

Do We Mind Other Minds When We Mind Other Minds' Actions? A Functional Magnetic Resonance Imaging Study

Moritz F. Wurm*, D. Yves von Cramon, and Ricarda I. Schubotz

Max Planck Institute for Neurological Research, Cologne, Germany



Abstract: Action observation engages higher motor areas, possibly reflecting an internal simulation. However, actions considered odd or unusual were found to trigger additional activity in the so-called theory of mind (ToM) network, pointing to deliberations on the actor's mental states. In this functional magnetic resonance imaging study, the hypothesis was tested that an allocentric perspective on a normal action, and even more so the sight of the actor's face, suffices to evoke ToM activity. Subjects observed short videos of object manipulation filmed from either the egocentric or the allocentric perspective, the latter including the actor's face in half of the trials. On the basis of a regions of interest analysis using ToM coordinates, we found increased neural activity in several regions of the ToM network. First, perceiving actions from an allocentric compared with the egocentric perspective enhanced activity in the left temporoparietal junction (TPJ). Second, the presence of the actor's face enhanced activation in the TPJ bilaterally, the medial prefrontal cortex (mPFC) and posterior cingulate cortex (PCC). Finally, the mPFC and PCC showed increased responses when the actor changed with respect to the preceding trial. These findings were further corroborated by zmap findings for the latter two contrasts. Together, findings indicate that observation of normal everyday actions can engage ToM areas and that an allocentric perspective, seeing the actor's face and seeing a face switch, are effective triggers. *Hum Brain Mapp* 32:2141–2150, 2011. © 2011 Wiley Periodicals, Inc.

Key words: action observation; object manipulation; perspective; egocentric; allocentric; theory of mind; face perception; motor system



INTRODUCTION

Action observation usually engages the higher motor system, including particularly premotor and parietal areas. These are often referred to as “mirror neuron system”, in reference to macaque mirror neurons [Rizzolatti and Craighero, 2004]. In contrast, unexpected or contextually inconsistent

actions were also reported to activate the so-called “Theory of Mind (ToM) network”, suggesting considerations about the actor's mental states [Brass et al., 2007; German et al., 2004]. However, actions are always performed by persons to whom we also spontaneously attribute mental states. Hence, the exact conditions for the recruitment of the ToM network during action observation remain unclear.

Among 151 studies on action observation [Van Overwalle and Baetens, 2009], only 13 report ToM activity. All of these 13 studies show videos of unusual or implausible actions, but many of them also use an allocentric perspective, i.e., actions are shown from the third person perspective (3pp). Importantly, 3pp implies that the action is caused by another agent. In contrast to the first person perspective (1pp), the 3pp may be particularly conducive to this impression as it first entails that right and left hand are flipped by a 180° rotation to the opposite perspective, and second, that we usually see the actor's face. One therefore may hypothesize that a 3pp suffices to engage ToM during observation even of normal action.

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Moritz F. Wurm, Max Planck Institute for Neurological Research, Motor Cognition Group, Gleueler Street 50, 50931 Cologne, Germany. E-mail: mwurm@nf.mpg.de

Received for publication 4 January 2010; Revised 18 June 2010; Accepted 30 August 2010

DOI: 10.1002/hbm.21176

Published online 21 January 2011 in Wiley Online Library (wileyonlinelibrary.com).

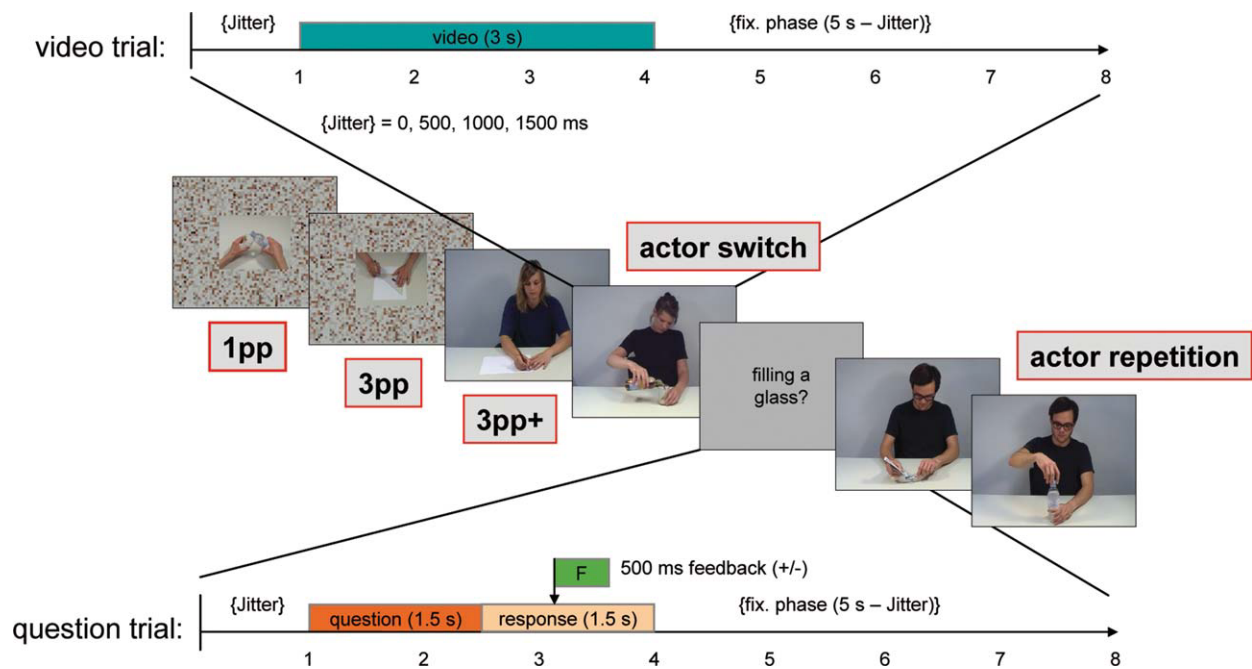


Figure 1.

