

Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition

Philipp Kanske^{1,†}, Anne Böckler^{1,2,†}, Fynn-Mathis Trautwein¹, Franca H. Parianen Lesemann¹, and Tania Singer¹

¹Department of Social Neuroscience, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany and ²Department of Psychology, Würzburg University, Würzburg, Germany.

Correspondence should be addressed to: Tania Singer, Department of Social Neuroscience, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany. E-mail: singer@cbs.mpg.de.

[†]These authors contributed equally to this work.

Abstract

Although the processes that underlie sharing others' emotions (empathy) and understanding others' mental states (mentalizing, Theory of Mind) have received increasing attention, it is yet unclear how they relate to each other. For instance, are people who strongly empathize with others also more proficient in mentalizing? And (how) do the neural networks supporting empathy and mentalizing interact? Assessing both functions simultaneously in a large sample ($N = 178$), we show that people's capacities to empathize and mentalize are independent, both on a behavioral and neural level. Thus, strong empathizers are not necessarily proficient mentalizers, arguing against a general capacity of social understanding. Second, we applied dynamic causal modeling to investigate how the neural networks underlying empathy and mentalizing are orchestrated in naturalistic social settings. Results reveal that in highly emotional situations, empathic sharing can inhibit mentalizing-related activity and thereby harm mentalizing performance. Taken together, our findings speak against a unitary construct of social understanding and suggest flexible interplay of distinct social functions.

Key words: social cognition; empathy; theory of mind; mentalizing; fMRI; dynamic causal modeling.

Introduction

Picture yourself in a conversation with your brother who just told you how his beloved decided that she 'wants to stay friends'. In order to act appropriately, to give adequate advice and support, you need to quickly grasp both what your brother's ex girl-friend actually means when she talks about staying friends and the emotional state your brother is in. This example illustrates that successful social interaction relies on our ability to share others' emotions (empathy) and comprehend their thoughts and intentions [mentalizing, Theory of Mind (ToM)] (Frith and Frith, 2005; Singer, 2012; Kanske *et al.*, 2015a). These two routes of social understanding have been associated with distinct neural networks. Meta-analyses yielded a core network

including anterior insula (AI), inferior frontal gyrus (IFG) and anterior cingulate cortex (ACC) for empathy for pain (Lamm *et al.*, 2011). When a broader range of empathy paradigms are considered, a more extensive neural network was identified, including AI, IFG and ACC but also more dorsal regions of the dorsolateral and medial prefrontal cortex (dmPFC and dlPFC), and anterior/dorsal parts of the temporoparietal junction (TPJ) including supramarginal gyrus (SMG; Bzdok *et al.*, 2012). ToM involves a distinct network including posterior/ventral TPJ, temporal poles, superior temporal sulcus, medial prefrontal cortex and precuneus [(Schurz *et al.*, 2014); direct tests of the dissociation confirm the distinction of empathy and ToM networks on the level of brain function (Kanske *et al.*, 2015b) and structure (Eres

Received: 12 November 2015; Revised: 11 April 2016; Accepted: 12 April 2016

© The Author (2016). Published by Oxford University Press. For Permissions, please email: journals.permissions@oup.com

et al., 2015; Valk et al., 2016)]. As the literature varies in the definition of the concepts empathy and ToM (Schurz and Perner, 2015), we want to clarify that we define empathy as an emotional state that is isomorphic to and elicited by observing or imagining another's affective state (de Vignemont and Singer, 2006). Hence, empathy can be described as 'feeling with' rather than 'feeling for' another, the latter referring to empathic concern and compassion (Batson et al., 1987). For instance, when confronted with others' suffering, empathy would refer to sharing the suffering and hence be associated to negative affect, whereas compassion would include feelings of warmth, loving-kindness and concern and a strong motivation to alleviate the others' suffering (Singer and Klimecki, 2014). ToM, in contrast, refers to cognitively representing and reasoning on others' mental (or affective) states (Premack and Woodruff 1978; Frith and Frith, 2005). The key difference between empathy and ToM is that the former entails embodied sharing of a sensory, affective or bodily state while the latter yields propositional knowledge of another's state. Though previous studies have compared cognitive and affective aspects of ToM (i.e. mentalizing on others' cognitive, perceptual or affective states; Vollm et al., 2006; Schulte-Ruther et al., 2007; Shamay-Tsoory and Aharon-Peretz, 2007; Schnell et al., 2011; Bruneau et al., 2012; Sebastian et al., 2012; Corradi-Dell'Acqua et al., 2014; Bruneau et al., 2015) or have studied empathy and ToM in separation (Dziobek et al., 2011), the interaction of empathy and ToM within individuals has received only little attention (c.f. Lockwood et al., 2013). We therefore addressed two critical questions: First, on the level of inter-individual differences, are people who strongly empathize with others also more proficient in ToM? This would point toward a general capacity underlying social understanding, analogously to the proposed g-factor in intelligence research (Jensen, 1998; Garlick, 2002). Second, on the level of the individual brain, (how) do the neural networks supporting empathy and ToM interact during complex real-life understanding of others? For example, highly emotional situations may lead to prioritized empathic responding at the cost of accurate mentalizing.

To address these questions, we assessed behavioral and neural responding in a validated functional magnetic resonance imaging (fMRI) paradigm [EmpaToM, see Figure 1 (Kanske et al., 2015b)] that allows the manipulation of empathy and ToM during realistically complex social understanding. Participants view videos in which individuals of different age, gender and race talk about autobiographic episodes. These videos can either be emotionally negative, eliciting empathic responding, or neutral. Participants are asked to rate their affective state after each video. ToM performance is assessed by questions that either require inferring the mental state of the narrator (ToM condition) or physical facts (factual reasoning, control condition). The EmpaToM was validated with existing behavioral and fMRI paradigms and separates empathy and ToM on a behavioral and neural level. Specifically, it elicits distinct activity in AI, dorsomedial prefrontal/ACC and IFG for empathy (when contrasting emotional vs neutral videos and when testing for activity that is parametrically modulated by the valence ratings after each video) and ventral TPJ and anterior and posterior midline regions for ToM (when contrasting ToM vs non-ToM questions or videos with ToM vs non-ToM demands) (Figure 2A; Kanske et al., 2015b).

In the present study, this paradigm was employed to directly investigate the interrelation of empathy and ToM on an inter-individual level and on the level of the individual brain. To address the question whether the proclivities to empathize and

mentalize are related on a behavioral and neural level, we first calculated correlations of inter-individual differences in empathy and ToM measures and secondly tested whether people with high empathic responding are distributed equally across people with high and low mentalizing capacities, and vice versa. To investigate how the two underlying neural networks work together on an intra-individual level during online social understanding, we used dynamic causal modeling (DCM) to investigate how activity in regions identified to be crucial for empathic responses and mentalizing (AI and TPJ) interacts under empathy and ToM demands.

