Modality-specific dysfunctional neural processing of social and non-social information in schizophrenia

Yifei He[1,4]*, Miriam Steines[1,4], Gebhard Sammer[2], Arne Nagels[3], Tilo Kircher[1,4], Benjamin Straube[1,4]

1. Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Marburg, Germany
2. Cognitive Neuroscience at Centre for Psychiatry, Justus-Liebig University Giessen, Giessen, Germany
3. Department of General Linguistics, Johannes-Gutenberg University Mainz, Mainz, Germany
4. Center for Mind, Brain and Behavior - CMBB, Hans-Meerwein-Straße 6, 35032 Marburg, Germany

*Corresponding address: Yifei He, Department of Psychiatry and Psychotherapy, Philipps University Marburg, Rudolf-Bultmann-Str. 8, 35039, Marburg, Germany

* Corresponding email: yifei.he@staff.uni-marburg.de, ORCID: 0000-0002-2826-4230

Abstract

Background: Schizophrenia (SZ) is characterized by marked social dysfunctions encompassing potential deficits in the processing of social and non-social information, especially in everyday settings where multiple modalities are present. To date, the neurobiological basis of these deficits remains elusive.

Methods: In a functional magnetic resonance imaging (fMRI) study, 17 patients with schizophrenia or schizoaffective disorder, and 18 matched controls watched videos of an actor speaking, gesturing (unimodal), and both speaking and gesturing (bimodal) about social or non-social events in a naturalistic way. Participants had to judge whether each video contains person-related (social) or object-related (non-social) information.

Results: When processing social content, controls activated the medial prefrontal cortex (mPFC) for both speech and gesture conditions; patients, in comparison to controls, showed no different activation in the speech condition but reduced activation in the mPFC in the gesture condition. For non-social content, across modalities, controls recruited the bilateral pre/postcentral gyrus, superior temporal gyrus, and insula, as well as the left occipitotemporal cortex; patients showed reduced activation of the left postcentral gyrus and the right insula only in the speech condition. Moreover, in the bimodal conditions, patients displayed improved task performance and comparable activation to controls in both social and non-social content.

Conclusions: Patients with SZ displayed modality-specific aberrant neural processing of social and non-social information, which is not present for the bimodal conditions. This finding provides novel insights into dysfunctional social cognition in SZ, and may have potential therapeutic implications.

Key words: social, multimodal processing, mPFC, gesture, speech, mentalizing, schizophrenia

Introduction

Dysfunctional social cognition is a hallmark feature of schizophrenia (SZ) (1, 2). Social cognition entails subdomains ranging from perception of social information to mentalizing and social interaction (3). Patients with SZ, however, commonly suffer from deficits in these processes. These deficits, correlated with negative symptoms (4), contribute directly to impairments in social functioning (5). Among these processes, an important field in SZ research is the processing of social-relevant stimuli and its underlying neurobiological mechanisms.

In daily social communication, individuals encounter a diverse spectrum of information from multiple modalities (6). These include non-linguistic inputs from others' facial expressions, body movements, e.g., postures and gestures, as well as linguistic stimuli from auditory speech and written texts. Importantly, the multimodal information comprises social and non-social information to individuals, and the processing of this information forms the basis to further mentalize social intentions, and to perform appropriate social interaction with others (7). To date, despite substantial reports of deficits in social processes in SZ (2), neuroscientific studies on SZ's social dysfunction have most extensively investigated aberrant emotional perception of faces (8), and to a lesser degree, voices (9), while limited research has directly examined potentially aberrant perception of social (person-related) and non-social (object-related) information. In basic research, a seminal fMRI (functional magnetic resonance imaging) study on social information processing, using linguistic stimuli, has identified distinct brain regions for processing socially and non-socially relevant information (10): when healthy participants were asked to judge whether visual word pairs are person- or object-related, person-related social stimuli activated the medial prefrontal cortex (mPFC), a crucial region forming the mentalizing network (7, 11-13). This functional relevance of the mPFC for processing social information has been replicated in later studies exploiting comparable tasks (14, 15). In addition, non-social vs. social content comparison elicited a network including the bilateral insula and the left parietal lobe, these are regions typically reported for processing concrete objects or tools (16, 17). Importantly, this line of research has primarily examined linguistic stimuli such as single written words, largely neglecting the multimodal nature of social and non-social information in everyday life: for instance, marked social or non-social features can be delivered by either a “be silent” emblem or a “hammering” pantomime via hand gestures, without providing any linguistic information. Moreover, it is common to use both gesture and speech together. For example, to ask someone else to stop, we often use a “stop” gesture (e.g., a raised hand) together with its verbal counterpart.

Interestingly, irrespective of encoding modality, processing of social information is shown to consistently activate a mentalizing network including the mPFC, at least in healthy individuals (13, 18). For example, It is reported that a left-lateralized network, including the mPFC, is activated when processing social-abstract information encoded in both auditory speech and visual gestures (18). This ‘supramodal’ nature of social information processing further concurs with the role of the mPFC (and the mentalizing network) for a wide range of social tasks based on linguistic and non-linguistic stimuli (19). For non-social concrete information, literature also suggests that humans may recruit brain networks that is modality independent (20, 21). These characteristics of social and non-social information processing may have profound implications in SZ research: given the well-documented deficits of SZ in social processes (2, 22), it remains unknown whether social information processing is impaired in SZ, and if so, whether this potential deficit is modality dependent. Similarly, for processing of concrete, object-related non-social information, previous literature suggests that patient’s processing may be impaired at least in visually presented linguistic form (23). In the form of hand action, however, the reports on SZ’s potential neural deficits are mixed (24, 25). Importantly, to date, no prior study has directly compared social/non-social information processing in speech and gesture in SZ.

