Neuroscience Letters 454 (2009) 176–181

Contents lists available at ScienceDirect

Neuroscience Letters

journal homepage: www.elsevier.com/locate/neulet

Online mentalising investigated with functional MRI

Tilo Kircher[a,*], Isabelle Blümel[c], Dominic Marjoram[d], Tineke Lataster[e], Lydia Krabbendam[e], Jochen Weber[f], Jim van Os[e], Sören Krach[a,b]

[a] Department of Psychiatry and Psychotherapy, Philipps-University Marburg, Rudolf-Bultmann-Str. 8, D-35039 Marburg, Germany
[b] Department of Psychiatry and Psychotherapy, Section of Neuroimaging, Philipps-University Marburg, Rudolf-Bultmann-Str. 8, D-35039 Marburg, Germany
[c] Department of Psychiatry and Psychotherapy, RWTH Aachen University Hospital, Pauwelsstr. 30, D-52076 Aachen, Germany
[d] Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, UK
[e] Department of Psychiatry and Neuropsychology, Maastricht University, PO BOX 616 (location DOT10), 6200 MD Maastricht, The Netherlands
[f] Social Cognitive Neuroscience Laboratory, Department of Psychology, Columbia University, 1190 Amsterdam Avenue, New York City, NY 10027, USA

ARTICLE INFO

Article history:
Received 18 July 2008
Received in revised form 9 March 2009
Accepted 10 March 2009

Keywords:
fMRI
Theory of Mind
Mentalising
Social interaction

ABSTRACT

For successful interpersonal communication, inferring intentions, goals or desires of others is highly advantageous. Increasingly, humans also interact with computers or robots. In this study, we sought to determine to what degree an interactive task, which involves receiving feedback from social partners that can be used to infer intent, engaged the medial prefrontal cortex, a region previously associated with Theory of Mind processes among others. Participants were scanned using fMRI as they played an adapted version of the Prisoner's Dilemma Game with alleged human and computer partners who were outside the scanner. The medial frontal cortex was activated when both human and computer partner were played, while the direct contrast revealed significantly stronger signal change during the human–human interaction. The results suggest a link between activity in the medial prefrontal cortex and the partner played in a mentalising task. This signal change was also present for to the computers partner. Implying agency or a will to non-human actors might be an innate human resource that could lead to an evolutionary advantage.

Humans are highly social beings and consequently are dependent on social interaction with others. For successful interpersonal communication, inferring intentions, goals or desires of others is highly advantageous. This ability has been referred to as mentalising or having a Theory of Mind (ToM) [24]. In everyday situations the ability to take the perspective of a partner one interacts with helps to prepare one's own behaviour. The neural correlates of mentalising have been investigated in recent years, using a number of different approaches [7,26,30]. Humans do interact not only with each other, but also with animals, machines and increasingly with computers and robots [11,12,17]. It is not yet completely clear whether or to which extent humans mentalise with non-human counterparts.

In commonly applied functional neuroimaging tasks investigating ToM, the participant is asked to infer the intention of various stimuli types e.g. cartoon characters [8,33], persons in a photograph [1], or even geometrical shapes chasing each other [3]. For these tasks, subjects are asked to evaluate ToM situations from an explicit point of view. In contrast, more recent imaging studies have focused on an implicit detection of ToM by using interactive games [4,6,9,16,17,20,25–27]. Here, tasks employed include variations of the Prisoner's Dilemma Game (PDG), and others such as the ultimatum [26], stone–paper–scissor [9], a coloured disc pattern game [4] or economic decision game (Iowa Gambling Task) [20].

In social psychology, the PDG has been widely applied for decades as a paradigm for investigating reciprocal altruistic vs. selfish behaviour. In this game, two players are faced with the same decision: cooperate with each other or defect. Both players may gain a previously defined sum of money depending on both their own as well as their counterpart's decision. The dilemma eventuates such that a unilateral (selfish) win of one player is maximised by defection from the cooperative counterpart, but punished bilaterally if both players defect. Hence, relying on mutual cooperation – yielding small shared earnings – comes along with the risk of being deceived. The classical real life situation involves two criminal suspects being offered reduced punishment by the prosecutor in the hope that one betrays the companion. However, if both suspects stick together (insisting on their innocence), mutual punishment would be minimal or absent. Again, the worst outcome appears whenever both suspects simultaneously defect, resulting in high mutual punishment. The PDG evokes a taking over of another's perspective and thereby implicitly measuring ToM processes. Different pay-off matrices can be selected to gear the decisions towards mutual cooperation or mutual non-cooperation [6,11,25,26].

