The role of the amygdala in naturalistic mentalising in typical development and in autism spectrum disorder

Gabriela Rosenblau, Dorit Kliemann, Benjamin Lemme, Henrik Walter, Hauke R. Heekeren and Isabel Dziobek

Background
The substantial discrepancy between mentalising in experimental settings v. real-life social interactions hinders the understanding of the neural basis of real-life social cognition and of social impairments in psychiatric disorders.

Aims
To determine the neural mechanisms underlying naturalistic mentalising in individuals with and without autism spectrum disorder.

Method
We investigated mentalising with a new video-based functional magnetic resonance imaging task in 20 individuals with autism spectrum disorder and 22 matched healthy controls.

Results
Naturalistic mentalising implicated regions of the traditional mentalising network (medial prefrontal cortex, temporoparietal junction), and additionally the insula and amygdala. Moreover, amygdala activity predicted implicit mentalising performance on an independent behavioural task. Compared with controls, the autism spectrum disorder group did not show differences in neural activity within classical mentalising regions. They did, however, show reduced amygdala activity and a reduced correlation between amygdala activity and mentalising accuracy on the behavioural task, compared with controls.

Conclusions
These findings highlight the crucial role of the amygdala in making accurate implicit mental state inferences in typical development and in the social cognitive impairments of individuals with autism spectrum disorder.

Declaration of interest
None.

Copyright and usage
© The Royal College of Psychiatrists 2016.

A great number of psychiatric disorders involve severe impairments in the ability to infer others' thoughts, emotions or intentions, i.e. mentalising or theory of mind (ToM).[1] The social cognitive deficits of individuals with autism spectrum disorder are particularly prominent in real-life settings.[2] Mentalising in real life requires the integration of multimodal information from faces, voices, body language and social context,[3] and can occur explicitly, i.e. consciously taking another person's perspective. Most often, however, we infer others' mental states implicitly, i.e. spontaneously, without explicit prompts.[4] Individuals with autism spectrum disorder seem to have more pronounced deficits in the latter set of processes.[5] Despite the importance of implicit mentalising for effective social functioning in naturalistic settings, most studies have investigated explicit mentalising processes with abstract and static stimuli such as written stories[6] or cartoon tasks.[7,8] Consequently, such paradigms might lack the sensitivity to detect subtle aberrant behaviour and brain function in individuals with social impairments in real life.

A common network of brain regions has been consistently implicated in mentalising,[9] whereby different regions within the network seem to mediate explicit and implicit social cognitive processes.[10] The amygdala, for instance, is not thought to be involved in belief inferences *per se*, but rather in the implicit evaluation of arousal-related and motivational properties of affective stimuli.[10–12] In dynamic, naturalistic settings, however, decoding the social emotional saliency of stimuli is essential to accurately infer others' intentions.[13] In the same vein, the social cognitive impairments of individuals with autism spectrum disorder have been tightly linked to a dysfunctional profile of the amygdala.[14,15] In sum, the amygdala has been consistently implicated in implicit social cognition, in particular in dynamic, more naturalistic settings (see Adolphs[16] and Schilbach et al[17] for a review). Its precise role in mentalising, however, has yet to be determined. Given the amygdala's high sensitivity to dynamic, context-dependent social stimuli,[13] there is a need to study its role in mentalising with dynamic and naturalistic tasks. Furthermore, despite the crucial role of accurate social cognition for successful social functioning, the relationship between behavioural performance and neural processing has been widely neglected in the field of social neuroscience.[18] Here, we assessed naturalistic mentalising while watching videos of social interactions with a newly developed functional magnetic resonance imaging (fMRI) task in individuals with autism spectrum disorder and healthy controls. In addition, we investigated the relationship between neural processing of mental states and behavioural accuracy of implicit mental state inferences, measured with independent naturalistic tasks. We expected reduced activation of mentalising regions in particular of the amygdala during naturalistic mentalising and a reduced relationship between amygdala activity and accuracy of implicit mental state inferences in individuals with autism spectrum disorder compared with controls.

Method

Participants

Twenty-eight adults with autism spectrum disorder (18 men, mean age 33, age range 19–47) and 22 controls (16 men, mean age 31, age range 19–46) participated in the study. The autism spectrum disorder group was recruited through the autism out-patient clinic for adults of the Charité – University Medicine Berlin, Germany or were referred to us by specialised clinicians. All of the participants were diagnosed according to the DSM-IV

criteria for Asperger syndrome and autism without intellectual disabilities.[19] Diagnostic instruments included the Autism Diagnostic Observation Schedule (ADOS)[20] and the Autism Diagnostic Interview – Revised (ADI-R)[21] if parental informants were available ($n = 15$). Diagnosis was confirmed using at least one of the gold-standard instruments ADOS or ADI-R. Additionally, the diagnosis of Asperger syndrome was confirmed with the Asperger Syndrome and High-Functioning Autism Diagnostic Interview (ASDI[22]). The two groups were matched according to gender, age and verbal IQ (Table 1), as measured by a German vocabulary test (Mehrfachwahl-Wortschatz-Test (MWT)[23]). Eight of the autism spectrum disorder group met exclusion criteria for the fMRI experiment (i.e. claustrophobia: $n = 3$; no normal or corrected-to-normal vision: $n = 1$, no current health insurance: $n = 1$; under psychotropic medication: $n = 3$). The remaining 20 participants in the autism spectrum disorder group included in the fMRI experiment did not differ from the control group with respect to age, gender, and IQ (Table 1). All participants received payment for participation and gave written informed consent in accordance with the requirements of the ethics committee of the German Society for Psychology.

Tasks and material

We developed two new film-based behavioural and fMRI mentalising tasks, comprising naturalistic and dynamic stimuli, i.e. film scenes displaying social interactions between two or three professional actors.

Stimuli

The film scenes used in both the behavioural and fMRI experiments displayed different social interactions between two or three protagonists. The scenes' scripts were written in cooperation with professional scriptwriters and cover a variety of real-life social settings and scenarios such as a dinner with friends or waiting in line. The interpersonal relationships between the protagonists in the film clips vary from that of strangers to close friends or romantic partners. The storylines of the film clips were written and selected to include different, often conflicting, beliefs and perspectives of the protagonists. We further included classical ToM concepts such as false belief (online Appendix, scene 2), deception (online Appendix, scene 7) and irony (online Appendix, scene 6). A total of 30 professional actors (18 women) of varying age took part in the film clips, which were produced in the film studio of the Humboldt-Universität zu Berlin, Germany in cooperation with its Computer and Media Service team.