Stimuli and experimental design. Video trials and question trials were interleaved in an event-related design. Video trials were composed of 3 s long video clips of everyday life actions filmed from an egocentric perspective (1pp) and two allocentric perspectives showing either hands only (3pp) or the whole actor including the actor's face (3pp+). 3pp+ trials entered a repeti-

tion suppression analysis when they were preceded by another 3pp+ trial showing either the same actor (actor repetition) or a different actor (actor switch). 20% of video trials were followed by question trials requiring participants to confirm or reject verbal action descriptions with respect to the preceding trial.

Interestingly enough, we are ignorant about the neural effects of an observer's perspective on everyday object manipulation. So far, grasping [Shmuelof and Zohary, 2008], placing [Hesse et al., 2009], and intransitive movements [Jackson et al., 2006], as well as static pictures of body parts [Chan et al., 2004; Saxe et al., 2006a] have been investigated with respect to the influence of the observer's perspective. These studies consistently report increased activation in premotor-parietal areas of the contralateral hemisphere for egocentric compared with allocentric perspective, whereas the reverse contrasts revealed activation in either occipital areas, including the lingual gyrus [Hesse et al., 2009; Jackson et al., 2006] and/or ipsilateral premotor-parietal regions [Hesse et al., 2009; Shmuelof and Zohary, 2008]. Notably, in all of these studies, the action goal was (i) known to the observer beforehand (e.g., grasping); (ii) largely invariant across the entire study (e.g., grasping whatever object is presented); and (iii) of low complexity (e.g., only grasp, then stop without further manipulation). Although these features were suitable for the purpose of the cited studies, they do not necessarily coincide with action interpretation and recognition in everyday life, where goals are often unknown, variant, and complex.

This functional magnetic resonance imaging (fMRI) study aimed to investigate whether action observation from an allocentric perspective recruits ToM, even when mentaliz-

ing is not explicitly required. To systematically disentangle the two characteristics featuring allocentric observation, i.e., perceiving body movements from a perspective other than my own, and seeing the actor's face, we conducted two analyses. First, everyday actions filmed from an allocentric perspective (3pp) were contrasted with those filmed from an egocentric perspective (1pp) to tap effects of the perspective. Second, to investigate effects due to the sight of the actor's face, we compared actions filmed from an allocentric perspective including the actor's face (3pp+) with the 3pp actions, which showed only the actor's hands. Finally, the effect of switching actors between successive movies was analyzed. To this end, 3pp+ actions performed by an actor who was new with respect to the preceding trial were compared with those performed by the same actor again.

METHODS

Participants

Twenty-one right-handed, healthy volunteers (16 female) participated in the study (age range = 22–27 years; mean age = 24.0 years). After being informed about potential risks and being screened by a physician of the institution,

subjects gave written informed consent before the fMRI measurements. The study was performed according to the Declaration of Helsinki. Data were handled anonymously.

Stimuli and Tasks

After a training phase of nine trials (six video trials and three question trials that were not used in the analysis), subjects were presented with movies showing actions (action trials) and with written action descriptions referring to these actions (question trials, Fig. 1). Each action trial (8 s) started with a movie (3 s) followed by a fixation phase. To enhance the temporal resolution of the BOLD response, a variable jitter (0, 500, 1000, or 1500 ms) was inserted before the movie. Fifty percent of the actions was performed on appropriate objects (e.g., pouring water from a bottle into a glass, altogether 60 everyday life actions) and 50% on inappropriate objects (e.g., making the same movement with a pen and a candle [cf. Schubotz and von Cramon, 2009]). However, in this study, we limit our analysis to trials performed with appropriate objects.

Subjects were instructed to attend to the presented movies. They were informed that some of the movies would be followed by an action description that either matched or did not match the content of the preceding movie and that they were to indicate whether the description matched the movie (accept) or not (reject). When a question trial was presented, subjects immediately delivered their responses on a two-button response box, using their index finger to accept and their middle finger to reject. 50% of the action descriptions matched and 50% did not. Great care was taken that the action descriptions used in the question trials referred merely to the low-level goals of actions (e.g., filling a glass), which were directly derivable from the object manipulations. Phrasings like “preparing a drink” were avoided not to refer to the intention of the actor ensuring that the task is implicit with respect to ToM.

Three experimental conditions were implemented by showing the actor’s hands from an egocentric perspective (1pp), the actor’s hands from an allocentric perspective (3pp), and the whole actor including the actor’s face from an allocentric perspective (3pp+). In addition, the trial succession was balanced such that each of the nine possible transitions occurred an equal number of times. 3pp+ trials were further analyzed by using a switching protocol: trials that were preceded by 3pp+ trials showing a different actor (actor switch) were compared with trials that were preceded by trials showing the same actor again (actor repetition).

