Abnormal Brain Activation during Theory of Mind Tasks in Schizophrenia: A Meta-Analysis

RUNNING TITLE: Theory of Mind Brain Activation in Schizophrenia

Lisa Kronbichler¹, Melanie Tschernegg², Anna Isabel Martin², Matthias Schurz² & Martin Kronbichler¹,²

¹ Neuroscience Institute, Paracelsus Medical University, Christian-Doppler Clinic, Salzburg, Austria

² Centre for Neurocognitive Research and Department of Psychology, University of Salzburg, Salzburg, Austria

Supplementary Data and Material
Materials and Methods

Included Studies and Characteristics

To assure the quality of included studies, only papers published in peer-reviewed journals were included in the meta-analysis. The initial MEDLINE® search with the terms “schizophrenia”, “psychosis”, “Theory of Mind”, “mentalizing”, “fMRI”, “functional magnetic resonance imaging”, “PET”, “Positron Emission Tomography” and “neuroimaging”\(^1\) revealed 89 candidate papers (last updated on 10.11.2016). By means of title and abstract screenings, we excluded 36 studies which clearly were not suitable for our purpose (including reviews, meta-analyses, methodological or theoretical papers, and gene studies). The remaining 53 papers were further addressed concerning their statistical and neuroimaging methods and the availability of the appropriate group and condition contrasts. When a sample was used in more than one publication, the one with the best fitting Tom contrasts (compared to other included studies) was selected. Details are presented in the Figure below. An additional search was done using paper citations and google scholar search. 21 studies that matched the a priori criteria remained for this meta-analysis.

---

1 The exact search term was (“schizophrenia” OR “psychosis”) AND (“Theory of Mind” OR “mentalizing” OR “perspective taking”) AND (“fMRI” OR “functional magnetic resonance imaging” OR “PET” OR “Positron Emission Tomography” OR “neuroimaging”).
No significant differences between patients and controls were found in terms of mean age and sex, since the included studies were already well matched in this regard. A great majority of studies reported symptom severity by means of PANSS and SAPS/SANS scores. For the sake of comparability, the latter were transformed into PANSS values, using the formulas for total scale values of Van Erp et al. (2014).

Medication could not be considered as meta-regressor in our analyses since 10 studies did not report the exact dose of medication. Data were last updated in November 2016.

**Task categorisation**

Task categorisation strongly relies on the criteria and descriptions Molenbergh provides in his meta-analysis. Task categorisation based on the following criteria was done by the first and the senior authors independently. Rare discrepancies were discussed and resolved in meetings with experienced colleagues. For details concerning theoretical background and detailed descriptions please refer to Molenbergh et al., (2016).
Presentation Modality

The modality of presentation was categorised as visual if the task did not include any written or spoken language content (like for example TOM cartoons, silent videos and photographs). Tasks that require verbal abilities (like TOM stories) were defined as verbal.

Mentalizing Inference

For the classification of mentalizing inference, we searched in the studies task description for key terms which Molenbergh used to describe affective vs cognitive mentalizing.

“Whereas “hot” or affective ToM requires an understanding of others’ emotions or feelings, “cold” or cognitive ToM requires an understanding of their beliefs, thoughts or intentions (Brothers and Ring, 1992).” (Molenbergh et al., 2016)

Accordingly, tasks were characterized as affective if the individual studies’ task description included key terms like emotions, feelings, affective states, emotional states.

By contrast, tasks were classified as cognitive if they were described by terms like belief, thoughts, intention, sarcasm, deception, (white) lie, social faux-pas or false-belief.

Instructional Cues

Tasks were categorised as explicit if participants were explicitly asked to attend to other’s mental states. If this was not the case, the task was defined as implicit.

