhibited weaker prior expectations, they relied strongly on sensory evidence to make their decisions. Furthermore, this pattern of performance predicted the severity of negative symptoms. Based on the absence of early preference for the TFT mode of interaction, we hypothesized that social situations may not prompt the same expectations in patients as those typically observed in healthy participants, leading to the formation of abnormal (unreliable) predictions about others' social intentions. Such abnormal predictions may result in an incapability to reduce the intrinsic uncertainty of social situations. We suggest that constant and pervasive uncertainty about other's social attitudes and behaviours could jeopardize patients' propensity to social interactions, and may ultimately account for some of the incapacitating features associated with negative symptoms of schizophrenia.

Acknowledgements

We are grateful to Mathilde Vernet, Juliette Sablier and Mariia Kaliuzhna for their help in preparing the stimuli. We also gratefully acknowledge Mrs Ghislaine Bailly, Drs Olivier Martin and Caroline Demily and Pr. Jean-Louis Terra for their help in the inclusion of patients. We thank Emma Kilford and Geeta Reddy for their help in proofreading the final draft of the article.

Funding

Agence Nationale de la Recherche obtained by C.F. (ANR-ZR81); Conseil Scientifique de Recherche, Le Vinatier (CSR U03); French Ministry of Research (to V.C.).

Supplementary material

Supplementary material is available at *Brain* online.

References

Abu-Akel A, Bailey AL. The possibility of different forms of theory of mind. Psychol Med 2000; 30: 735–8.

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, DSM-IV. 4th edn. Washington, DC: APA; 1994.

André JB, Day T. Perfect reciprocity is the only evolutionarily stable strategy in the continuous iterated prisoner's dilemnna. J Theor Biol 2007; 247: 11–22.

Andreasen NC. The scale for the assessment of negative symptoms (SANS). Iowa City: The University of Iowa; 1983.

Andreasen NC. The scale for the assessment of positive symptoms (SAPS). Iowa City: The University of Iowa; 1984.

Axelrod R. The complexity of cooperation. Princeton, NJ: Princeton University Press; 1997.

Baker CL, Tenenbaum JB, Saxe RR. Bayesian models of human action understanding. In: Weiss Y, Scholkopf B, Platt J, editors. Advances in neural information processing systems. Cambridge, MA: MIT Press; 2006. p. 99–106.

Baker CL, Saxe RR, Tenenbaum JB. Action understanding as inverse planning. Cognition 2009; 113: 329–49.

Bara BG, Ciaramidaro A, Walter H, Adenzato M. Intentional minds: a philosophical analysis of intention tested through fMRI experiments involving people with schizophrenia, people with autism, and healthy individuals. Front Hum Neurosci 2011; 5: 7.

Barbalat G, Chambon V, Franck N, Koechlin E, Farrer C. Organization of cognitive control within lateral prefrontal cortex in schizophrenia. Arch Gen Psych 2009; 66: 1–10.

Bentall RP, Corcoran R, Howard R, Blackwood N, Kinderman P. Persecutory delusions: a review and theoretical integration. Clin Psychol Rev 2001; 21: 1143–92.

Blakemore SJ, Frith U. How does the brain deal with the social world? Neuroreport 2004; 19: 119–28.

Blakemore SJ, Boyer P, Pachot-Clouard M, Meltzoff A, Segebarth C, Decety J. The detection of contingency and animacy from simple animations in the human brain. Cereb cortex 2003; 13: 837–44.

Bora E, Yucel M, Pantelis C. Theory of mind impairment in schizophrenia: meta-analysis. Schizophr Res 2009; 109: 1–9.

Brankovic SB, Paunovic VR. Reasoning under uncertainty in deluded schizophrenic patients: a longitudinal study. Eur Psychiatry 1999; 14: 76–83.

Brüne M, Lissek S, Fuchs N, Witthaus H, Peters S, Nicolas V, et al. An fMRI study of theory of mind in schizophrenia patients with 'passivity' symptoms. Neuropsychologia 2008; 46: 1992–2001.