Materials and methods

Participants

One-hundred ninety-one participants were tested in the context of a longitudinal study at baseline (ReSource study; Kanske et al., 2015b; Singer et al., 2015)¹. Thirteen participants were excluded due to technical problems during data acquisition. One-hundred seventy-eight participants (age mean = 40.9 years, s.d. = 9.5, 106 female, 176 right-handed) underwent the experimental procedures. All participants signed informed consent prior to participation. The study was approved by the local ethics committees of the University of Leipzig and the Humboldt Universität zu Berlin.

Tasks

For details of the EmpaToM task, see Kanske et al. (2015b; Figure 1). Videos were presented which were either emotionally neutral or negative and which were designed to later allow ToM or factual reasoning questions (2 × 2 factorial design). In ToM videos, the narrator's mental states (beliefs, intentions, thoughts about someone else's states, etc.) are not being spelled out and have to be inferred by the participant, while the non-ToM videos are not defined by the absence of any mental or affective states, but mental states do not have to be inferred as they are either not relevant or they are concretely spelled out (for exemplary stimuli, see Supplementary data S1). After each video, participants rated how they felt (valence rating) and how much compassion they felt for the person in the video (compassion rating). A subsequent multiple choice question was presented either demanding ToM reasoning or factual reasoning on the contents of the video. One of the three response options was always correct while the other two were incorrect. Finally, participants rated how confident they were to have chosen the correct response. Twelve trials per condition were presented. Three measures were of particular interest: First, empathic responding was assessed as the difference in valence ratings after negative minus neutral videos. Second, a further measure of empathic responding was derived, the degree to which performance was impaired by the emotionality of the preceding video. This measure was calculated as performance composite after emotional minus neutral videos. Finally, ToM performance was assessed by the composite of response time (RT) and accuracy in ToM questions. Specifically, responses were classified as incorrect when participants chose a wrong response option or when no response was provided within 14 s (note that the rate for omissions was very low, mean = 3.6%, median = 2.1%). As the time it

¹ Please note that this study is based on the same participant sample as described in Kanske et al. (2015b). Importantly however, all data, analyses and results described in this study are novel and have not been described or shown elsewhere.

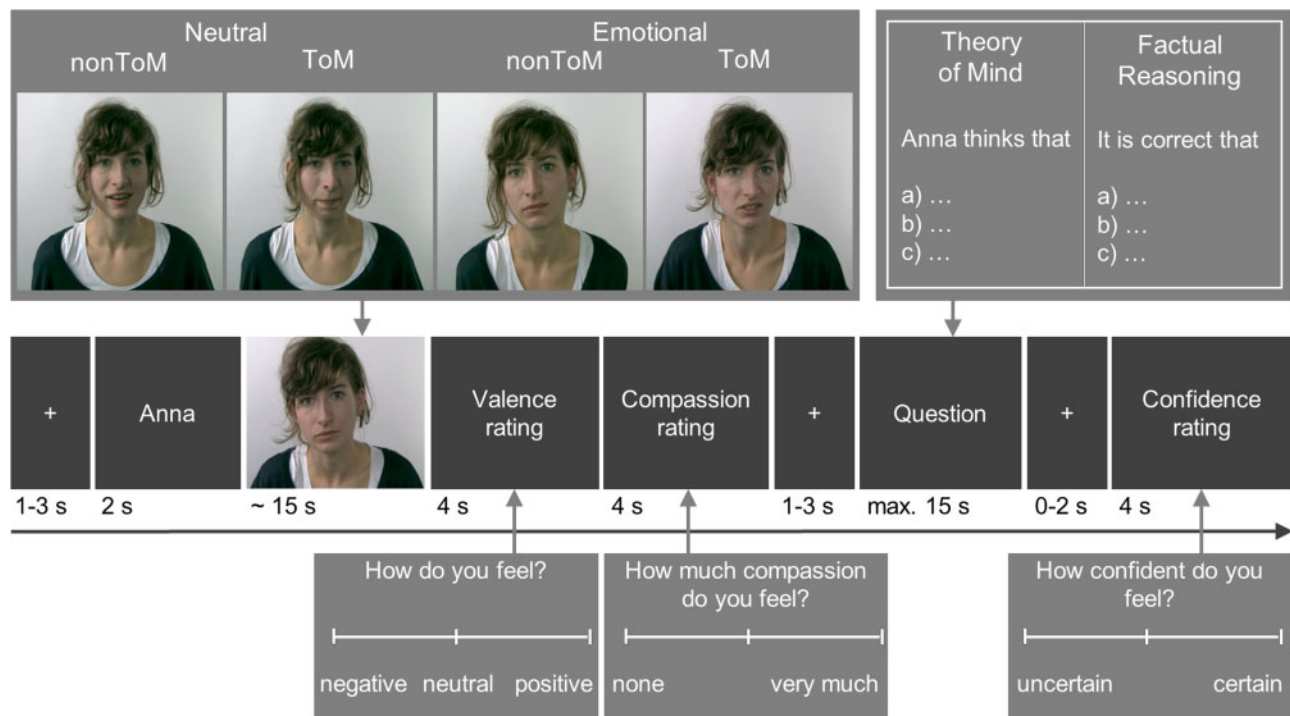


Fig. 1. EmpaToM trial sequence. Emotional and neutral videos with and without ToM demands are followed by valence and compassion ratings, ToM and factual reasoning questions, and a confidence rating (adapted from Kanske et al., 2015b).

takes participants to correctly respond to ToM questions may account for additional variance in ToM proficiency levels, RTs for correct trials were calculated. Speed and accuracy were combined by calculating z-scores for both measures and taking their unweighted mean.

Additionally, we behaviorally assessed two established tasks: In the empathy task [Socio-affective Video Task (SoVT; Klimecki et al., 2013)], participants are presented with silent videos depicting people in distress (high emotion) or performing everyday activities (low emotion) and are asked to rate how they feel (valence rating) and how they feel for the other (compassion rating). The empathy measure is derived by subtracting valence ratings after highly emotional from neutral videos. The perspective taking task [Samson Visual Perspective Taking Task (Samson et al., 2010)] measures people's ability to flexibly switch between their own and another's visuo-spatial perspective, and, specifically, the degree to which their own egocentric perspective interferes when judging the other's point of view. The inverse egocentricity bias was taken as an indication of perspective taking ability.

Note that in the course of validating the EmpaToM task based on the present and previous data sets [see Kanske et al., (2015b) for more details], we demonstrated (1) that emotional videos elicit affective responses in subjective (valence rating) and physiological measures, and that (2) the empathy measure (valence ratings) was significantly related to the empathy measure (valence ratings) of the SoVT while (3) the ToM measure of the EmpaToM showed significant relations to performance in the Imposing Memory Task (Kinderman et al., 1998), a verbal high-level mentalizing task and the reverse egocentricity bias in the Samson Task, a measure of ability to overcome an egocentric bias in visual perspective taking (Samson et al., 2010). Furthermore, (4) the empathy-related activations overlapped with activity elicited during the SoVT, tested

in the same participants, and with a meta-analysis of empathy studies (Bzdok et al., 2012). The (5) ToM-related activation overlapped with activity during a false-belief task (Dodell-Feder et al., 2011) and a meta-analysis on ToM studies (Bzdok et al., 2012).

