

## **Characterizing the Theory of Mind Network in Schizophrenia Reveals a Sparser Network Structure**

Florian Bitsch<sup>1\*</sup>, Philipp Berger<sup>1</sup>, Arne Nagels<sup>1,2</sup>, Irina Falkenberg<sup>1</sup>, & Benjamin Straube<sup>1</sup>

<sup>1</sup>Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Rudolf-Bultmann Str. 8, 35039 Marburg, Germany

<sup>2</sup>Department of English and Linguistics, Johannes Gutenberg-University Mainz, Jakob-Welder-Weg 18, 55128 Mainz, Germany

\*Correspondence to:

Florian Bitsch

Philipps-University Marburg

Department of Psychiatry and Psychotherapy

Rudolf-Bultmann-Straße 8, 35039 Marburg, Germany

Tel: +49-6421-58-63604

Email: [bitsch@med.uni-marburg.de](mailto:bitsch@med.uni-marburg.de)

## Abstract

Impaired social functioning is a hallmark of schizophrenia and altered functional integration between distant brain regions are expected to account for signs and symptoms of the disorder. The functional neuroarchitecture of a network relevant for social functioning, the mentalizing network, is however poorly understood. In this study we examined dysfunctions of the mentalizing network in patients with schizophrenia compared to healthy controls via dynamic causal modelling and an interactive social decision-making game. Network characteristics were analyzed on a single subject basis whereas graph theoretic metrics such as in-degree, out-degree and edge-connectivity per network node were compared between the groups. The results point to a sparser network structure in patients with schizophrenia and highlight the dorsomedial prefrontal cortex as a disconnected network hub receiving significantly less input from other brain regions in the network. Further analyses suggest that integrating pathways from the right and the left temporo-parietal junction into the dorsomedial prefrontal cortex were less frequently found in patients with schizophrenia. Brain and behavior analyses further suggest that the connectivity-intactness within the entire network is associated with functional interpersonal behavior during the task. Thus, the neurobiological alterations within the mentalizing network in patients with schizophrenia point to a specific integration deficit between core brain regions underlying the generation of higher-order representations and thereby provide a potential treatment target.

**Key words:** effective connectivity, disconnectivity, mentalizing, psychosis, social cognition

# **1. Introduction**

Discovering the pathophysiology of psychiatric disorders by examining communication deficits between distant brain regions is a promising trend in clinical neuroscience [1,2]. This perspective seems particularly appropriate to understand the psychopathology of schizophrenia, given that the symptoms of the disorder have been interpreted as signs of disconnections between neural circuits [3]. More than 100 years later, this idea still receives support [4,5,6]. Within this framework it is assumed that a dysfunctional integration between functionally specialized brain systems is linked with deficits in forming learning dependent representations. These are assumed to extend to the highest form of mental representations, the mental states of others [7]. These theoretical assumptions are underlined by recent empirical evidence pointing to altered functional [8,9,10,11,12] and structural network architecture in patients with schizophrenia [8,11,13]. Along these lines, the probability of alterations in a brain region increases with the region's interconnectedness to other brain regions. This led to the assumption that the connectivity-rich frontal and temporal hubs are particularly vulnerable in schizophrenia [13,14].

A hallmark of schizophrenia is impaired social functioning [15] which has been associated with dysfunctional mentalizing abilities, i.e. the inference and representation of other people's mental states [16]. Alterations in mentalizing processes have a high clinical relevance, since they occur before the onset of the disorder [17], persist over its course [15,18,19] and are linked to the patients' outcome [20]. Potential dysfunctions within the neural network that underlies these social-cognitive alterations are, however, unknown. Therefore, we will leverage on advances in network neuroscience by using an analysis-scheme that determines each subject's most likely network architecture while processing the task. Afterwards, the subject-wise networks will be condensed at the group-level to compare topological network alterations between patients with schizophrenia and healthy controls.

Previous neuroimaging studies revealed a set of core brain regions, which show enhanced functional activation during the inference of other people's mental states. This suggests that the right temporo-parietal junction (rTPJ), the left TPJ (lTPJ), the medial prefrontal cortex (mPFC) and the precuneus/posterior cingulate cortex (PCC) are critical nodes of a network underlying mentalizing processes [21,22]. Previous research has frequently referred to the parallel activation of these brain regions as a network. However, the examination of the regions' joint neural interactions during the mentalizing process is so far missing. Network approaches can thereby provide novel insight into the mutual influence of different brain regions within the network community. This can pave the way for a mechanistic understanding of mentalizing processes and potential alterations in psychiatric disorders, such as schizophrenia, to facilitate diagnostic procedures and treatment stratification.

Functionally, the regions within the mentalizing network have been found to be associated with social cognition and unconstrained thinking [23,24], leading to the assumption that the network might be relevant for a similar set of self – other computations [25,26]. This idea has been put forward by within-subject analysis, showing that the same regions are relevant for the processing of mental states and default mode cognition, which probably has been shaped by our species' social nature [27]. Meta-analytical evidence further suggests that the midline-structures within the mentalizing network (dmPFC and PCC) are critical for social cognitive processes in general and self-other-related processing in particular. The bilateral TPJ on the other hand has been associated particularly with mentalizing processes and self-other distinction [28], which is why the fluent integration between these higher-order cognitions are likely fundamental for the emergence of complex social representations.

Clinical studies which have examined the neural basis of altered mentalizing processes in patients with schizophrenia found reduced activity in the core regions of the mentalizing network, such as the rTPJ [29,30,31], the lTPJ [30,31], the mPFC [30,31,32,33] and the

precuneus/PCC [30, 34] during mentalizing tasks, indicating marked dysfunctions in regions relevant for mental states inferences. However, some studies found enhanced activation of these mentalizing brain regions in patients with positive symptoms, which could indicate a deficit in integrating self and other representations [32,34] or an ‘over-attribution’ of intentions [31] during mentalizing. A recent meta-analysis systemized these findings with converging evidence of functional activation alterations in two core regions of the mentalizing network, the TPJ and the mPFC, in patients with schizophrenia [35]. Although brain regions relevant for mentalizing processes have been consistently found to show altered neural responses in patients, the functional integration between them is, so far, less clear. This, however, is of high importance given that pathological states in the disorder can be understood as a dysfunctional integration of different cognitive processes [6,36], such as the attribution of an internal generated speech to an external source [7]. In this sense, it has been suggested that an online coding process, a dynamic information integration between distributed neural regions, by consistently adjusting edge strengths, i.e., their connectivity [37], might be altered in the disorder [6]. In this study, we empirically evaluate this hypothesis by examining, for the first time, specific network indices of the mentalizing network by means of graph theoretical methodologies [38,39]. In the terminology of graph theory, brain organization is considered as a network consisting of interconnections (edges) between spatially distributed neuronal populations (nodes). Accordingly, brain regions can be considered as hubs, which influence other brain regions within the network by their out-going connectivity (out-degree, the proportion of times a specific edge is present) or get influenced by other brain regions within the community (in-degree, the proportion of times a specific edge is present). Hubs are considered as important features of brain networks, given they distribute and integrate information within the community in a powerful way [40]. Their functional relevance increases with a higher amount of network connections. Hence, the elimination of highly connected brain regions can impact the global network function and

might be associated with alterations in mental states. Therefore, we will describe potential integration deficits at a global (in-degree, out-degree) and a local level (edge-connectivity) to precisely capture the mentalizing network in patients with schizophrenia.