All in all, four actresses and four actors performed a different set of actions, equally balanced with respect to the experimental conditions. 1pp videos were filmed over the shoulder of the actor resulting in an egocentric “like-me” perspective, whereas 3pp videos were filmed from the same distance and angle, but the actor sat on the other side of the table to show exactly the same detail of the hands differing solely in perspective. 3pp+ videos were filmed face to face with the actor

using a different angle and a different camera zoom factor. Actresses and actors were instructed not to show any facial expression and to focus their gaze on the objects to be manipulated. Moreover, the viewpoint of the camera did not allow for unambiguous detection of actors’ gaze direction.

The size of the hands and manipulated objects were kept constant across conditions by scaling down 1pp and 3pp videos to the same hand size as the 3pp+ videos to avoid confounding visibility effects. In addition, resized videos were placed on a scrambled background to provide an identical amount of visual information (Fig. 1, Supporting Information).

Twenty percent of the analyzed movies (i.e., 24 of 120 actions) were followed by a question trial that had the length of a regular trial (1.5 s description, 1.5 response phase, and a 5-s fixation phase). Accordingly, 120 trials (40 1pp trials, 40 3pp trials, and 40 3pp+ trials) entered the analyses contrasting 3pp > 1pp and 3pp+ > 3pp, and 11 of the 40 3pp+ trials were used for the contrast actor switch > actor repetition. Finally, 12 empty trials (fixation baseline) were presented after each second question trial.

In a post-fMRI session survey, subjects were presented with a questionnaire to measure the subjects’ ability to recognize the actors’ faces. To this end, subjects first guessed how many different actors had occurred in the video clips and subsequently were presented with 16 pictures of faces, eight of which were faces of the actresses and actors and eight were unrelated (new) faces. Subjects rated on a scale from 1 to 6 whether or not the faces occurred in the video clips. Behavioral performance in face recognition was assessed by a modified version of the Discrimination index $P(r)$, which is the difference between hit rate and false alarm rate [Snodgrass and Corwin, 1988]. The hit rate was defined as the sum of ratings of correctly recognized faces relative to the sum of the maximal rating score of all faces shown in the videos, and the false alarm rate as the sum of ratings to falsely indicated unrelated faces relative to the sum of the maximal rating score of all unrelated faces.

MRI Data Acquisition

Imaging was performed on a 3-T Siemens (München, Germany) Trio system equipped with a standard birdcage head coil. Participants were placed on the scanner bed in supine position with their right index and middle fingers positioned on the appropriate response buttons of a response box. Form-fitting cushions were used to prevent head, arm, and hand movements. Participants were provided earplugs to attenuate scanner noise. Twenty-six axial slices (192 mm field of view; 64 × 64 pixel matrix; 4 mm thickness; 1 mm spacing; in-plane resolution of 3 × 3 mm) covering the whole brain were acquired using a single-shot gradient EPI sequence (2000 ms repetition time; 30 ms echo time; 90° flip angle; 116 kHz acquisition bandwidth) sensitive to BOLD contrast. Before functional imaging, 26 anatomical T1-weighted MDEFT images [Norris, 2000; Ugurbil et al., 1993] were acquired. In a separate

session, high-resolution whole brain images were acquired from each subject to improve the localization of activation foci using a T1-weighted 3D-segmented MDEFT sequence.

MRI Data Analysis

After motion correction using rigid-body registration to the central volume, fMRI data were processed using the software package LIPSIA [Lohmann et al., 2001]. To correct for the temporal offset between the slices acquired in one image, a cubic-spline interpolation was used. Low-frequency signal changes and baseline drifts were removed using a temporal high-pass filter with a cutoff frequency of 1/99 Hz. Spatial smoothing was performed with a Gaussian filter of 5.65 mm FWHM. To align the functional data slices with a 3D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (three rotational, three translational) was performed. The rotational and translational parameters were acquired on the basis of the MDEFT and the EPI-T1 slices to achieve an optimal match between these slices and the individual 3D reference dataset. The MDEFT volume dataset with 160 slices and 1-mm slice thickness was standardized to the Talairach stereotactic space [Talairach and Tournoux, 1988]. The rotational and translational parameters were subsequently transformed by linear scaling to a standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system, thus generating output data with a spatial resolution of $3 \times 3 \times 3$ mm (27 mm^3). The statistical evaluation was based on a least-squares estimation using the general linear model for serially autocorrelated observations [Friston et al., 1995; Worsley and Friston, 1995]. The design matrix was generated with a gamma function, convolved with the hemodynamic response function and its first derivative. Brain activations were analyzed time-locked to onset of the movies, and the analyzed epoch comprised the full duration (3 s) of the presented movies, the duration of the null events (8 s), and the reaction time in action description trials (max. 3 s). The model equation, including the observation data, the design matrix, and the error term, was convolved with a Gaussian kernel of dispersion of 4 s FWHM to account for the temporal autocorrelation (Worsley and Friston, 1995). In the following, contrast images, i.e., beta value estimates of the raw-score differences between specified conditions, were generated for each participant. As all individual functional datasets were aligned to the same stereotactic reference space, the single-subject contrast images were entered into a second-level random effects analysis for each of the contrasts.

One-sample t tests were used for the group analyses across the contrast images of all subjects that indicated whether observed differences between conditions were significantly distinct from zero. The t values were subsequently transformed into Z scores.