Indirect behavioural task

The behavioural task, Arena of Emotions – indirect behavioural task, measures the capacity to implicitly infer others' mental states

and has been introduced in a separate study.[24] In a behavioural testing session participants performed the indirect behavioural task, which is a web-based application comprising 24 items. In the indirect task, participants first watch film scenes displaying everyday social interactions between two or three protagonists (range: 9–31 s). Each film scene is preceded by a short written introduction, describing the context and setting of the interaction. After watching an initial film clip (for example a couple having breakfast), participants subsequently watch four short film clips (4 s) displaying different options for how the scene might continue (for example couple starts to argue) and are then asked to select the correct option. In order to pick the correct continuation, participants have to infer the protagonists' mental states. Importantly, mentalising performance (comprising accuracy and reaction times) is assessed indirectly, i.e. participants are not explicitly instructed to infer others' mental states (see online Fig. DS1).

fMRI task

In a separate session, participants performed the following fMRI experiment in two runs of 9 min 20 s each. The experiment was presented using Presentation (Version 14.1, Neurobehavioral Systems Inc, Albany, California) and consisted of 16 blocks of two alternately presented conditions, a ToM condition (eight blocks) and a physical inference control condition (eight blocks). Eight independent film scenes were included in the fMRI task and presented twice, once in the ToM condition, and another time in the physical inference condition. A description of the social interaction in each film clip is displayed in the online Appendix. During the ToM condition, participants were asked to make inferences about changes in the protagonists' affective states, relying on affective mentalising or cognitive empathy (see for example Walter et al[25]). During the physical inference condition participants were asked to judge changes in the protagonists' body movements.

Each task block comprised film scenes that were split into three consecutive film clips (mean duration 6.9 s, s.d. = 1.7, range 3.7–12.5 s). Task blocks started with an introductory screen (6 s), which contained information about the setting of the social interaction (for example Mrs Hauser is going to the post office) and introduced the main protagonist (picture of the main protagonist with his/her name). The introductory screen was followed by a cue (4 s) indicating the type of task participants had to solve in the respective block (ToM or physical inference). In ToM blocks, participants subsequently watched three consecutive film clips (i.e. video phases) and judged clip-by-clip changes in the affective states of the protagonists (i.e. Does the main protagonist feel 'worse', 'equally well', or 'better' in (a) the second compared with the first and (b) in the third compared with the second film clip?). The task structure resembled a

Table 1 Demographical and symptom characteristics

	Controls		Autism spectrum disorder group, total sample			Autism spectrum disorder group, fMRI sample		
	n	Mean (s.d.)	n	Mean (s.d.)	P^a	n	Mean (s.d.)	P^a
Total, n	22		28			20		
Gender, female	6	–	10		0.525	6		0.845
Age, mean (s.d.)	22	31.3 (8.5)	28	33.1 (8.5)	0.471	20	31.8 (9.3)	0.878
Mehrfachwahl-Wortschatz-Test-IQ, mean (s.d.)	22	110.8 (14.8)	28	110.8 (14.8)	0.342	20	110.8 (14.8)	0.352
Autism Diagnostic Observation Schedule, mean (s.d.)	–	–	24	10.6 (3.4)	–	19	10.4 (3.5)	–

fMRI, functional magnetic resonance imaging.
a. P-values: two-tailed significance-value for F- and χ^2-tests in autism spectrum disorder v. control group.

previously published ToM cartoon task.[25] In physical inference blocks, participants were asked to judge clip-by-clip changes in the amount of different body movements of the protagonist (for example head-turning: does Mrs Hauser turn her head less often, equally often, or more often (a) in the in the second compared to the first and (b) in the third compared with the second film clip?). Participants indicated changes in the affective state (ToM) or the amount of movement (physical inference) of the protagonist after the second and third film clip with a button press with the right hand (index finger for less, middle finger for more, index and middle finger simultaneously for equal affective state or movement). The respective answer phases had a fixed response duration of 6 s (see online Fig. DS2).

In each block participants were asked to give two responses: (a) indication of changes from the first to the second film clip (answer frame 1) and (b) indication of changes from the second to the third film clip (answer frame 2). Since a correct response in answer frame 2 partly depended on responding correctly in answer frame 1, we only included accuracy and reaction time for answer frame 2 in the analysis, if participants had answered correctly in answer frame 1.

To test whether participants paid an equal amount of visual attention to the screen between groups and conditions, we tracked participants' eye movements during the experiment (please refer to the section on eye-tracking acquisition for more information).

To ensure that the physical inference and ToM conditions did not differ with respect to task demands, for example load-related differences between the two conditions, we reduced the attentional demands as well as the social saliency of the physical inference condition by blurring the faces of the protagonists in the film clips. Thus, participants were less distracted by the affective content of the social scenes while counting body movements. To alternate in which condition the film scene would be presented first (ToM or physical inference) and to ensure a maximal distance between the same two film scenes (displayed once in the ToM and the other time in the physical inference condition), we presented the blocks in different predetermined sequences, which were counterbalanced across runs and participants.

For both tasks, the cue screen containing the task instruction indicated whether participants had to track mental states in the ToM condition or body movements in the physical inference condition. Given that we were specifically interested in capturing spontaneously occurring mentalising in naturalistic settings, we modelled video (spontaneous mentalising) and answer phases (explicit mentalising) as separate regressors and only investigated mentalising during the video phases.