Castelli F, Happé F, Frith U, Frith CD. Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. Neuroimage 2000; 12: 314–25.

Chambon V, Domenech P, Pacherie E, Koechlin E, Baraduc P, Farrer C. What are they up to? The role of sensory evidence and prior knowledge in action understanding. Plos ONE 2011; 6: e17133.

Chambon V, Franck N, Koechlin E, Ciuperca G, Fakra E, Azorin J-M, et al. The architecture of cognitive control in schizophrenia. Brain 2008; 131: 962–70.

Chung D, Kim YT, Jeong J. Cognitive motivations of free riding and cooperation and impaired strategic decision making in schizophrenia during a public goods game. Schizophr Bull 2011; 24.

Ciaramidaro A, Adenzato M, Enrici I, Erk S, Pia L, Bara BG, et al. The intentional network: how the brain reads varieties of intentions. Neuropsychologia 2007; 45: 3105–13.

Corcoran R. Theory of mind in schizophrenia. In: Penn D, Corrigan P, editors. Social cognition in schizophrenia. Washington: APA; 2001.

Csibra G, Gergely G. Obsessed with goals': functions and mechanisms of teleological interpretation of actions in humans. Acta Psychol 2007; 124: 60–78.

Csukly G, Polgár P, Tombor L, Réthelyi J, Kéri S. Are patients with schizophrenia rational maximizers? Evidence from an ultimatum game study. Psychiatry Res 2011; 187: 11–7.

Cutting J, Murphy D. Impaired ability of schizophrenic, relative to manics or depressives, to appreciate social knowledge about their culture. Br. J. Psychiatry 1990; 157: 355–8.

Franck N, Farrer C, Georgieff N, Marie-Cardine M, Daléry J, d'Amato T, et al. Defective recognition of one's own actions in patients with schizophrenia. Am J Psychiatry 2001; 158: 454–9.

Fletcher PC, Frith CD. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. Nat Rev Neurosci 2009; 10: 48–58.

Frith CD. Theory of mind in Schizophrenia. In: David AS, Cutting JC, editors. The neuropsychology of schizophrenia. Hillsdale, NJ: Erlbaum Press; 1994. p. 147–61.

Frith CD, Corcoran R. Exploring 'theory of mind' in people with schizophrenia. Psychol Med 1996; 26: 521–30.

Frith CD. Schizophrenia and theory of mind. Psychol Med 2004; 34: 385–9.

Garety PA, Hemsley DR, Wessely S. Reasoning in deluded schizophrenic and paranoid patients. Biases in performance on a probabilistic inference task. J Nerv Ment Dis 1991; 179: 194–201.

Gradin VB, Kumar P, Waiter G, Ahearn T, Stickle C, Milders M, et al. Expected value and prediction error abnormalities in depression and schizophrenia. Brain 2011; 134: 1751–64.

Griffiths TL, Kemp C, Tenenbaum JB. Bayesian models of cognition. In: Sun R, editor. The Cambridge handbook of computational cognitive modelling. Cambridge: Cambridge University Press; 2008.

Hardy-Baylé MC, Sarfati Y, Passerieux C. The cognitive basis of disorganization symptomatology in schizophrenia and its clinical correlates: toward a pathogenetic approach to disorganization. Schizophr Bull 2003; 29: 459–71.

Harrington L, Siegert RJ, McClure J. Theory of mind in schizophrenia: a critical review. Cognit Neuropsychiatry 2005; 10: 249–86.

Heider F, Simmel M. An experimental study of apparent behaviour. Am J Psychol 1944; 57: 243–59.

Hemsley DR. The development of a cognitive model of schizophrenia: placing it in context. Neurosci Biobehav Rev 2005; 29: 977–88.

Horan WP, Green MF, Degroot M, Fiske A, Hellemann G, Kee K, et al. Social cognition in schizophrenia, Part 2: 12-month stability and prediction of functional outcome in first-episode patients. Schizophr Bull 2011 (in press).