The current study directly addresses these gaps. We presented videos of an actor communicating in a spontaneous and naturalistic manner. The actor performs either social (person-related) or non-social (object-related) content in different modalities, where social and non-social features are perceivable in gesture- and speech-only modalities. Similar to approaches from previous research (10, 18), we directly compared social vs. non-social videos, so as to identify neural perception of social and non-social information in both speech and gesture modalities. Besides, we also showed to participants videos with bimodal inputs (actor both speaking and gesturing). Based on previous basic research, we hypothesized that the mentalizing network supports the processing of social (person) information (10), and that non-social (object) information processing will activate left-lateralized regions including the lateral occipitotemporal cortex (LOTC), the superior temporal gyrus/sulcus (STG/STS), as well as pre/postcentral gyri forming the putative mirror neuron system (26-29). We focused on group differences between a group of patients suffering from schizophrenia or schizoaffective disorder, and their age- and education-matched controls: for processing social content, we expected patients to show reduced activation in the mentalizing regions, irrespective of encoding modality (22); for non-social content, despite mixed findings from previous neuroimaging research on hand action observation on SZ (24, 25), following previous report on dysfunctional processing of non-social linguistic stimuli in SZ (23), we hypothesized neural modulation of the object-related regions for patients with SZ for both gesture and speech modalities. Additionally, we hypothesized that the presence of bimodal content might compensate for those unimodal deficits, leading to improved performance and similar neural processing to the control group.

Methods

Participants

We summarized participants' demographic and clinical characteristics in **Table 1**. Seventeen patients were recruited at the Department of Psychiatry and Psychotherapy at the University of Marburg, and were diagnosed according to ICD-10 with schizophrenia (F20.0, n=13, and F20.3, n=1) or schizoaffective disorder (F25.0, n=2, and F25.3, n=1). Participants in both groups are native speakers of German, and have no knowledge of Russian language. All except one of the patients received antipsychotic treatment; six were additionally treated with antidepressive medication. Positive and negative symptoms were assessed with the Scale for the Assessment of Positive Symptoms (SAPS) (30), and the Scale for the Assessment of Negative Symptoms (SANS) (31). Eighteen age- and education-matched healthy participants with no history of any mental disorders were recruited from the same area. Exclusion criteria for both groups were brain injury and neurological or other medical diseases affected by brain physiology. In both groups, we conducted neuropsychological tests to assess working memory function, digital span, trail making (TMT), verbal IQ (MWT-B) (32), and metaphoric language processing (concretism, evaluated with the Proverb Interpretation Task) (33). These measures are reported in **Table 1**. We report, additionally, scores from the subscales of SAPS and SANS, word fluency test, as well as gesture production and perception (BAG, Brief Assessment of Gesture (34)) in the supplement (**Table S1**). All participants had normal or corrected-to-normal vision and hearing. Except for one control and one patient, all other participants are right-handed. All participants gave written informed consent prior to participation in the experiment and were compensated monetarily. The study was approved by the local ethics committee.

Table 1. Demographic, medication, symptom, and neuropsychological measures.

	Patients (n=17)	Controls (n=18)
Age (years)	33.12 (12.35)	31.94 (10.21)
Gender male/female	13/4	13/5
Education (years)	11.82 (1.77)	12.72 (1.36)
TMT A (seconds)	31.49 (10.73)	26.17 (9.89)
TMT B (seconds)	68.56 (37.8)	52.93 (19.58)

Digit Span forward	7.94 (1.75)	8.05 (2.43)
Digit Span backward	6.35 (1.93)	6.61 (2.50)
Verbal IQ	28.8 (5.25)	28.5 (3.79)
*Concretism	1.38 (0.45)	1.14 (0.19)
SAPS (global)	15 (6.89)	
SANS (global)	9 (6.02)	
CPZ Equivalent	562.52 (372.63)	

Values are presented as mean (SD). TMT: trail making test; CPZ: chlorpromazine. Asterisk * indicates significant difference between controls and patients ($p<0.05$, two-tailed t-test).

Materials and procedure

We employed a content judgement paradigm from a previous study (18), to investigate multimodal (speech and gesture) processing of social and non-social information. We showed to participants five-second videos of an actor spontaneously communicating both social (S) and non-social (N) events in the following modalities: 1) incomprehensible Russian sentences with gestures. This is considered as a gesture-only (G) condition because social feature is only available to participants in the gesture form. 2) comprehensible German sentences (S) without any gestures. Additionally, we also showed to participants 3) German sentences with accompanying gestures as a bimodal input condition (B). A filler condition is also included with videos of incomprehensible Russian sentences with meaningless gestures. An example of both a social (S) and non-social (N) bimodal videos is illustrated in **Figure 1A**. For a complete list of all videos, please refer to Appendix in (35).