To correct for false-positive results, in a first step, an initial voxel-wise z -threshold was set to $z = 2.576$ ($P = 0.005$). In a second step, the results were corrected for multiple comparisons using cluster-size and cluster-value thresholds obtained by Monte Carlo simulations at a significance level of $P = 0.05$, i.e., the reported activations are significantly activated at $P < 0.05$, corrected for multiple comparisons at the cluster level.

RESULTS

Behavioral Results

Performance was assessed by error rates and reaction times. Repeated-measures ANOVAs were performed for each of these measures with the levels 1pp, 3pp, and 3pp+.

Regarding reaction times, a main effect was found [$F_{(2,40)} = 10.293$, $P < 0.001$]. Paired samples t tests reflected that responses to the action descriptions were significantly slower for allocentric perspectives (mean \pm standard error; 3pp: 1102 ± 60 ms, 3pp+: 1062 ± 49 ms) compared with trials shown from an egocentric perspective [975 ± 48 ms; t_{20} (1pp-3pp) = -4.136 , $P < 0.001$; t_{20} (1pp-3pp+) = -2.984 , $P = 0.007$]. Regarding error rates, there was no significant effect [1pp: $7.2 \pm 2.7\%$, 3pp: $9.3 \pm 2.0\%$, 3pp+: $8.8 \pm 1.6\%$; $F_{(2,40)} = 26.578$, $P = 0.66$].

Face recognition was assessed by a postsession recognition test, i.e., subjects guessed the number of actors and actresses appearing in the experiment and performed a face recognition test by discriminating faces belonging to the actors and actresses, and faces of unfamiliar persons. On average, subjects spontaneously estimated that 4.7 ± 1 different actors and actresses were shown in the experiment. Subjects correctly recognized, on average, $60.2 \pm 4.9\%$ of the actors and actresses (hit rate = 0.60) and correctly rejected $86.1 \pm 5.2\%$ of the unrelated new faces (false alarm rate = 0.24). The average discrimination index (hits minus false alarms) was 0.46 ± 0.5 . A paired-samples t test showed that the discrimination index was significantly different from the chance level of 0 ($t_{20} = 7.561$, $P < 0.001$). These results indicate that, although not required for solving the task, subjects noticed that actions were performed by different actors and that they were able to remember them after the experiment.

FMRI Results

Perspective

To investigate whether allocentrically perceived actions recruit ToM in addition to the premotor-parietal network, we analyzed 3pp and 1pp trials. Compared with baseline, both perspectives revealed an extensive bilateral activation pattern of occipital, premotor-parietal, and temporal regions. However, the direct contrast 3pp > 1pp revealed an increase of neural activity in the lingual gyrus and ToM regions,

TABLE I. Anatomical specification, Brodmann area, hemisphere (R, right; L, left), Talairach coordinates (x, y, and z), and maximal Z scores (Z) of activations in perspective conditions (1pp > 3pp, and vice versa; for 1pp > 3pp corrected cluster threshold, $P < 0.05$)

Area	Brodmann area	Hemisphere	Talairach coordinates			Z
			x	y	z	
3pp > 1pp						
TPJ	39	L	-38	-48	24	2.95
		L	-47	-60	27	2.53
		R	49	-60	30	2.70
mPFC	32	R	13	36	0	2.59
PCC	31	R	1	-54	30	2.47
Lingual gyrus	17	R	1	-84	0	3.44
1pp > 3pp						
PMd	6	L	-28	-12	57	3.64
aIPS	7	L	-31	-42	54	5.26
SMG	40	L	-50	-27	36	4.02
pIPS	7/19	L	-22	-81	36	4.27
EBA/hMT	19/37	L	-43	-60	9	4.69
Cuneus	18	L	-10	-99	15	4.11
		R	17	-93	18	4.89
Cerebellum		R	16	-57	-36	3.75
		R	35	-39	-24	3.67

including posterior cingulate cortex (PCC), medial prefrontal cortex (mPFC, BA 32), and bilateral temporoparietal junction (TPJ). Because these activations did not survive the correction for multiple comparisons, we further analyzed activation sites via regions of interest (ROI) analyses using averaged coordinates of ToM belief studies (see below).

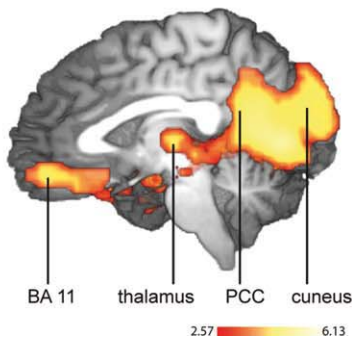
Regarding the opposite contrast, 1pp > 3pp revealed enhanced activity in the left dorsal premotor cortex (PMd), left supramarginal gyrus (SMG), and adjacent intraparietal sulcus (IPS), a left temporo-occipital region, bilateral

cuneus, and right cerebellum. According to the literature (Downing et al., 2007), we suggest that activation in the temporo-occipital region comprised the (hardly separable) extrastriate body area (EBA) and human motion area (hMT; Table I).