Jacob P, Jeannerod M. The motor theory of social cognition: a critique. Trends Cogn Sci 2005; 9: 21–5.

Jones E. The phenomenology of abnormal belief: a philosophical and psychiatric inquiry. Phil Psych Psychol 1999; 6: 1–16.

Kourtis D, Sebanz N, Knoblich G. Favouritism in the motor system: social interaction modulates action simulation. Biol Lett 2010; 6: 758–61.

Kilner JM. More than one pathway to action understanding. Trends Cogn Sci 2011; 15: 352–7.

Kilner JM, Friston KJ, Frith CD. Predictive coding: an account of the mirror neuron system. Cogn Process 2007a; 8: 159–66.

Kilner JM, Friston KJ, Frith CD. The mirror-neuron system: a Bayesian perspective. Neuroreport 2007b; 18: 619–23.

McCabe R. On the inadequacies of theory of mind explanations of schizophrenia. Alternative accounts of alternative problems. Theory Psychol 2004; 14: 738–52.

Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 1971; 9: 97–113.

Pacherie E. The content of intentions. Mind Language 2000; 15: 400–32.

Pacherie E. The phenomenology of action: a conceptual framework. Cognition 2008; 107: 179–217.

Scholl BJ, Tremoulet PD. Perceptual causality and animacy. Trends Cogn Sci 2000; 4: 299–309.

Sprong M, Schothorst P, Vost E, Hox J, van Engeland H. Theory of mind in schizophrenia: a meta-analysis. Br J Psychiatry 2007; 191: 5–13.

Taylor GJ. The alexithymia construct: conceptualization, validation, and relationship with basic dimensions of personality. New Trends Exp Clin Psychiatry 1994; 10: 61–74.

van Overwalle F. Social cognition and the brain: a meta-analysis. Hum Brain Mapp 2009; 30: 829–58.

van't Wout M, Sanfey AG. Interactive decision-making in people with schizotypal traits: a game theory approach. Psychiatry Res 2011; 185: 92–6.

Ventura J, Wood RC, Hellemann GS. Symptom domains and neurocognitive functioning can help differentiate social cognitive processes in schizophrenia: a meta-analysis. Schizophr Bull 2011 (in press).

Walter H, Ciaramidaro A, Adenzato M, Vasic N, Ardito RB, Erk S, et al. Dysfunction of the social brain in schizophrenia is modulated by intention type: an fMRI type. Scan 2009; 4: 166–76.

Woodward T S, Moritz S, Menon M, Klinge R. Belief inflexibility in schizophrenia. Cognit. Neuropsychiatry 2008; 13: 267–77.

Zalla T, Bouchilloux N, Labruyere N, Georgieff N, Bougerol T, Franck N. Impairment in event sequencing in disorganised and non-disorganised patients with schizophrenia. Br Res Bull 2006; 68: 195–202.