* Corresponding author. Tel.: +49 6421 58 66219; fax: +49 6421 58 68939.
E-mail address: kircher@med.uni-marburg.de (T. Kircher).

0304-3940/$ – see front matter © 2009 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.neulet.2009.03.026

Cerebral areas most consistently associated with taking someone else's perspective, such as in the PDG, are regions located in the proximity of the medial prefrontal cortex [5,7]. Frith and Frith assume that in the context of mentalising these regions play major roles in the anticipation of what and how a partner is feeling or thinking. Furthermore, activation of these structures supposedly enables us to predict what another person is intending to do. By means of switching the perspective to another person's "view of the world" we are able to imagine how we would feel and think being in the same situation. In turn this shift permits to assess and evaluate own feelings or thoughts on a highly self-reflective level [14,15].

In the present study we were interested in the question of whether human as well as computer game partners evoke similar mentalising processes which would be signified by medial prefrontal cortex activation [4,6,25,26]. Further, is has been proposed that putative human game partners would trigger stronger mentalising associated cortical activity as humans might simply be more engaged when facing real human partners as opposed to a soulless computer opponent. To test our hypothesis we applied an adapted version of the PDG with subjects instructed to play either a putative human partner or a computer partner (while actually both were programmed to "play" a random sequence). The subjects were able to play either more cooperatively or competitively, however the pay-off matrix chosen for the present study favoured competitive behaviour. Previously, it has been shown that the medial prefrontal cortex is activated most consistently by competitive rather than cooperative behaviour [4].

A questionnaire handed out after scanning revealed that 12 out of 14 participants had been completely convinced that they had played a "real" human contender in the "human condition" and for these the "deceit" aspect was validated (see Table 1). Two participants indicated, they had seen through the cover story and were therefore discarded from later data analyses. Reaction times and averaged accumulated pay-off differences are listed in Table 1. Paired samples *t*-tests revealed that reaction times between conditions did not differ significantly (RT differences human partner vs. computer partner: $t_{11} = 1.68$; $p = .12$). Further, irrespective of the condition being played participants reached similar pay-offs (games against computer vs. games against human partner: $t_{11} = 0.75$; $p = .47$). However, participants applied a rather competitive strategy during both conditions (one sample *t*-test against 50%; condition "computer partner": $t_{11} = 3.99$; $p < .0001$; condition "human partner": $t_{11} = 6.71$; $p < .0001$).

In a debriefing session after scanning, we asked for differences in the participants' perception of the response behaviour of their game partners (i.e. putative computer or human). Although participants indicated having noticed a somewhat different strategy used by either partner, none of the participants mentioned having treated the putative game partners differently. Further, participants mentioned not having seen through their game partners' strategies (see Table 1).

Table 1
Sociodemographic and behavioural data.

	♂ = 12	
	M	SD
Biographical and behavioural data		
Age	28.0	5.5
RT (playing against computer partner) (ms)	388.2	101.7
RT (playing against human partner) (ms)	405.2	101.7
Pay-off computer (playing against computer partner) [points]	360.0	172.4
Pay-off subject (playing against computer partner) [points]	803.3	72.4
Pay-off computer (playing against human partner) [points]	476.7	133.5
Pay-off subject (playing against human partner) [points]	775.0	111.3
Questionnaire: (no, not at all = 1; yes, very much = 7)		
Did you have the impression to play against another person?	5.0	1.7
Did you succeed in detecting the human partner's strategy?	3.0	1.9
Did you succeed in detecting the computer partner's strategy?	4.0	1.4

Regarding second-level group effects, brain activity differed with respect to the partner being played. Activity modulation during the simple contrast "human partner > baseline" comprised a wide-spread network of right middle frontal, superior medial frontal and bilateral inferior parietal regions. Areas involved during "computer partner > baseline" centred around the right middle frontal gyrus extending into the inferior parietal cortex bilaterally.