Sight of the Actor

Allocentric observation of actions, including faces compared with those restricted to the hands (3pp+ > 3pp),

A 3pp+ > 3pp



B actor switch > actor repetition

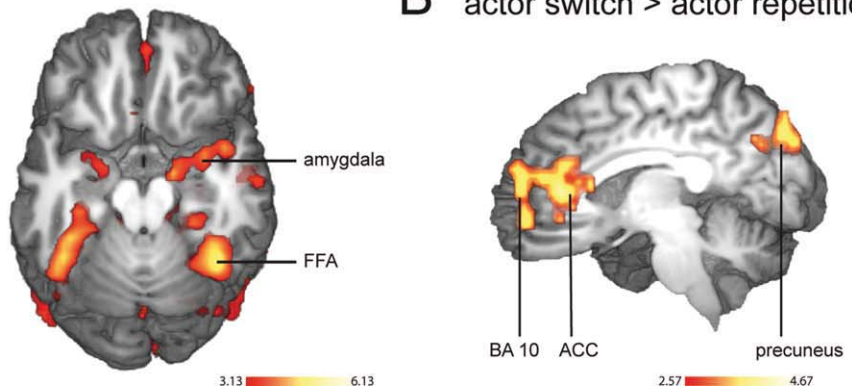


Figure 2.

Effects of the actor's face and their switches. **A**, areas activated for actions showing the upper body including the face (3pp+) compared with allocentrically perceived hands (3pp, corrected cluster threshold $P < 0.05$). For the axial view, a higher threshold ($z = 3.13$) was chosen to accentuate activation peaks. **B**, areas activated for actor switch compared with actor repetition trials (corrected cluster threshold $P < 0.05$).

TABLE II. Anatomical area, Brodmann area, hemisphere, Talairach coordinates (x, y, and z), and maximal Z scores of significant activations in face conditions (3pp+ > 3pp, actor switch > actor repetition; corrected cluster threshold $P < 0.05$)

Area	Brodmann area	Hemisphere	Talairach coordinates			Z
			x	y	z	
3pp+ > 3pp						
mPFC	11	R	2	48	-12	3.72
PCC	31	R	1	-56	33	3.33
Temporal pole	38	R	40	15	-21	4.10
aSTS	22	R	47	-12	-6	4.33
pSTS	39	L	-49	-66	18	5.24
		R	40	-63	18	5.64
IFG	45	R	47	30	3	4.96
FFA	37	L	-37	-48	-12	4.80
		R	34	-48	-9	5.44
Cuneus	31	L	-7	-69	6	6.13
Amygdala		L	-19	-6	-15	3.64
		R	14	-9	-12	4.46
Thalamus, Pulvinar		L	-19	-27	3	5.22
		R	20	-27	0	4.97
Actor switch > actor repetition						
Precuneus	7		2	-78	39	4.67
ACC	33		8	33	12	4.36
mPFC	10		-5	54	-3	3.33
			5	51	18	3.50
Actor repetition > actor switch						
pIPS	7	L	-28	-54	51	4.43
		R	14	-57	60	4.80
aIPS	7	L	-53	-33	48	4.63
		R	50	-30	48	4.61
Postcentral gyrus	40	L	-58	-21	27	3.21
EBA/hMT	19	L	-43	-66	-3	4.07
		R	44	-51	-9	4.25

yielded extended bilateral activation in the fusiform and parahippocampal gyrus, inferior temporal gyrus, right anterior and posterior superior temporal sulcus (STS), right temporal pole, EBA/hMT, bilateral amygdala, pulvinar nucleus of thalamus (PLV), mPFC (BA 11), cuneus extending into PCC and retrosplenial cortex, and right inferior frontal gyrus (IFG, BA 45) (Fig. 2A; Table II). Maximum activation within the region of the fusiform gyrus amounted to the coordinates reported for the fusiform face area (FFA) [Kanwisher et al., 1997; Spiridon et al., 2006].

Actor Switch versus Actor Repetition

Observing a different actor than in the preceding trial (actor switch) yielded enhanced activation in medial frontal areas (BA 10 and pregenual as well as subgenual ACC) and posterior precuneus. In contrast, repetition of the same actor increased activity in bilateral inferior postcentral gyrus and anterior IPS as well as in EBA/hMT (Fig. 2B; Table II).

ROI Analysis

To discuss the differential activation of the ToM network, we averaged the coordinates of 14 studies reported in Van Overwalle and Baetens [2009] using ToM belief tasks for the definition of ToM-ROIs [Abraham et al., 2008; Ferstl and von Cramon, 2002; Gallagher et al., 2000; Gobbin et al., 2007; Hynes et al., 2006; Kobayashi et al., 2007; Mitchell, 2008; Perner et al., 2006; Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Saxe et al., 2006b; Sommer et al., 2007; Vogeley et al., 2001; Wakusawa et al., 2007]: left TPJ: -51, -59, 26; right TPJ: 54, -49, 22; PCC: -1, -55, 33; mPFC: -3, 50, 20. Mean beta values were extracted from the averaged coordinate voxel plus six adjacent voxels of the following conditions versus rest: 1pp, 3pp, 3pp+, actor switch, and actor repetition. Mean beta values are shown in Figure 3. With regard to the effect of perspective (3pp > 1pp), the activation of the left TPJ was significantly higher in 3pp-R compared with 1pp-R ($F_{(1,20)} = -2.072$, $P = 0.026$; paired samples t test). The presence of the actor's face (3pp+ > 3pp) revealed significant