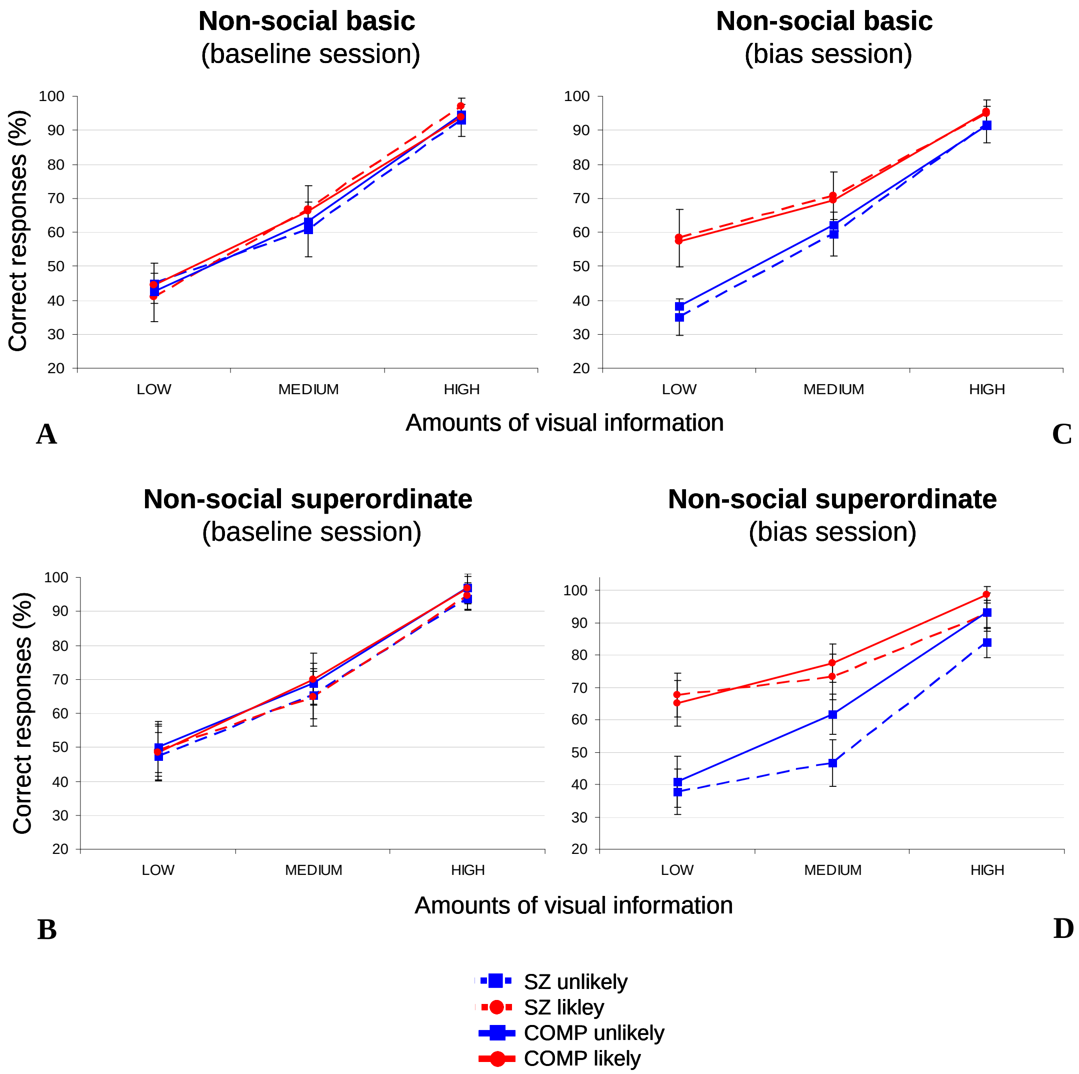
Supplementary Information

Mentalizing Under Influence: Abnormal Dependence on Prior Expectations in Patients with Schizophrenia