Meta-analytic Methods

Anisotropic kernels were used to account for spatial anisotropy of activation clusters due to anatomical constraints. All foci of under- and overactivation reported by the included studies were transformed to MNI space with a built-in feature using the icbm2tal transform (Lancaster et al., 2007). Meta-analyses were restricted to a specific built-in SDM gray matter
Effect size maps were recreated for each study by convolving reported foci with a fully anisotropic un-normalized Gaussian kernel ($\alpha=1$). The anisotropy of the kernel was based on the spatial correlations of the gray matter template. Within a study, values obtained by close anisotropic kernels were combined by square-distance-weighted averaging. To combine the data across study-specific effect size maps, a random effects general linear model was used. Statistical significance was examined by a permutation test ($n=50$) that randomizes the location of activation foci within the SDM gray matter template. Please note the recent publication of Joaquim Radua et al. (Radua, J., Mataix-Cols, D., Phillips, M. L., El-Hage, W., Kronhaus, D. M., Cardoner, N., & Surguladze, S. (2012). A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. European psychiatry, 27(8), 605-611.), demonstrating that as few as 20 randomisations return high statistical stability.

Replicability of meta-analytic results - Jackknife sensitivity Analysis

Systematic whole-brain voxel-based jackknife sensitivity analysis was used to evaluate the replicability of the meta-analytic findings. That means, the separate meta-analyses were repeated for the number of included studies while excluding each time a different study. The rationale of this procedure is that if a meta-analytic finding remains statistically significant in all or most of the combinations of studies, it can be concluded that this finding is robust against changes of the sample and thus highly replicable (Radua and Mataix-Cols, 2009).

Jackknife sensitivity analysis revealed that most of our meta-analytic results remained unchanged in the majority of recalculations, indicating strong robustness against changes in individual samples. Perfect replicability was found for right pre-motor cortex and left inferior parietal cortex. Strong replicability (19 or 20 of 21 leave-one-out recalculations) was found for medial prefrontal cortex, medial occipitoparietal, right lingual gyrus and right inferior parietal cortex. Left lateral occipitotemporal and left cingulated gyrus revealed under-activation in 18 out of 21 recalculations. Weaker robustness was found for left orbitofrontal cortex in 14 out of 21 recalculations.
**Publication Bias**

We checked for possible publication bias by means of two established methods widely used in SDM meta-analyses, namely a visual inspection of funnel plots and Egger test (e.g., Bora, Fornito, Yücel, & Pantelis, 2010; Bora, Fornito, Pantelis & Yücel, 2010; Fusar-Poli, Radua, McGuire, & Borgwardt, 2011). Estimated effect sizes and variance of each study were extracted from each meta-analytic cluster. Effect sizes were then plotted against the standard error in each region. Neither the visual inspection of funnel plots nor Egger regression (Egger, Smith, Schneider & Minder, 1997) revealed evidence for publication bias in the current dataset (all ps > .3).

**Meta-Regressions**

To assess the relation between meta-analytic effect sizes and patient sample characteristics like symptom severity and duration of illness, the factors positive symptoms (PANSS Positive score), negative symptoms (PANSS Negative score) and duration of illness (n(Years) since first diagnosis) were taken as regressors in separate Meta-Regressions. However, it is recommended to interpret these meta-regressions with caution and to handle them as exploratory analyses since a visual inspection of the data showed that significant results were caused by several strong outliers in brain activation estimates. Scatterplots depict the results in Figure SUP2. Statistical values are also provided in the figure.

**Overlap Task Subgroups with Main Analysis**

Overlaps between the statistical maps of the main analysis and the separate task analyses were calculated using the ImCalc function of SPM12. Here, the overlap was calculated by multiplying the two statistical maps thus resulting in a statistical map that indicates the overlapping clusters.

As evident from Table 3, some overlaps between the subgroup analyses with the main analysis seem rather small. However, the small overlap is mostly due to bigger clusters in the subgroup analyses as compared to their corresponding cluster in the main analysis. The percentage overlap in our study describes to what extent the cluster of the subgroup analysis lies within the corresponding cluster of the overall analysis. Therefore, most widespread cluster of the subgroup analyses reveal moderate to small overlap with the corresponding (smaller) clusters of the main analysis. This is not the case for the cingulate cortex cluster of the affective tasks subgroup.
the minimum overlap is caused by a more ventral location of the subgroup analysis cluster. Example images of these two minimum-overlap types are shown in Figure SUP4.