MRI data acquisition

Brain images were acquired on a 3T Siemens Verio scanner (Siemens Medical Systems, Erlangen), equipped with a 32-channel head coil. Structural images were acquired using a MPRAGE T1-weighted sequence (TR = 2300 ms; TE = 2.98 ms; TI = 900; flip angle = 9°; 176 sagittal slices; matrix size = 256 × 256; FOV = 256 mm; slice thickness = 1 mm), yielding a final voxel size of 1 × 1 × 1 mm. For the functional imaging, a T2*-weighted echo-planar imaging (EPI) sequence was used (TR = 2000 ms; TE = 27 ms, flip angle = 90°). Thirty-seven axial slices were acquired covering the whole brain with a slice thickness of 3 mm, in-plane resolution 3 × 3 mm, 1 mm interslice gap, field of view = 210 mm; matrix size 70 × 70. Each run began with three dummy volumes that were discarded from further analysis.

Data analysis

Images were analysed using SPM8 (Wellcome Department of Imaging Neuroscience, London, UK). All volumes were coregistered to the SPM single-subject canonical EPI image, slice-time corrected and realigned to the mean image volume. A high resolution anatomical image of each subject was first coregistered to the SPM single-subject canonical T1 image and then to the average functional image. The transformation matrix obtained by normalizing the anatomical image was then used to normalize functional images to Montreal Neurological Institute (MNI)

space. The normalized images (3 mm isotropic voxel) were spatially smoothed with a Gaussian kernel of full-width half-maximum at 8 mm. A high-pass temporal filter with cutoff of 128 s was applied to remove low-frequency drifts from the data.

Statistical analyses were carried out using the general linear model. Onset and duration of the four video types, their corresponding questions and the rating periods were modeled. These regressors were convolved with a canonical hemodynamic response function. The six motion parameters were modeled as effects of no interest. To further reduce influence of potential noise-artifacts, we used the RobustWLS Toolbox (Diedrichsen and Shadmehr, 2005), which down-weights images with higher noise variance through a weighted-least-squares approach. Contrast images for the empathy contrast (emotional vs neutral videos), the ToM contrast during questions (ToM vs non-ToM questions) and the ToM contrast during videos (ToM vs non-ToM videos) were calculated by applying linear weights to the parameter estimates and entered into one-sample t-tests for random effects analysis.

On the second level, one-sample t-tests for the empathy and ToM contrasts were performed. As described in (Kanske et al., 2015b), the specifics of ToM and empathy were analysed by entering the respective first-level contrast images [i.e. contrast images for empathy (emotional vs neutral videos) and ToM (ToM vs non-ToM questions)] into a one-factorial design with dependent levels. Specifics were then assessed with t-contrasts of 'ToM > Empathy' and 'Empathy > ToM', which were inclusively masked for significant voxels of the respective simple contrast (Figure 2A).

Resting state analysis. SPM8 and DPARSF were applied to analyse the resting state data (Chao-Gan and Yu-Feng, 2010). After discarding the first 10 volumes, all functional scans were slice-time corrected and realigned. T1 images were coregistered to the functional scans and a DARTEL template was created using the averaged T1 images from all subjects. Nuisance covariates including six head motion parameters, head motion scrubbing regressor, white matter signal and cerebro-spinal fluid (CSF) signal were removed from the functional data. To reduce very low-frequency drifts and high-frequency respiratory and cardiac noise the linear trend of time courses was removed and then temporally band-pass filtering (0.01–0.08 Hz).

To calculate functional connectivity, spheres (radius = 5 mm) around the peak regions in the left AI and left TPJ [as derived from the specific 'Empathy > ToM' ($x = -33$, $y = 21$, $z = -3$) and 'ToM > Empathy' ($x = -48$, $y = -57$, $z = 24$) contrasts] were defined as seed regions. Averaged time courses were obtained from the spherical region of interest (ROIs) and correlation analyses were performed in a voxel-wise way to generate a functional connectivity map. Correlation coefficient maps were then converted into z maps by Fisher's r -to- z transform to improve normality. These maps, calculated in original space were normalized into MNI space, re-sampled to 3-mm isotropic voxels and smoothed with a 4 mm FWHM kernel.

Dynamic causal modeling. Dynamic causal modeling was used to characterize effective connectivity between left AI and left TPJ as well as the modulatory influence of empathy and ToM on this connectivity (Friston et al., 2003). Time-series from 5 mm spheres around individual subject peaks in AI and TPJ were extracted. V1, defined anatomically, was included in the model as a sensory seed region (Lu et al., 2012; Chen et al., 2014), to model the direct influence of the stimuli onto the system (Friston et al., 2003). Because connectivity in DCM analysis reflects functional