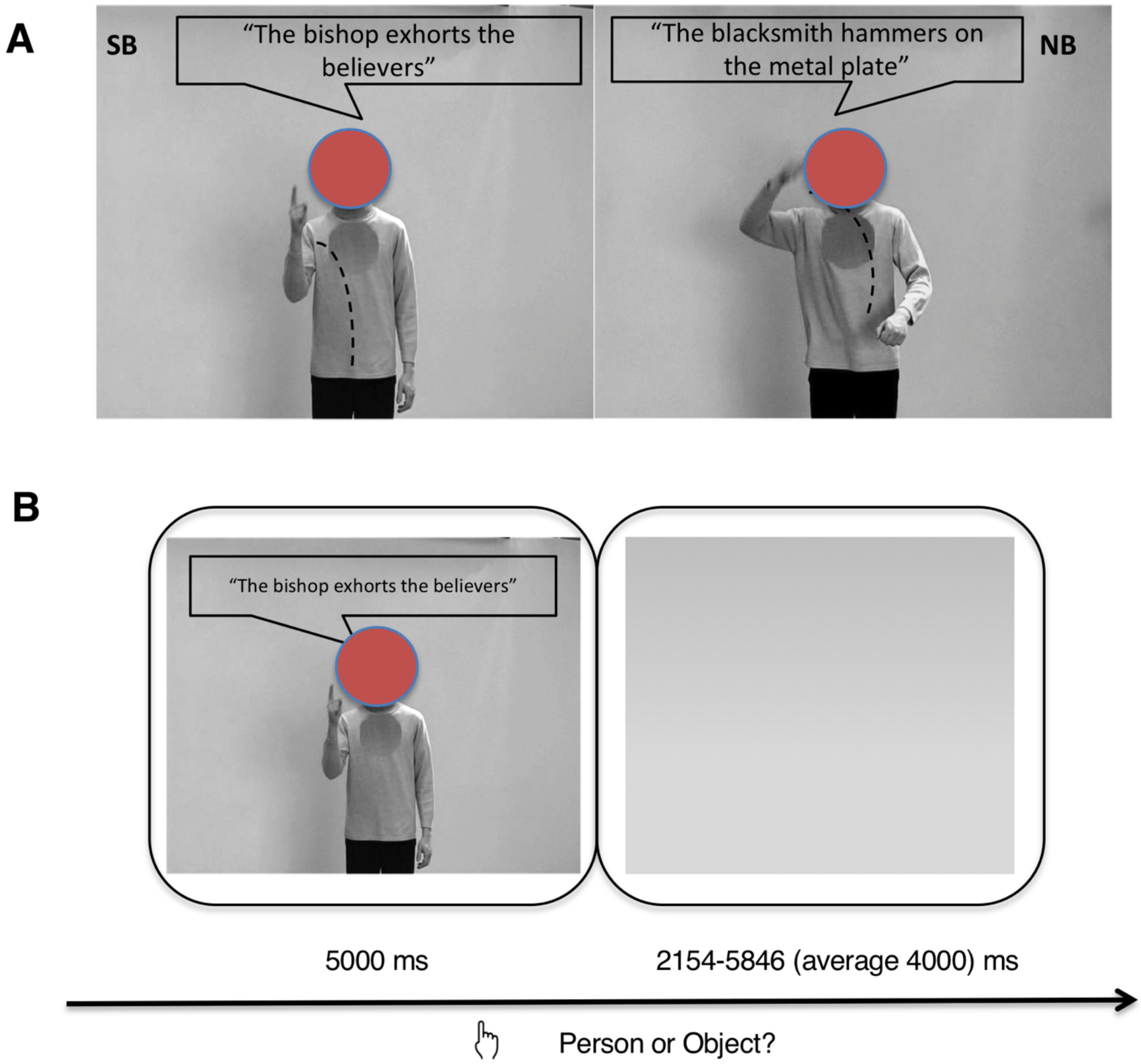

Figure 1. Panel A: Picture illustration for social (S) and non-social (N) videos in the bimodal condition (B). The same stimuli were also presented in two additional modalities: gestures with foreign Russian sentences (G) and German sentences without any gestures (S). For illustrative purposes, the spoken German sentences were translated into English, and all spoken sentences were written into speech bubbles. Panel B: Illustration of a sample trial. Participants performed a content judgment task for each video, indicating via button press whether a stimulus was either person- or object-related. The face of the actor is masked in the manuscript to avoid inclusion of identifying information of people. It is displayed to the participants during the experiment.

Experimental procedure

Altogether, 312 experimental video slips (26 videos per condition × 6 conditions × 2 sets) were included in the study. For each participant, an experimental session comprised

182 videos from one set of videos (156 critical videos and 26 filler videos), and consisted of two 14-min runs. Each run contained 91 trials with a matched number of items from each condition. The stimuli were presented in an event-related design in pseudo-randomized order, and were counterbalanced across participants. Within each trial, each video-clip was followed by a gray background with a variable duration of 2154-5846 ms (jitter average: 4000 ms), as illustrated in **Figure 1B**. Participants performed a content judgement task for each video (10, 18), indicating via button press (with their left hand) whether a stimulus was either person- or object-related. Participants were instructed to respond to the task as soon as they had decided on an answer.

fMRI acquisition and preprocessing

All images were acquired using a 3T MRI scanner (Siemens MRT Trio series). The functional images were obtained using a T2*-weighted echo-planar image sequence (TR = 2 s, TE = 30 ms, flip angle = 90°, slice thickness = 4 mm, interslice gap= 0.36 mm, field of view= 230 mm, matrix = 64 × 64, voxel size = 3.6 x 3.6 x 4.0 mm, 30 axial slices orientated parallel to the AC-PC line, ascending order). Two runs of 425 volumes each were acquired during the experiment. Additionally, simultaneous EEG data from the participants were also collected for other analyses not relevant for the current study, and are therefore not further discussed here. MR images were preprocessed using the SPM12 software package (Statistical Parametric Mapping, Welcome Trust Center for Neuroimaging, London, UK) based on Matlab R2017a (version 9.2.0; MathWorks): after discarding the first five volumes to minimize T1-saturation effects, all images were spatially and temporally realigned, and normalized into the MNI space using the MNI template (resulting voxel size 2 × 2 × 2 mm), smoothed (8 mm isotropic Gaussian filter), and high-pass filtered (cut-off period 128 s).

fMRI data analysis

We performed statistical whole-brain analysis in a two-level, mixed-effects procedure. On the first level, single-participant BOLD responses were modeled by a design matrix comprising the onset time points of each event (critical word of each sentence as used in the previous event-related fMRI and EEG studies, e.g., (18, 35-38)), with a duration of 5 seconds for all experimental conditions. The micro-time onset was set to the average time bin (8 of 16) to align the onset vector to the slice in the middle of the brain. For all conditions, the duration of speech or gesture was used as parameters of no interests on a single trial level. Six movement regressors (three rotations and three translations) were entered in the single participant's model to account for movement-induced effects on fMRI results. HRF was defined as the canonical HRF. Contrasts images against implicit baseline for all experimental conditions were used as summary measures and were included in the between-group analysis. We applied a flexible factorial analysis of variance using condition as main effect. To determine the cluster extent threshold to correct for multiple comparisons, we applied a Monte-Carlo simulation following Slotnick et al., (39, 40). For all statistical comparisons, the whole-brain activation was simulated assuming a voxel type-I error voxel activation of $p<.05$, this revealed a cluster extent of 2268 contiguous resampled voxels as sufficient to correct for multiple comparisons at $p<.0167$ (Bonferroni-corrected for three modalities). The reported voxel coordinates of activation peaks are located in MNI space. For the anatomical localization, functional data were referenced to the AAL toolbox (41).