We used a network discovery approach [41] which allows to determine effective connectivity, i.e. the causal influence a brain region exerts over another within a functional network [42]. We examined this by means of an interactive mentalizing task, a social decision-making game, in which the participants need to infer the intentions of the playing partners to improve interpersonal decision-making [43]. The functional activity analysis was aligned with a meta-analytical derived functional mask of previous mentalizing studies [44] in order to define the core nodes of the network based on meta-analytical information. Then, we optimized each participant's individual mentalizing network structure via dynamic causal modelling (DCM) [39,45,46] to identify the most likely functional mentalizing network architecture of each participant during the mentalizing task. Between-group comparisons focused on network indices such as the in-degree and out-degree of each hub and edge-connectivity between the regions (Fig. 1). Based on previous functional activation studies in patients with schizophrenia [31,32,34] and recent meta-analytical evidence [35], we expected to find integration deficits, particularly between the prefrontal cortex and temporo-parietal regions in patients during the mentalizing process.

## 2. Methods

### 2.1 Participants and Inclusion Criteria

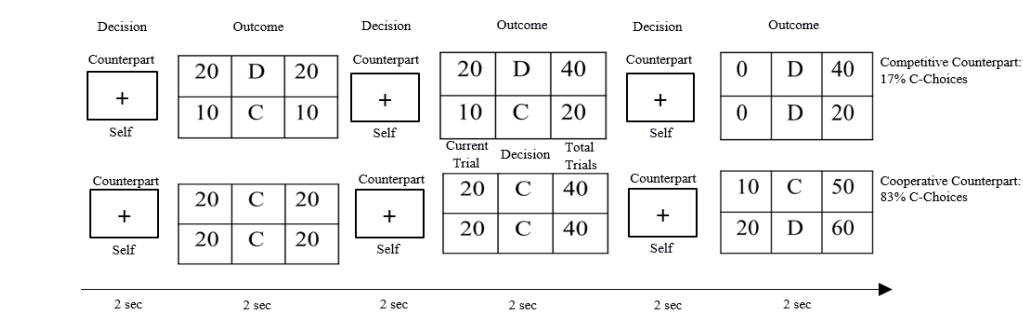
A total of 31 (10 females) patients with a diagnosis of schizophrenia or schizoaffective disorder (ICD-10: F20 or F25) in the range between 18 and 53 years of age ( $M \pm STD = 31.29 \pm 8.53$ ) and 20 healthy controls (10 females) between 23 and 45 years of age ( $M \pm STD = 29.55 \pm 6.17$ ) were included in the analysis (for further sample characteristic see the results section). Both groups participated in a previous study, with a different research

question [43]. A structured clinical interview for DSM-IV [47] was used to confirm the patients' diagnosis and to exclude any psychiatric disorder in healthy controls. The severity of patients' symptoms during the week before they participated in the study was assessed with the Scale for the Assessment of Positive (SAPS, [48]) and Negative Symptoms (SANS, [49]). Participants were only included by sufficient fMRI data quality (details in fMRI data preprocessing) and sufficient task engagement (not more than 25% misses of behavioral responses during the task). According to these criteria 9 patients (3 data-quality, 6 task-engagement) were excluded out of the original sample of 40 patients who participated in the study. All participants had normal or corrected-to-normal vision, reported no history of neurological disorders and were right-handed [50]. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration. The study was approved by the local Ethics Committee at Philipps-University Marburg and all participants gave written informed consent prior to the experiment.

## *2.2 Experimental task*

Before the experiment, participants were instructed that they will interact with three real playing partners in a social decision-making game in which the payoff depends on their own and the counterpart's decisions (iterated prisoner's dilemma game, for details: [43, 51, Fig.1]) this was practiced beforehand in a trainings session. In fact, the participants interacted with fictive playing partners who pursued a stable strategy either competitively (defecting in 83.3% of trials), cooperatively (cooperating in 83.3% of trials) or randomly (defecting/cooperating in 50% of trials respectively). The counterbalanced interaction blocks started with the presentation of a picture and a name of the current playing partner ( $3.5 \pm 1.25$ s), followed by a decision phase (2s) during which the participants had to decide to press the left (cooperate) or the right (defect) button, followed by the presentation (2s) of the decision and outcome for

both players determined by the payoff matrix [43,51, Fig.1]. Participants interacted with each counterpart in alternate order in 7 interaction blocks, each containing 6 decisions. Because the individual outcome is determined by the participant's and the counterpart's decision (Fig.1) an interdependence between both interaction partners exists. Accordingly, the subjects can accumulate evidence about the playing partners' intentions during the experiment and iteratively adapt their strategies accordingly. A condition in which the picture of the playing partner was replaced by a red cross and the hint that due to technical reasons no playing partner is connected, acted as control condition. During this condition, the participants had to alternately press the right and the left button with no outcomes being presented, i.e. no mentalizing processes were necessary. The control condition contained exactly the same sequences as the task conditions. The only difference between both conditions is that the outcome numbers are replaced by a hash in the control condition.



**Fig.1.** Examples of the interaction sequences with the intentionally different counterparts. Each interaction block contained six decision phases in which the participants had to decide to either defect (D) or cooperate (C) with the counterpart. Afterwards the partners' payoffs were presented. The COMP and COOP playing partner decided in 5 out of 6 decisions per block with either the D- or C-choice.

## 2.3 Behavioral Analysis

In order to examine differences in representing the different counterparts' intentions during the task, we examined between-group differences in behavioral adaptation to the different



counterparts ('Dynamic-Tom': task-dependent mentalizing index, which is the difference between defective decisions in the competitive vs. cooperative condition), according to our previous approach (see [43]). Additionally, we examined connectivity indices and mentalizing abilities with a task-independent measurement, the Reading the Mind in the Eyes Task (RMET), which we managed to collect for 33 participants.

## *2.4 fMRI data analysis*

### *2.4.1 fMRI data acquisition*

All images were acquired using a Siemens 3-Tesla Trio, A Tim scanner with a 12-channel head matrix receive coil. Functional images were acquired using a T2\* weighted single shot echo planar imaging (EPI) sequence (parallel imaging factor of 2 (GRAPPA), TE=30ms, TR=2000ms, flip angle 90°, slice thickness 3.6mm, matrix 64×64, in-plane resolution 3×3mm<sup>2</sup>, bandwidth 2232Hz/pixel, EPI factor of 64 and an echo spacing of 0.51ms). Data from 33 transversal slices oriented to the AC–PC line were gathered in descending order.

### *2.4.2 fMRI data preprocessing*

According to our previous approach [43] functional data preprocessing was performed using SPM8 (<http://www.fil.ion.ucl.ac.uk/spm>) implemented in MATLAB13a (MathWorks, MA). The first five volumes of each functional run were discarded from the analysis to account for T1 equilibration effects. Functional data were realigned and unwarped, corrected for slice timing, spatially normalized onto a common brain space (Montreal Neurological Institute, MNI) and spatially smoothed using a Gaussian filter with a 8mm full-width half maximum (FWHM). Prior to the analysis the fMRI-data was controlled for quality constraints. Participants were only included when less than 5% of the functional images were detected as outliers by an automated quality assurance protocol (Artifact Detection toolbox,

[https://www.nitrc.org/projects/artifact\\_detect/](https://www.nitrc.org/projects/artifact_detect/)), based on massive head motion (> 2mm) or aberrant signal intensity (global-signal z-value exceeded a threshold of 9) criteria.

