Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain–behavior relations for empathy and Theory of Mind

Philipp Kanske [1], Anne Böckler [1], Fynn-Mathis Trautwein [1], Tania Singer *

Department of Social Neuroscience, Max Planck Institute for Human Cognitive and Brain Sciences, 04103 Leipzig, Germany

ARTICLE INFO

ABSTRACT

Successful social interactions require both affect sharing (empathy) and understanding others' mental states (Theory of Mind, ToM). As these two functions have mostly been investigated in isolation, the specificity of the underlying neural networks and the relation of these networks to the respective behavioral indices could not be tested. Here, we present a novel fMRI paradigm (EmpaToM) that independently manipulates both empathy and ToM. Experiments 1a/b (N = 90) validated the task with established empathy and ToM paradigms on a behavioral and neural level. Experiment 2 (N = 178) employed the EmpaToM and revealed clearly separable neural networks including anterior insula for empathy and ventral temporoparietal junction for ToM. These distinct networks could be replicated in task-free resting state functional connectivity. Importantly, brain activity in these two networks specifically predicted the respective behavioral indices, that is, inter-individual differences in ToM related brain activity predicted inter-individual differences in ToM performance, but not empathic responding, and vice versa. Taken together, the validated EmpaToM allows separation of affective and cognitive routes to understanding others. It may thus benefit future clinical, developmental, and intervention studies on identifying selective impairments and improvement in specific components of social cognition.

© 2015 Published by Elsevier Inc.

Introduction

Understanding others – be it through sharing their emotions or reflecting on their thoughts – is a key component of successful social interaction. The ease with which we accomplish this task every day, readily makes us forget the complex computations and processes it entails. In the last decade, social neuroscience has investigated affective and cognitive routes to understanding others (Frith and Frith, 2005; Mitchell, 2005; Singer, 2006, 2012). Affective routes have mainly been studied under the term empathy, defined as sharing another's emotional state while being aware that the other is the source of the emotion (de Vignemont and Singer, 2006). Meta-analyses show that the anterior insula (AI) and middle anterior cingulate cortex (mACC) are core regions underlying empathic responding when witnessing others' suffering and when suffering oneself (Fan et al., 2011; Lamm et al., 2011). Therefore, 'shared' brain networks have been proposed as an underlying mechanism for our ability to empathize (Decety, 2010; Keysers and Gazzola, 2009; Singer et al., 2004). Complementarily to empathy,

others' suffering may also induce compassion, that is feelings of warmth and care and the wish to alleviate the other's suffering (Singer and Klimecki, 2014). Compassion relies on a different neural network than empathy, comprising areas linked to positive affect such as ventral striatum (Klimecki et al., 2014). Another line of research has focused on a cognitive route to understanding others that has been investigated under the terms Theory of Mind (ToM), mentalizing or cognitive perspective taking and comprises inferring and reasoning about the beliefs, thoughts or emotions of others (Frith and Frith, 2005; Mitchell et al., 2005; Premack and Woodruff, 1978). The neural network underlying ToM includes the temporoparietal junction (TPJ), temporal poles (TP), medial prefrontal cortex (MPFC) and precuneus/posterior cingulate (PCC) (Saxe and Kanwisher, 2003, for meta-analyses see Bzdok et al., 2012; Schurz et al., 2014). Importantly, ToM entails both, the reasoning about others' mental and affective states. Thus the crucial difference between ToM and empathy is that the first yields propositional knowledge of another's state, while the latter entails embodied sharing of a sensory, affective or bodily state (Singer, 2006).

Previous studies have compared cognitive and affective aspects of ToM (i.e. mentalizing on others' cognitive, perceptual, or affective states (Bruneau et al., 2012; Schnell et al., 2011; Schulte-Ruther et al., 2007; Shamay-Tsoory and Aharon-Peretz, 2007; Vollm et al., 2006) or have studied empathy and ToM in separation (Dziobek et al., 2011). Crucially, however, no paradigm has yet allowed investigation of both behavioral

* Corresponding author at: Department of Social Neuroscience, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany.
  E-mail address: singer@cbs.mpg.de (T. Singer).
  [1] Contributed equally.

indices and neural networks underlying empathy and ToM within the same individuals. In the current study, we intended to investigate these two capacities within one paradigm and in the same individuals, and thereby address two important questions. The first concerns brain–behavior relations, that is, the specific relation of neural activity observed during empathy and ToM conditions to the behavioral indices of both functions. Are empathy ratings related to neural activity elicited during empathy conditions, but not during ToM conditions, and vice versa for ToM performance? While empathy ratings have previously been shown to parametrically modulate brain activity observed during empathy paradigms (Klimecki et al., 2013), the relation of ToM performance to brain activity observed in mentalizing networks is yet unknown, and so is the dissociation of brain–behavior relations comparing empathy and ToM abilities within an individual. The second question concerns the characterization and distinction of the neural networks underlying empathy and ToM when assessed within an individual. Which brain regions are engaged specifically by one social capacitiy compared to the other? For example, meta-analytic findings suggest that TPJ is activated not only during ToM, but also during empathic responding (for a meta-analysis on studies in both domains see Bzdok et al., 2012). As TPJ is comprised of anatomically and functionally diverse subregions in temporal, parietal and occipital cortex (Mars et al., 2012; Silani et al., 2013) it is conceivable that distinct subregions of the TPJ subserve the two functions. A study that assesses empathy and ToM within one task and in the same individuals would allow directly contrasting the activity related to empathy and ToM and thereby delineating the specifics of each function. Finally, an open question is whether the neural networks underlying empathy and ToM replicate in task-free resting state functional connectivity within the same individuals. The striking similarity of the empathy and ToM related networks with the so called task-control and default mode network, respectively, would suggest that the task-based activation peaks are embedded in these domain-general networks (Buckner et al., 2011; Tops et al., 2014; Wen et al., 2012; Yeo et al., 2011).

In order to address these questions, we developed a new experimental paradigm, the EmpaToM, that specifically allows the simultaneous investigation of affective and cognitive understanding of others. We aimed at carefully validating both the brain and behavioral measures of the EmpaToM, because investigating the specificity and separability of empathy and ToM crucially relies on its solid and accurate assessment. Furthermore, thorough validation would allow application of this paradigm in future clinical, developmental and intervention research. The EmpaToM implements an orthogonal manipulation of empathy and ToM during an ongoing realistically complex and demanding situation requiring social understanding of others. The task probes empathy through naturalistic video stimuli depicting autobiographic narratives that are either emotionally negative (e.g. experiences of loss or threat) or neutral, the latter serving as control condition. Participants' subjective empathic response was assessed via valence ratings of their affective state (positive versus negative). A second rating asked for the degree of experienced compassion for the observed other (compassion ratings). ToM was assessed during subsequent questions asking for the thoughts, goals or intentions of the other (or for factual reasoning as control condition). A last rating concerned participants' confidence with their preceeding response to allow assessment of metacognitive abilities (Fleming et al., 2010).[2] In sum, the paradigm follows a two by two factorial design with videos depicting stories with (a) negative or neutral emotional valence (later giving rise to valence and compassion ratings) and (b) ToM (e.g., irony or deception) or nonToM related story contents (later giving rise to ToM or factual reasoning questions).

For an overview of the main goals and measures applied in each experiment see Fig. 1. In Experiment 1a and b, we validated the EmpaToM with existing behavioral and functional magnetic resonance imaging (fMRI) paradigms of empathy/compassion and ToM, including the Socio-affective Video Task (Klimecki et al., 2013), a False Belief Task (Dodell-Feder et al., 2011), and the Imposing Memory Task (Kinderman et al., 1998). In Experiment 2, the EmpaToM was administered to a large representative sample in the context of a large-scale longitudinal study, the ReSource project (Singer et al., in press). In order to investigate the specific link of inter-individual differences in the activation of empathy and ToM related neural networks to inter-individual differences in the behavioral indices of these two capacities (question 1), we calculated composite measures for neural responding during empathy and ToM and tested their relation to both specific and composite behavioral measures of empathy and ToM. To test for separability of the neural networks underlying empathy and ToM (question 2), we directly contrasted empathy and ToM related activity. Furthermore, we analyzed functional connectivity of resting state scans to probe whether the observed task related neural networks are coherent across situations.

Experiment 1

In order to validate the EmpaToM on the level of behavior and neural networks, two experiments were performed. In Experiment 1a, neuroimaging of the EmpaToM, the Socio-affective Video Task (empathy), and a False Belief Task (ToM) allowed validation of the related neural networks and of the behavioral empathy measure. As the False Belief Task yields no behavioral variability in ToM performance in adults, Experiment 1b, behaviorally tested the EmpaToM and the Imposing Memory Task in a different sample to validate the behavioral ToM measure.
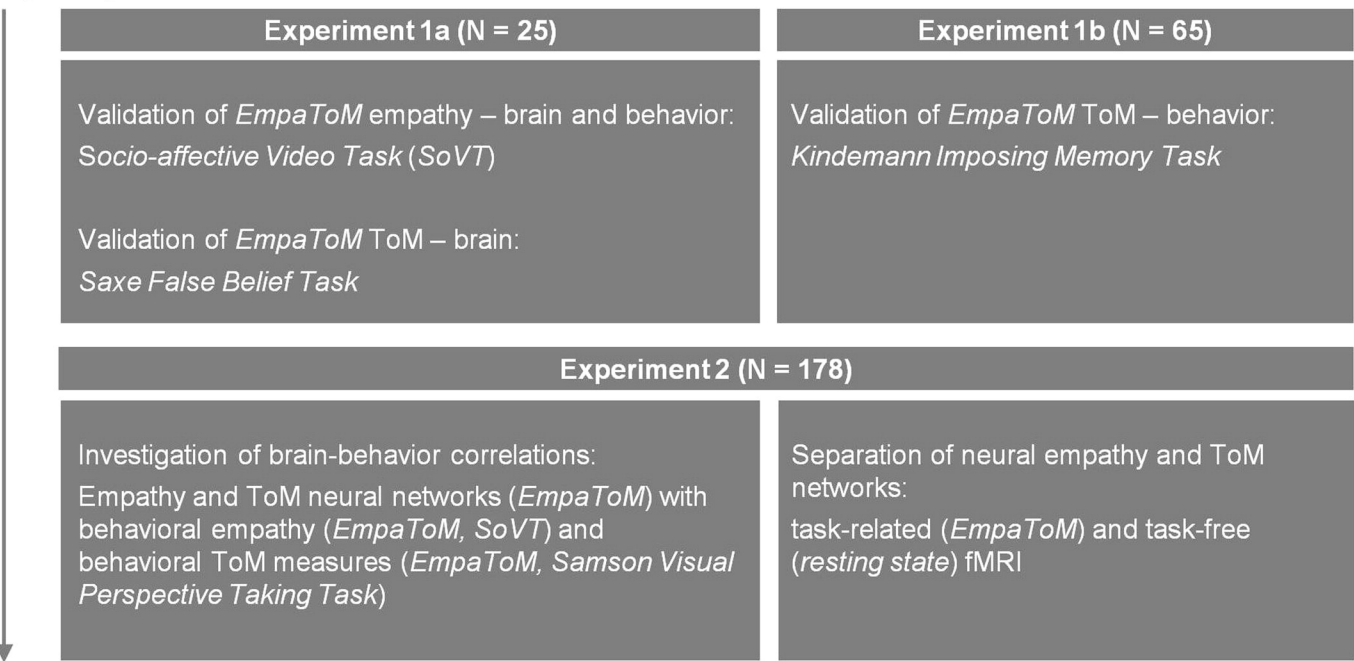
Experimental procedures