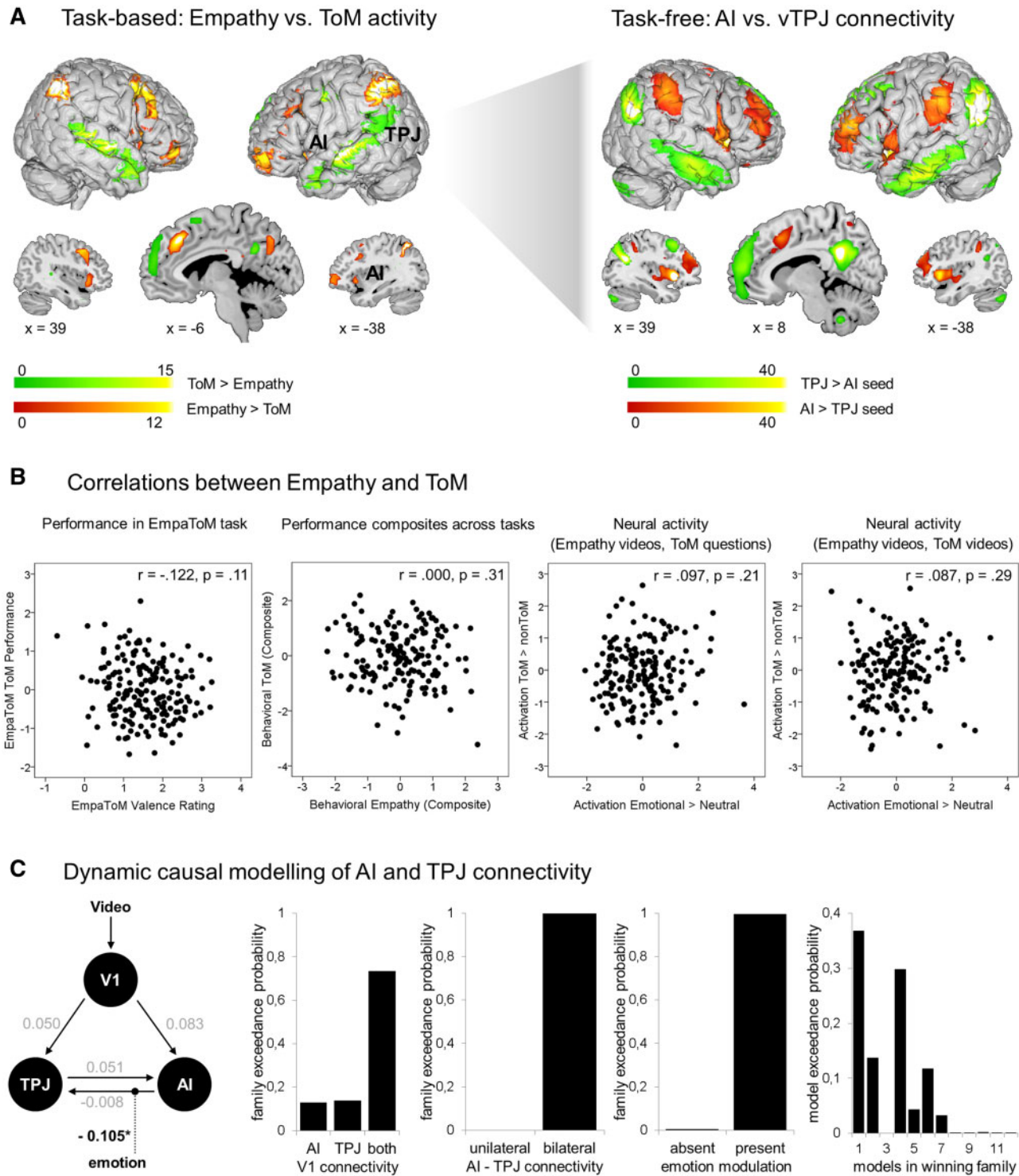
interactions between regions rather than anatomic connectivity (Friston et al., 2003), the information transfer from V1 to the other nodes of the network, can be estimated without including further relay regions during the information transfer (Eickhoff et al., 2009; Grefkes et al., 2010). Twenty-six participants did not show above threshold activation in one of the regions and analyses were run on the remaining 152 participants. DCMs were defined with video stimulation as driving input to V1 and connections from V1 to AI only, TPJ only or both (V1→AI, TPJ←V1, TPJ←V1→AI). Connections between AI and TPJ were either unilateral or mutual (AI→TPJ, AI←TPJ, AI↔TPJ). We also allowed emotion and ToM to influence all possible unilateral and mutual connections of AI and TPJ, either alone or jointly. Varying all possible combinations yielded 72 DCMs, which were fitted for each participant. Multi-step random-effects Bayesian model selection was used to identify the optimal model family and then compare the DCMs within this winning family. Model families were defined based on (1) whether V1 was connected to AI, TPJ or both, (2) whether connections between AI and TPJ were unilateral or bilateral and (3) whether emotion had an influence on any of the AI-TPJ connections or not. In a last step, significance of all parameters was tested for further interpretation (Stephan et al., 2010).

Results

Relating inter-individual differences in empathy and ToM

To investigate whether strong empathizers are better mentalizers, we tested the relation between individual differences in these social capacities on a behavioral and neural level (Figure 2B). First of all, behavior in the EmpaToM task revealed no correlation between empathizing (valence ratings) and ToM performance (see Table 1 for correlations between all behavioral measures of empathy and ToM). To generalize the observed independence on the behavioral level, we integrated the EmpaToM measures with the two additional empathy and ToM tasks [SoVT (Klimecki et al., 2013), Visual perspective taking task (Samson et al., 2010)] using principal component analysis (PCA; for details on all PCA analyses see Supplementary data S2). This analysis yielded two clearly distinct data-driven composites, a factor representing empathy (valence ratings in both the EmpaToM and the SoVT as well as performance impairments after emotionally negative videos in the EmpaToM) and a factor representing ToM (ToM performance in the EmpaToM and inverse egocentricity bias in the Samson task) (Supplementary data S3, Kaiser-Meyer-Olkin=0.54; Bartlett's Test of Sphericity= 42.4, $P < 0.001$; % variance explained = 54; see Supplementary data S4 for PCA including errors and RTs for ToM separately). Even though permitted in the statistical analysis (PCA with oblique rotation), these factors were also not correlated.

We then turned to ask whether independence of empathy and ToM also holds for the related neural activity. Subjecting all above threshold peak activations from the EmpaToM to PCAs yielded two composites that clearly distinguished neural activation elicited during empathy and during ToM, both when the ToM contrast was derived during the questions and when the ToM contrast was derived during the videos (Supplementary data S5 and S6, KMO \geq .84; Bartlett's Test of Sphericity \geq 1608, $P < 0.001$; % variance explained \geq 58.3). Mirroring behavioral findings, factor scores were uncorrelated; hence, individual differences in neural activity during empathizing were not correlated with differences in neural activity during ToM reasoning.



Downloaded from https://academic.oup.com/scan/article/11/9/1383/2223804 by guest on 21 August 2022

Fig. 2. (A) Neural activity for empathy > ToM and ToM > empathy (Kanske et al., 2015b), and resting state networks seeded from AI and ventral TPJ. (B) Correlations of empathizing (valence ratings) and ToM performance in the EmpaToM; correlations of the factor scores representing empathy and ToM across different behavioral tasks; correlations of neural activity related to empathy (emotional > neutral videos) and ToM (ToM > non-ToM questions and ToM > non-ToM videos). (C) DCM results depicting the winning model (significant coefficients depicted in bold); exceedance probabilities for the comparisons of model families and of models within the winning family.

Finally, in order to corroborate the correlational results, we statistically tested whether people with low and high empathic responding in behavior (based on a median split on scores on the empathy factor) were equally distributed across people with low and high mentalizing abilities (median split on scores on

the ToM factor). In addition, the same was done for scores on neural activations during empathy and ToM (either when the ToM contrast was derived during the videos or during the questions). Crucially, the absence of a relation revealed in correlations was corroborated by the finding that people with high

Table 1. Correlation matrix of all individual behavioral measures of empathy and ToM

Measure	SoVT vr	EmpaToM vr	EmpaToM ap	EmpaToM ToM	Samson eb
SoVT valence rating		0.36***	0.22**	-0.01	0.12
EmpaToM valence rating	0.36***		0.15*	-0.12	0.03
EmpaToM affect performance	0.22**	0.15		0.04	0.03
EmpaToM ToM performance	-0.01	-0.12	0.04		0.17*
Samson egocentricity bias	0.12	0.03	0.03	0.17*	

Correlation coefficients are reported. *** indicates significant correlations at $P < 0.001$ (2-tailed); ** $P < 0.01$ (2-tailed); * $P < 0.05$ (2-tailed).

empathic responding were distributed equally across people with high and low mentalizing capacities, and vice versa, both in behavioral and neural measures (Pearson's Chi Square Tests and Fishers Exact Tests: values < 2.0 , $P_s > 0.1$).