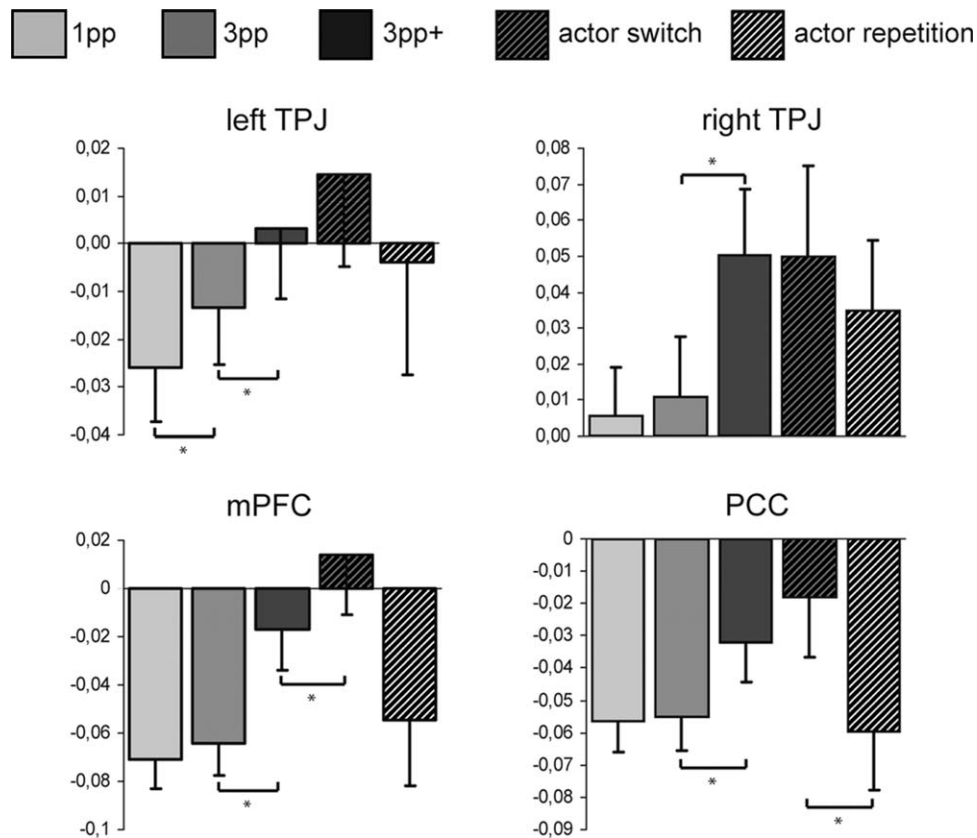


Figure 3.

ROI analysis in ToM regions defined by averaging coordinates of 14 ToM belief studies reported in Van Overwalle and Baetens (2009). Mean beta values were extracted from the contrasts 1pp > rest, 3pp > rest, 3pp+ > rest, actor switch > rest, and actor repetition > rest. Error bars indicate standard error of mean.

enhancement in all ToM regions (left TPJ [$F_{(1,20)} = -2.046$, $P = 0.028$], right TPJ [$F_{(1,20)} = -5.567$, $P < 0.001$], PCC [$F_{(1,20)} = -4.023$, $P < 0.001$], mPFC [$F_{(1,20)} = -2.227$, $P = 0.019$]; paired-samples t tests). Finally, the activation of ToM regions was enhanced for actor switches compared with actor repetitions, yielding significant effects in the PCC ($F_{(1,20)} = 2.446$, $P = 0.012$; paired-samples t test), and the mPFC ($F_{(1,20)} = 2.286$, $P = 0.017$; paired samples t test).

DISCUSSION

This fMRI study investigated whether perspective on the action and the visibility on the actor's face modulates an observer's mentalizing (ToM) network. An independent ROI analysis using coordinates extracted from classic ToM task studies yielded significant effects (i) in the left TPJ for 3pp > 1pp; (ii) in the TPJ bilaterally, the PCC and the mPFC for 3pp+ > 3pp; and (iii) in the mPFC and PCC for actor switch > actor repetition. Together, findings indicate that parts of the ToM network are differentially enhanced

during observation of normal everyday actions (a) when we see the action from the third person perspective, (b) when we can see the actor's face, and (c) when we see a new actor. These findings uncover the ToM network as being intimately involved in the perceptual analysis of ordinary action.

Observing Actions From the Allocentric Perspective (3pp)

Actions of others are typically perceived from an allocentric but rarely from an egocentric perspective, which is usually associated with own actions. Following our assumption that action observation uses not only the motor system but also draws on the ToM network, e.g., by considering the mental states of the actor, ToM regions were expected to be stronger activated for the allocentric perspective. Indeed, we found activations in regions associated with ToM for 3pp compared with 1pp, including the TPJ bilaterally, the mPFC, and the PCC. However,

these activations fell below a conservative statistical threshold, which may be indicative of spontaneously triggered mentalizing that is unconstrained and not explicitly required by the task. A ROI analysis using the mean of peak coordinates obtained by classic ToM tasks revealed a significant effect in the left TPJ (Fig. 3), thus we will limit our discussion of ToM effects to this region.