Eye-tracking acquisition

To assess visual attention during the fMRI task, eye movements were recorded during scanning using a 1000 Hz embedded infrared camera (Eyelink 1000, SR Research). For each video we defined a rectangular region of interest (ROI) comprising the entire screen (1024 × 768). We subsequently analysed the number of gaze points within that ROI and the total number of gaze points tracked by the eye tracker (valid gaze points) for the duration of the videos using a MATLAB algorithm (MathWorks, Natick, Massachusetts). The latter measure was used to assess data quality for further analysis. Participants' data were discarded if there were less than 50% valid gaze points per condition (ToM or physical inference) and run. As a result, we included 11 participants with autism spectrum disorder (9 men) and 18 control participants (13 men) in the gaze-duration analysis. To assess the amount of visual attention on the screen, we divided the amount of gaze points within the screen ROI by the total number of valid gaze

points for each video. We further conducted an ROI analysis to investigate group differences in the amount of visual attention to the heads and bodies of the actors in the scenes. Given that we only obtained valid gaze data from 9 individuals in the autism spectrum disorder group and 18 in the control group, we report the exploratory analysis of group differences in the online supplement.

fMRI data acquisition

MRI data were acquired on a 3 Tesla scanner (Tim Trio; Siemens, Erlangen, Germany) using a 12-channel head coil. Functional images were acquired using an echo-planar T_2^*-weighted gradient echo pulse sequence (repetition time (TR) = 2000 ms, echo time (TE) = 30 ms, flip angle 70, 64 × 64 matrix, field of view (FOV) = 192 mm, voxel size 3 × 3 × 3 mm^3). A total of 37 axial slices (3 mm thick, no gap) were sampled for whole-brain coverage. The first two volumes at the beginning of each run were discarded to allow for T_1 equilibration. Functional imaging data were acquired in two separate 280-volume runs of 9 min 20 s each. In the same scanning session, a high-resolution T_1-weighted anatomical whole-brain scan was acquired for each participant and was later used for registration of the fMRI data (256 × 256 matrix, voxel size 1 × 1 × 1 mm^3).

fMRI data analysis

Preprocessing

The fMRI data were preprocessed and analysed using FEAT (FMRI Expert Analysis Tool) within the FSL toolbox (FMRIB's Software Library, Oxford Centre of fMRI of the Brain, www.fmrib.ox.ac.uk/fsl).[26] Preprocessing included brain extraction, slice timing, motion correction and spatial smoothing (8 mm full-width at half maximum (FWHM) Gaussian kernel). To remove low frequency artefacts, we applied a high-pass temporal filter (Gaussian-weighted straight line fitting, sigma 100 s) to the data. Functional data were first registered to individuals' T_1-weighted structural image and then registered to standard space using the FMRIB Nonlinear Image Registration Tool.

fMRI single-participant analysis

We modelled the time series individually for each participant and run including nine epoch regressors, comprising the instruction, cue, video and answer phases (separate for the two experimental conditions ToM and physical inference), as well as one regressor for all button presses. Additionally, we included six regressors modelling head movement parameters. There were no differences between the two runs, conditions or groups and in the total amount of motion between functional volumes (mean relative displacement; run (run1 v. run2): $F(1,40) = 1.56$, $P > 0.2$; condition (ToM v. physical inference): group: $F(1,40) = 0.31$, $P > 0.5$). The regressors were then convolved with a Gamma haemodynamic response function. Contrast images were computed for each condition, each run and each participant. They were spatially normalised, transformed into standard space and then submitted to a second-order within-participant fixed-effects analysis across the two runs.

Group analysis

We performed higher-level mixed-effects analyses across participants with the between-participant factor group (autism spectrum disorder and control group) using the FMRIB local analysis of mixed-effects tool provided by FSL (FLAME, stage 1 and 2) based on single participants' contrast images. To identify the mentalising network (see Marr[27]) in the two groups we contrasted ToM with physical inference video phases (ToM contrast). Subsequently, we investigated group differences in the

ToM contrast in a whole-brain analysis. Given our specific hypothesis of group differences in amygdala activation, we additionally performed a ROI analysis using an anatomical mask of the bilateral amygdala for small volume correction. For all mentioned analyses we report clusters of maximally activated voxels that survived family-wise error correction for multiple comparisons at a statistical threshold of $P < 0.05$ and a Z-value > 2.3. To further visualise and characterise possible group × condition interactions for significantly activated clusters, we extracted and plotted the parameter estimates for each condition and group.

Covariate analysis

To determine whether amygdala activation during mentalising processing (ToM video phases v. physical inference video phases) correlated with implicit mentalising performance in healthy controls, we performed a higher-level mixed-effects analysis for the control and autism spectrum disorder groups separately and added accuracy rates for the Arena of Emotions task as a covariate in the model. Crucially, we ensured that this analysis was not prone to potential non-independence errors by taking two measures. First, we entered performance on an independent behavioural task as a covariate in the fMRI analysis. Second, we used an anatomically defined ROI of the amygdala (see description above). To further investigate whether individuals with autism spectrum disorder and controls exhibit different relationships between changes in individual blood oxygen level-dependent (BOLD) responses of the amygdala and accuracy scores in the behavioural task (group × accuracy interaction), respectively, we included individual accuracy scores as a covariate in the higher-level model for each group separately. To visualise the strength and direction of the correlations between changes in neural activity and the behavioural covariate we extracted parameter estimates from the activated clusters identified in the contrast of interest.

Results

Behavioural results

Indirect behavioural task performance

The behavioural results for the Arena of Emotions task have been previously reported in a separate study.[24] In summary, the indirect task was reliable and sensitive to the mentalising impairments of individuals with autism spectrum disorder. The total autism spectrum disorder group ($n = 28$) had significantly fewer correct responses than the control group and the autism spectrum disorder group's performance on the task was negatively correlated with autism symptomatology (see Rosenbau $et\ al$[24] for further details).

fMRI task – behavioural data

To test whether the two experimental conditions (ToM and physical inference) were comparable with regards to task difficulty across all participants, we performed repeated measures ANOVAs separately for accuracy rates and correct reaction time with the within-participant factor condition (ToM/physical inference) and the between-participant factor group (control/autism spectrum disorder). There were no significant differences in accuracy rates (condition: $F(1,40) = 2.22$, $P > 0.1$) or correct reaction time (condition: $F(1,40) = 0.76$, $P > 0.1$) between the two conditions. Furthermore, the autism spectrum disorder group showed a trend for lower accuracy on the fMRI task in general (group: $F(1,40) = 2.97$, $P < 0.10$, $\eta_p^2 = 0.07$), but no differences in reaction time (group: $F(1,40) = 0.38$, $P > 0.5$). There was no significant interaction between group and condition (ToM/physical inference) for accuracy rates and reaction times (accuracy: condition × group interaction: $F(1,40) = 0.58$, $P > 0.50$; correct reaction time: condition × group interaction: $F(1,40) = 0.14$, $P > 0.50$; Table 2). The lack of a significant group × condition interaction suggests that both groups tracked changes in the protagonists' emotional state to a similar extent. Thus, between-group differences in BOLD activation during the ToM condition reflect differences in processing mental states and cannot be accounted for by group differences in task difficulty. See online supplement for additional analysis on accuracy-corrected reaction times.