Directly contrasting both experimental conditions revealed circumscribed activations of the thalamic region and the medial frontal areas only for "human partner > computer partner" (see Table 2; Fig. 1). Based on previous findings by Rilling et al. [26] we applied a ROI-analysis approach yielding a highly significant activation of the medial frontal gyrus (SVC at coordinate $x = 4$, $y = 44$, $z = 20$; $p < .004$, FWE-corr.; $k = 80$). The reversed contrast "computer partner > human partner" did not elicit any significant activation, even by applying a more liberal threshold.

In the current study, subjects played an adapted version of the prisoners dilemma game (PDG), with putative counterparts either being another human or a computer. Whilst playing the putative human as opposed to the computer opponent, stronger activation was found in the medial prefrontal cortex. However, when contrasted with low level baseline, playing both counterparts elicited medial prefrontal as well as right temporo-parietal junction (TPJ) activations. As mentalising processes have been linked to signal changes of medial frontal regions (see [5]) we hypothesize that humans attribute something akin to "intentions" to non-human counterparts, such as a computer [11,17]. The attribution of agency or will might therefore be an innate human resource and occurs independently of whether we interact with real human partners or "just" machines [17].

However, there are a number of ToM studies focussing on the TPJ as the crucial structure for mentalising processes [28,29]. As the TPJ activity, detected in the baseline contrast, is similar during games with the human and the computer partner (with right hemisphere > left hemisphere), this activity was subtracted out in the direct comparison between human > computer partner, and vice versa. It is therefore argued that the TPJ activity is only secondary with respect to the game partner being played and rather displays a somewhat general attribution of behaviour to another agent (and the analysis of the goals and outcomes of such behaviours) [2,13,18,19].

The medial prefrontal cortex activation detected in the present study was highly consistent with findings of previous functional imaging results employing implicit ToM tasks similar to ours [4,6,7,17,25,26]. In order to test our main hypothesis, we chose the local maximum activation in the medial prefrontal cortex based on the study by Rilling et al. [26] to define the centre-of-ROI for the present study. The ROI in the medial prefrontal region of the present study proved to be highly significant (corrected for multiple comparisons) and therefore clearly replicated the findings by Rilling and colleagues. This replication is even more convincing as both study designs slightly differ in terms of the interaction triggered between the participants and their anticipated game partners. Opposed to Rilling et al. who investigated single-shot interactions, the present paradigm had an online and highly interactive character.

Thus, in line with the reasoning of Rilling and colleagues we assume higher engagement in human–human interactions as opposed to games against a soulless computer, which in the following might have yielded stronger activations of cortical structures important for mentalising processes.

The anterior cingulate cortex (ACC) is an anatomically highly variant structure often without a clear differentiation from other

Table 2
ToM relevant activation peaks with their local maxima coordinates. Significance level and the size of the respective activation cluster (number of voxels) for Human > Baseline, Computer > Baseline and Human > Computer. Only clusters of at least 10 voxels are depicted (uncorrected for multiple comparisons at $p < .001$. Coordinates are listed in [32] atlas space. BA is the Brodman area nearest to the coordinate and should be considered approximate.