Experiment 1a

Participants. Twenty-seven volunteers participated in Experiment 1a. Data from two of them had to be excluded because of technical difficulties with the scanner, leaving 25 participants (age mean = 32.6, SD = 9.9, 14 women, all right-handed). Participants in all experiments gave written informed consent and the study was approved by the Ethics Committee of the University of Leipzig, Germany.
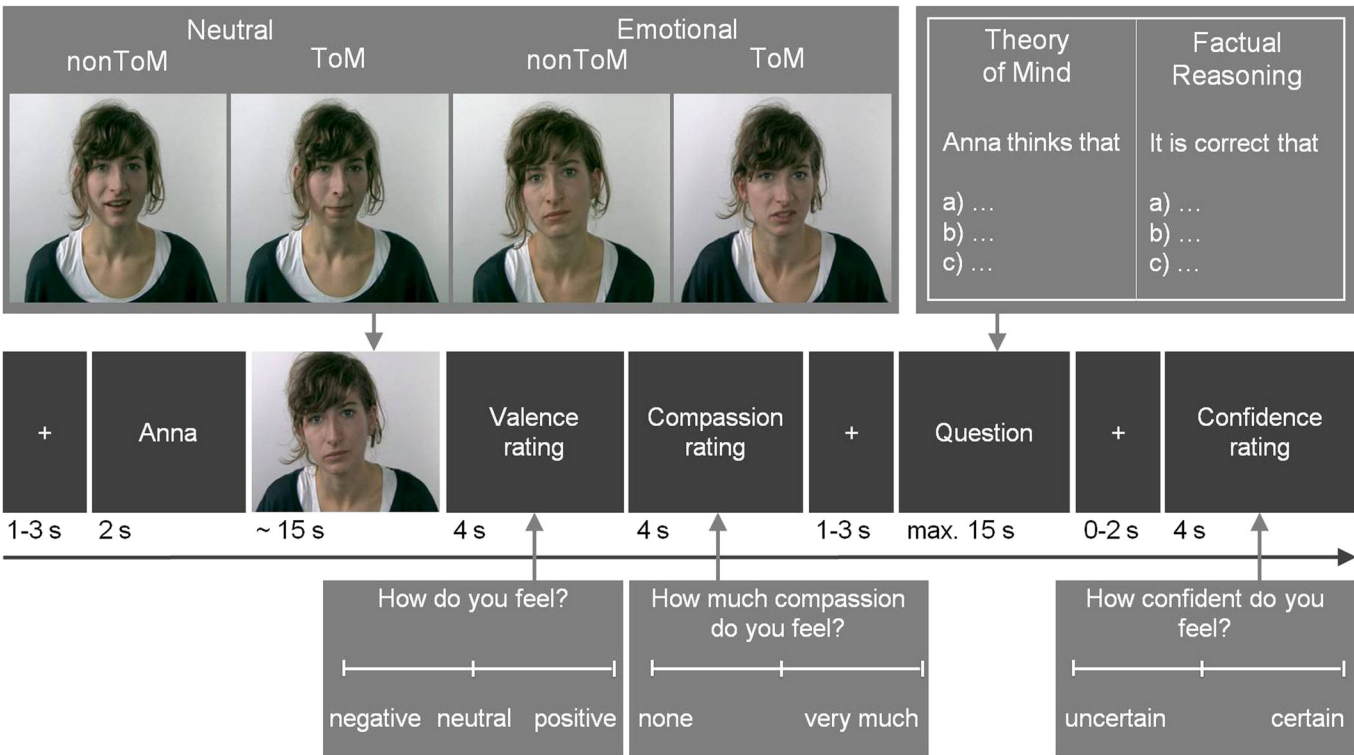
Tasks

EmpaToM. To allow measuring empathy, compassion, and ToM, the EmpaToM presented participants with a sequence of stimuli in each trial (see Fig. 1). After a fixation cross (1–3 s), the name of a person (1 s) who would subsequently be speaking in a short video (~15 s) was presented. The videos differed in emotionality (emotionally neutral vs. negative contents) and in what question they gave rise to (ToM vs. nonToM). After each video, participants were asked to rate how they felt (on a scale from negative to positive; 4 s) and how much compassion they felt for the person in the previous video (scale from none to very much; 4 s). After a fixation cross (1–3 s), a multiple choice question with three response options was presented. The questions either demanded a ToM-inference or factual reasoning on the contents of the previous video. Participants had a maximum of 14 s to select one of the response options, which was then highlighted and remained on the screen for another second. After a fixation cross (0–2 s), a confidence rating was presented asking participants how confident they were to have chosen the correct response in the previous question (4 s). Twelve trials per condition were presented. In order to control for possible effects of specific actor characteristics, each actor recounted one story per condition, thus 12 different actors were part of the

---

[2] This measure of metacognitive ability and its underlying neural network are beyond the scope of this manuscript and will be described in more detail elsewhere.

A) Study overview

Experiment 1a (N = 25)	Experiment 1b (N = 65)
Validation of *EmpaToM* empathy – brain and behavior: *Socio-affective Video Task (SoVT)* Validation of *EmpaToM* ToM – brain: *Saxe False Belief Task*	Validation of *EmpaToM* ToM – behavior: *Kindemann Imposing Memory Task*

Experiment 2 (N = 178)	
Investigation of brain-behavior correlations: Empathy and ToM neural networks (*EmpaToM*) with behavioral empathy (*EmpaToM, SoVT*) and behavioral ToM measures (*EmpaToM, Samson Visual Perspective Taking Task*)	Separation of neural empathy and ToM networks: task-related (*EmpaToM*) and task-free (*resting state*) fMRI

B) EmpaToM design

Fig. 1. (A) Overview of the conducted experiments, their specific aims and experimental tasks applied. (B) EmpaToM trial sequence. Following a 2 (Emotionality of the Video) × 2 (ToM Requirements) design, 4 different video types were presented for each actor: Emotionally negative and neutral videos; videos with and without ToM demands, thereby leading to ToM vs. factual reasoning questions. After each video, participants rated their own affect and their compassion for the person in the video. After each question, participants rated their confidence regarding their performance in the question.

stimulus set. Examplary video stories and questions can be found in Supplement S1.

Socio-affective video task (SoVT). The SoVT is an established empathy paradigm (Klimecki et al., 2013) in which participants are presented with silent video clips depicting people in distress (high emotion) or performing everyday activities (low emotion) and are asked to rate after each video how they feel themselves (valence rating) and how they feel for the other (compassion rating). The behavioral measures

derived for validation of the EmpaToM were participants' valence ratings (emotionally negative vs. neutral) and compassion ratings (mean across all conditions).

Saxe false belief task. To validate the ToM measure on the level of neural activation, the Saxe False Belief Task was applied (Dodell-Feder et al., 2011). Participants are presented with brief written statements about the beliefs of a person or about physical causality and are asked to decide if the statement was true or not.

MRI data acquisition. Brain images were acquired on a 3 T 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, FOV = 210 mm; matrix size 70 × 70. Each run began with three dummy volumes that were discarded from further analysis.

Data analysis

Behavioral data analysis. In the EmpaToM, ratings (affect, concern) and performance (reaction times (RTs) and error rates) were analyzed by means of a repeated measures analysis of variance (ANOVA). A 2 × 2 factorial design was applied with the within-subject factors Emotionality of Video (emotionally negative videos versus neutral videos) and ToM Requirement (ToM versus nonToM). Behavioral empathy was assessed with valence ratings (emotionally negative vs. neutral), behavioral compassion was assessed with compassion ratings (mean across all conditions).

The SoVT was analyzed by deriving a measure of empathy (valence ratings; emotionally negative vs. neutral) and compassion (compassion ratings; mean across all conditions). In order to validate the behavioral empathy measure of the EmpaToM, valence ratings and compassion ratings were correlated with the respective measures of the SoVT.

The Saxe False Belief Task does not provide a meaningful behavioral measure because of ceiling effects in adult populations.

fMRI data analysis. Images were analyzed using SPM8 (Wellcome Department of Imaging Neuroscience, London, UK). Preprocessing was identical for all tasks: All volumes were coregistered to the SPM single-subject canonical EPI image, slice-time corrected and realigned to the mean image volume in order to correct for head motion. 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 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.

After preprocessing, statistical analysis was carried out using the general linear model (Friston et al., 1994). For the EmpaToM, 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 (HRF). Effects of head motion were accounted for by modeling the six motion parameters for each subject as effects of no interest in the design matrix. 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' (emotionally negative vs. neutral videos) and the 'ToM contrast' (ToM vs. nonToM questions) were then calculated by applying linear weights to the parameter estimates and entered into one-sample t-tests for random effects

analysis. The SoVT was modeled with regressors for negative and neutral videos and the ToM localizer with regressors for physical and belief stories and questions. The same model estimation procedure and random effects analysis as in the EmpaToM was applied. For this first validation experiment, the more liberal threshold of p < .001 uncorrected, with a cluster threshold of k > 10 contiguous voxels was applied.

Experiment 1b

Participants. Sixty-five people (age mean = 26.6, SD = 7.0, 32 women, 63 right-handed) were recruited.

Tasks. The EmpaToM and the Kinderman Imposing Memory Task (IMT; (Kinderman et al., 1998) were assessed in order to validate the behavioral measure of ToM in the EmpaToM. The IMT measures complex and verbally based ToM performance in healthy adults. Stories were read to participants and they were asked to answer increasingly complex dual forced choice questions that either concerned ToM elements of the stories (expectations or beliefs of involved persons) or were memory questions.

Results

Experiment 1a: behavioral results

Ratings. As expected, in valence ratings, participants reported more negative affect after emotionally negative videos [$F(1, 24) = 161.9$, $p < .001$, $\eta^2 = 0.871$] (Fig. 2ηvalence ratings were also more negative after ToM videos [$F(1, 24) = 26.0$, $p < .001$, $\eta^2 = 0.529$]. The latter effect was only present in emotionally neutral videos [$t(24) = 5.4$, $p < .001$], but not in negative videos [$t(24) < 1$], reflected in a two-way interaction [$F(1, 24) = 25.8$, $p < .001$, $\eta^2 = 0.518$]. Ratings of compassion were significantly enhanced after emotionally negative videos [$F(1, 24) = 79.1$, $p < .001$, $\eta^2 = 0.767$]. Crucially, valence ratings in the EmpaToM correlated with valence ratings in the SoVT ($r = .37$, $p < 05$) and compassion ratings in the EmpaToM correlated with compassion ratings in the SoVT ($r = .59$, $p < .01$).