Eye-tracking analysis

To test whether in both groups participants paid an equal amount of attention to the screen in both runs and both conditions, we performed a repeated measures ANOVA on eye gazes with the within-participant factors run (run1 v. run2) and condition (ToM v. physical inference) and the between-participant factor group (controls v. autism spectrum disorder). In both runs ($F(1,27) = 2.45$, $P > 0.10$) participants paid an equal amount of attention to both conditions ($F(1,27) = 0.01$, $P > 0.90$). Furthermore, there was no significant difference in visual attention to the screen between groups ($F(1,27) = 0.321$, $P > 0.50$) and no significant interactions between the factors.

fMRI results

Mentalising-related neural activity

The contrast ToM v. physical inference video phases yielded robust activations ($P < 0.05$, $Z = 2.3$, cluster corrected) of mentalising regions including the medial prefrontal cortex, the left temporo-parietal junction, the left superior temporal sulcus, the bilateral temporal poles, the bilateral inferior frontal gyrus, the insula, and bilateral amygdala (Fig. 1(a) and Table 3).

Differences in mentalising-related neural activity

We did not find between-group differences in the whole-brain analysis. With respect to the ROI analysis, we found significantly increased activation of the bilateral amygdala in the control compared with the autism spectrum disorder group (left:

| **Table 2** Functional magnetic resonance imaging (fMRI) task performance		
	Mean	
	Control group ($n = 22$)	Autism spectrum disorder fMRI group ($n = 20$)
Accuracy in the theory of mind condition	0.72 (0.18)	0.66 (0.18)
Accuracy in the physical inference condition	0.67 (0.14)	0.59 (0.17)
Reaction times for correctly solved items in the theory of mind condition, s	1.25 (0.37)	1.34 (0.61)
Reaction times for correctly solved items in the physical inference condition, s	1.21 (0.34)	1.26 (0.42)
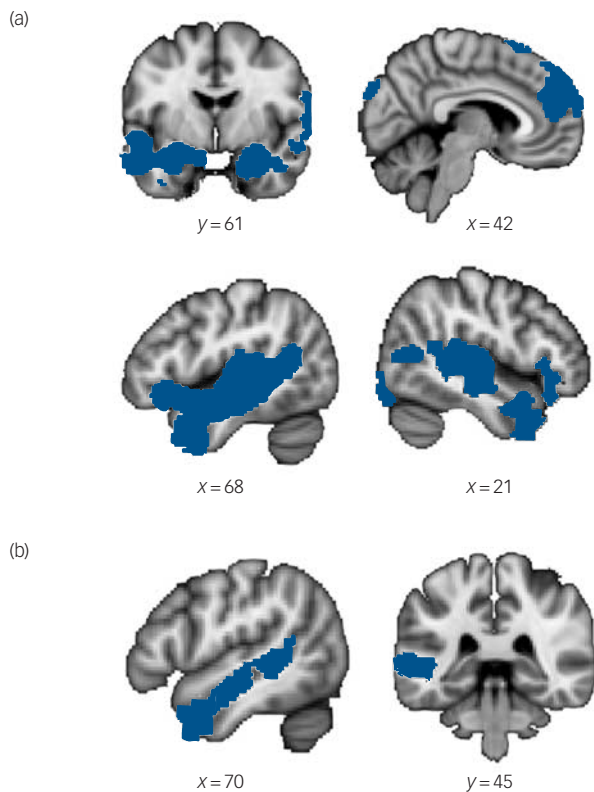
Fig. 1 Theory of mind (ToM)-related blood oxygen level-dependent (BOLD) signal in (a) the control group and (b) the autism spectrum disorder group.

Network of brain regions showing significantly greater activation during theory of mind video phases compared with physical inference video phases in (a) the control group and (b) in the autism spectrum disorder group.

$x = -16$, $y = -8$, $z = -16$, peak Z-score 3.29, 39 voxels; right: $x = 24$, $y = -6$, $z = -22$, peak Z-score 2.76, 25 voxels; Fig. 2(a)).

Brain behaviour relationship

Amygdala activation and performance on the indirect behavioural mentalising task

To identify regions within the ToM network that predict implicit ToM performance, we investigated the relationship between performance on the behavioural Arena of Emotions task and changes in BOLD signal during the video phases in the ToM v. physical inference condition. Activation in the left amygdala was positively correlated with performance on the behavioural task in the control group ($x = -26$, $y = -2$, $z = -24$, peak Z-score 4.07, 78 voxels) but not in the autism spectrum disorder group. Furthermore, a smaller cluster within the previously identified left amygdala correlated more strongly with performance on the Arena of Emotions task in the control than in the autism spectrum disorder group ($x = -26$, $y = -2$, $z = -24$, peak Z-score 3.06, 21 voxels; cluster corrected within anatomical mask, Fig. 2(b)). No significant correlation was found between superior temporal sulcus activation and performance on Arena of Emotions task in either group.

Discussion

Main findings

The goal of this study was to identify the neural underpinnings of naturalistic mental state inferences with a video-based task in typical and impaired social cognition. In addition, we investigated the relationship between neural processing and accuracy of mentalising. In typically developed controls, inferring mental states from dynamic, naturalistic videos of social interaction

Table 3 Significant activations in videos phases[a]