### 2.4.3 *fMRI data analysis*

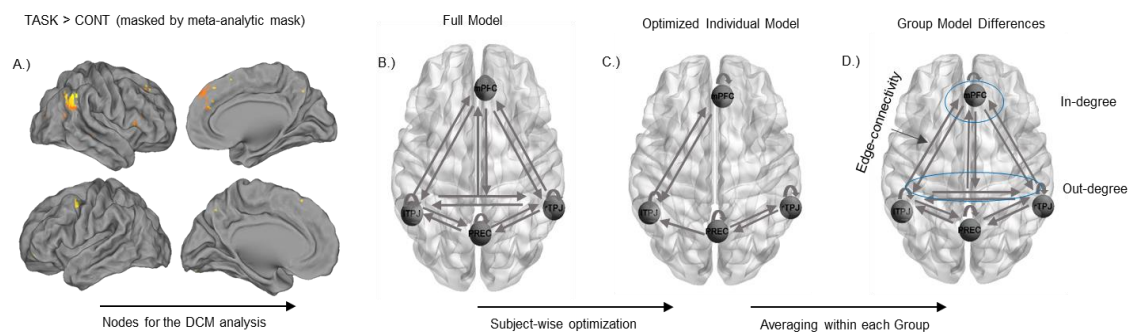
For each participant a general linear model (GLM) was calculated in which the decision and the outcome phase (4sec) were convolved with a hemodynamic response function (HRF) and implemented in the GLM with the presentation of the playing partners' pictures and the 6 - realignment parameters (rigid body) as nuisance regressors. The high-pass filter was adapted to the experimental design and set to 284-sec cut-off period.

### 2.4.4 *Specification of dynamic causal models (DCM)*

We used DCM12 as implemented in SPM12 (<http://www.fil.ion.ucl.ac.uk/spm>) to determine effective connectivity in the mentalizing network according to a network discovery approach [41]. In general, DCM allows to examine the causal influence one brain region exerts over another and how this coupling is influenced by changes in the experimental context [42]. We used the optimization procedure for post-hoc inferences [45,46] to discover each participant's most likely network architecture, indicated by the greatest conditional probability in comparison to alternative networks [41]. Therefore, a full bilinear DCM was specified, subject-wise estimated and the most likely model selected by the Bayesian model evidence [45,52]. Individual network models (existence of connections between the regions) were used to set up the between-group comparisons, which focused on each node's in-degree (the sum of present forward-connections in each subject of a group), out-degree (the sum of present backward-connections in each subject of a group) and edge-connectivity, i.e. the presence of an effective connection between the specific network nodes. Regions that were included in the dynamic causal modeling had to fulfill specific criteria in order to analyze commensurable functional areas [53]. Therefore, the region selection was guided by the activation peak of mentalizing brain regions found in the main effect (TASK>CONT) inclusively masked by a

meta-analytically derived functional mask, covering brain regions relevant for mentalizing processes from 140 previous mentalizing studies (neurosynth; [44]). This strategy allowed us to limit our analysis to the mentalizing network found in previous studies of mentalizing processes. The analysis revealed activation in the core regions (the rTPJ, the lTPJ, the mPFC and the precuneus) of the mentalizing network across both groups (corrected for multiple comparisons, FWE,  $p < .05$ ) which guided the extraction of each Volume-of-Interest's (VOI) first eigenvariate of the regions (details in the results section). Each region's first eigenvariate was extracted at a single subject level from all suprathreshold voxels ( $p = .05$ , uncorrected, main effect) of a 6mm sphere. The sphere was centered on the maximum of the local activation foci within 12mm search radius centered on the activation foci of the group analysis, adjusted for the effect of interest. Five patients and two controls showed no suprathreshold activation in one or more of the extracted mentalizing brain regions and were therefore excluded from the DCM analysis leading to a DCM sample of SZ:  $n = 26$  and HC:  $n = 18$ . For the DCM analysis a novel first-level general linear model (GLM) was specified, which contained two regressors, one for the exogenous input (B-matrix), specified by the mentalizing conditions and the driving input (C-Matrix) specified by the entire task. Endogenous connections (A-matrix) were fully coupled between all four brain regions and the subject-wise connections (present or not) were the main objective of statistical analysis. Based on previous DCM-studies reporting group differences between patients with schizophrenia and healthy control subjects ( $N < 20$ , e.g. 52, 54) we expected, that our fMRI sample size is sufficient to reveal the expected effects. Between-group analyses were conducted with Bonferroni-corrected (12 edge-connectivity, 8 in-/out-degree comparisons) chi-squared tests and Mann-Whitney U tests and considered as significant at  $\alpha < .0025$ . Furthermore, we report results as trends when the analyses were significant at an uncorrected threshold of  $\alpha < .05$ . Clinical characteristics were examined in patients with and without an absence of an edge-wise connection (please see supplementary material).

**Fig. 1. Analysis pipeline to define the network architecture**



**Fig. 2.** A) The main effect of mentalizing conditions was masked by a meta-analytic mask (140 mentalizing studies) to align the current study with evidence from further mentalizing studies and guided the selection of the nodes for the Dynamic Causal Model (DCM) B) Example of a fully connected network including the precuneus (PREC), the left and right temporo-parietal junction (ITPJ,rTPJ) and the dorsal medial prefrontal gyrus (mPFC) was optimized for each subject C) Example of an optimized model, which indicates the most likely network structure D) Edge-connectivity, in and out-degree per network node were calculated for each group and compared.

### 3. Results

#### 3.1 Sociodemographic and Clinical Characteristics

No significant age differences existed between the groups ( $t(49)=.788$ ,  $p=.434$ ). Educational level ( $\chi^2=7.65$ ,  $p=.054$ ) and gender ( $\chi^2=2.57$ ,  $p=.107$ ) differed slightly between the groups, which is why both variables were added as covariates in the functional activation analysis. The patient group had a mean SAPS score of  $12.35 \pm 10.57$  ( $M \pm STD$ , range:0-36) and a SANS score of  $10.35 \pm 11.74$  ( $M \pm STD$ , range:0-39) leading to a global score of  $22.70 \pm 19.44$  ( $M \pm STD$ , range:0-73). The duration of illness was  $9.35 \pm 8.40$  years ( $M \pm STD$ , range:0–30 years). Twenty-four patients received stable doses of atypical antipsychotic medication ( $M \pm STD=387.89 \pm 451.74$ mg/day Chlorpromazine equivalent [55], whereas seven patients did not receive any antipsychotic medication.

#### 3.2 Differences in Adaptive Social Behavior during the Task

Healthy controls showed a higher adaptation to the intentionally different counterparts during the task, as significant between-group differences in the Dynamic ToM value suggest,  $t(49)=2.35$ ,  $p=.023$ , HC:  $M\pm STD=5.5\pm 6.48$ , SZ:  $M\pm STD=1.39\pm 5.85$ .

### *3.3 Main effect TASK > CONT (Masked by the Mentalizing Network)*

The main effect of mentalizing against the control condition (masked by the meta-analytically derived functional mentalizing mask) revealed enhanced activation in the core regions of the mentalizing network (Fig.2) such as the rTPJ ( $x=60$   $y=-46$   $z=28$ ,  $t=5.81$ ,  $FWEp<.001$ ,  $k=404$ ), the lTPJ ( $x=-52$   $y=-50$   $z=36$ ,  $t=8.74$ ,  $FWEp<.001$ ,  $k=20$ ), the precuneus ( $x=2$   $y=-64$   $z=46$ ,  $t=7.01$ ,  $FWEp<.001$ ,  $k=33$ ) and the dmPFC ( $x=6$   $y=38$   $z=48$ ,  $t=8.74$ ,  $FWEp<.001$ ,  $k=110$ ) across groups.