ToM performance. Both error rates and response times have previously been used to assess mentalizing capacities (Kinderman et al., 1998; Samson et al., 2010) and the EmpaToM meaningfully assesses both measures. As it is possible that individual response strategies differentially emphasize one over the other, RTs and error rates to the questions were combined into one composite measure of performance by z-transforming and averaging both for each condition. Performance was decreased after emotionally negative videos [$F(1, 24) = 7.54$, $p < .05$, $\eta^2 = 0.239$] (Fig. 3), but enhanced for ToM than for nonToM questions [$F(1, 24) = 26.85$, $p < .001$, $\eta^2 = 0.528$], suggesting that ToM questions were easier. The latter effect was larger for emotionally negative [$t(24) = 7.1$, $p < .001$] than for neutral [$t(24) = 1.6$, $p = .13$] videos [$F(1, 177) = 15.94$, $p < .01$, $\eta^2 = 0.399$] (see Supplements S2 and S3 for results on RTs and errors separately as well as for the results of the confidence ratings). As, in line with previous research (Dodell-Feder et al., 2011; Saxe and Kanwisher, 2003), there was no behavioral variability in the Saxe False Belief task, the correlation with the ToM performance from the EmpaToM and this task could not be computed (see Experiment 1b for the behavioral validation).

Experiment 1a: fMRI results

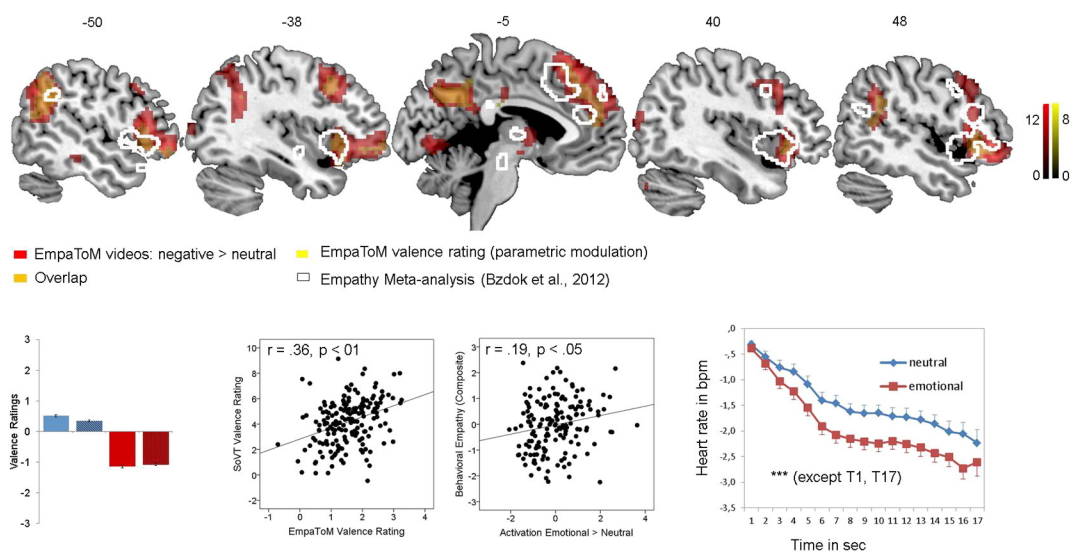Comparing emotionally negative with neutral videos in the EmpaToM ('Empathy contrast') yielded activation in bilateral AI and inferior frontal gyrus (IFG), in MPFC extending into dorsal ACC and in left TPJ with a relatively dorsal peak including supramarginal gyrus (SMG; Fig. 2; Table 1). These clusters largely overlapped with a meta-analysis of empathy studies (Bzdok et al., 2012) and partially with
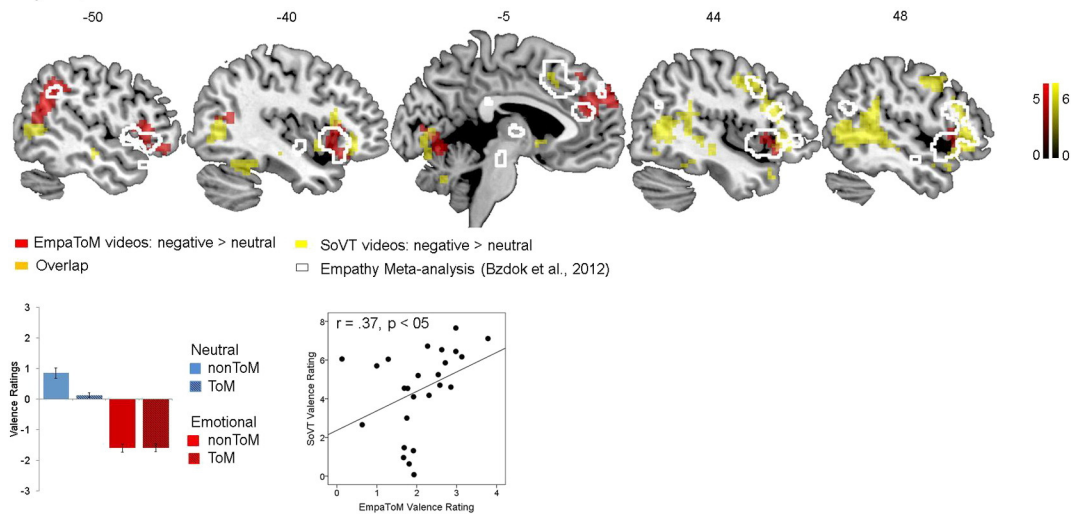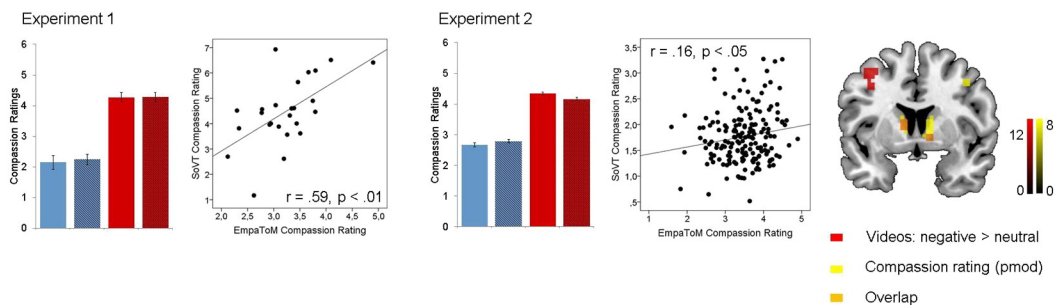
Fig. 2. Empathy and compassion in Experiments 1a and 2. (A) Experiment 1a: Brain activation for emotionally negative > neutral videos in the EmpaToM (red) and the SoVT (yellow). Meta-analytic masks are depicted as white outlines. Valence ratings of the EmpaToM and their correlation with the valence ratings in the Socio-affective Video Task are illustrated. (B) Experiment 2: Brain activation for emotionally negative > neutral videos (red) and parametric modulation with the valence ratings (yellow) in the EmpaToM. Meta-analytic masks are depicted as white outlines. Valence ratings of the EmpaToM and their correlation with the valence ratings in the Socio-affective Video Task are illustrated. Correlation of brain activation in Experiment 2 (from peak coordinates of Experiment 1a) with a composite score of affect related behavior. Heart rate deceleration in reaction to emotionally negative (red) vs. neutral (blue) videos. (C) Experiments 1a and 2 Compassion: Brain activation for emotionally negative > neutral videos (red) and parametric modulation with compassion ratings (yellow) in Experiment 3. Compassion ratings in the EmpaToM and their correlation with compassion ratings in the Socio-affective Video Task.

Theory of Mind

A) Experiment 1a and 1b
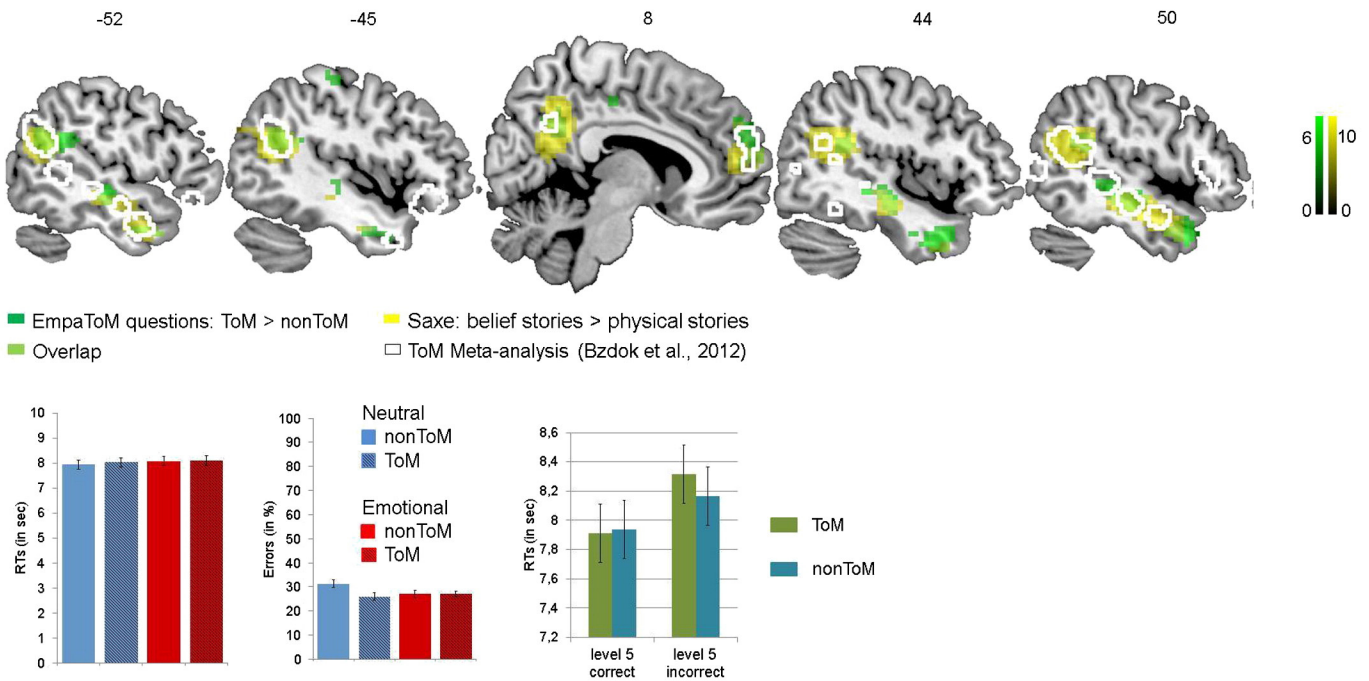
B) Experiment 2
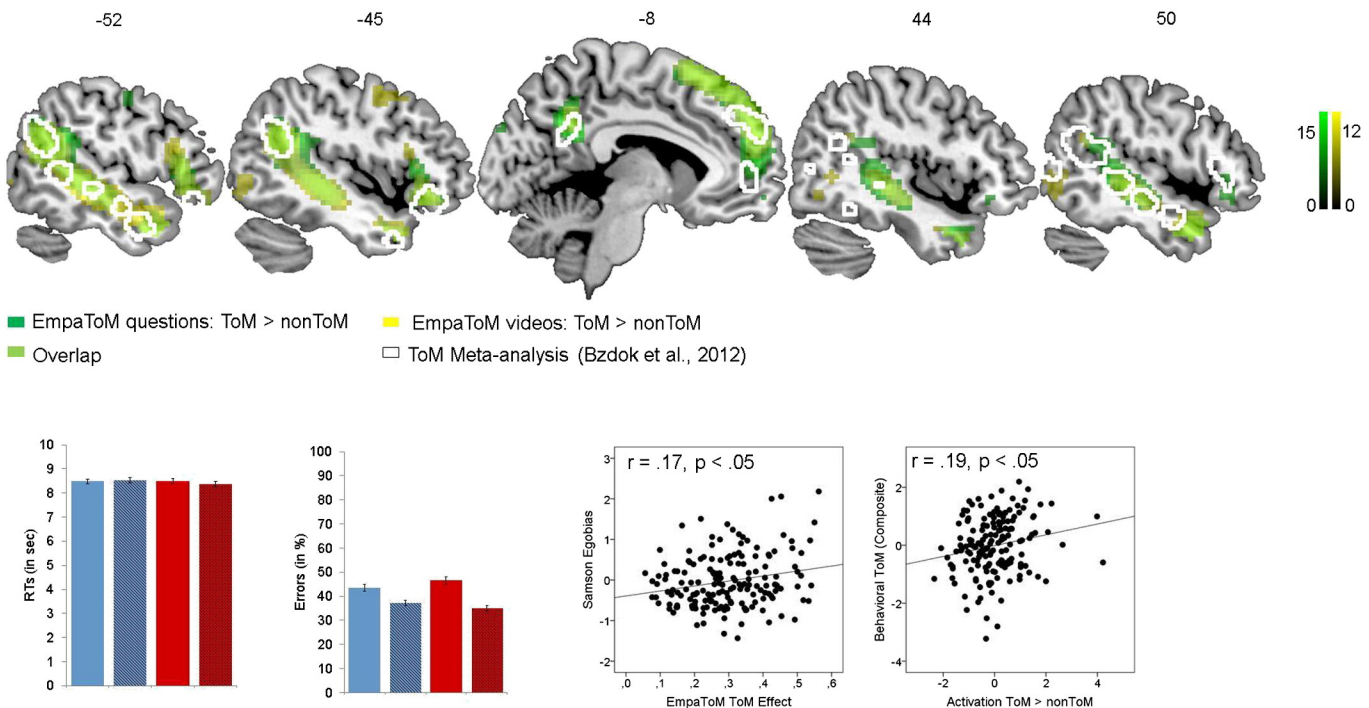
Fig. 3. ToM in Experiments 1a, 1b and 2. (A) Experiment 1a and 1b: Brain activation for ToM > factual reasoning during questions (green) in the EmpaToM and belief > physical stories in the Saxe False Belief Task (yellow). RTs and error rates in the EmpaToM and their relation with the Kinderman Imposing Memory Task (IMT). Performance in the IMT is displayed for the highest level of theory of mind (level 5). Participants with higher RTs in the ToM measure of the EmpaToM (green) performed worse in the IMT, while RT performance in the nonToM measure of the EmpaToM (blue) is not related to IMT performance. (B) Experiment 2: Brain activation for ToM > factual reasoning during questions (green) and during videos (yellow). Performance in the EmpaToM and the correlation of the composite score of ToM performance in the EmpaToM with the composite score of the egocentricity bias in the Samson Visual Perspective Taking Task are shown. The correlation of brain activation in Experiment 2 (from peak coordinates of Experiment 1a) with a composite score of ToM related behavior is illustrated.