For both groups, we firstly reported contrast images comparing the processing of social vs. non-social conditions (S>N and N>S) within each modality for each group. Secondly, for each modality, we performed interaction analyses to investigate group differences in the processing of social or non-social conditions. Lastly, we tested the three-way interaction of group*modality*content, and performed conjunction analyses between this contrast and contrasts from the last step. This step revealed modality-specific group differences for the processing of social and non-social content, which is reported in the results section. Additionally, for patients, we tested the interaction between modalities on social vs. non-social content processing, so as to reveal how bimodal stimuli might compensate potential neural processing deficits for patients with SZ. This is reported in **Supplement S2**.

Based on the literature showing a potential relationship between symptom severity (especially negative symptoms) and social/non-social cognition (4, 42), for patients with SZ, we explored the relationship between clinical measures and brain activation in areas that are relevant to social/non-social information processing. We conducted explorative correlation (spearman) analyses between 1) parameter estimates from clusters showing significant group difference for either social or non-social conditions, 2) behavioral measures (reaction times and accuracy) for each experimental condition, and 3) scores from sum/general and subscales of SAPS and SANS.

Results

Behavioral results

Descriptive and inferential statistics from the content judgement task are reported in detail in the supplement (**S1** and **Table S2**). In general, in the content judgement task, patients responded slower and were less accurate than controls. Additionally, task accuracy for the bimodal condition was higher than for other modalities in both patient and controls. However, we observed no group interaction with either content or modality manipulations.

fMRI results

Social>Non-social: we report whole-brain fMRI results for S>N comparisons in **Figure 2** and **Table S3** in the supplement. For the speech conditions (SS>NS), healthy controls activated an extensive fronto-temporal-parietal network including the bilateral inferior frontal gyrus (IFG) and the temporal lobe, the dorsolateral prefrontal cortex (dlPFC) and mPFC, and the left supramarginal gyrus; patients revealed similar regions for this comparison, and we observed no group difference for social > non-social speech. For the gesture conditions (SG>NG), controls activated the bilateral PFC and IFG; patients activated the bilateral prefrontal cortex. Group interaction (Control (SG>NG)>Patient (SG>NG)) suggests that patients showed reduced activation in the mPFC and the anterior cingulate cortex for the social gesture condition when compared to controls (**Figure 2B**). In the bimodal condition (SB>NB), both controls and patients activated regions similar to that of the bimodal condition. For patients, we additionally reported modality*content interaction in the **Supplement S2**, which shows that patients' aberrant processing of social gestures is enhanced in the bimodal modality.

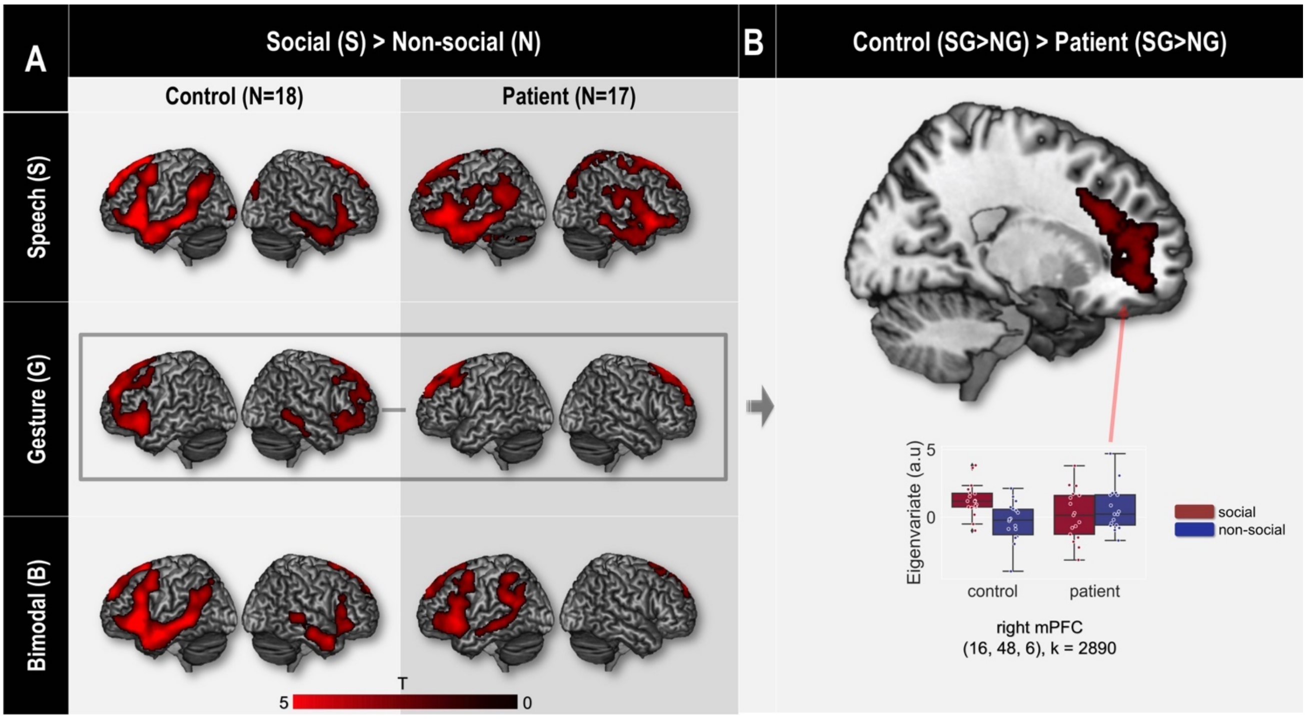

Figure 2. Activation maps for social>non-social videos (S>N). Panel A: S>N contrasts within each modality (S: Speech, G: Gesture, B: bimodal) for controls and patients. Panel B: interaction analysis (Control>Patient) in the gesture modality (SG>NG) together with box- and swarm-plots of eigenvariates for selected clusters.