![Image of brain with TPJ clusters](image.png)

**Figure SUP4.** The TPJ cluster of the current meta-analysis on renown TPJ parcellations in the literature. The left panel depicts the TPJ parcellation of Mars et al. (2012), the right panel depicts the TPJ parcellation of Igelström et al. (2015). (Dark) blue = Inferior parietal lobe / TPJd as defined in the literature; red = TPJp as defined in the literature; green = inferior parietal lobe as shown in our analysis; pink = TPJp as shown in our analysis. Please note that the atlas of Mars was originally done for the right hemisphere. For illustration purposes, we flipped the atlas horizontally.

### Results

**TPJ cluster**

To compare the TPJ cluster of the current analyses with prior TPJ parcellations, we mapped our results on the TPJ atlases of Mars et al. (2012) and Igelström et al. (2015).

*Meta-Regression: Positive Symptoms*

Increased Positive Symptoms were related to convergent over-activation across studies in patients compared to healthy controls in the left inferior frontal gyrus.

*Meta-Regression: Negative Symptoms*

Increased Negative Symptoms were related to convergent under-activation in right supramarginal gyrus. Increased Negative Symptoms were found to go along with over-activation in the right frontal pole.

*Meta-Regression: Duration of Illness*
A longer duration of illness was related to convergent under-activation in patients compared to controls in the left superior frontal gyrus. Put differently, in this region studies with an increased duration of illness of their patients group show greater under-activation for patients compared to healthy controls.

Analyses of Task-Subgroups

The extent of overlap of the individual clusters with the corresponding cluster of the main analysis are listed in table 3 of the main document.

Cognitive/Affective
The analysis of studies using a cognitive Theory of Mind task revealed convergent under-activation for patients in MPFC, right premotor cortex, left lateral occipito-temporal cortex and right lingual gyrus. Convergent over-activation for patients was shown in left inferior parietal cortex. Under-activation that did not overlap with any of the main cluster was localized in the middle temporal gyrus (MTG).

Affective Theory of Mind studies depict convergent under-activation for patients compared to controls in left MPFC, left medial occipitoparietal and left cingulate. Convergent over-activation was reported in right inferior parietal cortex. Under-activation that did not overlap with any of the main cluster was localized in the right thalamus.

Verbal/Visual

The analysis of studies using verbal Theory of Mind tasks shows convergent under-activation in MPFC and in right premotor cortex. Convergent over-activation was found in left inferior parietal cortex. Under-activation that did not overlap with any of the main cluster was localized in the left thalamus, right superior temporal gyrus and the right cerebellum and non-overlapping over-activation was found in the left precentral gyrus.

Visual Theory of Mind studies revealed convergent under-activation for patients in left orbito-frontal cortex, left cingulate gyrus and right premotor cortex. Convergent over-activation was found in left inferior parietal cortex.

Implicit/Explicit

The analysis of studies using implicit Theory of Mind tasks show convergent under-activation in right premotor cortex and in left orbito-frontal cortex. Convergent over-activation for patients compared to healthy controls was found in bilateral inferior parietal cortex. Under-activation that did not overlap with any of the main cluster was localized in right MTG, left angular gyrus and left precentral gyrus.

Explicit Theory of Mind studies revealed convergent under-activation for patients in MPFC, right lingual gyrus, medial occipitoparietal cortex and cingulate cortex. Convergent over-activation
was found in left inferior parietal cortex. Under-activation that did not overlap with any of the main cluster was localized in left MTG and non-overlapping over-activation was found in the right orbito-frontal cortex and in the left angular gyrus.

Separate Group Analyses

In order to explore on brain regions that may not differ between schizophrenic patients and controls, we computed separate analyses for both groups. However, since the main focus of the included individual studies was on a direct comparison between groups, the required contrast were merely provided in 11 out of 21 studies. Those studies are marked with an asterisk in the Supplementary Table 1. Due to the relatively small amount of included studies, these analyses should be interpreted cautiously. Healthy controls reveal increased activation during Theory of Mind tasks compared to control tasks in medial prefrontal cortex, left inferior parietal cortex and in medial occipito-parietal cortex. Schizophrenic patients show convergent overactivation in left orbito-frontal cortex, bilateral inferior parietal cortex, right premotor cortex, medial prefrontal cortex and in medial occipito-parietal cortex.