Table 1

Activation peaks for empathy and ToM in Experiment 1a during the EmpaToM.

	H	MNI coordinates			T	Z	Cs
		x	y	z			
EmpaToM: emotionally negative > neutral video							
Middle frontal gyrus	R	24	57	21	4.92	4.05	21
Superior frontal cortex	L	−18	57	27	5.06	4.13	336
Superior medial frontal cortex	L	−9	57	30	4.99	4.09	
Anterior cingulate cortex	R	6	45	24	4.98	4.08	
Superior medial frontal cortex	L	−3	33	51	3.8	3.33	18
Supplementary motor area	L	−3	24	54	3.58	3.18	
Inferior frontal gyrus	L	−48	33	−9	5	4.1	257
Anterior insula	L	−36	18	−6	4.97	4.08	
Anterior insula	R	36	22	−9	4.25	3.63	81
Inferior frontal gyrus	R	54	21	0	4.25	3.63	
Middle frontal gyrus	L	−48	18	45	3.99	3.46	14
TPJ–angular/supramarginal gyrus	L	−51	−51	30	5.86	4.57	201
TPJ–supramarginal gyrus	L	−51	−42	39	4.85	4.01	
Lingual gyrus	L	−9	−66	−9	5.63	4.45	
Lingual gyrus		0	−69	3	5.88	4.58	132
Cerebellum	R	24	−81	−36	5.28	4.26	131
EmpaToM: ToM > nonToM question							
Superior medial frontal cortex	L	−3	66	27	4.82	3.99	260
Superior medial frontal cortex	L	−9	57	24	4.65	3.89	
Inferior frontal gyrus	L	−30	21	−18	4.71	3.92	11
Temporal pole	R	48	9	−36	7.01	5.12	106
Temporal pole	R	48	15	−27	6.79	5.02	
Temporal pole	L	−54	3	−30	5.76	4.52	51
Putamen	L	−24	−12	12	4.47	3.78	14
Superior temporal cortex	R	51	−12	−9	5.88	4.59	155
Superior temporal cortex	R	48	−30	−6	5.47	4.36	
Superior temporal cortex	L	−51	−18	−9	5.78	4.53	47
Middle cingulate cortex	R	6	−18	45	4.14	3.56	11
Middle cingulate cortex	L	−3	−18	45	3.76	3.3	
Central opercular cortex	L	−57	−21	21	4.59	3.85	18
Postcentral gyrus	L	−39	−27	66	4.62	3.87	57
TPJ–supramarginal/superior temporal gyrus	R	66	−27	27	5.46	4.36	83
TPJ–supramarginal/angular gyrus	R	54	−42	21	5.12	4.17	84
TPJ–angular/supramarginal gyrus	R	45	−45	21	4.93	4.06	
TPJ–angular gyrus	R	63	−45	21	3.88	3.38	
Precuneus	L	−6	−51	36	6.17	4.73	228
Posterior cingulate cortex	L	−6	−51	27	5.98	4.64	
Posterior cingulate cortex	R	6	−51	33	5.76	4.52	
TPJ–angular gyrus	L	−45	−57	24	7.25	5.23	213
TPJ–supramarginal/angular gyrus	L	−51	−48	24	5.29	4.27	
Cerebellum	L	−21	−75	−39	4.73	3.94	60
Cerebellum	R	30	−81	−36	5.5	4.38	63

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

328 activation for emotionally negative > neutral videos in the SoVT in left
329 and right AI/IFG (Table S4).
330 Comparing ToM with nonToM questions in the EmpaToM ('ToM
331 contrast') yielded activation in bilateral TPJ with more ventral peaks
332 than in the empathy contrast, superior temporal sulcuc (STS), TP,
333 MPFC and precuneus/PCC (Fig. 3 and Table 1). These clusters largely
334 overlapped with a meta-analysis of ToM studies (Bzdok et al., 2012)
335 and with the Saxe False Belief task (Table S4).

336 *Experiment 1b: behavioral results*

337 Valence and compassion ratings and ToM performance replicated
338 the main findings of Experiment 1 (Supplement S5 and S6). Crucially,
339 the composite ToM performance in the EmpaToM correlated with per-
340 formance on the most difficult level of ToM in the Imposing Memory
341 Task, a verbal high-level mentalizing task (point biserial correlation
342 $r = .28$, $p < 05$).

343 *Discussion*

344 Experiments 1a and b demonstrate the validity of the empathy and
345 ToM measures of the newly developed task on a behavioral and neural
346 level. Specifically, empathy and compassion ratings in the EmpaToM

347 were related to the respective measures of the established SoVT and
348 ToM performance in the EmpaToM correlated with high-level ToM
349 performance in the IMT. These findings were paralleled by the substan-
350 tial overlap of the empathy and ToM related neural networks with
351 activity observed during established tasks (SoVT, Saxe False Belief
352 Task) and with the regions identified in a recent meta-analysis (Bzdok
353 et al., 2012).
354 In Experiment 2 we aimed at using a larger sample of participants to
355 directly investigate the relation of behavioral empathy and ToM mea-
356 sures to the respective neural network activity as observed in the
357 EmpaToM. Further, we probed the separability of the neural networks
358 related to empathy and ToM in task-related and task-free fMRI.

359 **Experiment 2**

360 *Experimental procedures*

361 *Participants*

362 191 participants participated in the experiment. Thirteen partici-
363 pants were excluded due to technical problems during data acquisition.
364 178 participants (age mean = 40.9 years, SD = 9.5, 106 female, 176
365 right-handed) were included in the final data set.

366 *Tasks*

367 *EmpaToM task.* The task was similar to Experiment 1. Five parallel
368 versions of the task were created and randomly applied to five
369 subgroups of the total sample. Each set was composed of 48 videos
370 from 12 actors (each actor contributing one video per condition of the
371 2×2 design, emotionally negative versus neutral videos, ToM versus
372 nonToM). Sets were created based on iterative behavioral pilot data so
373 that the five final task sets did not differ in terms of valence and compas-
374 sion ratings, RTs, errors, confidence ratings as well as duration of the
375 videos (for each condition). The following semantic characteristics of
376 the questions were matched to be constant across conditions: number
377 of words, number of characters, number of predicates, number of
378 changes in tense, complexity of the sentences (number of main and
379 subordinate clauses), number of passive sentence constructions, and
380 number of conjunctives (Table S7).

381 *Socio-affective video task (SoVT) and Samson Visual Perspective Taking*
382 *Task.* The SoVT (see above) and the Samson Visual Perspective Taking
383 Task (Samson et al., 2010) were assessed behaviorally. The Samson
384 Visual Perspective Taking Task requires participants to judge a three
385 dimensional visual scene either from their own or an avatar's perspec-
386 tive by delivering speeded dual choice responses to questions
387 concerning how many objects either they themselves or the avatar
388 can see (Samson et al., 2010). The measure of relevance to us was the
389 egocentricity bias, that is, the tendency to implicitly calculate one's
390 own perspective when judging the avatar's perspective. The ability to
391 overcome one's egocentricity bias when required to select the other's
392 perspective has been argued to be a cognitively demanding component
393 of mentalizing (Qureshi et al., 2010).

394 *MRI data acquisition*

395 Data acquisition for the EmpaToM was identical to Experiment 1.
396 Furthermore, 6.7 min of resting state (eyes opened with instruction to
397 focus on a fixation cross) data were recorded on the same day with
398 the same EPI sequence.

399 *Physiological data acquisition*

400 The electrocardiogram (ECG), skin conductance response (SCR),
401 and respiration were measured with a Brainamp ExG MR compatible
402 amplifier and Brain Vision Recorder 1.20 (Brain Products). Signals
403 were acquired unfiltered and sampled at 5000 Hz. The ECG was record-
404 ed using three Easycap electrodes (20kOhm) with Ten20 conductive

paste (12.5% NaCl) that were located on the back under the seventh cervical vertebra, on the left dorsal side at the height of the tenth rib and on the lumbar part of the spine. Two skin conductance electrodes were placed adjacently on the left middle and index fingertips, using GSR-MR-electrodes (Ag/AgCl, Brain Products) filled with skin conductance electrode paste. Respiration was acquired by a movement sensitive belt attached to the abdomen (3D acceleration sensor MR, Brain Products). By means of this belt it was possible to record movements in three dimensions.