### *3.4 Network Characteristics*

Network analysis indicated that in patients with schizophrenia the dmPFC received overall lower input from other brain regions within the mentalizing network compared to healthy controls,  $U=98.0$ ,  $p=.001$  (Tab.1). Further edge-wise analysis indicates a significantly lower proportion of bilateral rTPJ to the dmPFC connections,  $\chi^2=10.47$ ,  $p=.001$  and of the lTPJ to the dmPFC,  $\chi^2=10.18$ ,  $p=.001$  in patients with schizophrenia compared to healthy controls (Tab.2, Fig. 2). Reduced overall out-going network connections of the dmPFC ( $U=132.0$ ,  $p=.011$ ), the lTPJ ( $U=119.0$ ,  $p=.004$ ) and the rTPJ ( $U=143.0$ ,  $p=.023$ ) in patients didn't reach the Bonferroni corrected threshold ( $\alpha<.0025$ ). The same applies to the reduced lTPJ in-going connections from other brain regions within the network in patients with schizophrenia ( $U=152.5$ ,  $p=.043$ ) (Tab.2).

Tab. 1. In-degree and out-degree of each network hub

| Direction/Region | Schizophrenia<br><i>M (STD)</i> | Controls<br><i>M (STD)</i> | Mann-Whitney-U | <i>Z</i> | <i>p</i> -value |
|------------------|---------------------------------|----------------------------|----------------|----------|-----------------|
| In_ITPJ          | 0.5128 (0.329)                  | 0.7222 (0.366)             | 152.5          | -2.026   | .043            |
| In_rTPJ          | 0.6026 (0.377)                  | 0.7593 (0.358)             | 173.5          | -1.536   | .125            |
| In_PREC          | 0.5897 (0.380)                  | 0.7222 (0.357)             | 185.0          | -1.235   | .217            |
| In_mPFC          | 0.3462 (0.358)                  | 0.7593 (0.339)             | 98.0           | -3.368   | .001*           |
| Out_ITPJ         | 0.5000 (0.316)                  | 0.7778 (0.228)             | 119.0          | -2.884   | .004            |
| Out_rTPJ         | 0.4872 (0.301)                  | 0.7037 (0.300)             | 143.0          | -2.278   | .023            |
| Out_PREC         | 0.5641 (0.336)                  | 0.7222 (0.307)             | 171.0          | -1.576   | .115            |
| Out_mPFC         | 0.5000 (0.341)                  | 0.7593 (0.375)             | 132.0          | -2.538   | .011            |

In = Number of effective connections a network node receives from other brain regions. Out = Number of effective connections a network node integrates into other brain regions. \* Significant Bonferroni corrected between-group differences are indicated.

Tab. 2. Edge-wise effective connectivity between the brain regions and group differences

| From region → region | Proportion Present<br>Schizophrenia | Proportion Present<br>Controls | $\chi^2$ | <i>p</i> -value |
|----------------------|-------------------------------------|--------------------------------|----------|-----------------|
| ITPJ → mPFC          | 0.3462                              | 0.8333                         | 10.182   | .001*           |
| ITPJ → rTPJ          | 0.5769                              | 0.7778                         | 1.91     | .167            |
| ITPJ → PREC          | 0.5769                              | 0.7222                         | .970     | .325            |
| rTPJ → mPFC          | 0.2308                              | 0.7222                         | 10.471   | .001*           |
| rTPJ → ITPJ          | 0.4231                              | 0.7222                         | 0.970    | .325            |
| rTPJ → PREC          | 0.6538                              | 0.6667                         | .008     | .930            |
| PREC → mPFC          | 0.4615                              | 0.7222                         | 2.946    | .086            |
| PREC → rTPJ          | 0.7308                              | 0.7778                         | .125     | .723            |
| PREC → ITPJ          | 0.5000                              | 0.6667                         | 1.204    | .272            |
| mPFC → rTPJ          | 0.5000                              | 0.7222                         | 2.713    | .140            |
| mPFC → ITPJ          | 0.4615                              | 0.7778                         | 4.40     | .036            |
| mPFC → PREC          | 0.5385                              | 0.7778                         | 2.632    | .105            |

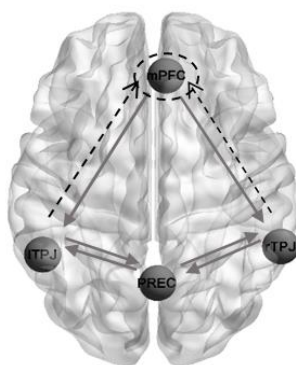
Effective endogenous connectivity between two brain regions (edge-connectivity) and their direction (indicated by the arrow) between the groups. \* Significant Bonferroni corrected between-group differences are indicated.

### *3.5 Network Characteristics and Task Dependent / Independent ToM-Indices*

Additionally, we examined potential associations between the Reading the Mind in the Eyes Task (RMET), as a task-independent index for mentalizing abilities and network characteristics. Notably, no significant between-group differences were observed for test scores in RMET ( $t(31)=1.69$ ,  $p=.101$ ; HC:  $M=26.56\pm3.56$ , Range: 33-21; SZ  $M=24.53\pm3.37$ , Range: 30-19). However, exploratory analyses revealed that the network characteristic is associated with task-independent and task-dependent indices of mentalizing, as suggested by significant higher RMET values in patients with vs. without an edge-connection in every network node (ITPJ-rTPJ ( $t(31)=3.04$ ,  $p=.005$ ), MPFC-ITPJ ( $t(31)=2.42$ ,  $p=.022$ ), rTPJ-PREC ( $t(31)=2.44$ ,  $p=.021$ )). Given that these analyses were not corrected for multiple comparisons, they should be considered as exploratory. Along this line, the subjects' understanding of the counterpart's intentions (Dynamic ToM) during the task correlated positively with the entire network structure (richness of edge-connections within the network;  $r=.257$ ,  $p=.046$ , one-tailed). Hence, internal (task-dependent) and external (task-independent) associations indicate an association between the ToM network structure and mentalizing processes. Given that the network characteristic and ToM-indices analysis are not corrected for multiple comparisons, they should be considered as exploratory.

### *3.6 Network Alterations and Medication*

In order to exclude that network alterations are influenced by medication, we examined a potential relation between-group network differences and the antipsychotic dose in patients (chlorpromazine equivalent). These analyses showed no significant associations between medication and the in-degree of the dmPFC in patients ( $r=-.068$ ,  $p=.74$ ), nor a higher dose in patients with vs without edge-wise connectivity between the rTPJ - dmPFC ( $t(24)=.66$ ,  $p=.51$ ) and the ITPJ - dmPFC ( $t(24)=-.28$ ,  $p=.78$ ). Additional control analysis can be found in the supplementary material.



**Fig. 3.** Differences in the network architecture between the groups, point on significantly reduced effective connections from the right and left temporo-parietal junction (rTPJ,lTPJ) to the dorsal medial prefrontal cortex (dmPFC, dashed lines) in patients with schizophrenia compared to healthy controls. Furthermore, in patients with schizophrenia the dmPFC received less frequently input from other brain regions in the mentalizing network (dashed circle).