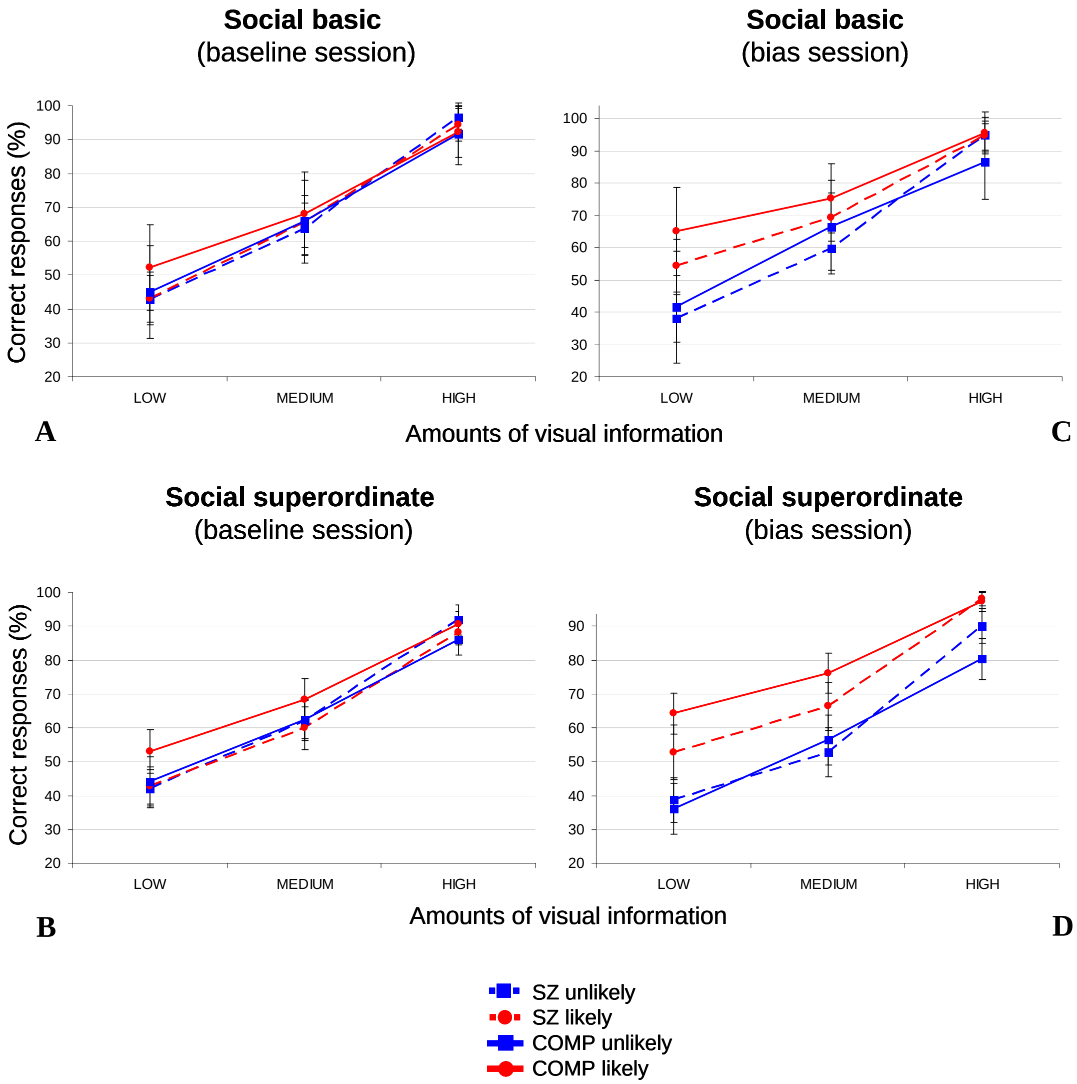
Valerian Chambon, Elisabeth Pacherie, Guillaume Barbalat, Pierre Jacquet, Nicolas Franck, and Chlöé Farrer

Address correspondence to: Valerian Chambon, Centre de Neuroscience Cognitive - UMR 5229 CNRS ; 67, bd Pinel, 69675 BRON Cedex, France; Email: **valerian.chambon@isc.cnrs.fr**; Phone: +33 (0)4 37 91 12 10

Supplementary figures



Figures S1. Non-Social tasks (COVERT blocks): Mean percentage of correct responses (± SD) for likely (red) vs. unlikely (blue) intentions for each amount of visual information (LOW, MODERATE, HIGH). SZ: patients with schizophrenia (dashed lines); COMP: comparison participants (solid lines).



Figures S2. Social tasks (COVERT blocks): Mean percentage of correct responses (± SD) for likely (red) vs. unlikely (blue) intentions for each amount of visual information (LOW, MODERATE, HIGH). SZ: patients with schizophrenia (dashed lines); COMP: comparison participants (solid lines).