Data analysis

Behavioral data analysis. Behavioral data analysis was identical to Experiment 1.

Physiological data analysis. An in house Matlab function was used to correct scanner gradient artifacts in the ECG signal. For the correction of the GSR Signal, a 3rd order butterworth 0.5 Hz low pass filter and a 2 Hz high pass filter was applied (Figner and Murphy, 2011). The respiration signal was corrected with a 3rd order butterworth 0.3 Hz low pass filter and a 0.05 high pass filter. R-peaks were detected by an in house Python 2.7 routine, that marks a peak that exceeds a threshold in the signal within an individually adjustable time window. Reactions in heart rate during the videos were determined by subtracting activity 1 s before video onset from that occurring each second after video onset (Bradley et al., 2001). Heart rate waveform scores were computed by determining, for each participant and each trial, the mean deceleration during the videos, and the maximum deceleration from baseline across the duration of the videos (Hodes et al., 1985). For skin conductance, the number of significant (= above-threshold 0.01 μS) SCRs, Area (i.e. time integral) of phasic driver (equals SCR multiplied by size of response window [muS*s]) and mean tonic activity (of decomposed tonic component) were calculated during each video by means of Ledalab V3.4.5. The response windows in which the GSR-signal was analyzed, corresponded to the mean video length of each video category. Reactions in respiration were assessed by calculating the amount of respirations (inhale and exhale) during each video. Heart rate deceleration, skin conductance response, and respiration were analyzed by means of a repeated measures one factor ANOVA on Emotionality of Video (emotionally negative vs. neutral videos).

fMRI data analysis. Preprocessing and first-level model estimation were the same as in Experiment 1. Two additional models were estimated with only one regressor for all videos and with regressors for the parametric modulation of video-related activity by valence and compassion rated after the video, respectively. On the second level, simple t-tests for the 'Empathy contrast' (emotionally negative–neutral videos) and the 'ToM contrast' (ToM–nonToM questions and ToM–nonToM videos) were performed. Specifics of ToM and empathy were analyzed by entering the respective first-level contrast images into a factorial design. The specific contrasts 'ToM > Empathy' and 'Empathy > ToM' were then inclusively masked for significant voxels of the respective simple contrast. All contrasts were thresholded at a p < .05 FWE-corrected level and an extent threshold of k > 10 contigous voxels was applied.

Resting state analysis. Resting state data was analyzed with SPM8 and DPARSF (Chao-Gan and Yu-Feng, 2010). The first 10 volumes were discarded. The remaining 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, the head motion scrubbing regressor, white matter signal and the CSF signal were removed from the functional data. The linear trend of time courses were removed and then temporally band-pass filtering (0.01–0.08 Hz) to reduce the very low-frequency drift and high-frequency respiratory and cardiac noise.

For functional connectivity calculation, spheres (radius = 5 mm) around the peak regions observed in the specific contrasts were defined as seed regions. The averaged time course was then obtained from the sphere ROI and the correlation analysis was performed in a voxel-wise way to generate the FC. The correlation coefficient map was 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 and re-sampled to 3-mm isotropic voxels as well as smoothed with a 4 mm FWHM kernel. All contrasts were thresholded at a p < .05 FWE-corrected level and an extent threshold of k > 10 contigous voxels was applied.

Results

In this section, we will first report the main behavioral, physiological and fMRI findings of Experiment 2 and then address the two main questions in succession: First, are increased empathy ratings and ToM performance specifically linked to higher activity in the respective neural networks. Second, what are the specifics of the neural networks related to empathy and ToM and can these networks also be separated in task-free resting state connectivity.

Behavioral results

Valence and compassion ratings and ToM performance replicated the main findings of Experiments 1a/b (Figs. 2 and 3; Supplement S8 and S9). Like in Experiment 1, valence ratings in the EmpaToM correlated with valence ratings in the SoVT (r = .36, p < 01), and compassion ratings in the EmpaToM correlated with compassion ratings in the SoVT (r = .16, p < .05). The composite of ToM performance in the EmpaToM correlated with the composite of the egocentricity bias in the Samson Task (r = .17, p < .05), indicating that participants who were better able to overcome the egocentricity bias are also more successful in solving ToM questions.

Physiological responses

In addition to behavioral measures we assessed parameters of the autonomic system to get more implicit measures of emotional involvement. As expected, heart rate deceleration was enhanced by emotionally negative videos [F(1, 158) = 12, p < .01, η^2 = 0.218] (Fig. 2). Skin conductance response and respiration however yielded no significant effects of Emotionality of Video [Fs(1, 159) < 1].

fMRI results

Similar to Experiment 1, comparing emotional with neutral videos ('Empathy contrast') activated bilateral AI and IFG, MPFC extending into ACC, as well as dorsal TPJ including SMG (Fig. 2, Table 2). Activity in these regions varied parametrically with the subjective valence ratings that participants gave after each video. Testing for parametric variation with the compassion ratings yielded a similar picture with two main differences; AI activity, while varying with the amount of experienced negative affect, did not vary with compassion ratings. Conversely, activity in a cluster in the ventral striatum, which was observed when comparing emotionally negative with neutral videos, did vary with compassion ratings, but not with the valence ratings (Table S10).

As in Experiment 1, comparing ToM with nonToM questions ('ToM contrast') yielded activation in bilateral TPJ, STS, TP, MPFC and precuneus/PCC (Fig. 3 and Table 2). The same pattern of activation was found when comparing BOLD responses when watching ToM versus nonToM videos, that is before explicit ToM or factual reasoning judgments where required (Table S10).

Linking behavioral empathy and ToM parameters to the respective neural activation

To obtain meaningful composite scores for all behavioral ToM and empathy measures assessed in Experiment 2, a principal component

Table 2

Activation peaks for empathy and ToM in Experiment 2 during the EmpaToM.

	H	MNI coordinates			T	Z	Cs
		x	y	z			
EmpaToM: emotionally negative > neutral video							
Inferior frontal gyrus	L	−48	39	−9	10.68	>8.21	1027
Middle frontal	L	−42	15	45	10.04	>8.21	
Anterior insula	L	−36	21	−6	8.58	7.82	
Superior medial frontal cortex	L	−3	33	51	10.53	>8.21	1257
Superior medial frontal	R	9	21	57	8.69	>8.21	
Inferior frontal gyrus	R	51	30	−6	10.01	>8.21	737
Middle frontal	R	42	21	39	6.96	6.53	
Anterior insula	R	30	24	−15	6.64	6.27	
Ventral striatum	R	9	3	0	6.29	5.97	153
Ventral striatum	L	−6	−3	0	6.16	5.86	
Caudate	L	−12	6	12	6.12	5.82	
Caudate	R	12	6	12	6	5.72	
Middle cingulate		0	−18	39	8.25	7.58	82
Middle temporal cortex	L	−54	−30	−12	6.08	5.79	26
TPJ–angular/supramarginal gyrus	R	63	−48	33	9.71	>8.21	448
Middle temporal cortex	R	60	−57	9	7.44	6.93	
TPJ–angular/supramarginal gyrus	L	−54	−51	33	12.49	>8.21	599
Precuneus		0	−63	36	12.01	>8.21	614
Lingual gyrus	L	−6	−75	−3	8.07	7.43	162
Middle occipital	R	42	−84	18	5.98	5.7	30
Middle occipital	L	−39	−90	9	5.1	4.92	13
Cerebellum	L	−15	−78	−30	9.88	>8.21	186
Cerebellum	R	18	−81	−33	10.04	>8.21	219
EmpaToM: ToM > nonToM questions							
Rectus	R	3	57	−18	7.71	7.15	38
Superior medial frontal	L	−9	54	24	13.72	>8.21	1185
Superior frontal	L	−9	54	33	12.34	>8.21	
Superior medial frontal	R	9	57	21	11.73	>8.21	
Inferior frontal gyrus	R	54	30	3	6.24	5.92	52
Inferior frontal gyrus	L	−51	24	6	10.32	>8.21	226
Inferior frontal gyrus	L	−45	27	−9	9.93	>8.21	
Temporal pole	R	51	9	−33	14.68	>8.21	121
Temporal pole	L	−51	3	−30	12	>8.21	79
Postcentral	L	−54	−6	48	6.05	5.76	13
Middle cingulate		0	−15	39	8.56	7.81	50
Supplementary motor area	R	6	−24	57	5.37	5.17	10
TPJ–middle temporal	R	51	−30	−3	10.61	>8.21	640
TPJ–superior temporal	R	48	−18	−9	9.81	>8.21	
TPJ–angular gyrus	R	63	−45	21	7.76	7.19	
Posterior cingulate/precuneus	L	−6	−51	30	16.38	>8.21	328
TPJ–angular gyrus	L	−51	−57	24	15.81	>8.21	1019
TPJ–middle temporal	L	−48	−30	−3	10.49	>8.21	
TPJ–superior temporal	L	−60	−18	−6	9.53	>8.21	
Cuneus	L	−9	−93	30	5.7	5.45	10
Cuneus	R	15	−87	39	6.11	5.82	24
Cerebellum	R	27	−78	−33	15.82	>8.21	145
Cerebellum	L	−27	−81	−36	14.65	>8.21	101

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

analysis (PCA) was performed on the critical measures of empathy and ToM from the EmpaToM, the SoVT, and the Samson Visual Perspective Taking Task. The following measures were included: EmpaToM ToM performance composite, Samson egobias composite, EmpaToM valence rating, SoVT valence rating, and EmpaToM performance in the questions (composites of errors and reaction times in both ToM and nonToM questions) following emotional videos minus following neutral videos. The latter measure reflects the degree to which participants' overall performance is impaired by the emotionality of the videos. A PCA with oblique rotation was performed, factors were derived according to the Kaiser Criterion. This analysis yielded two independent factors (KMO = .54; Bartlett's Test of Sphericity = 42.4, p < .001; % variance explained = 54) (Table 3). The first factor, 'Behavioral Empathy', entailed the valence ratings of the EmpaToM and the SoVT, and the degree to which performance in the EmpaToM was negatively influenced by the emotionality of the previous video. The second factor entailed ToM performances in the EmpaToM and in the Samson Task and was termed 'Behavioral ToM'.

Table 3

Factor solution and loadings of the variables for the PCA on the behavioral measures on empathy and ToM from the EmpaToM, the Socio-affective Video Task (SoVT) and the Samson Visual Perspective Taking Task (EmpaToM performance in the questions (affect performance represents composites of errors and reaction times in both ToM and nonToM questions following emotional videos minus following neutral videos).

	F1 (empathy in Tasks)	F2 (ToM in Tasks)
SoVT: valence rating	0.79	
EmpaToM: valence rating	0.75	−0.25
EmpaToM: affect performance	0.53	
EmpaToM: ToM performance		0.77
Samson: egocentricity bias		0.71

In order to relate the observed brain activation patterns to behavior on an inter-individual level, we extracted the percent signal change in Experiment 2 from those peak activations (5 mm spheres) observed for the respective contrasts in the independent Experiment 1a within meta-analytically identified brain regions for empathy and ToM (Bzdok et al., 2012). In this way, we avoided statistical overestimation of brain–behavior links (Kriegeskorte et al., 2010). In order to obtain data-driven composites, the extracted activations for the contrast of empathy (emotionally negative vs. neutral videos) and ToM (ToM vs. nonToM questions) were subjected to a factor analysis. A PCA with oblique rotation was performed, factors were derived according to the Kaiser Criterion. This analysis yielded two distinct factors: 'Brain Empathy' and 'Brain ToM' (KMO = .84; Bartlett's Test of Sphericity = 1608.1, p < .001; % variance explained = 58) (Table 4).