Among the ToM regions, TPJ is suggested to play a particular role in perspective taking [Ruby and Decety, 2001, 2004; Vogeley et al., 2004] and is considered to reflect that we mentally put ourselves in someone else's shoes [Abraham et al., 2008]. Recent findings suggest a role of the TPJ in control of shared motor representations to keep apart self- and other-caused actions [Brass et al., 2009]. TPJ was also reported in the context of visuospatial reorienting [Corbetta and Shulman, 2002]. The linkage between mentalizing and attentional reorienting implied by TPJ as the common node of the related networks is still puzzling, but both functions seem reconcilable [Corbetta et al., 2008]. Along these lines, TPJ activation in this study may reflect a visuospatial transformation during perspective taking to match the actor's spatial orientation.

Interestingly, activation of the TPJ was not found in other studies using 1pp and 3pp conditions in their experimental designs. Their foci of interest were the effect of perspective on static hands and body parts [Chan et al., 2004; Saxe et al., 2006a], and modulation of the premotor-parietal network during observation of placing [Hesse et al., 2009], grasping [Shmuelof and Zohary, 2008], and intransitive movements [Jackson et al., 2006]. All studies report a common overlap of activation in contralateral somatosensory and motor areas for 1pp, which is also consistent with our results. Regarding 3pp, increased activation was found in occipital regions [cuneus, lateral occipital [Chan et al., 2004], lingual gyrus [Hesse et al., 2009]] and/or in ipsilateral motor areas (right superior parietal lobe [Hesse et al., 2009; Shmuelof and Zohary, 2008], right precentral gyrus, and right EBA [Hesse et al., 2009]). Besides activation of the lingual gyrus, which is also activated in our study and might be explained by increased effort with regard to the analysis of the visual input, the differential activation of the motor system on the one hand, and the TPJ on the other, can be plausibly explained by two crucial aspects. First, grasping and placing can be performed with both hands equally well. Thus, corresponding motor representations are probably not lateralized to one hemisphere. Second, actions directed to a target object in space (often referred to as goal-directed actions [Bekkering et al., 2000]) are suggested to be remapped to the effector that would most efficiently replicate the action toward the relevant object so that a transformation into an observer-congruent reference frame is not required [cf. Shmuelof and Zohary, 2008]. In accordance, goal-directed actions observed from an allocentric perspective tend to be imitated in a mirrored fashion [Bekkering et al., 2000; Wohlschlagel et al., 2003], in line with the right-hemispheric activation found for grasping

and placing observation from 3pp. In contrast, bimanual object manipulation, which was used in this study, explicitly reveals the actor's handedness. Thus, actually taking the perspective of the actor may be a more suitable strategy for the analysis of bimanual object manipulation.

Observing and Recalling Actors

In a next step, we tested whether the opportunity to perceive not only the action itself but also the actor's face is a potential trigger for ToM activity as it provides person information. Contrasting 3pp+ with 3pp yielded extended activations in occipital and temporal areas as well as in the mPFC (BA 11) and in the right IFG (BA 45). The ROI analysis revealed significant enhancement in bilateral TPJ, mPFC (BA 10), and PCC (Fig. 3).

Other occipital and temporal activations were simply due to the sight of additional body parts (EBA, [Taylor et al., 2007]) and faces (pSTS [Allison et al., 2000; Puce and Perrett, 2003] and FFA [Kanwisher et al., 1997; Spiridon et al., 2006]). Coactivation of FFA with the amygdala and the PLV reflected components of the amygdalo-fusiform pathway [Smith et al., 2009] and suggests that our subjects dealt with the actors' faces, although this was not explicitly required by the task [Kouider et al., 2009; Pasley et al., 2004]. The right lateralized activation of the IFG, temporal pole, and aSTS supports this notion because these areas are suggested to belong to the so-called "extended" system for face perception [Barbeau et al., 2008; Ishai et al., 2005]. Moreover, temporal pole and aSTS were also found to play a role in ToM [Gallagher and Frith 2003], as well as mPFC and PCC. Interestingly, several studies on face recognition report activation of the precuneus, PCC, and mPFC when comparing recognition of familiar faces with recognition of unfamiliar faces [Gobbini et al., 2004; Leibenluft et al., 2004; Trinkler et al., 2009]. The precuneus and PCC are related to retrieval from long-term memory, whereas the mPFC was suggested for encoding information about personality traits of a familiar individual [Gobbini et al., 2004; Leibenluft et al., 2004]. It is possible that activation in the PCC was due to acquisition of visual familiarity over the course of the experiment (Gobbini and Haxby, 2006; Kosaka et al., 2003; Trinkler et al., 2009). Our post-session survey supports the assumption that the subjects acquired visual familiarity with the actors. Similarly, the mPFC may reflect the attempt of trait inference, as was reported for familiarity with faces [Gobbini et al., 2004; Trinkler et al., 2009]. The association of information about behaviour with faces was found to increase activity in the mPFC and pSTS [Todorov, 2007].

These findings and their interpretation were further substantiated by the analysis of trials with switched actors compared with trials with actors that were already presented in the preceding trial ($n-1$). Perceiving a new actor yielded enhanced activity in the mPFC, more precisely BA 10 and the ACC, as well as in the posterior precuneus.

Using ToM ROI coordinates of these regions yielded significant effects as well (Fig. 3).