Causal influences between empathy- and ToM-related neural activations

To investigate how empathy and ToM are coordinated in an individual brain in the process of understanding another person, we used DCM. DCM allows the characterization of temporal causal influences between core brain regions and their modulation by experimental conditions (Friston et al., 2003; Stephan et al., 2007). As core regions we chose AI and TPJ peaks observed in the EmpaToM because (1) meta-analyses identified these as the most consistently activated regions for empathy and ToM (Lamm et al., 2011; Schurz et al., 2014) and because (2) seeding from these peaks in resting state scans corroborated their embedding in separate networks in task-free functional connectivity (Figure 2A and Table 2). Families of DCMs were selected stepwise. Of the 72 tested DCMs (Stephan et al., 2007), highest probability was found for model families with input connections from V1 to both AI and TPJ (exceedance probability $\Phi = 0.733$, Figure 2C), with mutual connections between AI and TPJ (exceedance probability $\Phi = 1$) and with a modulatory influence of emotionality of the video (exceedance probability $\Phi = 0.997$). The winning model (exceedance probability $\Phi = 0.378$) showed that AI inhibited TPJ activity, particularly in conditions of high negative emotionality ($t(151) = 2.3$, $P < 0.05$). This finding suggests that the AI down-regulates ongoing processes in TPJ when salient emotional events demand prioritization.

An imminent question raised by this finding is whether ToM performance also suffers in situations of high empathic distress ('empathy costs'). We therefore calculated the degree to which individual ToM performance was impaired in emotionally negative vs neutral conditions (as a measure of empathy costs) in the EmpaToM task and grouped participants accordingly. Specifically, empathy cost was calculated as the difference between EmpaToM performance (composite score of reaction times and error rates) in ToM questions after negative minus neutral videos. The distribution of this variable was normal ($P > 0.20$; Supplementary data S7). Participants were grouped according to a median split (median = -0.157) into people with and people without empathy costs. We then asked whether the down regulation of TPJ by AI in situations of emotional distress would be found specifically in the group with empathy costs. Indeed, group-specific DCM analyses replicated the inhibiting influence of AI on TPJ during emotional conditions in those individuals who showed empathy costs, whereas no inhibition was found for individuals without

empathy costs. Specifically, when only participants with empathy costs were included, highest probability was found for model families with input connections from V1 to both AI and TPJ (exceedance probability $\Phi = 0.904$, Supplementary data S8), with mutual connections between AI and TPJ (exceedance probability $\Phi = 1$) they with a modulatory influence of emotionality of the video (exceedance probability $\Phi = 0.943$). In the winning model (exceedance probability $\Phi = 0.469$), AI inhibited TPJ activity, particularly in conditions of high negative emotionality ($t(75) = 3.0$, $P < 0.01$). No other coefficients were significant ($P_s > 0.20$); they are therefore not interpreted (Stephan et al., 2010). By contrast, when only participants without empathy costs were included, highest probability was found for model families with input connections from V1 to both AI and TPJ (exceedance probability $\Phi = 0.898$, Supplementary data S8), with mutual connections between AI and TPJ (exceedance probability $\Phi = 1$), and with a modulatory influence of emotionality of the video (exceedance probability $\Phi = 0.734$). In the winning model (exceedance probability $\Phi = 0.566$) none of the coefficients were significant ($P_s > 0.20$), including the influence of AI on TPJ in emotional situations, they can therefore not be interpreted (Stephan et al., 2010).

Discussion

This study investigated the relation of affective and cognitive routes to understanding others by addressing two questions: First, are empathy and ToM linked on an inter-individual level, that is, are strong empathizers also good mentalizers? Second, how are the neural networks underlying empathy and ToM working together during the online processing of social information? Data show that inter-individual differences in behavioral and neural measures of empathy and ToM were independent. During online processing of social situations, however, the two networks interacted; specifically, we found inhibitory causal influences among key nodes of the empathy- and ToM-related neural networks and impaired mentalizing performance in highly emotional situations.

Using different convergent methods and analyses, we observed neither significant relationships between empathy- and ToM-related behavioral measures (within the EmpaToM and when including independent empathy and ToM tasks) nor between the respective underlying neural networks. To corroborate these findings, we statistically demonstrated that high and low empathizers were equally distributed across high and low mentalizers, and vice versa. This finding suggests that the tendency to empathize does not contribute to and is not dependent on the capacity to infer others' mental states. Selective impairments of empathy and ToM have been suggested in psychopathology. In autism spectrum disorder, for instance, deficient ToM abilities have been reliably reported (Frith, 2001;

Table 2. Activation peaks for resting state connectivity from AI and TPJ

	H	MNI coordinates			T	Z	Cs
		x	y	z			
Resting state: AI seed > TPJ seed							
AI	L	-33	21	-3	43.18	>8.21	3132
Middle frontal gyrus	L	-42	45	27	14.29	>8.21	
AI	L	-45	9	-6	14.19	>8.21	
TPJ-SMG	L	-63	-24	24	11.21	>8.21	
AI	R	36	21	6	18.73	>8.21	3719
AI	R	33	27	0	18.63	>8.21	
Anterior cingulate	R	6	15	45	13.24	>8.21	
Anterior cingulate	L	-6	18	36	9.65	>8.21	
TPJ-SMG	R	60	-27	39	12.97	>8.21	1379
TPJ-SMG	R	60	-27	48	12.14	>8.21	
Inferior orbitofrontal	R	21	36	-18	9.41	>8.21	114
Frontal pole	R	24	45	-15	8.88	>8.21	
Inferior orbitofrontal	L	-27	39	-15	8.63	>8.21	54
Middle temporal	L	-54	-66	-3	8.07	7.72	175
Inferior temporal	L	-51	-60	-9	7.95	7.62	
Inferior temporal	L	-45	-45	-15	6.76	6.55	
Posterior cingulate	L	-15	-33	42	7.48	7.2	49
Middle temporal	R	57	-51	-6	6.73	6.52	69
Lateral occipital	R	36	-81	30	5.62	5.49	15
Lateral occipital	R	42	-72	21	5.19	5.09	
Cerebellum	L	-21	-69	-51	9.77	>8.21	75
Cerebellum	R	21	-72	-51	8.28	>8.21	33
Cerebellum	L	-36	-48	-30	6.57	6.38	40
Resting state: TPJ seed > AI seed							
TPJ-angular gyrus	L	-48	-57	24	44.62	>8.21	896
Precuneus	L	-3	-54	30	24.96	>8.21	1376
Posterior cingulate	L	-9	-48	33	23.33	>8.21	
TPJ-angular gyrus	R	54	-57	27	22.22	>8.21	711
Lateral occipital	R	51	-63	42	15.79	>8.21	
Superior frontal	L	-6	51	39	18.15	>8.21	3632
Superior frontal	L	-12	42	45	17.35	>8.21	
Superior frontal	L	-12	30	57	15.5	>8.21	
Cerebellum	R	21	-84	-39	17.36	>8.21	670
Cerebellum	L	-30	-81	-33	17.05	>8.21	482
Middle temporal	L	-60	-9	-18	17.05	>8.21	1065
Middle temporal	L	-60	-21	-15	15.64	>8.21	
Temporal pole	L	-45	12	-36	11.99	>8.21	
Middle temporal	R	63	-12	-21	15.06	>8.21	889
Middle temporal	R	66	-33	-3	10.57	>8.21	
Temporal pole	R	48	15	-36	9.97	>8.21	
Cerebellum	R	6	-51	-45	14.68	>8.21	167
Inferior orbitofrontal	L	-45	30	-12	9.14	>8.21	148
IFG	L	-54	24	9	8.95	>8.21	
Inferior orbitofrontal	R	45	33	-15	6.89	6.67	24
Cerebellum	L	-6	-51	-24	6.87	6.66	23
Subcallosal cortex	L	-3	9	-12	6.82	6.61	18
Hippocampus	L	-24	-15	-21	6.34	6.17	34

H, hemisphere; Cs, cluster size in number of voxels.