		BA	Coordinates			*t*-value	No. voxels
			x	*y*	*z*		
Human > Baseline							
R	**Middle Frontal Gyrus**	**6/8/10**	**32**	**19**	**−8**	**10.94**	**852**
	Superior Medial Frontal Gyrus		51	17	29	10.48	
			48	25	25	10.39	
R	**Inferior Parietal Cortex**	**7/40**	**55**	**−44**	**46**	**8.68**	**381**
	Temporo-Parietal Junction		44	−52	50	8.46	
			40	−52	43	7.46	
L	**Angular Gyrus**	**39/40**	**−28**	**−56**	**47**	**7.80**	**132**
	Superior Parietal Cortex		−40	−41	35	7.75	
			−28	−63	58	4.83	
Computer > Baseline							
R	**Middle Frontal Gyrus**	**8**	**36**	**55**	**5**	**10.72**	**419**
			48	25	39	9.64	
			51	13	25	9.19	
	Superior Frontal Gyrus (medial part)	**6/8/9**	**8**	**29**	**35**	**10.50**	**129**
L/R	**Angular Gyrus**	**39/40**	**−44**	**−41**	**39**	**9.71**	**655**
	Temporo-Parietal Junction		40	−56	47	8.25	
			40	−48	50	8.19	
R	**Inferior Frontal Gyrus (orbital part)**	**11/44/45**	**40**	**23**	**−15**	**9.67**	**64**
	Superior Temporal Pole		32	19	−4	7.93	
			12	15	−18	5.64	
L	**Middle Frontal Gyrus**	**8/10**	**−40**	**50**	**−13**	**7.96**	**78**
Human > Computer							
	Superior Frontal Gyrus (medial part)	**6/8**	**4**	**52**	**38**	**9.13**	**113**
	Anterior Cingulate Cortex		4	35	2	7.39	
			4	55	12	6.97	
	Thalamus		**12**	**−27**	**1**	**7.95**	**96**
			4	−23	1	6.15	
			−8	−20	−6	5.94	
	Olfactory Cortex		**4**	**3**	**−14**	**7.41**	**18**
			−24	−87	−23	6.03	
			−36	−79	−23	5.59	

medial prefrontal cortical structures (for a more detailed discussion the interested reader is referred to [23,34]. It is an ancient structure containing spindle cells only found in humans and other primates (pongids and hominids), suggesting that it has undergone recent evolutionary changes [21]. Patients with lesions in this area are impaired in understanding materials requiring attribution of mental states to others [5,10,31].

Decety and colleagues suggest a differentiation between tasks triggering competitive or cooperative behaviour. In their study, the orbital part of the frontal gyrus was associated with cooperation,

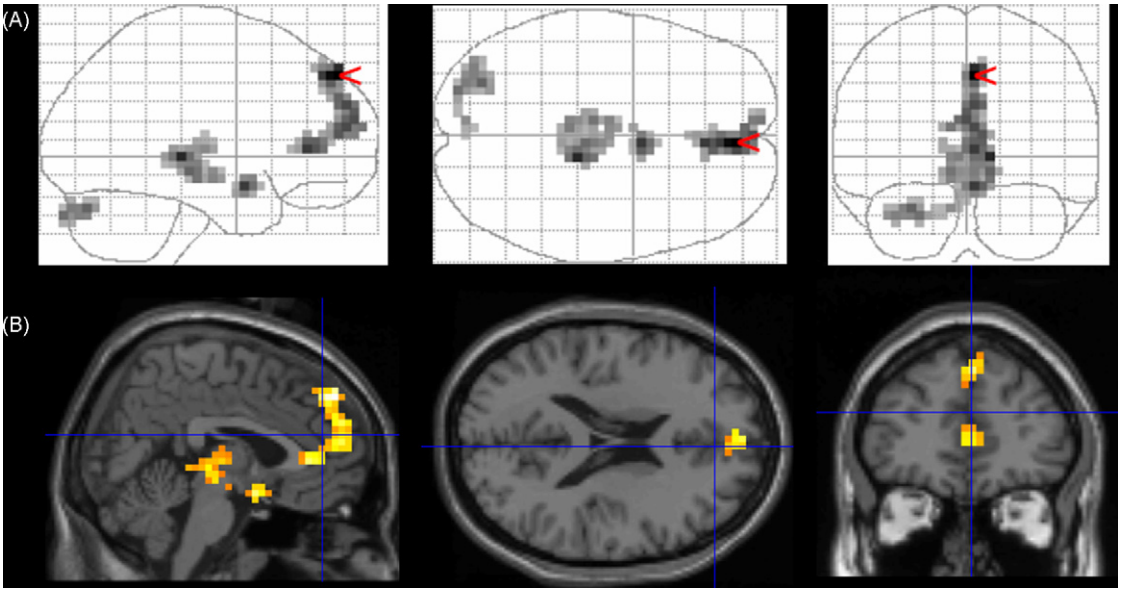

Fig. 1. (A) Human Partner > Computer Partner, cross-hair located at local maximum activation ($x = 4$, $y = 52$, $z = 38$; $p > .001$ unc.); (B) cross-hair located at coordinates $x = 4$, $y = 44$, $z = 20$ (derived from [26]). These coordinates were used as the centre-of-ROI for the SVC-analyses.

whereas the medial prefrontal cortex was activated when subjects played competitively [4]. In the current study a pay-off matrix more inclined to trigger competitive behaviour was chosen. In line with the hypothesis, strong medial prefrontal activation for both the computer and the human condition was found. For future studies, however, it would be interesting to systematically manipulate the pay-off outcomes and evaluate neural activity changes in this regard.