## 4. Discussion

In this study we characterized the effective connectivity between brain regions forming the mentalizing network in patients with schizophrenia for the first time. The DCM network discovery approach indicates a sparser network architecture in patients and highlights the dmPFC as a disconnected network hub, since the region's activation changes are less influenced by neural activity changes of the right and the left TPJ. These findings provide first evidence that alterations within the core mentalizing network arise due to disconnections between at least three core regions of mentalizing processes in patients with schizophrenia and pin-point dysfunctions particularly of the forward TPJ connectivity.

### *4.1 A network perspective can facilitate the understanding of alterations in patients' mentalizing processes*

The examination of functional networks can facilitate the understanding of phenotypical alterations in mental disorders [1,2]. Similarly, alterations in brain networks have been found in depression [52] and autism [56]. Schizophrenia has been associated with disturbances in the interaction of distant brain regions, indicating that neural 'disconnections' contribute to the disorder's pathophysiology [3,4,6] underlined by findings of disruptions in functional



networks in childhood-onset schizophrenia [57]. One characteristic epiphenomenon of schizophrenia is impaired social functioning [58], which is linked to impairments in mentalizing processes [15] and might be affected by disrupted information processing between distant brain regions [59]. By investigating effective connectivity during a social decision-making game, the current DCM analysis [41] allowed us to pin-point how brain regions within the mentalizing network influence each other and whether their forward or backward effective connectivity is altered under restricted task conditions.

Our findings reveal that in patients with schizophrenia the dmPFC arises as a disconnected network hub, due to a dysfunctional effective forward connectivity from the bilateral TPJ, indicating an impaired information integration within the mentalizing network. Further analyses suggest that the integration between the regions within the network is associated with enhanced general mentalizing abilities, indicated by significant associations with the Reading the Mind in the Eyes task, a task-independent index for mentalizing abilities. On a system level our findings indicate that the connection intactness within the entire network is linked with an enhanced ability to generate representations of different interaction partners during the task, pointing to a link between the plethora of interconnections within the mentalizing network and beneficial social behavior during the task.

The question arises how these network dysfunctions translate into alterations in the mentalizing process in schizophrenia. Previous research suggests that the brain regions in the mentalizing network are implicated in mnemonic, prospective and default mode processes, indicating a common core function, such as the projection into another constructed perspective beyond the current environment [60]. Along this line, it is assumed that some forms of mentalizing processes draw on the own mental experience to model other people's mental states [60,61], probably by guiding mentalizing processes through self-knowledge [62], which has been associated with an intact dmPFC function [60,62,63]. Accordingly, it

has been found that the processing of introspective information and inference processes of other people's mental states converge in the dmPFC [64,65], which might be related to the region's role in social interactions [65]. Furthermore, it has been suggested that the dmPFC is implicated in the generation of higher-order, particularly sensory-independent social cognitive processes, which is reflected in the region's interconnection with associative and integrative brain regions, such as the TPJ [67]. Hence, the current findings of reduced neural integration between the TPJ and the dmPFC in patients might indicate a diminished incorporation of other peoples' mental states in higher-order, stimulus-independent processing modes.

Previous findings suggest that in patients with schizophrenia impairments in generating mental representations of others' mental states are associated with dysfunctions in the functional connectivity of the rTPJ to brain regions in the temporal lobe [43]. This points to alterations in the liaison between mentalizing and memory processes [51] which is in line with behavioral findings [68]. Notably, these previous results might be associated with the present findings of disconnections in patients' mentalizing network. During the experiment, the subjects can accumulate evidence about the playing partners' intentions and iteratively adapt their strategies, leading to a continuous updating of their mental representations, particularly in the TPJ [51]. Previously it has been shown that impaired TPJ updating results from a dysfunctional integration of memory information from previous experiences in the hippocampus and other temporal lobe regions [43]. This might be associated with the TPJ's impaired effective forward connectivity and could lead to an incomplete integration of mental models of others' intentions into dmPFC processes in patients with schizophrenia. Since the TPJ [69] and the dmPFC [67] are neuroanatomically connected with brain regions relevant for memory processes, a consequent analysis how this cognitive mechanism relies on episodic memory information requires further examination. So far, there is evidence from neural [43] and behavioral studies [68] that episodic memory impairments account for deficits in mental state inferences in patients with schizophrenia. Furthermore, our findings of a disconnected

frontal hub in the mentalizing network is in line with previous findings, which indicate that the frontal cortex receives less information from distant brain regions in patients with schizophrenia [13, 70] and begs the question whether these findings extend to alterations in further cognitive and symptomatic domains.

#### *4.2 Dysconnectivity between the ITPJ and the dmPFC has been associated with psychosis risk*

Remarkably, a genome-wide association study has previously found that healthy individuals with a high risk for psychosis show reduced connectivity between the dmPFC and the ITPJ during a mentalizing task [71]. Taking our and previous findings together, the dysfunctional connectivity between the TPJ and the dmPFC could be a valuable biomarker that might be present before the onset of the disorder and can also be found in patients with a diagnosis. Importantly, the ITPJ has been found to be associated with hallucinatory experiences. Accordingly, when direct electrical stimulation is applied to the region the emergence of a visual hallucination has been reported in a non-psychiatric person [72]. Vice versa, a reduction of auditory hallucinations has been reported following a cathodic tDCS-application in patients with schizophrenia [73], indicating that the TPJ might be a critical target region in the pathophysiology of schizophrenia. Evidence from tDCS-studies suggest that the rTPJ is associated with an abstract process of taking a perspective to view another person's mind [74,75,76], and that the dmPFC facilitates the integration of these processes into a higher order representation of a social constellation [65,74]. Future brain-stimulation studies have therefore to show whether the TPJ modulation can facilitate the dmPFC's integration and thereby improve mentalizing abilities in patients with schizophrenia.

#### *4.3 Limitations and future perspectives*

The findings should be interpreted in the light of some limitations. First, DCM is a hypothesis driven analysis strategy in which the selection of brain regions is conducted a-priori. Therefore, it should be noted that alterations in the complex cognitions which occur during

social interactions, such as mentalizing, might be further facilitated by secondary cognitive processes (e.g. memory, attention or self-processing), or might be considered their complex conglomerate. Accordingly, recent research suggests that mentalizing can be considered a multidimensional construct, subsuming and integrating different sets of cognitive mechanisms [22,77,78], which might be particularly relevant for patients with schizophrenia [43]. Along this line, it should be considered that in total seven participants failed to engage specific nodes within the specified network, which is a common issue in DCM studies [79]. Although several reasons can account for an individual's missing activation of a specific brain region [80], it might be that the excluded participants used an alternative strategy to perform the task and thus recruit different brain regions. Future studies should therefore investigate whether the joint interaction of other brain regions or networks with the mentalizing network can compensate or potentially account for the topological alterations we have found in this study. Another emerging question is how our findings impact real-life interactions in healthy participants and patients with schizophrenia. Therefore, future studies might assess how the amount and the quality of an individual's daily social life shapes the neurobiology of the mentalizing network. Hence, it could be clinically relevant to examine whether the outcome of psychological (group) interventions, which can be considered as structured and goal-oriented social interactions, are modulated by the intactness of the neural substrates underlying social interactions or whether therapeutic processes can lead to a reorganization of previously altered network structures.