Non-social>Social: we report whole-brain fMRI results for N>S comparisons in **Figure 3** and **Table S4** in the supplement. For the speech conditions (NS>SS), healthy controls activated the left pre/postcentral gyrus, supramarginal gyrus, and the left insula, whereas patients did not reveal any significant activations for this comparison. The group interaction (Control (NS>SS)>Patient (NS>SS)) suggests that, when compared to controls, patients showed reduced activation in the left postcentral gyrus and the right insula for the processing of non-social content in the speech-only modality (**see Figure 3B**). For gesture conditions (NG>SG), controls showed increased activation for the non-social content in the bilateral posterior temporal gyrus, supramarginal gyrus, and occipital cortices, as well as the left pre/postcentral gyrus and the left insula. Patients also activated the bilateral posterior temporal gyrus and occipital lobe, as well as the left pre/postcentral gyrus. The group interaction revealed no significant clusters. For bimodal conditions (NB>SB), both controls and patients activated regions that are comparable to that of the gesture conditions. Additionally, for patients, bimodal input seems to enhance their aberrant processing of non-social speech, as reported in **Supplement S2**.

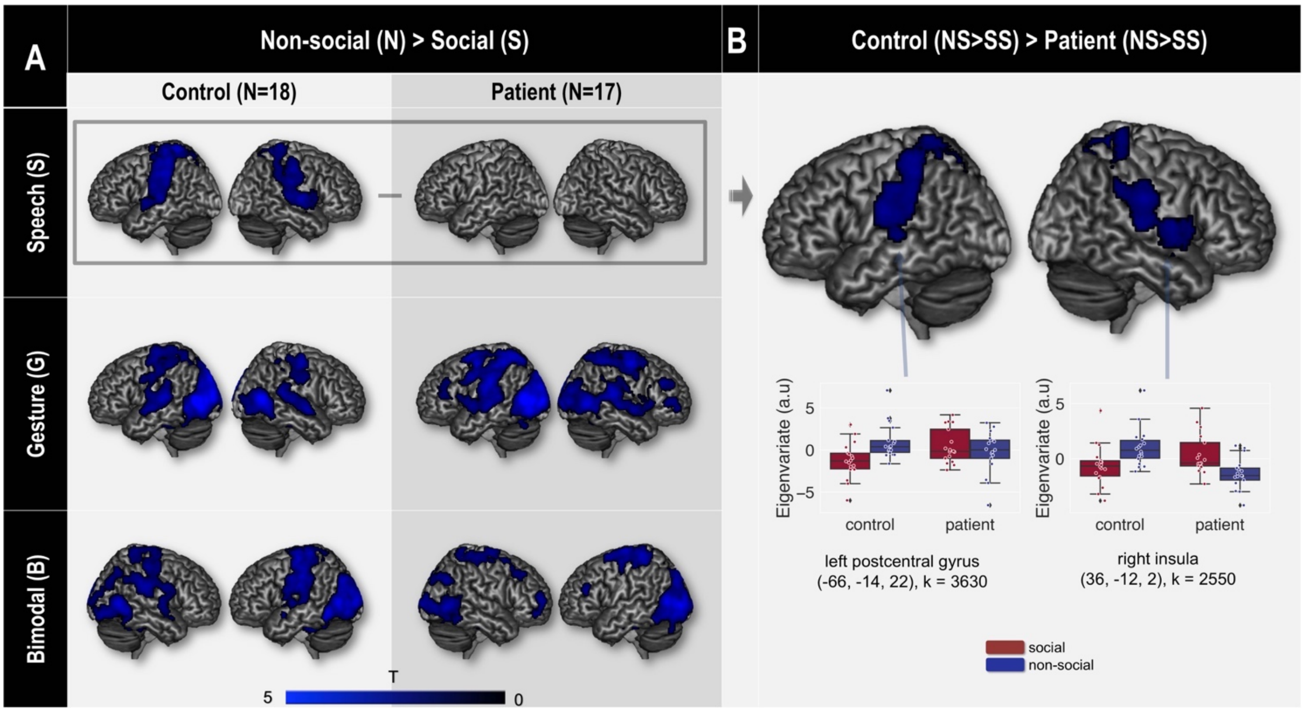

Figure 3. Activation maps for non-social> social videos (N>S). Panel A: N>S contrasts within each modality (S: Speech, G: Gesture, B: Bimodal) for controls and patients. Panel B: interaction analysis (Control>Patient) in the speech modality (NS>SS) together with box- and swarm-plots of eigenvariates for selected clusters.

Correlation analyses

In patients, for the NS condition, we found that the accuracy for the NS condition correlate negatively with the SANS composite scores of the patients ($r = -0.52$, $p = 0.03$, power = 0.63; **Figure 4A**). Additionally, SANS 1 (flat affect) and SANS 2 (alogia) scores correlate negatively with the accuracy for the NS condition (SANS 1: $r = -0.62$, $p = 0.008$, power = 0.79, **Figure 4B**; SANS 2: $r = -0.63$, $p = 0.006$, power = 0.82, **Figure 4C).**

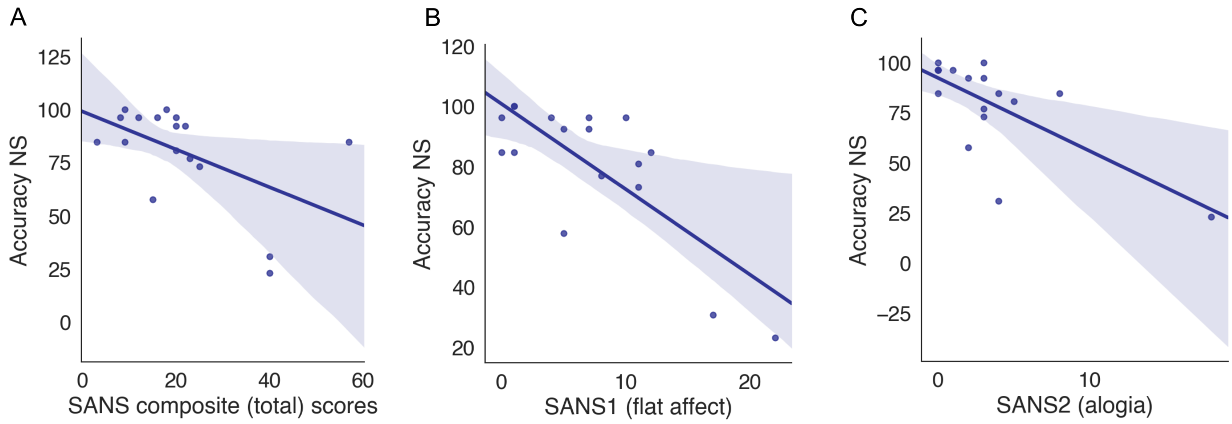

Figure 4. Significant negative correlations between patients' accuracy for the NS condition and A) patients' SANS composite scores, B) their SANS 1 (flat affect) scores, and C) their SANS 2 (alogia) scores.