Activation in the thalamus was more strongly present for the human compared to the computer partner. The thalamus has also been found to be activated in other studies probing implicit mentalising [26]. The thalamus has been implicated in coordinating cortical and subcortical regions, as well as in processes where emotion and cognition interact. Therefore, this region might be important in complex processes such as mentalising that implicate a wide cortico-subcortical network.

Regarding subjective impressions, all subjects included in the analysis reported that they genuinely had the impression playing against a real human counterpart. Reaction times did not differ between the two conditions, suggesting a similar processing time. Regarding pay-off outcomes, participants appeared to play rather competitively as intended with this particular construction of design matrix. When we asked participants for their perception of the partners' response behaviour (i.e. putative computer or human), they indicated having "noticed" different strategies used by either, the computer partner or the human partner. However, these subjective perceptions were not consistent. Some participants witnessed the computer to be more cooperative, others regarded the human play as more cooperative. Notably, although different strategy usages were ascribed to each game partner, none of the participants mentioned having treated the putative game partners differently.

We chose a block design for our study with the idea of minimising set shifting processes between conditions and maximising task engagement. We could therefore maximise differential BOLD effects between conditions, which initially were hypothesised to be small.

In conclusion, we engaged subjects in a real life social reciprocal task between another human and a computer. We found medial prefrontal activation in both conditions but signal changes were significantly stronger when subjects were confronted with an alleged human partner.

We believe that attributing agency to non-human entities might be an innate human capability and occurs independently of whether we interact with real human partners or "just" machines [17].

Fourteen consenting healthy male participants with an average age of 27.4 years were recruited from within the RWTH University Hospital Aachen and were paid a fee for participation. All participants had normal or corrected-to-normal vision and were right-handed according to the Edinburgh Handedness Index [22]. Participants were excluded if they had been diagnosed with a past or present psychiatric, neurological, or medical disease. The study was approved by the local ethics committee.

Prior to scanning, participants were familiarized with the decision matrix by completing three tutorial rounds. The decision matrix resembled matrices already applied by other research groups and is considered as a variant of the PDG [6,25,26]. In short, participants were informed that if both contenders (participant vs. human partner or participant vs. computer partner) pressed the left button, both of them would receive 10 points each (CC). If the participant pressed the left button (cooperate) while the partner pressed the right button at the same time (defect), the participant would receive zero points for this game and the partner would receive 20 points (CD). In the inverse condition, the participant (defect) would gain 20 points, whilst the partner would get zero points (DC). In case both contenders chose to defect, the dilemma would eventuate with both sides receiving zero points (DD). CC implies mutual cooperation, while DD involves mutual non-cooperation [26].

The setting of the briefing was as follows: each participant was seated face-to-face with a confederate (always the same male person) with both having a commercial notebook laptop located at their side of the table. Both notebooks were linked by a connecting cable. The experimenter introduced participant and confederate and explained the upcoming task design. One condition comprised a series of nine single games (equalling one round) with the participant playing against the confederate (*human partner*). For each single game the participant had to make a decision about cooperating or defecting with the partner. Cooperation was signalled by pressing the left button (←) on the computer keyboard, defection by pressing the right button (→), respectively. During the other condition participants were instructed to play against the computer, again consisting of nine single games (*computer partner*). No hint was given about the computer's response selection. During the tutorial both conditions were presented twice in random order, interspersed by a low level baseline condition that enforced participants to alternately press the right and left button when a central cross appeared on the computer screen (see below). Furthermore, participants were confronted with two goals: on the one hand participants where enforced to win the series, while on the other hand participants had to reach a virtual highscore. As, per definition, these two converse goals could not be reached by solely pressing one button, this instruction pushed the idea: finding a decision based upon the reasoning about the opponent's last decisions ('I think that you think that I think. . .'), i.e. triggering mentalising processes. In a pre-testing (with subjects not participating in the fMRI study) involving four different winning matrices the selected scenario proved to be best in enforcing participants to vary their responses with respect to their accumulated pay-off. During the entire briefing the experimenter was standing aside the participant, "helpfully" indicating aloud at the beginning of each series which partner/condition will be encountered. By using this scenario the confederate was unofficially informed when to press the buttons (human condition) and when to relax (computer condition and baseline).