However, by interpreting the findings it should be considered that, in the current task, mainly the cognitive aspects (as opposed to affective components) of the mentalizing process are assessed. Hence, future studies might examine whether alterations in the dmPFC-integration extend to alterations in affective mentalizing processes in patients with schizophrenia. A further limitation of our study is the medication intake in patients. Although, we statistically could not detect a significant association between alterations in the mentalizing network and

medication intake, it should be considered that the healthy and patients' group did differ in this relevant variable. Accordingly, it should be considered that the duration of illness in our sample was quite variable. Further large-scale studies might therefore examine in specific sub-samples whether the duration of illness influences the mentalizing network architecture in patients with schizophrenia. These aspects should be kept in mind during the interpretation of the results. The aim of this study, however, is to examine alterations in the architecture of the core network for mentalizing processes in patients, which might be an initial starting point for therapeutic approaches.

## **5. Conclusion**

The characterization of the mentalizing network in patients with schizophrenia revealed a sparser structure in patients, particularly due to a disconnected dmPFC, which is less driven by bilateral TPJ computations. Our findings highlight the relevance of the connectivity profile of the TPJ for social cognitive processes in schizophrenia, rendering it to a potential target region for neurostimulation.

## **Acknowledgments**

This work was supported by the Core Facility Brain Imaging, Faculty of Medicine, University of Marburg, Else Kröner-Fresenius-Stiftung (grant number: 2014\_A136) and Deutsche Forschungsgemeinschaft (grant numbers: STR-1146/4-1, STR-1146/8-1, STR-1146/9-1, to BS)

## **Competing interests**

The authors report no competing interests.

## References

- [1] V. Menon, Large-scale brain networks and psychopathology: a unifying triple network model, *Trends Cogn Sci.* 15(10) (2011) 483–506. <https://doi.org/10.1016/j.tics.2011.08.003>
- [2] O. Sporns, Structure and function of complex brain networks, *Dialogues Clin Neurosci.* 15(3) (2013) 247–262.
- [3] E. Bleuler, *Dementia Praecox oder Gruppe der Schizophrenien*, Deuticke, Leipzig, 1911.
- [4] K. J. Friston, The disconnection hypothesis, *Schizophr Res.* 30(2) (1998) 115–125. [https://doi.org/10.1016/S0920-9964\(97\)00140-0](https://doi.org/10.1016/S0920-9964(97)00140-0)
- [5] K.J. Friston, Schizophrenia and the disconnection hypothesis, *Acta Psychiatr Scand Suppl.* 395 (1999) 68–79.
- [6] K. Friston, Brown, H. R., Siemerikus, J., & K. E. Stephan, The dysconnection hypothesis (2016). *Schizophr Res.* 176(2-3) (2016) 83–94. <https://doi.org/10.1016/j.schres.2016.07.014>
- [7] K. Friston, Dysfunctional connectivity in schizophrenia, *World Psychiatry.* 1(2), 2002, 66–71.
- [8] J. Camchong, A.W. MacDonald, C. Bell, B.A. Mueller, K.O. Lim, Altered functional and anatomical connectivity in schizophrenia, *Schizophr Bull.* 37(3) (2011) 640–650. <https://doi.org/10.1093/schbul/sbp131>
- [9] E. Damaraju, E.A. Allen, A. Belger, J.M. Ford, S. McEwen, D.H. Mathalon, B.A. Mueller, G.D. Pearlson, S.G. Potkin, A. Preda, et al, Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia, *Neuroimage Clin.* 5 2014 298–308. <https://doi.org/10.1016/j.nicl.2014.07.003>
- [10] M.-E. Lynall, D.S. Bassett, R. Kerwin, P.J. McKenna, M. Kitzbichler, U. Muller, E. Bullmore, Functional connectivity and brain networks in schizophrenia, *J Neurosci.* 30(28) 2010 9477–9487. <https://doi.org/10.1523/JNEUROSCI.0333-10.2010>
- [11] P. Skudlarski, K. Jagannathan, K. Anderson, M.C. Stevens, V.D. Calhoun, B.A. Skudlarska, G. Pearlson, Brain Connectivity Is Not Only Lower but Different in Schizophrenia: A Combined Anatomical and Functional Approach, *Biological Psychiatry.* 68(1) 2010 61–69. <https://doi.org/10.1016/j.biopsych.2010.03.035>
- [12] B. Straube, A. Green, K. Sass, T. Kircher, Superior Temporal Sulcus Disconnectivity During Processing of Metaphoric Gestures in Schizophrenia, *Schizophr Bull.* 40(4) 2013 936–944. <https://doi.org/10.1093/schbul/sbt110>
- [13] M.P. van den Heuvel, R. C.W. Mandl, C.J. Stam, R.S. Kahn, H.E. Hulshoff Pol, Aberrant frontal and temporal complex network structure in schizophrenia: a graph theoretical analysis, *J Neurosci.* 30(47) 2010 15915–15926. <https://doi.org/10.1523/JNEUROSCI.2874-10.2010>
- [14] N.A. Crossley, A. Mechelli, J. Scott, F. Carletti, P.T. Fox, P. McGuire, E.T. Bullmore, The hubs of the human connectome are generally implicated in the anatomy of brain disorders, *Brain.* 137(8) 2014 2382–2395.
- [15] J.M. Hooley, Social factors in schizophrenia, *Curr Dir Psychol Sci.* 19(4) 2010 238–242. <https://doi.org/10.1177/0963721410377597>
- [16] M. Brüne, "Theory of mind" in schizophrenia: a review of the literature, *Schizophr Bull.* 31(1) 2005 21–42. <https://doi.org/10.1093/schbul/sbi002>

- [17] M. Sprong, P. Schothorst, E. Vos, J. Hox, H. van Engeland, Theory of mind in schizophrenia: meta-analysis, *Br J Psychiatry*. 191 2007 5–13.  
<https://doi.org/10.1192/bjp.bp.107.035899>
- [18] K.M. Healey, C.F. Bartholomeusz, D.L. Penn, Deficits in social cognition in first episode psychosis: A review of the literature, *Clin Psychol Rev*. 50 2016 108–137. <https://doi.org/10.1016/j.cpr.2016.10.001>
- [19] A.K. Martin, G. Robinson, I. Dzaif, D. Reutens, B. Mowry, Theory of mind and the social brain: implications for understanding the genetic basis of schizophrenia, *Genes Brain Behav*. 13(1) 2014 104–117. <https://doi.org/10.1111/gbb.12066>
- [20] S.M. Couture, E.L. Granholm, S.C. Fish, A Path Model Investigation of Neurocognition, Theory of Mind, Social Competence, Negative Symptoms and Real-World Functioning in Schizophrenia. *Schizophr Res*. 125(2-3) 2010 152–160.  
<https://doi.org/10.1016/j.schres.2010.09.020>
- [21] P. Molenberghs, H. Johnson, J.D. Henry, J.B. Mattingley, Understanding the minds of others: A neuroimaging meta-analysis, *Neurosci Biobehav Rev*. 65 (2016) 276–291.  
<https://doi.org/10.1016/j.neubiorev.2016.03.020>
- [22] M. Schurz, J. Radua, M. Aichhorn, F. Richlan, J. Perner, Fractionating theory of mind: a meta-analysis of functional brain imaging studies, *Neurosci Biobehav Rev*. 42 (2014) 9–34.  
<https://doi.org/10.1016/j.neubiorev.2014.01.009>
- [23] R.B. Mars, F.X. Neubert, M.P. Noonan, J. Sallet, I. Toni, M.F. Rushworth, On the relationship between the “default mode network” and the “social brain”, *Front hum neurosci*. 6 (2012) 189.
- [24] L. Schilbach, D. Bzdok, B. Timmermans, P.T. Fox, A.R. Laird, K. Vogeley, S.B. Eickhoff, Introspective minds: using ALE meta-analyses to study commonalities in the neural correlates of emotional processing, social & unconstrained cognition, *PloS one*. 7(2) (2012) e30920.
- [25] L. Schilbach, S.B. Eickhoff, A. Rotarska-Jagiela, G.R. Fink, K. Vogeley, Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the “default system” of the brain, *Conscious and Cog*. 17(2) (2008) 457-467.
- [26] R.N. Spreng, R.A. Mar, I remember you: A role for memory in social cognition and the functional neuroanatomy of their interaction, *Brain Res*. 1428C (2010) 43–50.  
<https://doi.org/10.1016/j.brainres.2010.12.024>
- [27] R.P. Spunt, M.L. Meyer, M.D. Lieberman, The default mode of human brain function primes the intentional stance, *J Cog Neurosci*. 27(6) (2015) 1116-1124.
- [28] M. Amft, D. Bzdok, A.R. Laird, P.T. Fox, L. Schilbach, S.B. Eickhoff, Definition and characterization of an extended social-affective default network, *Brain Struct. Funct*. 220(2) (2015) 1031-1049.
- [29] P. Das, J. Lagopoulos, C.M. Coulston, A.F. Henderson, G.S. Malhi, Mentalizing impairment in schizophrenia: a functional MRI study, *Schizophr Res*. 134(2-3) (2012) 158–164. <https://doi.org/10.1016/j.schres.2011.08.019>