Schwenck et al., 2012; Kana et al., 2014), while empathy seems to be preserved when controlling for comorbid alexithymia (Bird et al., 2010; Bird and Viding, 2014). Psychopathy, on the other hand, is characterized by preserved ToM, but reduced empathic responding (Blair, 2008), at least if empathic responding is not explicitly encouraged (Meffert et al., 2013). Also in the neuro-typical population, autistic and psychopathic traits are selectively correlated with reduced capacities in perspective-taking and empathy, respectively (Lockwood et al., 2013). While

clinical studies hint at selective disruption, this study shows that the capacities to empathize and mentalize vary independently from one another across a wide range of individuals within a healthy population. Thus, unlike in some lines of intelligence research, where a general capacity to perform well across tasks has been proposed (Jensen, 1998; Garlick, 2002), the socio-affective capacity of sharing feelings with another and the socio-cognitive capacity of inferring mental states may be independent.

However, it is interesting to note, first, that a recent study by (Lockwood et al., 2013) did find a moderate positive correlation between a social animations task and an affective resonance task. As their sample was smaller, this cannot indicate that the null-correlation in the present study is due to a lack of power. Rather, it suggests that the specific operationalization of empathy and ToM gave rise to the difference. Meta-analyses demonstrate that social animations consistently activate a subset of the regions in the full ToM network, possibly indicating that processing social animations involves also a specific subset of the cognitive processes that constitute ToM (Schurz et al., 2014; Schurz and Perner, 2015). The ToM contrast in the EmpaToM (i.e. ToM > non-ToM questions), in comparison, activates the entire mentalizing network (Kanske et al., 2015b) suggesting that this task requires the involvement of all computations involved in full-blown ToM. Interestingly, social animations activate the left insular cortex (Schurz et al., 2014) and the amygdala (Martin and Weisberg, 2003) and also elicit subjective emotional responses since they depict affective scenarios (e.g. tricking, coaxing; Rimé et al., 1985). Such emotional responses in the observer may contribute to the ascription of intentionality to the animated characters, thereby precipitating a correlation with the affective resonance task. Furthermore, there was a considerable age difference between the two studies [21.9 years in Lockwood et al. (2013) and 40.9 years in this study]. Previous research indicates that socio-cognitive skills change over the lifespan, showing reduced ToM capacities in younger, but also in older age (Moran, 2013). Whether the relation between socio-cognitive and socio-affective skills is also altered over the course of one's lifetime (e.g. whether empathy and ToM are correlated in younger, but not in middle age) will need to be addressed in future research. Second, studies on self-reported social skills also paint a different picture. Employing questionnaires such as the Interpersonal Reactivity Index (Davis, 1983) or the Questionnaire of Affective and Cognitive Empathy (Reniers et al., 2011), researchers showed correlations between participants' tendencies to describe and/or perceive themselves as empathic and as willing to take other people's perspectives (Davis, 1983; Reniers et al., 2011; Georgi et al., 2014; Lockwood et al., 2014). A possible reason for the discrepancy may be that people's implicit and explicit conceptualizations of empathy and ToM are not as distinct as researchers' operationalizations are, making questionnaires somewhat more susceptible to subjective conceptions of these capacities, folk psychology and social desirability (Stone et al., 2000). It will be an interesting future endeavor to elucidate the link between self-reported and behavior-based assessments of empathy and ToM. In addition, future studies may aim at addressing the relation between empathy and ToM and further aspects of social understanding and social interaction, ranging from compassion and prosocial motivation to the ability to negotiate with or manipulate others.

The absence of a general capacity underlying behavioral empathy and ToM has implications, in particular for psychopathology and plasticity research. Across different mental disorders, deficits in social interaction have been described, for example

in personality disorders and depression (Arntz et al., 2009; Wolkenstein et al., 2011). Comprehensively characterizing socio-affective and -cognitive deficits could help delineating specific profiles of different disorders. From this clinical as well as from a basic neuroscience perspective, an important question is the malleability of social functioning. The simultaneous assessment of empathy and ToM could benefit intervention studies, for example on specific meditation practices or psychotherapy, to differentiate plasticity in affective and cognitive social capacities. Attempts to train social skills need to take into account that improving empathy may not necessarily enhance mentalizing capacities, and vice versa. On a more general level, it may prove interesting to relate specific capacities of social cognition to other interpersonal characteristics such as prosocial behavior, communication or partnership satisfaction. Interestingly, recent evidence suggests that individual differences in people's tendency to rather engage empathy or mentalizing in complex social settings predict differences in prosocial decision-making (Tusche et al., 2016). In a similar vein, the assessment of empathy and ToM needs to also include more balanced interactions with symmetrical contributions of all parties (Schilbach et al., 2013). This would allow investigation of social cognitive processes elicited by observing another person, but also the processes that result from being empathized with or understood by another (Seehausen et al., 2012).

The second goal was to investigate how the brain networks underlying empathy and ToM work together during complex social situations that require online understanding of others. To characterize the interplay of the empathy- and ToM-related networks, we used DCM, which allows investigating the temporal causal influences between the AI as core empathy-related region (Lamm et al., 2011) and the ventral TPJ as critical part of the ToM-related network (Schurz et al., 2014). Results revealed that the interaction of AI and TPJ was best described by a model assuming mutual influence between the two brain areas. The AI exerted inhibiting influence on the TPJ particularly during emotional videos, suggesting that the AI has a role in modulating ongoing processes in TPJ, such as taking others' perspectives, reasoning about others' mental states or, more generally, shifting between different sources of information (Decety and Lamm, 2007). Indeed, only in those participants who showed a TPJ inhibition through AI activity in emotional situations, performance in ToM questions was hampered in the emotional condition. These results suggest inter-individual variability in the degree to which empathizing is prioritized over mentalizing in real-life like situations that call for both social capacities simultaneously. As the down-regulation of TPJ can come at the cost of impaired mentalizing, the question is whether chronified variants may be a precursor for mental disorders [as has been hypothesized in borderline personality disorder (Fonagy and Luyten, 2009)]. The adaptive function of this down-regulation, on the other hand, may be to efficiently draw attention to those aspects of a social situation that require immediate action, such as others' emotions (Menon and Uddin, 2010). Of course, future research is necessary to probe the role of other regions within the empathy- and ToM-related neural networks in order to draw stronger conclusions on the interplay of the two social functions. A comprehensive understanding of the functional significance of the present findings furthermore depends on testing the causal role of the observed inhibition between AI and TPJ on the specifics of subsequent behavior. On a more general level, an interesting question will be whether inhibitory relations characterize the influence of highly emotional

states on socio-cognitive functioning (Kanske et al., 2013; Wessa et al., 2013).