	Side	Cluster size, voxel	Peak voxel MNI coordinates, mm			Peak Z-score
			x	y	z	
Theory of mind videos > physical inference videos in the control group						
Cluster 1		10 932				
Superior temporal sulcus	Left		−48	−28	−4	7.69
Middle temporal gyrus	Left		−52	−2	−20	6.27
Temporal pole	Left		−52	4	−22	6.06
Temporoparietal junction	Left		−54	−44	8	5.83
Cluster 2		8812				
Superior temporal gyrus/Heschl's gyrus	Right		44	−30	8	6.28
Amygdala	Right		30	0	−26	4.76
Temporal pole	Right		52	14	−36	4.69
Cluster 3		4720				
Medial prefrontal cortex	Right		2	54	28	5.78
	Left		−2	58	22	5.45
Superior frontal gyrus	Right		6	58	32	5.39
	Left		−6	24	62	5.26
Left inferior frontal gyrus	Left		−46	26	−8	4.81
Amygdala	Left		−20	−6	−14	3.90
Right inferior frontal gyrus	Right		42	30	−14	3.75
Insula	Left		−32	−28	10	3.72
Theory of mind videos > physical inference videos in autism spectrum disorder group						
Cluster		2696				
Temporal pole	Left		−46	10	−38	3.98
Inferior frontal gyrus	Left		−38	24	−22	3.79
Superior temporal sulcus[b]	Left		−60	−40	6	3.71
			−52	−12	−10	3.65

MNI, Montreal Neurological Institute.
a. All reported clusters are family-wise error cluster corrected for multiple comparisons at $P < 0.05$ and $z = 2.3$.
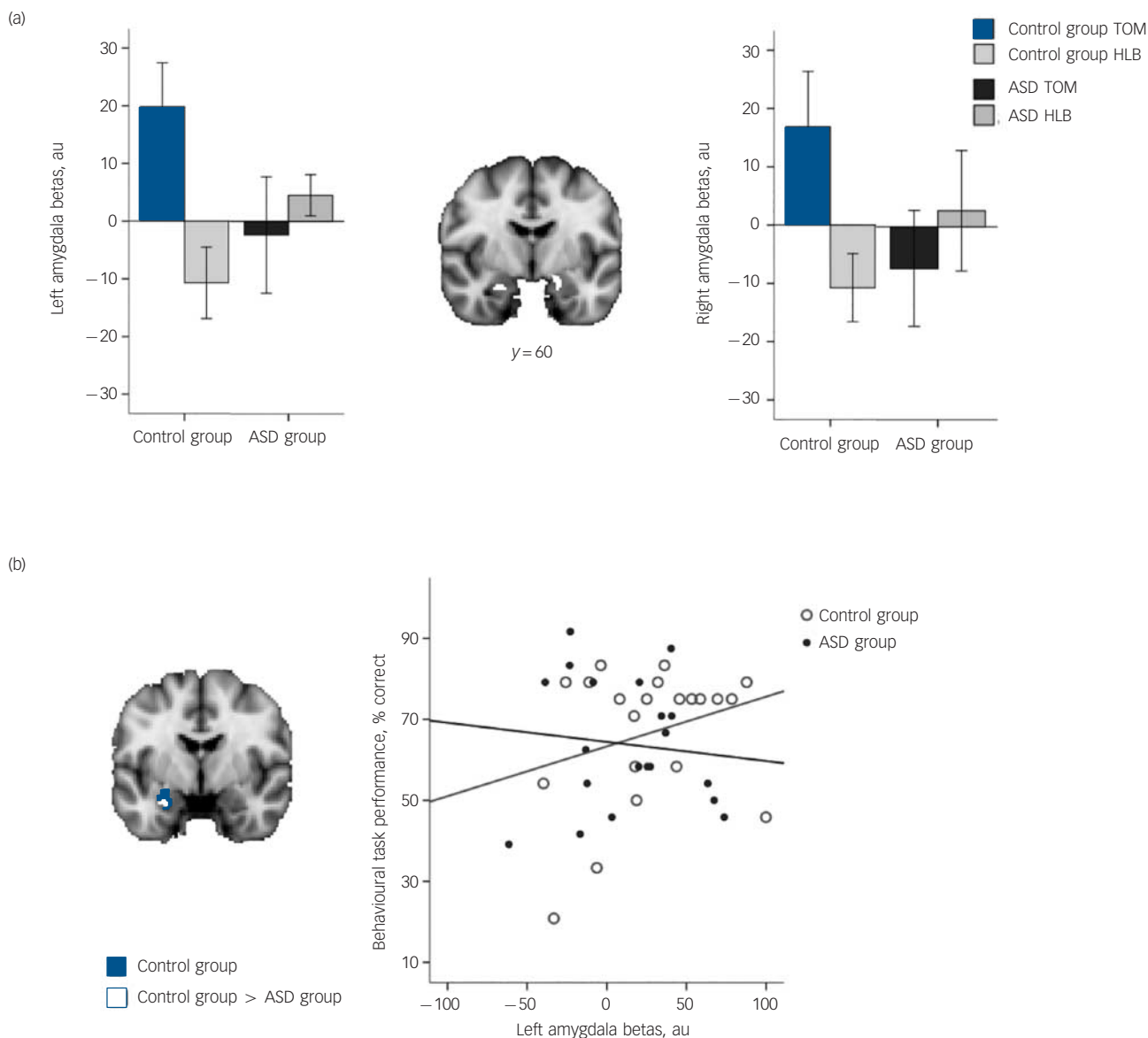b. Two peaks were observed for the superior temporal sulcus.

(a)

(b)

Fig. 2 Group differences in theory of mind (ToM)-related blood oxygen level-dependent (BOLD) signal and the relationship between BOLD signal changes and implicit ToM performance.

(a) Higher bilateral amygdala activations in the control group than in the autism spectrum disorder (ASD) group during ToM videos compared with physical inference videos. Parameter estimates (arbitrary units (au)) extracted from the bilateral amygdala are illustrated in bar graphs. Error bars indicate standard error of mean. (b) Changes in BOLD signal in the left amygdala during ToM videos compared with physical inference videos were correlated with implicit ToM performance on an independent behavioural ToM task in the control group (blue). Changes in BOLD signal of a cluster within the left amygdala (white) and implicit ToM performance were more strongly correlated in the control group than in the autism spectrum disorder group. The correlation plot displays the strength and direction of correlation between parameter estimates (arbitrary units (au)) extracted from the left amygdala (white) and accuracy on the independent behavioural ToM task.

yielded activity of regions within the well-established mentalising network (see for example Mar[27]). Additionally the network comprised the amygdala and insula, which play a prominent role in emotion processing.[16] This suggests greater emotional engagement during mentalising assessed with naturalistic as compared with traditional, static tasks. There were no significant differences between individuals with autism spectrum disorder and controls in neural activity of regions typically assigned to the mentalising network. The study further revealed the crucial role of the amygdala during spontaneous online mental state processing in typically developed individuals. Amygdala activity significantly predicted accurate implicit mental state inferences on an independent task in controls. Finally, reduced amygdala activity and a reduced relationship between amygdala activity and accuracy of implicit mentalising in individuals with autism spectrum disorder highlight the role of the amygdala in their social deficits.