---

2 Please note that separate analyses must provide the exact same contrasts as the group comparison contrasts in order to be included. For example, if the group comparison schizophrenia versus healthy controls shows the contrast false belief > false photograph, the separate analyses for controls and schizophrenia patients must - both - also provide the false belief > false photograph contrast. Contrasts between the TOM condition and another control condition were not accepted as substitute.
Figure SUP3. Results of separate meta-analyses for healthy controls’ and patients’ brain activation during Theory of Mind tasks. Convergent activation for healthy controls is depicted in violet. Convergent brain activation of patients is shown in yellow.

Supplementary Table 1. Separate meta-analyses of functional brain activation during Theory of Mind Tasks in patients and healthy controls.

| Label                        |   x  |   y  |   z  | SDM-Z | Voxel |
|------------------------------|------|------|------|-------|-------|
| Controls                     |      |      |      |       |       |
| R Inferior Parietal Cortex   | 64   | -30  | -2   | 5.35  | 2557  |
| L Inferior Parietal Cortex   | -60  | -62  | 24   | 4.73  | 2521  |
| Medial Occipitoparietal Cortex | -6  | -56  | 28   | 3.2   | 2183  |
| Medial Prefrontal Cortex     | -6   | 56   | 16   | 3.8   | 1606  |
| Patients                     |      |      |      |       |       |
| L Inferior Parietal Cortex   | -52  | -62  | 8    | 3.83  | 2281  |
| R Inferior Parietal Cortex   | 58   | -26  | 0    | 2.87  | 1810  |
| Medial Occipitoparietal Cortex | -12 | -52  | 34   | 2.26  | 281   |
| Medial Prefrontal Cortex     | -8   | 48   | 40   | 2.50  | 119   |
| R Premotor Cortex            | 48   | 4    | 30   | 2.17  | 84    |
| L Orbitofrontal Cortex       | -42  | 22   | -4   | 2.24  | 71    |

Note: L = left, R = right, all listed cluster revealed more for ToM tasks compared to control tasks.
Figure SUP4. Examples of two different overlapping patterns of task subgroup analyses and the overall analysis. Red shows example activation in the overall analysis. Green depicts example activation in the affective tasks subgroup analysis. A) Small overlap by dislocation: The small amount of overlap is due to a more ventral location of the cingulate gyrus cluster in the subgroup compared to the overall analysis. B) Small overlap by cluster size: The small amount of overlap is due to a more widespread IPL activation in the affective tasks compared to the overall analysis.
### Supplementary Table 2. Task description of the individual studies included in the meta-analysis.

| Author      | Imaging | Experimental Task                                                                 | Control Task                                                                                     | Contrast                                | Task Classification                      |
|-------------|---------|------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|-----------------------------------------|------------------------------------------|
| Brüne (2008)* | fMRI    | Participants see a cartoon in correct order and subsequently answer two questions concerning the Intention and expectations of the protagonist. | Participants see a cartoon with randomly arranged pictures and subsequently answer questions about objects characteristics in the cartoon. | mentalizing judgment > factual judgment | cognitive/affective explicit/implicit visual/verbal |
|             |         |                                                                                    |                                                                                                  |                                         |                                          |
| Walter (2009)* | fMRI    | Participants see a cartoon and select a correct ending. Three cartoon categories are presented: private Intention, prospective social Intention and communicative Intention. | Participants see a cartoon and select a correct ending with respect to physical causality.        | social intention > physical causality   | cognitive implicit visual               |
|             |         |                                                                                    |                                                                                                  |                                         |                                          |
| Benedetti (2009) | fMRI    | Participants see a cartoon and select a correct ending. Two cartoon categories are presented: Theory of Mind and Empathy. | Participants see a cartoon and select a correct ending with respect to physical causality.        | Theory of Mind > physical causality     | cognitive implicit visual               |
|             |         |                                                                                    |                                                                                                  |                                         |                                          |
| Lee (2010)* | fMRI    | Participants see a cartoon and select a correct ending. Two cartoon categories are presented: cognitive empathy, emotional empathy and inhibitory empathy. | Participants see a cartoon and select a correct ending with respect to physical causality.        | cognitive empathy > physical causality  | cognitive implicit visual               |
|             |         |                                                                                    |                                                                                                  |                                         |                                          |
| Lee (2011) | fMRI    | Participants read false belief stories and select one of two words that fits best in the subsequently sentence. | Participants read false photography stories and select one of two words that fits best in the subsequently sentence. | false belief > false photograph         | cognitive explicit verbal               |