Taken together, an individual's inclination to empathize is generally independent from her ability to understand others' mental states. However, in real life social interactions, highly emotional situations can lead to impaired mentalizing. Thus, social understanding is subserved by a flexible interplay of distinct social functions rather than monolithic social intelligence.

Acknowledgments

We are thankful to all the members of the Department of Social Neuroscience involved in the ReSource study, in particular, to Astrid Ackermann, Christina Bochow, Matthias Bolz, and Sandra Zurborg for study management and to Manuela Hofmann, Sylvie Neubert, and Nicole Pampus for help with data collection.

Funding

Tania Singer, as principal investigator, received funding for the ReSource Project from a) the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013/ ERC Grant Agreement Number 205557 to T.S.), and b) from the Max Planck Society.

Supplementary data

Supplementary data are available at SCAN online.

Conflict of interest. None declared.

References

- Arntz, A., Bernstein, D., Oorschot, M., Schobre, P. (2009). Theory of mind in borderline and cluster-c personality disorder. *The Journal of Nervous and Mental Disease*, *197*(11), 801–7.
- Batson, C.D., Fultz, J., Schoenrade, P.A. (1987). Distress and empathy: two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of Personality* *55*(1), 19–39.
- Bird, G., Silani, G., Brindley, R., White, S., Frith, U., Singer, T. (2010). Empathic brain responses in insula are modulated by levels of alexithymia but not autism. *Brain* *133*(Pt 5), 1515–25.
- Bird, G., Viding, E. (2014). The self to other model of empathy: providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience and Biobehavioral Reviews*, *47*, 520–32.
- Blair, R.J. (2008). Fine cuts of empathy and the amygdala: dissociable deficits in psychopathy and autism. *Quarterly Journal of Experimental Psychology*, *61*(1), 157–70.
- Bruneau, E.G., Jacoby, N., Saxe, R. (2015). Empathic control through coordinated interaction of amygdala, theory of mind and extended pain matrix brain regions. *Neuroimage*, *114*, 105–19.
- Bruneau, E.G., Pluta, A., Saxe, R. (2012). Distinct roles of the 'shared pain' and 'theory of mind' networks in processing others' emotional suffering. *Neuropsychologia*, *50*(2), 219–31.
- Bzdok, D., Schilbach, L., Vogeley, K., et al. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on

- morality, theory of mind, and empathy. *Brain Structure and Function*, **217**(4),783–96.
- Chao-Gan, Y., Yu-Feng, Z. (2010). DPARSF: a MATLAB toolbox for “pipeline” data analysis of resting-state fMRI. *Frontiers in Systems Neuroscience*, **4**, 13.
- Chen, Y.C., Chen, C.C., Decety, J., Cheng, Y.W. (2014). Aging is associated with changes in the neural circuits underlying empathy. *Neurobiology of Aging*, **35**(4),827–36.
- Corradi-Dell’Acqua, C., Hofstetter, C., Vuilleumier, P. (2014). Cognitive and affective theory of mind share the same local patterns of activity in posterior temporal but not medial prefrontal cortex. *Social Cognitive and Affective Neuroscience*, **9**(8),1175–84.
- Davis, M.H. (1983). Measuring individual-differences in empathy—evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, **44**(1),113–26.
- Decety, J., Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *Neuroscientist*, **13**, 580–93.
- de Vignemont, F., Singer, T. (2006). The empathic brain: how, when and why? *Trends. Cogn Sci* **10**, 435–41.
- Diedrichsen, J., Shadmehr, R. (2005). Detecting and adjusting for artifacts in fMRI time series data. *Neuroimage*, **27**(3),624–34.
- Dodell-Feder, D., Koster-Hale, J., Bedny, M., Saxe, R. (2011). fMRI item analysis in a theory of mind task. *Neuroimage*, **55**(2), 705–12.
- Dziobek, I., Preissler, S., Grozdanovic, Z., Heuser, I., Heekeren, H.R., Roepke, S. (2011). Neuronal correlates of altered empathy and social cognition in borderline personality disorder. *Neuroimage*, **57**(2),539–48.
- Eickhoff, S.B., Heim, S., Zilles, K., Amunts, K. (2009). A systems perspective on the effective connectivity of overt speech production. *Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences*, **367**(1896), 2399–421.
- Eres, R., Decety, J., Louis, W.R., Molenberghs, P. (2015). Individual differences in local gray matter density are associated with differences in affective and cognitive empathy. *Neuroimage*, **117**, 305–10.
- Fonagy, P., Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and Psychopathology*, **21**(4),1355–81.
- Friston, K.J., Harrison, L., Penny, W. (2003). Dynamic causal modelling. *Neuroimage*, **19**(4),1273–302.
- Frith, C.D., Frith, U. (2005). Theory of mind. *Current Biology*, **15**(17),R644–6.
- Frith, U. (2001). Mind blindness and the brain in autism. *Neuron*, **32**(6),969–79.
- Garlick, D. (2002). Understanding the nature of the general factor of intelligence: the role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review*, **109**(1),116–36.
- Georgi, E., Petermann, F., Schipper, M. (2014). Are empathic abilities learnable? Implications for social neuroscientific research from psychometric assessments. *Social Neuroscience*, **9**(1),74–81.
- Grefkes, C., Nowak, D.A., Wang, L.E., Dafotakis, M., Eickhoff, S.B., Fink, G.R. (2010). Modulating cortical connectivity in stroke patients by rTMS assessed with fMRI and dynamic causal modeling. *Neuroimage*, **50**(1),233–42.
- Jensen, A. (1998). *The G Factor: The Science of Mental Ability*. Westport, CT: Greenwood Publishing.
- Kana, R.K., Libero, L.E., Hu, C.P., Deshpande, H.D., Colburn, J.S. (2014). Functional brain networks and white matter underlying theory-of-mind in autism. *Social Cognitive and Affective Neuroscience*, **9**(1),98–105.
- Kanske, P., Böckler, A., Singer, T. (2015a). Models, mechanisms and moderators dissociating empathy and Theory of Mind. In: Wöhr, M., Krach, S., editors. *Current Topics in Behavioral Neurosciences – Social Behavior from Rodents to Humans: Neural Foundations and Clinical Implications*. Luxemburg: Springer, 1–14.
- Kanske, P., Bockler, A., Trautwein, F.M., Singer, T. (2015b). Dissecting the social brain: introducing the EmpaToM to reveal distinct neural networks and brain-behavior relations for empathy and theory of mind. *Neuroimage*, **122**, 6–19.
- Kanske, P., Heissler, J., Schonfelder, S., Forneck, J., Wessa, M. (2013). Neural correlates of emotional distractibility in bipolar disorder patients, unaffected relatives, and individuals with hypomanic personality. *American Journal of Psychiatry*, **170**(12), 1487–96.
- Kinderman, P., Dunbar, R., Bentall, R.P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, **89**(2),191–204.
- Klimecki, O.M., Leiberg, S., Lamm, C., Singer, T. (2013). Functional neural plasticity and associated changes in positive affect after compassion training. *Cerebral Cortex*, **23**(7),1552–61.
- Lamm, C., Decety, J., Singer, T. (2011). Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *Neuroimage*, **54**(3),2492–502.
- Lockwood, P.L., Bird, G., Bridge, M., Viding, E. (2013). Dissecting empathy: high levels of psychopathic and autistic traits are characterized by difficulties in different social information processing domains. *Frontiers in Human Neuroscience*, **7**,760.
- Lockwood, P.L., Seara-Cardoso, A., Viding, E. (2014). Emotion regulation moderates the association between empathy and prosocial behavior. *PLoS One*, **9**(5), e96555.
- Lu, Q., Li, H.R., Luo, G.P., et al. (2012). Impaired prefrontal-amygdala effective connectivity is responsible for the dysfunction of emotion process in major depressive disorder: a dynamic causal modeling study on MEG. *Neuroscience Letters*, **523**(2), 125–30.
- Martin, A., Weisberg, J. (2003). Neural foundations for understanding social and mechanical concepts. *Cognitive Neuropsychology*, **20**(3-6),575–87.
- Meffert, H., Gazzola, V., den Boer, J.A., Bartels, A.A.J., Keysers, C. (2013). Reduced spontaneous but relatively normal deliberate vicarious representations in psychopathy. *Brain*, **136**, 2550–62.
- Menon, V., Uddin, L.Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, **214**(5-6),655–67.
- Moran, J.M. (2013). Lifespan development: the effects of typical aging on theory of mind. *Behavioral Brain Research*, **237**, 32–40.
- Premack, D., Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral Brain Science*, 515–26.
- Reniers, R.L.E.P., Corcoran, R., Drake, R., Shryane, N.M., Vollm, B.A. (2011). The QCAE: a questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, **93**(1),84–95.
- Rimé, B., Boulanger, B., Laubin, P., Richir, M., Stroobants, K. (1985). The perception of interpersonal emotions originated by patterns of movement. *Motivation and Emotion*, **9**(3),241–60.
- Samson, D., Apperly, I.A., Braithwaite, J.J., Andrews, B.J., Bodley Scott, S.E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of*