Comparison with findings from other studies

Naturalistic mentalising, while watching complex social interactions, robustly activated regions assigned to the mentalising network in healthy controls. Besides activations in the classical mentalising regions such as the medial prefrontal cortex, temporoparietal junction, superior temporal sulcus, temporal poles and the inferior frontal gyrus,[27] naturalistic mentalising yielded activity of the bilateral insula and amygdala, regions strongly implicated in affective processing and empathy (for

example Carr *et al*[28]). The engagement of these regions in mentalising has been previously observed during judgements of others' affective states[29] or processing of highly dynamic and context-dependent stimuli.[30] Amygdala and insula activation is likely a result of a combination of both inferring the protagonists' affective states and processing dynamic naturalistic stimuli, which requires a multisensory integration of social information. As Schillbach and colleagues[17] point out, the amount of emotional engagement with the task (passive observer reads about others' mental states *v.* observer engages emotionally in a social interaction) fundamentally changes the social cognitive processes as well as the underlying neural networks. Our study thus extends previous research by showing that amygdala and insula are involved in mental state processing in emotionally engaging naturalistic settings.

As previously described,[24] our indirect behavioural task sensitively reflected the social deficits of individuals with autism spectrum disorder in implicit mentalising. Task accuracy and visual attention to the screen did not differ between groups in the fMRI adaptation of the task. Furthermore, our supplemental analyses indicate that there were no differences between conditions or groups in speed-accuracy trade-off effects. Thus, differences in neural activation between groups represent differences in mental state processing and are unlikely to represent more general load-related differences.

As expected, the control group showed greater changes in BOLD signal in the amygdala during naturalistic mental state processing than the autism spectrum disorder group. Our results are in line with a vast body of literature linking the amygdala to implicit processing of dynamic social information[31] and to the typical development of mentalising abilities.[15,32] Our findings extend the literature by showing that the amygdala not only supports the development of mentalising, but also plays a central role in mental state inferences in complex, unstructured social settings. Importantly, the social deficits of individuals with autism spectrum disorder have also been tightly linked to aberrant amygdala activation.[14,16,33,34] In this study we further specified the role of these brain regions for the pervasive mentalising deficits of individuals with autism spectrum disorder in naturalistic settings.

Contrary to our expectations and in contrast to previous studies (for example Ciaramidaro *et al*[35]), we did not find group differences in temporoparietal junction and medial prefrontal cortex activation for the ToM condition. One possible explanation for this could be that processing abstract and static ToM tasks relies to a higher degree on classical ToM regions[6,10] than our naturalistic task, and the control group might be able to engage these regions to a higher degree than individuals with autism spectrum disorder. In contrast, processing naturalistic stimuli as in the current study relies to a much larger extent on the online integration of a variety of sources of social information including facial expressions, prosody and biological motion,[3] which relies to a great extent on the social perception system, in particular on the amygdala.[33,34] The task used here could thus be more sensitive in revealing a dysfunctional profile of the amygdala in response to naturalistic social cognition in individuals with autism spectrum disorder. In sum, it is possible that the observed ToM impairments of individuals with autism spectrum disorder in naturalistic tasks might not be associated with aberrant activity in the ToM network but instead with dysfunctional activity of the social perception network. That being said, it is still a matter of debate whether individuals with autism spectrum disorder show reduced activity of ToM regions. A recent study by Dufour and colleagues[36] investigating neural activity during a story-based false-belief task in individuals with autism spectrum disorder compared with a large sample of typically developed controls did not find group differences. Following the authors' interpretations, it is possible that the group differences in ToM would be more pronounced when assessing spontaneous ToM or participants with autism spectrum disorder that are lower functioning.

Finally, our results highlight the specific role of the amygdala in making implicit social cognitive judgements by showing that in typical development amygdala activation is related to the accuracy of implicit social inferences, whereas in autism this relationship is reduced. This finding is in accordance with previous studies, that showed that the magnitude of amygdala activation was related to the implicit evaluation of social information. Amygdala activity correlated with implicitly judging another person's intention as deceptive or not,[37] adjusting to a competent advisor's opinion,[38] as well as to accurate emotion recognition.[39] Our study thus provides evidence for the crucial role of the amygdala for accurate implicit mental state inferences in typical development and in the pathophysiology of autism.

Implications

Most research aiming at capturing implicit aspects of mentalising typically operationalises mentalising in an abstract fashion (for example slower response times during decision-making tasks when having to take into account another protagonist's visual perspective). Whether this interference effect provides evidence for participants' tendency to implicitly take into account another's perspective is a matter of debate. An alternative explanation is that these effects are as are result of domain general attentional demands that are not specific to social settings.[40,41] In this study, implicit mentalising is defined in a narrower and more naturalistic sense. The indirect mentalising does not explicitly prompt participants to infer others' mental states but does include an objective measure of participants' mentalising accuracy. Participants watch film clips portraying everyday social interactions and have to select the correct continuation of the initial clips from four different film clip options. The four film clip options do not differ with respect to non-social features (number or position of actors, background etc.) Thus, to select the correct option, participants have to correctly infer the protagonists' mental states and intentions in the initial clip.

In the present study autism spectrum disorder provides a model of impaired social cognition. By comparing neural processing of naturalistic ToM in controls *v.* individuals with autism spectrum disorder, we can conclude that the amygdala is important for unimpaired higher-level social cognition in naturalistic settings. Our findings also have important implications for diagnosing and treating social impairments of individuals with autism spectrum disorder. We showed that naturalistic tasks are a sensitive means of assessing social cognitive impairments of individuals with autism spectrum disorder on the behavioural and neural level. Furthermore, given that an aberrant profile of the amygdala, which is crucial to detecting socially salient cues in early childhood, is related to mentalising impairments in adulthood, early social cognitive interventions addressing deficits in social orientation of individuals with autism spectrum disorder might alter social perception and amygdala activity and thus possibly mitigate later ToM impairments.