At the beginning of each series of the main experiment, participants were informed via the computer screen about the condition to be followed via the words: human, computer or baseline. Immediately after the relevant word, a central cross on the computer screen was shown that indicated the start of a series and prompted the participants to make their decision (left or right button press; see above, as in the briefing it was explained to the participants that whenever they saw a fixation cross they had make their decision via a left or right button press). The central cross disappeared after 1500 ms and was followed by an accumulated pay-off feedback for the current series (1000 ms). The accumulated pay-off feedback enabled participants to draw exact inferences about the partner's (i.e. human or computer) response selection. The participant's pay-off was indicated by the lower numbers and the partner's pay-off by the upper numbers. During the low level baseline no numeral response feedback was given. Instead two crosses replaced the numbers on the upper and lower side of the bar.

Unknowingly, participants always played against random choice "partners", never allowing participants to really cooperate or find "a best way". This deceit enables the possibility of calculating the hemodynamic changes related to differences in the instruction (human or computer partner) only, ruling out possible interaction effects of scattered strategic alliances during single participant vs. human partner interactions relative to others. Hence, the present paradigm offered the possibility to uniquely measure brain activity related to the simple supposition made by the participants about

the intentions, goals and ambitions of the partner independent of his behavioural response [9].

After the briefing the experimenter, the confederate and the participant passed on to the MR-environment after giving last instructions to the participant and verifying that participants understood the winning matrix as well as the converse requirement to both "win a series *and* reach a highscore". In the MR-scanner a condensed summary of the instructions from the earlier briefing session were projected onto MR-compatible video goggles (Resonance Technology). Participants indicated their decision (cooperation or defection) by pressing one of two buttons with index finger of their right hand which rested between both buttons on a fiberoptic custom-made response box. Prior to each series, participants were informed about the condition to be followed (human, computer or baseline). With the beginning of the functional imaging recording a randomized script file (the experiment was performed using Presentation® software; Version 10.7, www.neuro-bs.com) was started. The behavioural outcomes of each single game were recorded and saved to a log file. A series of nine games per condition completed one block. Overall, participants played ten blocks per condition (human partner, computer partner and low level baseline). After scanning participants were asked to fill out a last questionnaire about their impressions of the task and partners.

All scans were performed on a 1.5 T whole body scanner (Phillips Medical Systems, Achieva, Best, Netherlands) using standard gradients and a standard quadrature head coil. Participants lay in a supine position, while head movement was limited by foam padding within the head coil. In order to ensure optimal visual acuity, participants were offered fMRI-compatible glasses that could be fixed to the video goggles. For each participant, a series of 304 EPI-scans, lasting approximately 15 min, was acquired. Stimuli were presented in a blocked design fashion, with ten blocks per condition and a block length of nine single games.

Scans covered the whole brain, including five initial dummy scans parallel to the AC/PC line with the following parameters: number of slices (NS): 31; slice thickness (ST): 4 mm; interslice gap (IG): 4.4 mm; matrix size (MS): 64×64; field of view (FOV): 192 mm × 192 mm; repetition time (TR): 2.9 s; echo time (TE): 50 ms; flip angle (FA): 90°. For anatomical localization, we acquired high resolution images with a T1-weighted 3D FFE sequence (TR = 25 ms; TE = 4.59 ms; NS = 170 (sagital); ST = 2 mm; IG = 1 mm; FOV = 256 × 256 mm; voxel size = 1 × 1 × 2 mm).