Discussion

Social information processing in schizophrenia

In the current study, patients showed dissociable neural modulation during the processing of social content in speech and gesture modalities. In the speech modality, both controls and patients activated a left-lateralized set of brain regions, including the dlPFC, mPFC, the IFG, the temporal lobe, and the angular/supramarginal gyrus, without any group difference. This finding replicates results from our previous study showing supramodal social-abstract processing of healthy individuals (18), and is consistent with earlier studies in basic research on the role of the mPFC in both perceiving social-related stimuli and mentalizing social intentions (10, 13, 14). The observed left IFG and temporal lobe activation is also in line with the literature on the neural substrates of abstract vs. concrete semantics (17, 43), as the videos in the social condition, irrespective of modality, are more abstract than the non-social, object-related condition. The fact that we did not find any group differences in social speech processing suggests that patients with SZ exhibit intact neural processing of social content presented in this modality. This finding complies with a previous language study in SZ, in which patients also activated a comparable left fronto-temporal network to controls when they processed abstract vs. concrete visual sentences (23). Together, these data might suggest that, although SZ patients often show marked social cognition deficits (2, 3), they exhibit intact neural processing of social features encoded with linguistic stimuli (either in the written or verbal form). In the gesture modality, however, although patients activated the mPFC for the social vs. non-social stimuli, this activation was reduced when compared to controls. Notably, such modality-specific neural modulation is, for the first time, reported for social information processing in SZ. This finding is further discussed below.

Non-social information processing in schizophrenia

For the processing of non-social (object-related) information, again, the neural modulation in patients showed an apparent dissociation. In the gesture modality, both controls and patients activated the bilateral occipital-parietal cortices, STG, LOTC, insula, as well as the pre/postcentral gyrus. In the speech modality, although these regions were similarly activated in controls, their brain activation was significantly reduced for patients. The group comparison suggests reduced activation in SZ patients in the left insula and the left postcentral gyrus for non-social speech. Of note, the observed regions for non-social and object-related information processing overlap with part of the mirror neuron network, which is not only important for action observation and imitation, but also for the understanding of object- and motor-related features in verbal form (44-47). This process would require mental simulation of sensorimotor experience (20, 21, 48). Additionally, the LOTC is also crucially involved in the perception of biological motion, object, as well as tool-use (26, 49, 50). Our data from the control group suggest that these regions directly support the processing of non-social object-related features, irrespective of encoding modality. This finding is in line with the embodiment view of action and language processing (21, 51). With regard to the patients, we observed normal neural processing of non-social content in the gesture modality, supporting a previous study (24), which reported intact mirror neuron activity in SZ (but see (25)). However, as we also observed reduced bilateral postcentral gyrus and right insula activity for patients vs. controls for non-social speech this would, in turn, imply that motor simulation, as required for processing object-related features from auditory speech, might still be impaired in SZ (23, 25). This impairment concurs with the reported deficits of SZ in action imitation (25, 42, 52, 53), where certain degrees of motor simulation is required. Moreover, in the NS condition, we also observed negative correlation between patients' SANS composite and subscores and their task accuracy. This evidence converges with previous research, corroborating the potential role of the mirror neuron system during embodiment of non-social information (e.g., action imitation and observation), as well as its relation to the development and persistence of negative symptoms (42, 54)

Enhancing modality-specific social and non-social information processing deficits with bimodal input

The novelty of our findings lies in the dissociable modality-specificity concerning dysfunctional neural processing of social and non-social features. Social and non-social features are functionally and neurally dissociable at the representational level (10, 14). Besides, they might be differentially processed through either linguistic (speech) or non-linguistic (gesture) channels. It has been proposed that social-abstract concepts may be preferentially represented in speech, and that non-social concrete concepts are preferentially delivered in hand action and gesture (55, 56). Despite this theoretical proposal, however, during comprehension, healthy participants seem to be able to process both types of information in a supramodal manner (e.g., semantic processing with unitary core systems, irrespective of encoding modality, as in (18, 20)). For patients with SZ, as they exhibit similar neural activations when processing social speech and non-social gestures to controls, this might be an indication that they are at least intact in processing these contents at representational level. But, they might show activation reduction in relevant regions when these features are conveyed in a 'non-preferred' modality, as the processing of these features would require some form of mental simulation: In the case non-social information, patients are impaired in the simulation of motor-related experience from action to language (20); In contrast when patients are presented with social information, they might be impaired when simulating social features encoded by hand gestures (but not with speech), as shown in their reduced mPFC activation. This observed modality-specific processing deficit might also suggest that patients, unlike controls, are not capable of processing social/non-social information in a supramodal manner like healthy participants, as reported in previous studies (18, 57). More importantly, extending previous studies on aberrant processing of social/non-social content in SZ, our results indicate that this neural deficit is not universally present for either a specific modality or content, but rather appears only in specific combinations of these two factors.

Despite reduced neural processing of both social and non-social content in gesture and speech modalities, patients displayed intact neural processing of these features, as well as improved task accuracy in the bimodal conditions. This enhancement effect concurs with a line of proposals (6), who argue for a bi-directional facilitative relation between speech and gesture (for empirical evidence, see (58-62)). More importantly, our finding extends previous basic research, suggesting the translational implication of this mechanism. In SZ research, the past decade has witnessed substantial progress in the development of social cognitive training in SZ (63, 64), with recent innovation regarding the incorporation of social stimuli from a broader range of modalities (65). Our findings extend these approaches, proposing potential therapeutic implications of deploying naturalistic and multimodal stimuli during social cognitive training, as they might be able to normalize processing of both social and non-social information, at least at a neural level. Future research is expected to further explore whether the neural enhancements can be linked to functional outcome after social cognitive training in a multimodal setting.