Directions for future research

Our findings may be extended to other psychiatric disorders characterised by social impairments, such as schizophrenia or borderline personality disorder. These disorders involve an aberrant functional profile of the amygdala[42,43] as well as ToM

impairments.[8,44,45] Future studies need to investigate whether the relationship between amygdala dysfunction and ToM in naturalistic settings holds true for schizophrenia and borderline personality disorder and if so whether their symptomatology can be mitigated by treatments that focus on mentalising (for example Dimaggio et al[46] and Lysaker & Dimaggio[47]). In sum, amygdala functioning might represent a biomarker that predicts response to social cognitive treatments in a number of psychiatric conditions characterised by social impairments.

That being said, we cannot rule out that increased amygdala activation during the ToM v. physical inference condition in controls as well as between-group differences in amygdala activation could also be partly explained by the fact that the stimuli included in the ToM and physical inference condition differ with regards to the degree of social saliency (the protagonists' faces were blurred in the physical inference but not in the ToM condition). In everyday social settings, however, facial affect recognition represents an important part of understanding intentions[3] and that again makes it difficult to disentangle these processes in general. Also, the correlation between amygdala activation and accuracy on the external behavioural task in the control group may be in part explained by the fact that facial affect recognition plays an important role in both the behavioural mentalising task and the fMRI ToM condition. It is somewhat unlikely, however, that the correlation between performance on an external mentalising task and group differences in mental state processing, both assessed with naturalistic social stimuli including facial expressions, body language, tone of voice and specific context information, merely reflect differences in facial affect recognition. Furthermore, the quality of the eye-tracking data was significantly lower in the autism spectrum disorder group compared with the control group. Thus, this study cannot reliably address the question of whether differences in gaze behaviour influence the observed group differences in brain function or the relationship between mentalising accuracy and amygdala response. These questions need to be addressed in future studies.

In conclusion, the pervasive deficits of individuals with autism spectrum disorder in spontaneous mental state inference are associated with aberrant amygdala activation, which has been shown to play a crucial role in the development of mentalising abilities. Whereas in typically developed controls amygdala activation during spontaneous mentalising mediates accurate implicit mental state inferences on an independent task, in atypical development this coupling between brain response and behaviour is significantly reduced. Our results thus highlight the important role of the amygdala for implicit mental state inferences in naturalistic settings and point to the important role of this region for disorders characterised by social cognitive deficits. Finally, our study stresses the importance of investigating mentalising with accuracy-based naturalistic tasks in typical development as well as in clinical disorders characterised by social cognitive deficits.

Gabriela Rosenblau, PhD, Cluster of Excellence 'Languages of Emotion' and Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany and Center for Translational Developmental Neuroscience, Yale Child Study Center, Yale University, New Haven, Connecticut, USA; **Dorit Kliemann**, PhD, Cluster of Excellence 'Languages of Emotion' and Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany, McGovern Institute for Brain Research at Massachusetts Institute of Technology, Cambridge, Massachusetts and Department of Neurology, Massachusetts General Hospital/Harvard Medical School, Charlestown, Massachusetts, USA; **Benjamin Lemme**, BA, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany; **Henrik Walter**, MD, PhD, Department of Psychiatry, Division of Mind and Brain Research, Charité Universitätsmedizin Berlin, Berlin, Germany; **Hauke R. Heekeren**, MD, Cluster of Excellence 'Languages of Emotion', Department of Education and Psychology and Dahlem Institute for Neuroimaging of Emotion, Freie Universität Berlin, Berlin, Germany; **Isabel Dziobek**, PhD, Cluster of Excellence 'Languages of Emotion', Freie Universität Berlin, Berlin and School of Mind and Brain, Humboldt-Universität zu Berlin, Germany

Correspondence: Gabriela Rosenblau, PhD, Yale Child Study Center, 230 S. Frontage Road, New Haven, CT 06519, USA. Email: gabriela.rosenblau@yale.edu

First received 16 Apr 2014, final revision 22 Dec 2015, accepted 23 Feb 2015

References

1 Frith U. *Autism: Explaining the Enigma*. Blackwell Publishing, 1989.

2 Volkmar FR, Lord C, Bailey A, Schultz RT, Klin A. Autism and pervasive developmental disorders. *J Child Psychol Psychiatry* 2004; **45**: 135–70.

3 Zaki J, Ochsner K. The need for a cognitive neuroscience of naturalistic social cognition. *Ann N Y Acad Sci* 2009; **1167**: 16–30.

4 Frith CD, Frith U. Implicit and explicit processes in social cognition. *Neuron* 2008; **60**: 503–10.

5 Senju A, Southgate V, White S, Frith U. Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science* 2009; **325**: 883–5.

6 Saxe R, Powell LJ. It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol Sci* 2006; **17**: 692–9.

7 Walter H, Adenzato M, Ciaramidaro A, Enrici I, Pia L, Bara BG. Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *J Cogn Neurosci* 2004; **16**: 1854–63.

8 Walter H, Schnell K, Erk S, Arnold C, Kirsch P, Esslinger C, et al. Genetic risk for schizophrenia impacts Theory-of-Mind-related brain activation. *Mol Psychiatry* 2011; **16**: 353.

9 Mar RA. The neural bases of social cognition and story comprehension. *Annu Rev Psychol* 2011; **62**: 103–34.

10 Olsson A, Ochsner KN. The role of social cognition in emotion. *Trends Cogn Sci* 2008; **12**: 65–71.

11 Adolphs R. Recognizing emotion from facial expressions: psychological and neurological mechanisms. *Behav Cogn Neurosci Rev* 2002; **1**: 21–62.

12 Gobbini MI, Leibenluft E, Santiago N, Haxby JV. Social and emotional attachment in the neural representation of faces. *Neuroimage* 2004; **22**: 1628–35.

13 Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex* 2003; **13**: 1034–43.

14 Dziobek I, Bahnemann M, Convit A, Heekeren HR. The role of the fusiform-amygdala system in the pathophysiology of autism. *Arch Gen Psychiatry* 2010; **67**: 397–405.

15 Pelphrey KA, Shultz S, Hudac CM, Vander Wyk BC. Research review: constraining heterogeneity: the social brain and its development in autism spectrum disorder. *J Child Psychol Psychiatry* 2011; **52**: 631–44.