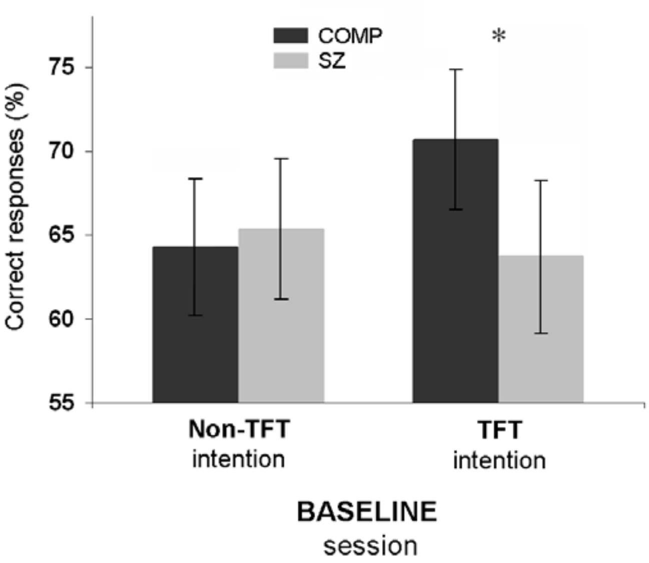


Figure S3. **SOCIAL SUPERORDINATE task (COVERT blocks)**: Mean percentage of correct responses (± SD) toward tit-for-tat (TFT) vs. non-TFT intentions in the *baseline* session. COMP: comparison participants; SZ: patients with schizophrenia. *: p<.05.

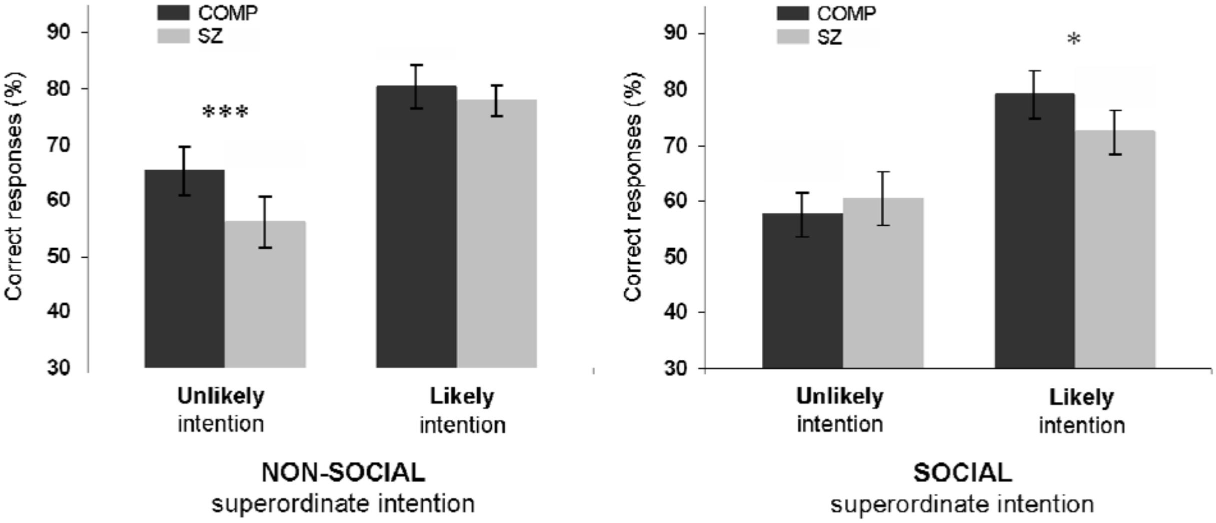


Figure S4. **Mean percentage of correct responses** (± SD) **for unlikely vs. likely intentions in the bias session (**COVERT blocks). Left panel: NON-SOCIAL SUPERORDINATE task; Right panel: SOCIAL SUPERORDINATE task. COMP: comparison participants; SZ: patients with schizophrenia. *: p<.05; ***: p<.001