Acknowledgements

This research project is supported by a grant from the 'Von-Behring-Röntgen-Stiftung' (project no. 59-0002 and 64-0001) and by the 'Deutsche Forschungsgemeinschaft' (project no. DFG: STR1146/11-2 & KI588/6-2 and CRC/TRR 135/2 project A3). Y.H. is supported by the 'Von-Behring- Röntgen-Stiftung' (project no. 64-0001). B.S. is supported by the DFG (project no. STR 1146/15-1). The study was also supported by the Core Facility Brain Imaging, Faculty of Medicine, University of Marburg, Rudolf-Bultmann-Str. 9, 35039, Marburg, Germany.

Conflict of interests

All authors declare no financial conflict of interests.

References

1. Savla GN, Vella L, Armstrong CC, Penn DL, Twamley EW (2013): Deficits in domains of social cognition in schizophrenia: a meta-analysis of the empirical evidence. *Schizophr Bull*. 39:979-992.
2. Green MF, Horan WP, Lee J (2015): Social cognition in schizophrenia. *Nature Reviews Neuroscience*. 16:620.
3. Ochsner KN (2008): The social-emotional processing stream: five core constructs and their translational potential for schizophrenia and beyond. *Biological Psychiatry*. 64:48-61.
4. Sergi MJ, Rassovsky Y, Widmark C, Reist C, Erhart S, Braff DL, et al. (2007): Social cognition in schizophrenia: relationships with neurocognition and negative symptoms. *Schizophrenia research*. 90:316-324.
5. Couture SM, Penn DL, Roberts DL (2006): The functional significance of social cognition in schizophrenia: a review. *Schizophrenia bulletin*. 32:S44-S63.
6. Holler J, Levinson SC (2019): Multimodal Language Processing in Human Communication. *Trends in Cognitive Sciences*.
7. Amodio DM, Frith CD (2006): Meeting of minds: the medial frontal cortex and social cognition. *Nature reviews neuroscience*. 7:268.
8. Anticevic A, Van Snellenberg JX, Cohen RE, Repovs G, Dowd EC, Barch DM (2012): Amygdala recruitment in schizophrenia in response to aversive emotional material: a meta-analysis of neuroimaging studies. *Schizophrenia bulletin*. 38:608-621.
9. Witteman J, Van Heuven VJ, Schiller NO (2012): Hearing feelings: a quantitative meta-analysis on the neuroimaging literature of emotional prosody perception. *Neuropsychologia*. 50:2752-2763.
10. Mitchell JP, Heatherton TF, Macrae CN (2002): Distinct neural systems subserve person and object knowledge. *Proceedings of the National Academy of Sciences*. 99:15238-15243.
11. Saxe R, Powell LJ (2006): It's the thought that counts: specific brain regions for one component of theory of mind. *Psychological science*. 17:692-699.
12. Frith CD, Frith U (2006): The neural basis of mentalizing. *Neuron*. 50:531-534.
13. Van Overwalle F (2009): Social cognition and the brain: a meta‐analysis. *Human brain mapping*. 30:829-858.
14. Mitchell JP, Banaji MR, Macrae CN (2005): General and specific contributions of the medial prefrontal cortex to knowledge about mental states. *Neuroimage*. 28:757-762.
15. Ma N, Vandekerckhove M, Van Overwalle F, Seurinck R, Fias W (2011): Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: spontaneous inferences activate only its core areas. *Social neuroscience*. 6:123-138.
16. Chao LL, Haxby JV, Martin A (1999): Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature neuroscience*. 2:913-919.
17. Binder JR, Westbury CF, McKiernan KA, Possing ET, Medler DA (2005): Distinct Brain Systems for Processing Concrete and Abstract Concepts. *Journal of Cognitive Neuroscience*. 17:905-917.
18. Straube B, He Y, Steines M, Gebhardt H, Kircher T, Sammer G, et al. (2013): Supramodal neural processing of abstract information conveyed by speech and gesture. *Front Behav Neurosci*. 7:120.
19. Mar RA (2011): The neural bases of social cognition and story comprehension. *Annual review of psychology*. 62:103-134.
20. Pulvermüller F, Shtyrov Y, Ilmoniemi R (2005): Brain signatures of meaning access in action word

recognition. *Journal of Cognitive Neuroscience*. 17:884-892.