- [30] J. Lee, J. Quintana, P. Nori, M.F. Green, Theory of mind in schizophrenia: exploring neural mechanisms of belief attribution, *Soc Neurosci.* 6(5-6) (2011) 569–581. <https://doi.org/10.1080/17470919.2011.620774>
- [31] H. Walter, A. Ciaramidaro, M. Adenzato, N. Vasic, R.B. Ardito, S. Erk, B.G. Bara, Dysfunction of the social brain in schizophrenia is modulated by intention type: an fMRI study, *Soc Cogn Affect Neurosci.* 4(2) (2009) 166–176. <https://doi.org/10.1093/scan/nsn047>
- [32] M. Brüne, S. Özgürdal, N. Ansorge, H.G. von Reventlow, S. Peters, V. Nicolas, M. Tegenthoff, G. Juckel, S. Lissek, An fMRI study of “theory of mind” in at-risk states of psychosis: Comparison with manifest schizophrenia and healthy controls, *Neuroimage.* 55(1), (2011) 329–337. <https://doi.org/10.1016/j.neuroimage.2010.12.018>
- [33] D. Dodell-Feder, L.M. Tully, S.H. Lincoln, C.I. Hooker, The neural basis of theory of mind and its relationship to social functioning and social anhedonia in individuals with schizophrenia, *Neuroimage Clin.* 4 (2014) 154–163. <https://doi.org/10.1016/j.nicl.2013.11.006>
- [34] M. Brüne, S. Lissek, N. Fuchs, H. Witthaus, S. Peters, V. Nicolas, G. Juckel, M. Tegenthoff, An fMRI study of theory of mind in schizophrenic patients with "passivity" symptoms, *Neuropsychologia.* 46(7) (2008) 1992–2001. <https://doi.org/10.1016/j.neuropsychologia.2008.01.023>
- [35] L. Kronbichler, M. Tschernegg, A.I. Martin, M. Schurz, M. Kronbichler, Abnormal brain activation during theory of mind tasks in schizophrenia: a meta-analysis. *Schizophr Bull.* 43(6) (2017) 1240-1250.
- [36] W. Pettersson-Yeo, P. Allen, S. Benetti, P. McGuire, A. Mechelli, Dysconnectivity in schizophrenia: where are we now?, *Neurosci Biobehav Rev.* 35(5) (2011) 1110-1124.
- [37] H.J. Park, K. Friston, Structural and functional brain networks: from connections to cognition, *Science.* 342(6158) (2013) 1238411.
- [38] O. Sporns, Contributions and challenges for network models in cognitive neuroscience, *Nature neurosci.* 17(5) (2014) 652.
- [39] O. Sporns, R.F. Betzel, Modular brain networks, *Annual review of psychology.* 67 (2016) 613-640.
- [40] J. D. Power, B.L. Schlaggar, C.N. Lessov-Schlaggar, S.E. Petersen, Evidence for hubs in human functional brain networks. *Neuron.* 79(4) (2013) 798-813.
- [41] K.J. Friston, B. Li, J. Daunizeau, K.E. Stephan, Network discovery with DCM, *Neuroimage.* 56(3) (2011) 1202–1221. <https://doi.org/10.1016/j.neuroimage.2010.12.039>
- [42] K.J., Friston, L. Harrison, W. Penny, Dynamic causal modelling, *Neuroimage.* 19(4) (2003) 1273–1302. [https://doi.org/10.1016/S1053-8119\(03\)00202-7](https://doi.org/10.1016/S1053-8119(03)00202-7)
- [43] F. Bitsch, P. Berger, A. Nagels, I. Falkenberg, B. Straube, Impaired Right Temporoparietal Junction–Hippocampus Connectivity in Schizophrenia and Its Relevance for Generating Representations of Other Minds, *Schizophr Bull.* 16(25) (2018) 559. <https://doi.org/10.1093/schbul/sby132>
- [44] T. Yarkoni, R.A. Poldrack, T.E. Nichols, D.C. van Essen, T.D. Wager, Large-scale automated synthesis of human functional neuroimaging data, *Nat. Methods.* 8(8) (2011) 665. <https://doi.org/10.1038/nmeth.1635>