Interestingly, when 'Brain Empathy' and 'Brain ToM' were entered stepwise as predictors in a multiple linear regression analysis with valence ratings in the EmpaToM (emotionally negative vs. neutral videos) as dependent variable, a specific relation of 'Brain Empathy' to valence ratings was revealed ($R^2 = .046$, standardized beta = .214, p < .01), while 'Brain ToM' did not explain additional variance ($R^2 = .004$, standardized beta = .062, p > .30). The same selective relation held true when including 'Behavioral Empathy' as dependent variable ('Brain Empathy': $R^2 = .033$, standardized beta = .181, p < .05; 'Brain ToM': $R^2 = .034$, standardized beta = −.028, p > .30). Also, when 'Brain ToM' and 'Brain Empathy' were entered stepwise as predictors in a multiple linear regression analysis with ToM performance in the EmpaToM as dependent variable, a specific relation of 'Brain ToM' to ToM performance was revealed ($R^2 = .022$, standardized beta = −.150, p < .05), while 'Brain Empathy' did not explain

Table 4

Factor solution and loadings of the variables for the PCA on the peak activations of the 'Empathy contrast' (emotionally negative vs. neutral videos) and the 'ToM contrast' (ToM vs. nonToM questions).

Contrast	Region	F1 (emo > neutral videos)	F2 (ToM > nonToM questions)
Empathy	Superior medial frontal	0.84	
	r AI	0.83	
	l middle frontal	0.82	
	l TPJ	0.82	
	r middle frontal	0.80	
	l AI	0.79	
	r TPJ	0.73	
	precuneus	0.70	
ToM	l superior temporal		0.82
	l TPJ		0.78
	precuneus/posterior cingulate		0.76
	l temporal pole		0.72
	r superior temporal		0.72
	Superior medial frontal		0.69
	r TPJ		0.67
	r temporal pole		0.62

572 additional variance ($R^2 = .004$, standardized beta $= -.064$, $p > .30$).
573 The same selective relation held true when including 'Behavioral ToM'
574 as dependent variable ('Brain ToM': $R^2 = .033$, standardized beta $=$
575 .181, $p < .05$; 'Brain Empathy': $R^2 = .039$, standardized beta $= -.080$,
576 $p > .30$).
577   In sum, this pattern clearly shows that the relation of brain
578 and behavioral parameters of empathy and ToM are specific and
579 dissociable.

Separability of the empathy and ToM related neural networks   580

Finding specific activation patterns for empathy and ToM   581
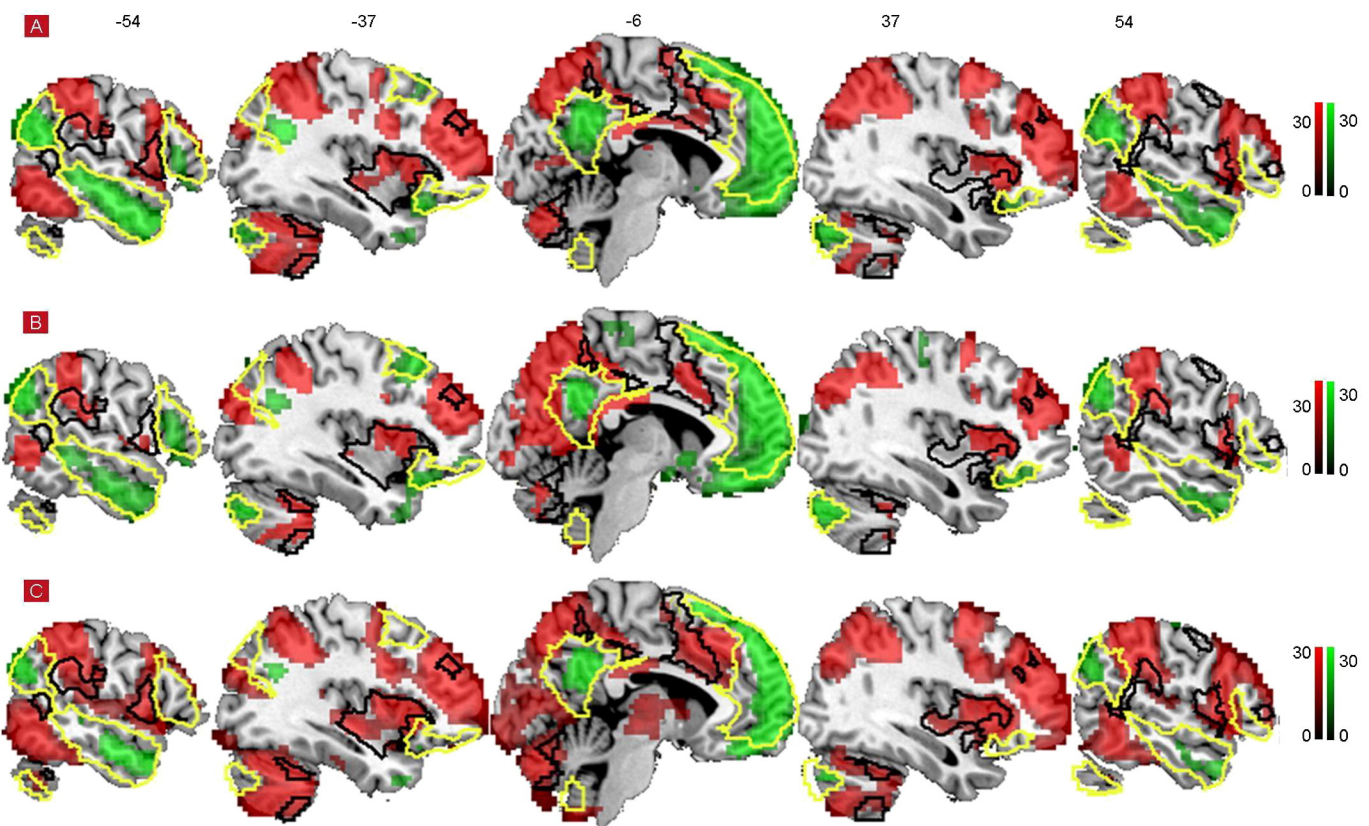
582   After having established two brain networks underlying the abilities
583 to empathize and mentalize, we asked whether we could identify
584 specific brain regions that are activated more by one or the other
585 route of social cognition. We therefore contrasted 'emotionally
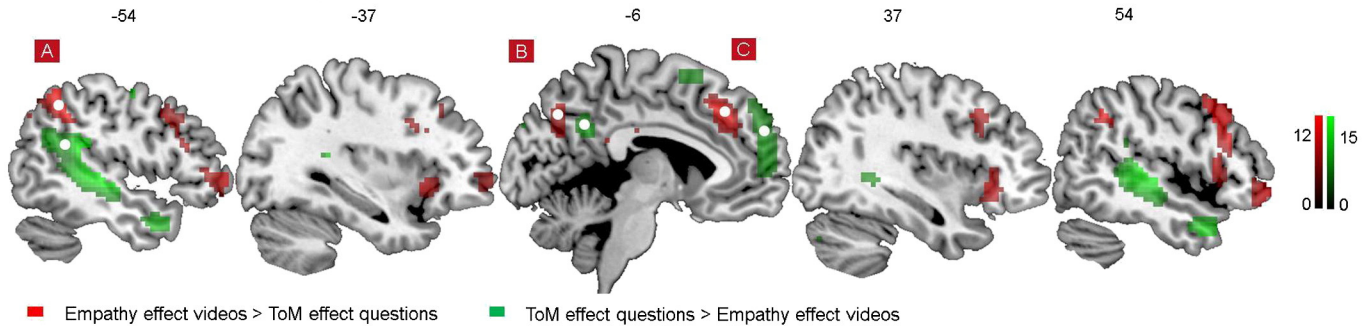
Fig. 4. Specifics of the social cognition networks (Experiment 2). (A) Contrast of empathy related (emotionally negative > neutral videos) > ToM related (ToM > nonToM questions) activations (red) and vice versa (green) in the EmpaToM. (B) Resting state connectivity seeded from peaks depicted in panel (A) for Empathy > ToM and vice versa. As outlines, the networks observed in Buckner et al. (2011) and Yeo et al. (2011) are displayed (networks 4 (as black outline) and 7 (yellow outline)).

Table 5

Activation peaks for the differential contrasts for empathy and ToM in Experiment 2 during the EmpaToM.

			MNI coordinates			T	Z	Cs
			x	y	z			
EmpaToM: (emotional > neutral video) > (ToM > nonToM questions)								
Middle frontal (IFG)	L		−45	51	3	8.79	>8.21	155
Inferior frontal triangularis	L		−45	39	6	7.91	7.58	
Superior medial frontal (incl. anterior cingulate)			0	30	45	12.37	>8.21	235
Anterior insula	L		−33	21	−3	7.97	7.63	58
Middle frontal	R		42	18	51	11.99	>8.21	424
Middle frontal	L		−45	12	36	7.56	7.27	108
Middle frontal	L		−36	12	33	7.26	7	
Inferior frontal triangularis	L		−48	27	21	7.11	6.87	
Anterior insula	R		30	22	−6	9.17	>8.21	
Inferior frontal operculum	R		48	18	36	9.92	>8.21	
Inferior orbitofrontal	R		45	48	−9	9.42	>8.21	
Middle cingulate	R		3	−24	30	7.62	7.33	24
Precuneus	R		9	−45	39	7.15	6.91	13
TPJ–angular/supramarginal gyrus	R		51	−48	45	13.64	>8.21	115
TPJ–angular/supramarginal gyrus	L		−39	−54	48	12.21	>8.21	217
TPJ–supramarginal gyrus	L		−48	−48	45	11.83	>8.21	
Cuneus	L		−9	−66	27	7.48	7.2	107
Precuneus	L		−6	−69	45	7.02	6.79	
Precuneus	R		12	−63	30	6.7	6.5	
Fusiform gyrus	L		−27	−69	−9	5.63	5.51	32
Cerebellum	L		−12	−75	−30	10.69	>8.21	45
Cerebellum	R		12	−81	−33	6.3	6.13	32
EmpaToM: (ToM > nonToM questions) > (emotionally negative > neutral video)								
Superior frontal	L		−9	54	36	8.72	>8.21	386
Superior medial frontal	L		−9	57	21	7.91	7.58	
Superior medial frontal	R		6	57	21	7.34	7.08	
SMA	L		−6	12	63	6.86	6.64	34
Temporal pole	L		−51	12	−24	10.11	>8.21	72
Temporal pole	L		−54	6	−15	8.32	>8.21	
Temporal pole	R		51	12	−30	11.52	>8.21	117
Temporal pole	R		57	9	−12	10.31	>8.21	
Superior temporal	R		60	0	−6	16.78	>8.21	393
Superior temporal	R		54	−18	0	16.15	>8.21	
Superior temporal	R		42	−33	12	9.42	>8.21	
Postcentral	L		−54	−6	48	12.07	>8.21	13
TPJ–superior temporal	L		−60	−15	0	18.75	>8.21	692
TPJ–planum temporale	L		−48	−39	21	12.6	>8.21	
TPJ–angular gyrus	L		−48	−57	24	8.98	>8.21	
SMA	R		6	−24	57	6.55	6.36	10
Posterior cingulate/precuneus	L		−6	−51	30	8.85	>8.21	93
Superior occipital	R		15	−87	36	6.68	6.48	24
Cerebellum	L		−27	−81	−36	8.18	7.81	27
Cerebellum	R		27	−81	−36	7.93	7.6	41

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

negative > neutral videos' with 'ToM > nonToM questions' ('Empathy > ToM') and vice versa ('ToM > Empathy'). These analyses yielded distinct networks with AI and IFG, MPFC extending into dorsal ACC, and TPJ (dorsal region including SMG) for 'Empathy > ToM' and ventral TPJ, STS, TP, MPFC and precuneus/PCC for 'ToM > Empathy' (Fig. 4, Table 5). These networks largely matched the previous patterns identified for the main effects of Empathy and ToM. Importantly, in those regions where some overlap was present for Empathy and ToM in the main contrasts, specific peaks for each capacity could be identified. This included the temporoparietal region (with a more ventral TPJ peak for ToM and a more dorsal TPJ peak (including SMG) for Empathy), as well as posterior and anterior midline regions (with more anterior ventral peaks for ToM and dorsal posterior peaks for Empathy in precuneus/posterior cingulate and MPFC).