MR images were analyzed using Statistical Parametric Mapping (SPM2, www.fil.ion.ucl.ac.uk) implemented in MATLAB 6.5 (Mathworks Inc., Sherborn, MA, USA). After discarding the first five volumes, all images were realigned to the first image to correct for head movement. Unwarping was used to correct for the interaction of susceptibility artefacts and head movement. After realignment and unwarping, the signal measured in each slice was shifted relative to the acquisition time of the middle slice using a sinc interpolation in time to correct for their different acquisition times. Volumes were then normalized into standard stereotaxic anatomical MNI-space by using the transformation matrix calculated from the first EPI-scan of each participant and the EPI-template. Afterwards, the normalized data with a resliced voxel size of $4 \times 4 \times 4$ mm were smoothed with an 8-mm FWHM isotropic Gaussian kernel to accommodate inter-participant variation in brain anatomy. The time series data were band-pass filtered to remove artefacts due to cardio-respiratory and other cyclical influences.

A general linear model (GLM) comprising three conditions (human partner, computer partner and baseline) was specified for each participant. On the first level, contrasts of main interest were human partner vs. computer partner or baseline and vice versa. An SPM2 group analysis was performed by entering these contrast images into random effects analyses using one-sample *t*-tests. The resulting group contrasts comprised computer partner > human partner, human partner > computer partner and both conditions vs. baseline. For all group analyses, we applied a voxel-wise threshold of $p < .001$. The reported voxel coordinates of activation peaks were transformed from MNI space to Talairach & Tournoux atlas space [32] by non-linear transformations (www.mrc-cbu.cam.ac.uk).

In order to control for multiple comparisons we applied a small volume correction (SVC) of the data by reducing the number of tests performed, i.e. only for voxels within this predefined region. Therefore, a sphere of 20 mm radius (at the coordinates $x = 4$, $y = 44$, $z = 20$), which we derived from a previous publication with a similar design, functioned as the region of interest [26].

Acknowledgements

The study was supported by a grant from the Interdisciplinary Centre for Clinical Research "BIOMAT" within the Faculty of Medicine at the RWTH Aachen University and BMBF project 01GW0751 "Mirror Neurons".

References

[1] S. Baron-Cohen, H.A. Ring, S. Wheelwright, E.T. Bullmore, M.J. Brammer, A. Simmons, S.C. Williams, Social intelligence in the normal and autistic brain: an fMRI study, The European Journal of Neuroscience 11 (1999) 1891–1898.
[2] S.J. Blakemore, C. Frith, Self-awareness and action, Current Opinion in Neurobiology 13 (2003) 219–224.
[3] F. Castelli, F. Happe, U. Frith, C. Frith, Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns, NeuroImage 12 (2000) 314–325.
[4] J. Decety, P.L. Jackson, J.A. Sommerville, T. Chaminade, A.N. Meltzoff, The neural bases of cooperation and competition: an fMRI investigation, NeuroImage 23 (2004) 744–751.
[5] C.D. Frith, U. Frith, The neural basis of mentalizing, Neuron 50 (2006) 531–534.
[6] H. Fukui, T. Murai, J. Shinozaki, T. Aso, H. Fukuyama, T. Hayashi, T. Hanakawa, The neural basis of social tactics: an fMRI study, NeuroImage 32 (2006) 913–920.
[7] H.L. Gallagher, C.D. Frith, Functional imaging of 'theory of mind', Trends in Cognitive Sciences 7 (2003) 77–83.
[8] H.L. Gallagher, F. Happe, N. Brunswick, P.C. Fletcher, U. Frith, C.D. Frith, Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks, Neuropsychologia 38 (2000) 11–21.
[9] H.L. Gallagher, A.I. Jack, A. Roepstorff, C.D. Frith, Imaging the intentional stance in a competitive game, NeuroImage 16 (2002) 814–821.
[10] F. Happe, H. Brownell, E. Winner, Acquired 'theory of mind' impairments following stroke, Cognition 70 (1999) 211–240.
[11] F. Hegel, S. Krach, T. Kircher, B. Wrede, G. Sagerer, Theory of mind (ToM) on robots: a functional neuroimaging study, in: Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction (HRI), 2008, pp. 335–342.
[12] F. Hegel, S. Krach, T. Kircher, B. Wrede, G. Sagerer, Understanding social robots: a user study on anthropomorphism, Robot and Human Interactive Communication (RO-MAN 2008). (2008).
[13] P.L. Jackson, J. Decety, Motor cognition: a new paradigm to study self-other interactions, Current Opinion in Neurobiology 14 (2004) 259–263.
[14] T. Kircher, A.S. David, The Self in Neuroscience and Psychiatry, Cambridge University Press, Cambridge, UK, 2003.
[15] T.T. Kircher, D.T. Leube, Self-consciousness, self-agency, and schizophrenia, Consciousness and Cognition 12 (2003) 656–669.
[16] S. Krach, I. Blümel, D. Marjoram, T. Lataster, L. Krabbendam, J. Weber, J. van Os, T. Kircher, Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI, BMC Neuroscience 10 (2009).
[17] S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, T. Kircher, Can machines think? Interaction and perspective taking with robots investigated via fMRI, PLoS ONE 3 (2008) e2597.
[18] D.T. Leube, G. Knoblich, M. Erb, W. Grodd, M. Bartels, T.T. Kircher, The neural correlates of perceiving one's own movements, NeuroImage 20 (2003) 2084–2090.
[19] D.T. Leube, G. Knoblich, M. Erb, T.T. Kircher, Observing one's hand become anarchic: an fMRI study of action identification, Consciousness and Cognition 12 (2003) 597–608.
[20] K. McCabe, D. Houser, L. Ryan, V. Smith, T. Trouard, A functional imaging study of cooperation in two-person reciprocal exchange, Proceedings of the National Academy of Sciences of the United States of America 98 (2001) 11832–11835.
[21] E.A. Nimchinsky, B.A. Vogt, J.H. Morrison, P.R. Hof, Spindle neurons of the human anterior cingulate cortex, Journal of Comparative Neurology 355 (1995) 27–37.
[22] R.C. Oldfield, The assessment and analysis of handedness: the Edinburgh inventory, Neuropsychologia 9 (1971) 97–113.