- [45] K. Friston, W. Penny, Post hoc Bayesian model selection, *Neuroimage*. 56(4) (2011) 2089–2099. <https://doi.org/10.1016/j.neuroimage.2011.03.062>
- [46] M.J. Rosa, K. Friston, W. Penny, Post-hoc selection of dynamic causal models, *J Neurosci Methods*. 208(1) (2012) 66–78. <https://doi.org/10.1016/j.jneumeth.2012.04.013>
- [47] H.U. Wittchen, M. Zaudig, T. Fydrich, Structured Clinical Interview for DSM-IV Axis I and II - SCID. Hogrefe, Göttingen, 1997.
- [48] N.C. Andreasen, Scale for the assessment of positive symptoms (SAPS). University of Iowa, Iowa City, 1984.
- [49] N.C. Andreasen, The Scale for the Assessment of Negative Symptoms (SANS), Conceptual and Theoretical Foundations, *Br J Psychiatry*. 155(S7) (1989) 49–52. <https://doi.org/10.1192/S0007125000291496>
- [50] R.C. Oldfield, The assessment and analysis of handedness: The Edinburgh inventory, *Neuropsychologia*. 9(1) (1971) 97–113. [https://doi.org/10.1016/0028-3932\(71\)90067-4](https://doi.org/10.1016/0028-3932(71)90067-4)
- [51] F. Bitsch, P. Berger, A. Nagels, I. Falkenberg, B. Straube, The role of the right temporo-parietal junction in social decision-making, *Hum Brain Mapp*. 39(7) (2018) 3072–3085. <https://doi.org/10.1002/hbm.24061>
- [52] M.P. Hyett, M.J. Breakspear, K.J. Friston, C.C. Guo, G.B. Parker, Disrupted effective connectivity of cortical systems supporting attention and interoception in melancholia, *JAMA Psychiatry*. 72(4) (2015) 350–358. <https://doi.org/10.1001/jamapsychiatry.2014.2490>
- [53] K.E. Stephan, N. Weiskopf, P.M. Drysdale, P.A. Robinson, K.J. Friston, Comparing hemodynamic models with DCM, *Neuroimage*. 38(3) (2007) 387–401. <https://doi.org/10.1016/j.neuroimage.2007.07.040>
- [54] A. Wroblewski, Y. He, B. Straube, Dynamic Causal Modelling suggests impaired effective connectivity in patients with schizophrenia spectrum disorders during gesture-speech integration. *Schizophrenia Research*. (2019).
- [55] N.C. Andreasen, M. Pressler, P. Nopoulos, D. Miller, B.C. Ho. Antipsychotic dose equivalents and dose-years: a standardized method for comparing exposure to different drugs, *Biol Psychiatry*. 67 (2010) 255–262.
- [56] S.J. Gotts, W.K. Simmons, L.A. Milbury, G.L. Wallace, R.W. Cox, A. Martin, Fractionation of social brain circuits in autism spectrum disorders, *Brain*. 135(9) (2012) 2711–2725. <https://doi.org/10.1093/brain/aws160>
- [57] R.A. Berman, S.J. Gotts, H.M. McAdams, D. Greenstein, F. Lalonde, L. Clasen, R.E. Watsky, L. Shora, A.E. Ordonez, A. Raznahan, et al, Disrupted sensorimotor and social-cognitive networks underlie symptoms in childhood-onset schizophrenia, *Brain*. 139(1) (2016) 276–291. <https://doi.org/10.1093/brain/awv306>
- [58] American Psychiatric Association, Diagnostic and statistical manual of mental disorders (DSM-5®). American Psychiatric Pub, 2013.
- [59] M.F. Green, W.P. Horan, J. Lee, Social cognition in schizophrenia, *Nature Rev Neurosci*. 16(10) (2015) 620. <https://doi.org/10.1038/nrn4005>
- [60] R.L. Buckner, D.C. Carroll, Self-projection and the brain, *Trends Cogn Sci*. 11(2) (2007) 49–57. <https://doi.org/10.1016/j.tics.2006.11.004>

- [61] A.C. Jenkins, J.P. Mitchell, Medial prefrontal cortex subserves diverse forms of self-reflection, *Soc Neurosci.* 6(3) (2011) 211–218.  
<https://doi.org/10.1080/17470919.2010.507948>
- [62] A. Waytz, J.P. Mitchell, Two Mechanisms for Simulating Other Minds, *Curr Dir Psychol Sci.* 20(3) (2011) 197–200. <https://doi.org/10.1177/0963721411409007>
- [63] D.I. Tamir, J.P. Mitchell, Neural correlates of anchoring-and-adjustment during mentalizing, *Proc Natl Acad Sci.* 107(24) (2010) 10827–10832.  
<https://doi.org/10.1073/pnas.1003242107>
- [64] J.K. Rilling, J.E. Dagenais, D.R. Goldsmith, A.L. Glenn, G. Pagnoni, Social cognitive neural networks during in-group and out-group interactions, *Neuroimage.* 41(4) (2008) 1447–1461. <https://doi.org/10.1016/j.neuroimage.2008.03.044>
- [65] R. Saxe, Uniquely human social cognition, *Curr Opin Neurobiol.* 16(2) (2006) 235–239.  
<https://doi.org/10.1016/j.conb.2006.03.001>
- [66] T. Vanderwal, E. Hunyadi, D.W. Grupe, C.M. Connors, R.T. Schultz, Self, mother and abstract other: an fMRI study of reflective social processing, *Neuroimage.* 41(4) (2008) 1437–1446. <https://doi.org/10.1016/j.neuroimage.2008.03.058>
- [67] D. Bzdok, R. Langner, L. Schilbach, D.A. Engemann, A.R. Laird, P.T. Fox, S.B. Eickhoff, Segregation of the human medial prefrontal cortex in social cognition, *Front Hum Neurosci.* 7:232 (2013). <https://doi.org/10.3389/fnhum.2013.00232>
- [68] R. Corcoran, C.D. Frith, Autobiographical memory and theory of mind: evidence of a relationship in schizophrenia, *Psychol Med.* 33(5) (2003) 897–905.  
<https://doi.org/10.1017/S0033291703007529>
- [69] R.M. Carter, S.A. Huettel, A nexus model of the temporal–parietal junction, *Trends Cogn Sci.* 17(7) (2013) 328–336. <https://doi.org/10.1016/j.tics.2013.05.007>
- [70] M.P. van den Heuvel, O. Sporns, G. Collin, T. Scheewe, R.C.W Mandl, W. Cahn, J. Goñi, H.E. Hulshoff Pol, R.S. Kahn, Abnormal rich club organization and functional brain dynamics in schizophrenia, *JAMA Psychiatry.* 70(8) (2013) 783–792.  
<https://doi.org/10.1001/jamapsychiatry.2013.1328>
- [71] H. Walter, K. Schnell, S. Erk, C. Arnold, P. Kirsch, C. Esslinger, D. Mier, M.M. Schmitgen, M. Rietschel, S.H. Witt et al., Effects of a genome-wide supported psychosis risk variant on neural activation during a theory-of-mind task, *Mol Psychiatry.* 16(4) (2011) 462.  
<https://doi.org/10.1038/mp.2010.18>
- [72] S. Arzy, M. Seeck, S. Ortigue, L. Spinelli, O. Blanke, Induction of an illusory shadow person, *Nature.* 443(7109) (2006) 287. <https://doi.org/10.1038/443287a>
- [73] M. Mondino, R. Jardri, M.-F. Suaud-Chagny, M. Saoud, E. Poulet, J. Brunelin. Effects of Fronto-Temporal Transcranial Direct Current Stimulation on Auditory Verbal Hallucinations and Resting-State Functional Connectivity of the Left Temporo-Parietal Junction in Patients With Schizophrenia, *Schizophr Bull.* 42(2) (2016) 318–326. <https://doi.org/10.1093/schbul/sbv114>
- [74] A.K. Martin, J. Huang, A. Hunold, M. Meinzer, Dissociable Roles Within the Social Brain for Self–Other Processing: A HD-tDCS Study, *Cereb. Cortex.* 29(8) (2019) 3642–3654.
- [75] I. Santiesteban, M.J. Banissy, C. Catmur, G. Bird, Enhancing social ability by stimulating right temporoparietal junction, *Curr Biol.* 22(23) (2012) 2274–2277.

- [76] H. Wang, E. Callaghan, G. Gooding-Williams, C. McAllister, K. Kessler, Rhythm makes the world go round: An MEG-TMS study on the role of right TPJ theta oscillations in embodied perspective taking, *Cortex*. 75 (2016) 68-81.
- [77] S.M. Schaafsma, D.W. Pfaff, R.P. Spunt, R. Adolphs, Deconstructing and reconstructing theory of mind, *Trends Cogn Sci*. 19(2) (2015) 65-72.
- [78] K.R. Warnell, E. Redcay, Minimal coherence among varied theory of mind measures in childhood and adulthood, *Cognition*. 191 (2019) 103997.
- [79] S.B. Zaghlool, C.L. Wyatt, Missing data estimation in fMRI dynamic causal modeling. *Frontiers in neuroscience*. 8 191 (2014).
- [80] P. Zeidman, A. Jafarian, N. Corbin, M.L. Seghier, A. Razi, C.J. Price, K.J. Friston, A guide to group effective connectivity analysis, part 1: First level analysis with DCM for fMRI. *Neuroimage*. 200 (2019) 174-190.