- Experimental Psychology. Human Perception and Performance*, **36**(5),1255–66.
- Schilbach, L., Timmermans, B., Reddy, V., et al. (2013). Toward a second-person neuroscience. *The Behavioral and Brain Sciences*, **36**(4),393–414.
- Schnell, K., Bluschke, S., Konradt, B., Walter, H. (2011). Functional relations of empathy and mentalizing: an fMRI study on the neural basis of cognitive empathy. *Neuroimage*, **54**(2),1743–54.
- Schulte-Ruther, M., Markowitsch, H.J., Fink, G.R., Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *Journal of Cognitive Neuroscience*, **19**(8),1354–72.
- Schurz, M., Perner, J. (2015). An evaluation of neurocognitive models of Theory of Mind. *Frontiers in Psychology*, **6**, 1610.
- Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J. (2014). Fractionating theory of mind: a meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, **42**C, 9–34.
- Schwenck, C., Mergenthaler, J., Keller, K., et al. (2012). Empathy in children with autism and conduct disorder: group-specific profiles and developmental aspects. *Journal of Child Psychology and Psychiatry*, **53**(6),651–9.
- Sebastian, C.L., Fontaine, N.M., Bird, G., et al. (2012). Neural processing associated with cognitive and affective theory of mind in adolescents and adults. *Social Cognitive and Affective Neuroscience*, **7**(1),53–63.
- Seehausen, M., Kazzer, P., Bajbouj, M., Prehn, K. (2012). Effects of empathic paraphrasing - extrinsic emotion regulation in social conflict. *Frontiers in Psychology*, **3**, 482.
- Shamay-Tsoory, S.G., Aharon-Peretz, J. (2007). Dissociable pre-frontal networks for cognitive and affective theory of mind: a lesion study. *Neuropsychologia*, **45**(13),3054–67.
- Singer, T. (2012). The past, present and future of social neuroscience: a European perspective. *Neuroimage*, **61**(2), 437–49.
- Singer, T., Klimecki, O.M. (2014). Empathy and compassion. *Current Biology*, **24**(18),R875–8.
- Singer, T., Kok, B.E., Bornemann, B., Bolz, M., Bochow, C.A. (2015). *The ReSource Project: Background, Design, Samples, and Measurements*. Leipzig: Max Planck Institute for Human Cognitive and Brain Sciences.
- Tusche, A., Böckler, A., Kanske, P., Trautwein, F.M., Singer, T. (2016). Decoding the charitable brain: Empathy, perspective taking and attention shifts differentially predict altruistic giving. *Journal of Neuroscience* **36**(17),4719–32.
- Stephan, K.E., Harrison, L.M., Kiebel, S.J., David, O., Penny, W.D., Friston, K.J. (2007). Dynamic causal models of neural system dynamics: current state and future extensions. *Journal of Biosciences*, **32**(1),129–44.
- Stephan, K.E., Penny, W.D., Moran, R.J., den Ouden, H.E., Daunizeau, J., Friston, K.J. (2010). Ten simple rules for dynamic causal modeling. *Neuroimage*, **49**(4),3099–109.
- Stone, A.A., Turkkan, J.S., Bachrach, C.A., Jobe, J.B., Kurtzman, H.S., Cain, V.S. (2000). *The Science of Self-Report: Implications for Research and Practice*. Mahwah, NJ: Erlbaum.
- Valk, S.L., Bernhardt, B.C., Böckler, A., Trautwein, F.M., Kanske, P., Singer, T. (2016). Socio-cognitive phenotypes differentially modulate large-scale structural covariance networks. *Cerebral Cortex*, pii,bhv319.
- Vollm, B.A., Taylor, A.N., Richardson, P., et al. (2006). Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. *Neuroimage*, **29**(1),90–8.
- Wessa, M., Heissler, J., Schonfelder, S., Kanske, P. (2013). Goal-directed behavior under emotional distraction is preserved by enhanced task-specific activation. *Social Cognitive and Affective Neuroscience*, **8**(3),305–12.
- Wolkenstein, L., Schönenberg, M., Schirm, E., Hautzinger, M. (2011). I can see what you feel, but I can't deal with it: impaired theory of mind in depression. *Journal of Affective Disorders*, **132**(1–2),104–11.