Resting state functional connectivity

To test how the specific regions identified when contrasting empathy and ToM related activity are embedded in task-free networks, we compared resting state functional connectivity from seed regions related to 'ToM > Empathy' or 'Empathy > ToM' in TPJ, MPFC, and precuneus/PCC (Fig. 4, Table 3). This analysis indeed yielded distinct networks that resembled the task-based networks (Table S11). Bilateral AI, ACC/MPFC, DLPFC and more dorsal regions in TPJ were connected more strongly to empathy related seed regions. More ventral regions of TPJ and MPFC, precuneus/PCC, STS, TP showed stronger connectivity to ToM related seed regions. The two networks strongly overlapped with previous descriptions of large-scale resting state circuits, specifically the default mode network and the task control network (also referred to as salience network, cingulo-opercular network or ventral attention network) (Buckner et al., 2011; Yeo et al., 2011).

Discussion

Replicating the findings of Experiment 1a and b and further validating the EmpaToM, Experiment 2 revealed correlations of empathy and compassion ratings with established behavioral measures of empathy and compassion (SoVT), and a correlation of ToM performance with a well-known measure of perspective taking (Samson Visual Perspective Taking Task). In addition, physiological and imaging results are in line with the literature, indicating sound assessment of all these measures. In line with the literature, the degree of empathy that participants reported in behavioral assessments was increased with increasing activation of the empathy network. Crucially, inter-individual differences in the activation of the empathy related neural network, but not the ToM related neural network, predicted behavioral empathy indices across tasks. Vice versa, differences in activation of the ToM related network, but not of the empathy related network, predicted ToM performance across tasks. These specific brain behavior relations provide further evidence for a dissociation of empathy and ToM (question 1). Finally, based on task related and task-free analyses of functional brain activation, results of Experiment 2 revealed separable networks underlying empathy and ToM (question 2).

General discussion

The present study validated a novel behavioral and fMRI paradigm, the EmpaToM, that assesses both the neural networks and behavioral markers related to empathy and compassion on the one hand and ToM ability on the other hand within one individual. We utilized this novel task to demonstrate (1) specific brain–behavior relations for both empathy and ToM and (2) distinct neural networks underlying empathy and ToM in task related and task-free fMRI within the same individuals.

In two experiments, the EmpaToM consistently induced empathy as evident in subjective valence ratings, heart rate deceleration, and activation of empathy related brain regions (including dorsal AI, dorsal ACC/MPFC, IFG, SMG/dorsal TPJ). Trial-by-trial variations in valence ratings modulated activation within this network, as did inter-individual variations in affective responding across subjects. Importantly, both the behavioral and neural markers of empathy were validated with an external established empathy task, the SovT (Klimecki et al., 2014). The activation cluster for empathy in MPFC and ACC was more dorsal than reported in empathy for pain studies (for a meta-analysis see Lamm et al., 2011), but overlapped with a recent meta-analysis on a broad range of empathy paradigms (Bzdok et al., 2012).

Interestingly, the ventral striatum, a region typically associated with positive emotion and reward (Cardinal et al., 2002), was activated during presentation of emotionally negative videos. Activation in this region varied parametrically with participants' subjective ratings of compassion, while it was independent of the valence ratings. This is in line with findings of enhanced ventral striatum activity after compassion but not empathy training (Engen and Singer, 2015; Klimecki et al., 2013, 2014) and suggests that the feeling of a caring affection for others is related to ventral striatum activity also in untrained individuals. The EmpaToM, thus, allows differentiating empathic, rather

negatively valenced, and compassionate, rather positive valenced, responses when exposed to the suffering of others.

Concerning the cognitive route to understanding others, ToM questions in the EmpaToM induced activation in regions typically observed for mentalizing (including bilateral ventral TPJ, STS, TP, precuneus and MPFC), which overlapped with activation in an established false-belief task (Dodell-Feder et al., 2011) assessed within the same sample and with regions observed in a recent meta-analysis on diverse ToM tasks (Bzdok et al., 2012). On a behavioral level, ToM performance in the verbal and high-level EmpaToM was correlated not only with performance in another verbal high-level ToM task (Kinderman et al., 1998), but also with a visual perspective-taking task (Samson et al., 2010). Besides validating the EmpaToM, this finding is of interest to the mentalizing literature (Apperly and Butterfill, 2009; Böckler and Zwickel, 2013) because it points towards shared mechanisms that underlie taking others' visual or cognitive perspectives (cf. Apperly and Butterfill, 2009). Interestingly, the same neural network that was involved in explicit mentalizing during the ToM questions in the EmpaToM was also observed to be active during previous watching of the ToM videos. As participants were not asked to think about the mental states of the observed others at this point, this finding may reflect spontaneous and rather implicit mentalizing. The overlap of activity during videos and questions suggests that implicit and explicit mentalizing processes may be closely related in a healthy population. During development and in clinical populations, however, implicit and explicit mentalizing capacities may dissociate (Frith and Frith, 2008; Kovács et al., 2010), a question that could be directly tested using the EmpaToM. Two important aspects should be considered before applying the EmpaToM to populations in clinical or developmental contexts. First, the ToM measure is relatively difficult as reflected in high error rates. While this makes the EmpaToM well-suited for investigations in the adult population and in plasticity research because it provides enough inter-individual variance, the task may be too difficult for younger participant samples or for clinical samples suffering from severe cognitive disabilities. Second, future research needs to specifically address the diagnostic value of the EmpaToM by investigating test parameters such as internal consistency and re-test reliability.

Having established external validaty of the EmpaToM on a behavioral and neural level, we addressed the two main questions of this study. Specifically, we aimed at investigating the relation between inter-individual differences in neural activity underlying empathy and ToM to the respective behavioral indices (question 1) and at separating the neural networks related to empathy and ToM requirements and embedding the observed peak activations in task-free resting state networks (question 2). First, concerning the brain–behavior relations of the two routes of social cognition, we asked whether the degree to which participants subjectively rate their experienced levels of empathy was specifically related to the degree of activation in neural network activated during emotional videos but not to differences in the degree of brain activation elicited during the ToM conditions, and vice versa? Indeed, results of Experiment 2 revealed clear-cut and specific brain–behavior relations: participants with higher activation in the empathy related network reported more negative affect after the emotional videos, while no relation to ToM performance was found. In contrast, enhanced activation of the ToM related network was linked to better ToM performance, but not to behavioral measures of empathy. This pattern was replicated when using composite scores of empathy and ToM performance derived from multiple tasks, which corroborates and generalizes the specificity of the brain–behavior relations of the two social capacities. This finding is highly interesting given that such brain–behavior relations, to our knowledge, have not yet been reported due to behavioral ceiling effects in most ToM fMRI paradigms in healthy adult populations. Thus, the EmpaToM is the first task allowing for the reliable assessment of individual differences in ToM in healthy adult populations on a neural and behavioral level. This extends previous findings by revealing a direct link between activity in the network that is typically reported for conditions requiring ToM and the inter-individual differences in ToM performance. Furthermore, and in contrast to the neural efficiency hypothesis, claiming that expertise in a specific task comes with reduced activation of the underlying neural network (Neubauer and Fink, 2009) and behavioral impairments relate to activation increase (Kanske et al., 2013a; Wessa et al., 2013), our results clearly show that enhanced performance in mentalizing tasks demands increased activation of the respective neural network. Taken together, the specific relation of brain and behavioral markers provides strong evidence for selective contributions of the neural networks to empathic responding on the one hand and Theory of Mind performance on the other.

Second, we aimed at investigating the separability of the neural networks related to empathy and ToM when assessed in the same individuals with a single well-controlled task. Directly contrasting empathy and ToM related brain activation patterns yielded highly specific networks that correspond to, but are much more circumscribed than the activation patterns revealed in the main contrasts for empathy and ToM, respectively. In the temporo-parietal cortex, two neighboring, but distinct peaks were identified for empathy (dorsal TPJ including SMG) on the one hand and ToM (ventral TPJ) on the other hand. This suggests that even though both empathy and ToM conditions activated the TPJ (see also (Bzdok et al., 2012), the specific regions engaged by each one of the two routes of social cognition differ. In line with this finding, a similar differentiation between TPJ and SMG had recently been observed when overcoming emotional egocentricity during empathic judgements (SMG) versus cognitive egocentricity during ToM tasks (ventral TPJ (Silani et al., 2013; Steinbeis et al., 2014). The observed activations in dorsal TPJ/SMG for empathy in the EmpaToM might therefore suggest involvement of a similar process of separating one's own emotional state from the emotional state of another when watching the emotional videos. In contrast, activation peaks in the ToM related ventral TPJ may serve similar self-other distinction in the cognitive domain (Decety and Lamm, 2007). Thus, the results suggest that both affective and cognitive understanding of others seem to rely on processes related to distinguishing one's own from others' psychological states: for empathy, the involvement of the dorsal TPJ/SMG may suggest differentiation of others' and own *affective* states and for ToM, the ventral TPJ indicates the differentiation of others' and own *cognitive* states.

Importantly, the differentiation between the two task-based functional networks was corroborated by task-free functional connectivity analyses during resting state. Thus, using the two distinct ventral and dorsal peaks in TPJ as seeds yielded distinct networks that closely resembled the task-based networks. Similar patterns emerged when exploring resting state connectivity from the neighboring peaks of empathy and ToM in posterior and anterior midline regions (precuneus/posterior cingulate and MPFC). The functional significance of the differentiation of empathy and ToM in these regions will need to be clarified in future research. Crucially, however, the findings of distinct patterns of empathic responding and ToM in behavioral markers, task-based neural activation, and task-free connectivity strongly support the existence of two separate routes to social understanding. Noticeably, such a distinction of functional networks has also been discussed in other domains, for example in the interplay of pro- and reactive cognitive control functions (Tops et al., 2014) or, relatedly, processing of internally and externally generated information (Wen et al., 2012).