16 Adolphs R. The social brain: neural basis of social knowledge. *Annu Rev Psychol* 2009; **60**: 693–716.

17 Schilbach L, Timmermans B, Reddy V, Costall A, Bente G, Schlicht T, et al. Authors' response: a second-person neuroscience in interaction. *Behav Brain Sci* 2013; **36**: 441–62.

18 Zaki J, Ochsner K. Reintegrating the study of accuracy into social cognition research *Psychol Inq* 2011; **22**: 159–82.

19 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (4th edn, revised) (DSM-IV-TR)*. APA, 2000.

20 Lord C, Rutter M, Dilavore P, Risi S. *Autism Diagnostic Observation Schedule*. Western Psychological Services, 2002.

21 Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* 1994; **24**: 659–85.

22 Gillberg C, Rastam M, Wentz E. The Asperger Syndrome (and high-functioning autism) Diagnostic Interview (ASDI): a preliminary study of a new structured clinical interview. *Autism* 2001; **5**: 57–66.

23 Lehrl S. *Mehrfachwahl-Wortschatz-Intelligenztest*. [Multiple-Choice Vocabulary Test.] Medizinische Verlagsgesellschaft, 1989.

24 Rosenblau G, Kliemann D, Heekeren HR, Dziobek I. Approximating implicit and explicit mentalizing with two naturalistic video-based tasks in typical development and autism spectrum disorder. *J Autism Dev Disord* 2014; **45**: 953–65.

25 Walter H, Schnell K, Erk S, Arnold C, Kirsch P, Esslinger C, et al. Effects of a genome-wide supported psychosis risk variant on neural activation during a theory-of-mind task. *Mol Psychiatry* 2011; **16**: 462–70.

26 Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, et al. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* 2004; **23** (Suppl 1): S208–19.

27 Mar RA. The neural bases of social cognition and story comprehension. *Annu Rev Psychol* 2011; **62**: 103–34.

28 Carr L, Iacoboni M, Dubeau MC, Mazziotta JC, Lenzi GL. Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proc Natl Acad Sci U S A* 2003; **100**: 5497–502.

29 Baron-Cohen S, Ring HA, Bullmore ET, Wheelwright S, Ashwin C, Williams SC. The amygdala theory of autism. *Neurosci Biobehav Rev* 2000; **24**: 355–64.

30 Adolphs R. What does the amygdala contribute to social cognition? *Ann NY Acad Sci* 2010; **1191**: 42–61.

31 Adolphs R, Gosselin F, Buchanan TW, Tranel D, Schyns P, Damasio AR. A mechanism for impaired fear recognition after amygdala damage. *Nature* 2005; **433**: 68–72.

32 Heberlein AS, Adolphs R. Impaired spontaneous anthropomorphizing despite intact perception and social knowledge. *Proc Natl Acad Sci USA* 2004; **101**: 7487–91.

33 Baron-Cohen S, Ring HA, Wheelwright S, Bullmore ET, Brammer MJ, Simmons A, et al. Social intelligence in the normal and autistic brain: an fMRI study. *Eur J Neurosci* 1999; **11**: 1891–8.

34 Kliemann D, Dziobek I, Hatri A, Baudewig J, Heekeren HR. The role of the amygdala in atypical gaze on emotional faces in autism spectrum disorders. *J Neurosci* 2012; **32**: 9469–76.

35 Ciaramidaro A, Bolte S, Schlitt S, Hainz D, Poustka F, Weber B, et al. Schizophrenia and autism as contrasting minds: neural evidence for the hypo-hyper-intentionality hypothesis. *Schizophr Bull* 2015; **41**: 171–9.

36 Dufour N, Redcay E, Young L, Mavros PL, Moran JM, Triantafyllou C, et al. Similar brain activation during false belief tasks in a large sample of adults with and without autism. *PloS One* 2013; **8**: e75468.

37 Grezes J, Frith C, Passingham RE. Brain mechanisms for inferring deceit in the actions of others. *J Neurosci* 2004; **24**: 5500–5.

38 Schilbach L, Eickhoff SB, Schultze T, Mojzisch A, Vogeley K. To you I am listening: perceived competence of advisors influences judgment and decision-making via recruitment of the amygdala. *Soc Neurosci* 2013; **8**: 189–202.

39 Derntl B, Habel U, Windischberger C, Robinson S, Kryspin-Exner I, Gur RC, et al. General and specific responsiveness of the amygdala during explicit emotion recognition in females and males. *BMC Neurosci* 2009; **10**: 91.

40 Santiesteban I, Catmur C, Hopkins SC, Bird G, Heyes C. Avatars and arrows: implicit mentalizing or domain-general processing? *J Exp Psychol Hum Percept Perform* 2014; **40**: 929–37.

41 Dolk T, Hommel B, Colzato LS, Schutz-Bosbach S, Prinz W, Liepelt R. How ''social'' is the social Simon effect? *Front Psychol* 2011; **2**: 84.

42 Donegan NH, Sanislow CA, Blumberg HP, Fulbright RK, Lacadie C, Skudlarski P, et al. Amygdala hyperreactivity in borderline personality disorder: implications for emotional dysregulation. *Biol Psychiatry* 2003; **54**: 1284–93.

43 Rasetti R, Mattay VS, Wiedholz LM, Kolachana BS, Hariri AR, Callicott JH, et al. Evidence that altered amygdala activity in schizophrenia is related to clinical state and not genetic risk. *Am J Psychiatry* 2009; **166**: 216–25.

44 Semerari A, Cucchi M, Dimaggio G, Cavadini D, Carcione A, Battelli V, et al. The development of the Metacognition Assessment Interview: instrument description, factor structure and reliability in a non-clinical sample. *Psychiatry Res* 2012; **200**: 890–5.

45 Lysaker PH, Leonhardt BL, Brune M, Buck KD, James A, Vohs J, et al. Capacities for theory of mind, metacognition, and neurocognitive function are independently related to emotional recognition in schizophrenia. *Psychiatry Res* 2014; **219**: 79–85.

46 Dimaggio G, Salvatore G, Fiore D, Carcione A, Nicolo G, Semerari A. General principles for treating personality disorder with a prominent inhibitedness trait: towards an operationalizing integrated technique. *J Person Disord* 2012; **26**: 63–83.

47 Lysaker PH, Dimaggio G. Metacognitive capacities for reflection in schizophrenia: implications for developing treatments. *Schizophr Bull* 2014; **40**: 487–91.