21. Barsalou LW (2008): Grounded cognition. *Annu Rev Psychol*. 59:617-645.
22. Frith CD (2004): Schizophrenia and theory of mind. *Psychological medicine*. 34:385-389.
23. Kuperberg GR, West WC, Lakshmanan BM, Goff D (2008): Functional Magnetic Resonance Imaging Reveals Neuroanatomical Dissociations During Semantic Integration in Schizophrenia. *Biological Psychiatry*. 64:407-418.
24. Horan WP, Iacoboni M, Cross KA, Korb A, Lee J, Nori P, et al. (2014): Self-reported empathy and neural activity during action imitation and observation in schizophrenia. *NeuroImage: Clinical*. 5:100-108.
25. Thakkar KN, Peterman JS, Park S (2014): Altered Brain Activation During Action Imitation and Observation in Schizophrenia: A Translational Approach to Investigating Social Dysfunction in Schizophrenia. *American Journal of Psychiatry*. 171:539-548.
26. Lingnau A, Downing PE (2015): The lateral occipitotemporal cortex in action. *Trends in Cognitive Sciences*. 19:268-277.
27. Johnson-Frey SH (2004): The neural bases of complex tool use in humans. *Trends in Cognitive Sciences*. 8:71-78.
28. Järveläinen J, Schuermann M, Hari R (2004): Activation of the human primary motor cortex during observation of tool use. *Neuroimage*. 23:187-192.
29. Johnson-Frey SH, Maloof FR, Newman-Norlund R, Farrer C, Inati S, Grafton ST (2003): Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neuron*. 39:1053-1058.
30. Andreasen NC (1984): *Scale for the Assessment of Positive Symptons:(SAPS)*. University of Iowa.
31. Andreasen NC (1981): Scale for the Assessment of Negative Symptoms (SANS).
32. Lehrl S (1999): *Mehrfachwahl-Wortschatz-Intelligenztest: MWT-B*. Spitta.
33. Barth A, Küfferle B (2001): Development of a proverb test for assessment of concrete thinking problems in schizophrenic patients. *Der Nervenarzt*. 72:853-858.
34. Nagels A, Kircher T, Steines M, Grosvald M, Straube B (2015): A brief self-rating scale for the assessment of individual differences in gesture perception and production. *Learning and Individual Differences*. 39:73-80.
35. He Y, Gebhardt H, Steines M, Sammer G, Kircher T, Nagels A, et al. (2015): The EEG and fMRI signatures of neural integration: An investigation of meaningful gestures and corresponding speech. *Neuropsychologia*. 72:27-42.
36. He Y, Steines M, Sommer J, Gebhardt H, Nagels A, Sammer G, et al. (2018): Spatial-temporal dynamics of gesture-speech integration: a simultaneous EEG-fMRI study. *Brain Struct Funct*. 223:3073-3089.
37. He Y, Nagels A, Schlesewsky M, Straube B (2018): The Role of Gamma Oscillations During Integration of Metaphoric Gestures and Abstract Speech. *Front Psychol*. 9:1348.
38. Straube B, Wroblewski A, Jansen A, He Y (2018): The connectivity signature of co-speech gesture integration: The superior temporal sulcus modulates connectivity between areas related to visual gesture and auditory speech processing. *NeuroImage*. 181:539-549.
39. Slotnick SD, Moo LR, Segal JB, Hart Jr J (2003): Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research*. 17:75-82.
40. Slotnick SD (2017): Resting-state fMRI data reflects default network activity rather than null data: A defense of commonly employed methods to correct for multiple comparisons. *Cognitive neuroscience*. 8:141-143.
41. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, et al. (2002): Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI

MRI single-subject brain. *Neuroimage*. 15:273-289.

42. Mehta UM, Thirthalli J, Aneelraj D, Jadhav P, Gangadhar BN, Keshavan MS (2014): Mirror neuron dysfunction in schizophrenia and its functional implications: a systematic review. *Schizophrenia research*. 160:9-19.

43. Binder JR, Desai RH, Graves WW, Conant LL (2009): Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*. 19:2767-2796.

44. Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, et al. (2001): Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European Journal of Neuroscience*. 13:400-404.

45. Di Pellegrino G, Fadiga L, Fogassi L, Gallese V, Rizzolatti G (1992): Understanding motor events: a neurophysiological study. *Experimental Brain Research*. 91:176-180.

46. Iacoboni M, Woods RP, Brass M, Bekkering H, Mazziotta JC, Rizzolatti G (1999): Cortical mechanisms of human imitation. *Science*. 286:2526-2528.

47. Rizzolatti G, Craighero L (2004): The mirror-neuron system. *Annual Review of Neuroscience*. 27:169-192.

48. Pulvermüller F (2005): Brain mechanisms linking language and action. *Nature Reviews Neuroscience*. 6:576-582.

49. Bi Y, Wang X, Caramazza A (2016): Object Domain and Modality in the Ventral Visual Pathway. *Trends in Cognitive Sciences*. 20:282-290.

50. Higuchi S, Imamizu H, Kawato M (2007): Cerebellar Activity Evoked By Common Tool-Use Execution And Imagery Tasks: An Fmri Study. *Cortex*. 43:350-358.

51. De Stefani E, De Marco D (2019): Language, Gesture, and Emotional Communication: An Embodied View of Social Interaction. *Frontiers in Psychology*. 10.

52. Park S, Matthews N, Gibson C (2008): Imitation, simulation, and schizophrenia. *Schizophrenia Bulletin*. 34:698-707.

53. Matthews N, Gold BJ, Sekuler R, Park S (2013): Gesture imitation in schizophrenia. *Schizophrenia Bulletin*. 39:94-101.

54. Buccino G, Amore M (2008): Mirror neurons and the understanding of behavioural symptoms in psychiatric disorders. *Current opinion in psychiatry*. 21:281-285.

55. Paivio A (2010): Dual coding theory and the mental lexicon. *The Mental Lexicon*. 5:205-230.

56. Perlovsky LI, Ilin R (2013): Mirror neurons, language, and embodied cognition. *Neural networks*. 41:15-22.

57. Straube B, Green A, Weis S, Kircher T (2012): A Supramodal Neural Network for Speech and Gesture Semantics: An fMRI Study. *Plos One*. 7:e51207.

58. Wang L, Chu M (2013): The role of beat gesture and pitch accent in semantic processing: an ERP study. *Neuropsychologia*. 51:2847-2855.

59. Drijvers L, Özyürek A, Jensen O (2018): Hearing and seeing meaning in noise: Alpha, beta, and gamma oscillations predict gestural enhancement of degraded speech comprehension. *Human Brain Mapping*. 59:617-613.

60. Krahmer E, Swerts M (2007): The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language*. 57:396-414.

61. He Y, Steines M, Sammer G, Nagels A, Kircher T, Straube B (2018): Action-Related Speech Modulates Beta Oscillations During Observation of Tool-Use Gestures. *Brain Topogr*. 31:838-847.

62. Cuevas P, Steines M, He Y, Nagels A, Culham J, Straube B (2019): The facilitative effect of gestures on

the neural processing of semantic complexity in a continuous narrative. *NeuroImage*. 195:38-47.
63. Kurtz MM, Gagen E, Rocha NB, Machado S, Penn DL (2016): Comprehensive treatments for social cognitive deficits in schizophrenia: A critical review and effect-size analysis of controlled studies. *Clinical psychology review*. 43:80-89.
64. Kurtz MM, Richardson CL (2011): Social cognitive training for schizophrenia: a meta-analytic investigation of controlled research. *Schizophrenia bulletin*. 38:1092-1104.
65. Nahum M, Fisher M, Loewy R, Poelke G, Ventura J, Nuechterlein KH, et al. (2014): A novel, online social cognitive training program for young adults with schizophrenia: A pilot study. *Schizophrenia Research: Cognition*. 1:e11-e19.