[23] D. Öngur, A. Ferry, J. Price, Architectonic subdivision of the human orbital and medial prefrontal cortex, The Journal of Comparative Neurology 460 (2003) 425–449.
[24] D. Premack, G. Woodruff, Does the chimpanzee have a theory of mind? Behavioural and Brain Science 1 (1978) 515–526.
[25] J.K. Rilling, D. Gutman, T. Zeh, G. Pagnoni, G. Berns, C. Kilts, A neural basis for social cooperation, Neuron 35 (2002) 395–405.
[26] J.K. Rilling, A.G. Sanfey, J.A. Aronson, L.E. Nystrom, J.D. Cohen, The neural correlates of theory of mind within interpersonal interactions, NeuroImage 22 (2004) 1694–1703.
[27] J.K. Rilling, A.G. Sanfey, J.A. Aronson, L.E. Nystrom, J.D. Cohen, Opposing BOLD responses to reciprocated and unreciprocated altruism in putative reward pathways, Neuroreport 15 (2004) 2539–2543.
[28] R. Saxe, N. Kanwisher, People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind", NeuroImage 19 (2003) 1835–1842.
[29] R. Saxe, A. Wexler, Making sense of another mind: the role of the right temporo-parietal junction, Neuropsychologia 43 (2005) 1391–1399.
[30] T. Singer, The neuronal basis of empathy and fairness, Novartis Foundation Symposium 278 (2007) 20–30, discussion 30–40, 89–96, 216–221.
[31] D.T. Stuss, G.G. Gallup Jr., M.P. Alexander, The frontal lobes are necessary for 'theory of mind', Brain 124 (2001) 279–286.
[32] J. Talairach, P. Tournoux, Co-planar stereotaxic atlas of the human brain, Thieme, Stuttgart, Germany (1988).
[33] K. Vogeley, P. Bussfeld, A. Newen, S. Herrmann, F. Happe, P. Falkai, W. Maier, N.J. Shah, G.R. Fink, K. Zilles, Mind reading: neural mechanisms of theory of mind and self-perspective, NeuroImage 14 (2001) 170–181.
[34] B.A. Vogt, E.A. Nimchinsky, L.J. Vogt, P.R. Hof, Human cingulate cortex: surface features, flat maps, and cytoarchitecture, Journal of Comparative Neurology 359 (1995) 490–506.