n_SZ = 9 n_HC = 13

n_SZ = 12 n_HC = 12

n_SZ = 24 n_HC = 20

n_SZ = 15 n_HC = 18
| Study | Method | Participants | Description | Control | Comparison |
|-------|--------|--------------|-------------|---------|------------|
| Pedersen (2012) | fMRI | n_{SZ}=12  n_{HC}=13 | Participants see short movies of two interacting triangles. Classify their interaction by means of given descriptions. | Theory of Mind animation > non Theory of Mind animation | cognitive  explicit  visual |
| Eack (2013) | fMRI | n_{SZ}=15  n_{HC}=14 | Participants see short movies of two randomly moving triangles. Classify their interaction by means of given descriptions. | perspective taking > non perspective taking | cognitive  explicit  visual |
| Varga (2013)* | fMRI | n_{SZ}=20  n_{HC}=20 | Participants hear a context phrase followed by an ironic statement. A subsequent question must be answered with yes or no. | irony Comprehension > physical causality | cognitive  explicit  verbal |
| Harvey (2013)* | fMRI | n_{SZ}=21  n_{HC}=24 | Participants see positively and negatively valenced videos and rate how positive or negative the social target felt on a 9-point scale. | mental state judgment > eye location judgment | affective  explicit  visual |
| | | n_{SZ}=15  n_{HC}=15 | Participants see positively and negatively valenced videos and rate how much to the left or right the eyes of the target were directed on a 9-point scale. | | |
Bedford (2012) *fMRI* Participants were shown trait adjectives and were asked to judge whether they applied to themselves or another person. The trait adjectives were categorised as positive, negative, mental-illness related or physical-illness related.

Rapp (2013) *fMRI* Participants read a context phrase followed by an ironic statement. Classify statement as ironic or literal.

Russell (2000) *fMRI* Participants select one of two words that describes best the mental state reflected in photographed eyes.

Brüne (2011) *fMRI* Participants see a cartoon in correct order and subsequently answer two questions concerning the Intention and expectations of the protagonist.

Brunet (2003)*PET* Participants see a cartoon and select a correct ending. Cartoon category: Attribution of Attention