While the presently observed empathy related activity resembles the reactive/externally oriented network (variably referred to as task control network, salience network, or cingulo-opercular network), the ToM related activity conforms with the proactive/internally oriented network (default mode network) (Buckner et al., 2011; Yeo et al., 2011). This overlap suggests that empathic responding may require functions ascribed to externally driven task control or salience network such as rapid detection of and reactive orienting to salient external events for immediate and adequate adaptation to others' emotional

states. Mentalizing, on the other hand, necessitates stimulus-independent process such as the generation and introspection on thoughts as well as distinguishing internally generated and externally provided information, processes ascribed to the default mode network.

While the neural networks related to empathy and ToM can be clearly distinguished, an open question is how they can influence one another. Previous research indicates, for example, that explicitly trying to take another person's perspective can modulate the degree of empathic responding (Lamm et al., 2007; Perry et al., 2010) and that inter-individual differences in perspective taking correlate with early emotion detection (Kanske et al., 2013b). Because the EmpaToM assesses and independently manipulates both processes, it could be used to expand our knowledge on the natural interplay of empathy and ToM in the future.

In conclusion, we introduce a novel behavioral and fMRI paradigm that enables the valid and independent assessment of socio-affective (empathy, compassion) and socio-cognitive processes (ToM) within an individual in a controlled but naturalistic setting in only 30 min. Future studies could furthermore utilize the EmpaToM for the assessment of metacognitive acuity by relating the confidence ratings to actual performance as well as interoceptive awareness by relating physiological arousal during the videos to self-reported subjective affect. Based on the successful validation, the EmpaToM will allow addressing questions ranging from characterizing specific impairments of social understanding in psychopathology, identifying different developmental trajectories of empathy and ToM to differential plasticity in intervention studies. A particular advantage is that five parallel sets of the EmpaToM are available, allowing for multiple testing within the same individuals.

Utilizing this paradigm, we demonstrated specific brain–behavior correlations for individual differences in empathy and ToM and provided evidence for clearly distinguishable neural networks underlying empathy and ToM in both task related and task-free fMRI. These findings are of importance to social neuroscience because they further characterize two distinct routes to understanding others and relate these to well-known resting state networks associated with internally generated processing of propositional knowledge about the world on the one hand and more externally driven saliency networks associated to emotional reactivity and regulation on the other hand. The present research, thus, lays the foundation for future understanding of how social cognition draws on and informs more general cognitive processes.

Author contributions

PK, AB, FMT, TS conceived the experiments, discussed the data and wrote the paper; PK, AB, FMT developed the task, and acquired and analyzed the data, TS fully provided for funding of the project.

Acknowledgments

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.07.082.

References

Apperly, I.A., Butterfill, S.A., 2009. Do humans have two systems to track beliefs and belief-like states? Psychol. Rev. 116, 953–970.

Böckler, A., Zwickel, J., 2013. Influences of spontaneous perspective taking on spatial and identity processing of faces. Soc. Cogn. Affect. Neurosci. 8, 735–740.

Bradley, M.M., Codispoti, M., Cuthbert, B.N., Lang, P.J., 2001. Emotion and motivation I: defensive and appetitive reactions in picture processing. Emotion 1, 276–298.

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, 219–231.

Buckner, R.L., Krienen, F.M., Castellanos, A., Diaz, J.C., Yeo, B.T., 2011. The organization of the human cerebellum estimated by intrinsic functional connectivity. J. Neurophysiol. 106, 2322–2345.

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A.R., Langner, R., Eickhoff, S.B., 2012. Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. Brain Struct. Funct. 217, 783–796.

Cardinal, R.N., Parkinson, J.A., Hall, J., Everitt, B.J., 2002. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci. Biobehav. Rev. 26, 321–352.

Chao-Gan, Y., Yu-Feng, Z., 2010. DPARSF: a MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. Front. Syst. Neurosci. 4, 13.

de Vignemont, F., Singer, T., 2006. The empathic brain: how, when and why? Trends Cogn. Sci. 10, 435–441.

Decety, J., 2010. To what extent is the experience of empathy mediated by shared neural circuits? Emot. Rev. 2, 204–207.

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, 580–593.

Diedrichsen, J., Shadmehr, R., 2005. Detecting and adjusting for artifacts in fMRI time series data. NeuroImage 27, 624–634.

Dodell-Feder, D., Koster-Hale, J., Bedny, M., Saxe, R., 2011. fMRI item analysis in a theory of mind task. NeuroImage 55, 705–712.

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, 539–548.

Engen, H.G., Singer, T., 2015. Compassion-based emotion regulation up-regulates experienced positive affect and associated neural networks. Soc. Cogn. Affect. Neurosci. **Q4**

Fan, Y., Duncan, N.W., de Greck, M., Northoff, G., 2011. Is there a core neural network in empathy? An fMRI based quantitative meta-analysis. Neurosci. Biobehav. Rev. 35, 903–911.

Figner, B., Murphy, R.O.A., 2011. Using skin conductance in judgment and decision making research. In: Schulte-Mecklenbeck, M., Kuehberger, A., Ranyard, R. (Eds.), Handbook of Process Tracing Methods for Decision Research: A Critical Review and User's Guide. Psychology Press, New York.

Fleming, S.M., Weil, R.S., Nagy, Z., Dolan, R.J., Rees, G., 2010. Relating introspective accuracy to individual differences in brain structure. Science 329, 1541–1543.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.P., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Mapp. 2, 189–210.

Frith, C., Frith, U., 2005. Theory of mind. Curr. Biol. 15, R644–R646.

Frith, C.D., Frith, U., 2008. Implicit and explicit processes in social cognition. Neuron 60, 503–510.

Hodes, R.L., Cook, E.W., Lang, P.J., 1985. Individual differences in autonomic response: conditioned association or conditioned fear? Psychophysiology 22, 545–560.

Kanske, P., Heissler, J., Schonfelder, S., Forneck, J., Wessa, M., 2013a. Neural correlates of emotional distractibility in bipolar disorder patients, unaffected relatives, and individuals with hypomanic personality. Am. J. Psychiatry 170, 1487–1496.

Kanske, P., Schonfelder, S., Wessa, M., 2013b. Emotional modulation of the attentional blink and the relation to interpersonal reactivity. Front. Hum. Neurosci. 7, 641.

Keysers, C., Gazzola, V., 2009. Expanding the mirror: vicarious activity for actions, emotions, and sensations. Curr. Opin. Neurobiol. 19, 666–671.

Kinderman, P., Dunbar, R., Bentall, R.P., 1998. Theory-of-mind deficits and causal attributions. Br. J. Psychol. 89, 191–204.

Klimecki, O.M., Leiberg, S., Lamm, C., Singer, T., 2013. Functional neural plasticity and associated changes in positive affect after compassion training. Cereb. Cortex 23, 1552–1561.

Klimecki, O.M., Leiberg, S., Ricard, M., Singer, T., 2014. Differential pattern of functional brain plasticity after compassion and empathy training. Soc. Cogn. Affect. Neurosci. 9, 873–879.

Kovács, Á.M., Téglás, E., Endress, A.D., 2010. The social sense: susceptibility to others' beliefs in human infants and adults. Science 330, 1830–1834.

Kriegeskorte, N., Lindquist, M.A., Nichols, T.E., Poldrack, R.A., Vul, E., 2010. Everything you never wanted to know about circular analysis, but were afraid to ask. J. Cereb. Blood Flow Metab. 30, 1551–1557.

Lamm, C., Batson, C.D., Decety, J., 2007. The neural substrate of human empathy: effects of perspective-taking and cognitive appraisal. J. Cogn. Neurosci. 19, 42–58.

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, 2492–2502.

Mars, R.B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., Rushworth, M.F.S., 2012. Connectivity-based subdivisions of the human right "temporoparietal junction area": evidence for different areas participating in different cortical networks. Cereb. Cortex 22, 1894–1903.

Mitchell, J.P., 2005. The false dichotomy between simulation and theory-theory: the argument's error. Trends Cogn. Sci. 9, 363–364 (author reply 364).

Mitchell, J.P., Banaji, M.R., Macrae, C.N., 2005. General and specific contributions of the medial prefrontal cortex to knowledge about mental states. NeuroImage 28, 757–762.

Neubauer, A.C., Fink, A., 2009. Intelligence and neural efficiency. Neurosci. Biobehav. Rev. 33, 1004–1023.

Perry, A., Bentin, S., Bartal, I.B., Lamm, C., Decety, J., 2010. "Feeling" the pain of those who are different from us: modulation of EEG in the mu/alpha range. Cogn. Affect. Behav. Neurosci. 10, 493–504.

Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? Behav. Brain Sci. 515–526.

Qureshi, A.W., Apperly, I.A., Samson, D., 2010. Executive function is necessary for perspective selection, not Level-1 visual perspective calculation: evidence from a dual-task study of adults. Cognition 117, 230–236.

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. J. Exp. Psychol. Hum. Percept. Perform. 36, 1255–1266.

Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". NeuroImage 19, 1835–1842.

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, 1743–1754.

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. J. Cogn. Neurosci. 19, 1354–1372.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., Perner, J., 2014. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. Neurosci. Biobehav. Rev. 42C, 9–34.

Shamay-Tsoory, S.G., Aharon-Peretz, J., 2007. Dissociable prefrontal networks for cognitive and affective theory of mind: a lesion study. Neuropsychologia 45, 3054–3067.

Silani, G., Lamm, C., Ruff, C.C., Singer, T., 2013. Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. J. Neurosci. 33, 15466–15476.

Singer, T., 2006. The neuronal basis and ontogeny of empathy and mind reading: review of literature and implications for future research. Neurosci. Biobehav. Rev. 30, 855–863.

Singer, T., 2012. The past, present and future of social neuroscience: a European perspective. NeuroImage 61, 437–449.

Singer, T., Klimecki, O.M., 2014. Empathy and compassion. Curr. Biol. 24, R875–R878.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R.J., Frith, C.D., 2004. Empathy for pain involves the affective but not sensory components of pain. Science 303, 1157–1162.

Singer, T., Kok, B.E., Bornemann, B., Bolz, M., Bochow, C.A., 2015. The ReSource Project: Background, design, samples, and measurements. Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig (in press).

Steinbeis, N., Bernhardt, B.C., Singer, T., 2014. Age-related differences in function and structure of rSMG and reduced functional connectivity with DLPFC explains heightened emotional egocentricity bias in childhood. Soc. Cogn. Affect. Neurosci. **Q5**

Tops, M., Boksem, M.A., Quirin, M., H. I.J., Koole, S.L., 2014. Internally directed cognition and mindfulness: an integrative perspective derived from predictive and reactive control systems theory. Front. Psychol. 5, 429.

Vollm, B.A., Taylor, A.N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., Deakin, J.F., Elliott, R., 2006. Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. NeuroImage 29, 90–98.

Wen, X., Yao, L., Liu, Y., Ding, M., 2012. Causal interactions in attention networks predict behavioral performance. J. Neurosci. 32, 1284–1292.

Wessa, M., Heissler, J., Schonfelder, S., Kanske, P., 2013. Goal-directed behavior under emotional distraction is preserved by enhanced task-specific activation. Soc. Cogn. Affect. Neurosci. 8, 305–312.

Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., et al., 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106, 1125–1165.