---

trait judgment > lexical judgment

affective explicit verbal

trait judgment > lexical judgment

affective explicit verbal

ironic statements > literal statements
cognitive implicit verbal

mental state judgment > sex judgment

affective explicit visual

mentalizing judgment > factual judgment
cognitive explicit visual

social intention > physical causality
cognitive implicit visual

---

n_{SZ}=11 n_{HC}=8

n_{SZ}=15 n_{HC}=15

n_{SZ}=5 n_{HC}=7

n_{SZ}=22 n_{HC}=26

n_{SZ}=7 n_{HC}=8
| Study           | Method | Description                                                                                                                                                                                                 | Task                                                                 | Stimulation Type | Task | Stimulation Type | Stimulation Type |
|-----------------|--------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------|------------------|------|------------------|------------------|
| Mier (2010)     | fMRI   | Participants see pictures of emotional faces. In the Theory of Mind condition they have to judge whether the shown person was going to carry out a certain action. In the Emotion Recognition condition subjects have to recognize the emotion of the shown person. | Participants have to decide whether a particular physical feature is present in the depicted person. | Theory of Mind > physical feature | affective | implicit | visual |
|                 |        |                                                                                             |                                                                                                                                  |                  |      |                  |                  |
|                 |        |                                                                                             |                                                                                                                                  |                  |      |                  |                  |
| Andreasen       | PET    | The experimental task examines the ability to attribute mental states to others by having the participants compose a story about the thoughts and internal experiences of another person. The narratives are scored to identify the number of mental state attributions. | Participants read aloud a neutral story which is presented on the video monitor. | mental state attribution > story reading | cognitive | implicit | verbal |
| (2008)          |        |                                                                                             |                                                                                                                                  |                  |      |                  |                  |
|                 |        |                                                                                             |                                                                                                                                  |                  |      |                  |                  |
| Ciaramidaro     | fMRI   | Participants see a cartoon and select a correct ending. Cartoon category: Private Intention and Communicative Intention.                                                                                      | Participants see a cartoon and select a correct ending concerning the physical causality of objects. | social intention > physical causality | cognitive | implicit | visual |
| (2015)*         |        |                                                                                             |                                                                                                                                  |                  |      |                  |                  |
|                 |        |                                                                                             |                                                                                                                                  |                  |      |                  |                  |
| Das (2012)*     | fMRI   | Participants see short movies of two interacting triangles. Classification of interaction outside the scanner.                                                                                              | Participants see short movies of two randomly moving triangles. Classification of interaction outside the scanner.            | Theory of Mind animation > non Theory of Mind animation            | cognitive | implicit | visual |
|                 |        |                                                                                             |                                                                                                                                  |                  |      |                  |                  |
|                 |        |                                                                                             |                                                                                                                                  |                  |      |                  |                  |

| n_{SZ} | n_{HC} |
|--------|--------|
| 16     | 16     |
| 18     | 13     |
| 18     | 23     |
| 20     | 19     |
| Study          | fMRI | Task Description                                                                 | Task Description                                                                 | Contrast           | Modality | Type  | Mode |
|---------------|------|----------------------------------------------------------------------------------|----------------------------------------------------------------------------------|--------------------|----------|--------|-------|
| Dodell-Feder  |      | Participants read false belief stories, followed by a true/false question.       | Participants read false photography stories, followed by a true/false question.  | false belief > false photograph | cognitive | implicit | verbal |
| (2014)*       |      |                                                                                  |                                                                                  |                    |          |        |       |
| nSZ=20 nHC=18 |      |                                                                                  |                                                                                  |                    |          |        |       |
| Pauly (2013)* |      | Participants were shown trait adjectives and were asked to judge whether they   | Participants were shown trait adjectives and were asked to judge whether they   | trait judgment > lexical judgment | affective | explicit | verbal |
|               |      | applied to themselves or another person. The trait adjectives were categorised   | contained the letter ‘r’. The trait adjectives were categorised as positive or   |                    |          |        |       |
|               |      | as positive or negative.                                                        | negative.                                                                        |                    |          |        |       |
| nSZ=13 nHC=13 |      |                                                                                  |                                                                                  |                    |          |        |       |
| Lee (2006)*   |      | Empathy scenarios described another person in need or distress (e.g., “You      | Baseline social reasoning scenarios related to social situations (e.g., “You are  | empathy scenario > baseline social reasoning | affective | explicit | verbal |
|               |      | come home. Your friend has had an unpleasant experience that day. Which is the  | approaching a large traffic jam on the motorway. It is not rush hour. There have |                    |          |        |       |
|               |      | more likely explanation for your friend’s state of mind?”)                      | been no roadwork signs. Which is the more likely explanation?”).                |                    |          |        |       |
| nSZ=14 nHC=14 |      |                                                                                  |                                                                                  |                    |          |        |       |

Note: SZ = Schizophrenic Patients, HC = Healthy Controls. Studies marked with an asterisk (*) provided the contrasts we used for group comparisons (schizophrenia vs healthy controls) also for their separate analyses.