Trinkler et al. [2009] reported activation of the retrosplenial cortex for the acquisition of visual familiarity and of the precuneus for the retrieval of personal knowledge. Thus, switching to a foreseen actor may trigger a recall of the visual image of the actor and episodic memory about the actions associated with this particular actor (e.g., this actor has squeezed an orange and sharpened a pencil before). Recalling characteristics of an actor could also include the retrieval of idiosyncrasies of movements or facial expressions, e.g., a person looks tired, ambivalent, or nervous, contributing to the formation of an actor-related knowledge, as reflected by activation of the mPFC. Moreover, the mPFC is often associated with the attempt to understand the reasons for a particular action [Van Overwalle, 2009]. In our study, actions were selected so that they did not imply some kind of long-term goal beyond or across several trials. However, coming up with a potential coherent global goal achieved by several single actions may be a spontaneous tendency during action observation. In contrast, the recollection of knowledge from remote trials showing the same actor is not required in actor repetitions and would be attenuated in comparison to actor switch trials.

CONCLUSIONS

Present findings suggest that perceiving actions from an allocentric perspective evokes mental inferences even when not required by the task. The particular selection of activated ToM areas reflected that the observers engaged in perspective taking as well as in the formation and retrieval of actor-related familiarity and knowledge. The latter aspect is especially elicited by both augmenting the “hands manipulating objects” scenario with the face of the actor, as well as by switching actors with regard to the preceding trial. Results indicate that the ToM network is intimately involved in the perception of ordinary actions.

REFERENCES

Abraham A, Werning M, Rakoczy H, von Cramon DY, Schubotz RI (2008): Minds, persons, and space: An fMRI investigation into the relational complexity of higher-order intentionality. *Conscious Cogn* 17:438–450.

Allison T, Puce A, McCarthy G (2000): Social perception from visual cues: Role of the STS region. *Trends Cogn Sci* 4:267–278.

Barbeau EJ, Taylor MJ, Regis J, Marquis P, Chauvel P, Liegeois-Chauvel C (2008): Spatiotemporal dynamics of face recognition. *Cereb Cortex* 18:997–1009.

Bekkering H, Wohlschlagel A, Gattis M (2000): Imitation of gestures in children is goal-directed. *Q J Exp Psychol A* 53:153–64.

Brass M, Ruby P, Spengler S (2009): Inhibition of imitative behaviour and social cognition. *Philos Trans R Soc Lond B Biol Sci* 364:2359–2367.

Brass M, Schmitt RM, Spengler S, Gergely G (2007): Investigating action understanding: Inferential processes versus action simulation. *Curr Biol* 17:2117–2121.

Chan AW, Peelen MV, Downing PE (2004): The effect of viewpoint on body representation in the extrastriate body area. *Neuroreport* 15:2407–2410.

Corbetta M, Patel G, Shulman GL (2008): The reorienting system of the human brain: From environment to theory of mind. *Neuron* 58:306–324.

Corbetta M, Shulman GL (2002): Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215.

Downing PE, Wiggett AJ, Peelen MV (2007): Functional magnetic resonance imaging investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. *J Neurosci* 27:226–233.

Ferstl EC, von Cramon DY (2002): What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? *Neuroimage* 17:1599–1612.

Gallagher HL, Frith CD (2003): Functional imaging of ‘theory of mind’. *Trends Cogn Sci* 7:77–83.

Gallagher HL, Happe F, Brunswick N, Fletcher PC, Frith U, Frith CD (2000): Reading the mind in cartoons and stories: An fMRI study of ‘theory of mind’ in verbal and nonverbal tasks. *Neuropsychologia* 38:11–21.

German TP, Niehaus JL, Roarty MP, Giesbrecht B, Miller MB (2004): Neural correlates of detecting pretense: Automatic engagement of the intentional stance under covert conditions. *J Cogn Neurosci* 16:1805–1817.

Gobbini MI, Haxby JV (2006): Neural response to the visual familiarity of faces. *Brain Res Bull* 71:76–82.

Gobbini MI, Koralek AC, Bryan RE, Montgomery KJ, Haxby JV (2007): Two takes on the social brain: A comparison of theory of mind tasks. *J Cogn Neurosci* 19:1803–1814.

Gobbini MI, Leibenluft E, Santiago N, Haxby JV (2004): Social and emotional attachment in the neural representation of faces. *Neuroimage* 22:1628–1635.

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS (1995): Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* 2:189–210.

Hesse MD, Sparing R, Fink GR (2009): End or means—The “what” and “how” of observed intentional actions. *J Cogn Neurosci* 21:776–790.

Hynes CA, Baird AA, Grafton ST (2006): Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia* 44:374–383.

Ishai A, Schmidt CF, Boesiger P (2005): Face perception is mediated by a distributed cortical network. *Brain Res Bull* 67:87–93.

Jackson PL, Meltzoff AN, Decety J (2006): Neural circuits involved in imitation and perspective-taking. *Neuroimage* 31:429–439.

Kanwisher N, McDermott J, Chun MM (1997): The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.

Kobayashi C, Glover GH, Temple E (2007): Children’s and adults’ neural bases of verbal and nonverbal ‘theory of mind’. *Neuropsychologia* 45:1522–1532.

Kosaka H, Omori M, Iidaka T, Murata T, Shimoyama T, Okada T, Sadato N, Yonekura Y, Wada Y (2003): Neural substrates participating in acquisition of facial familiarity: An fMRI study. *Neuroimage* 20:1734–1742.

- Kouider S, Eger E, Dolan R, Henson RN (2009): Activity in face-responsive brain regions is modulated by invisible, attended faces: Evidence from masked priming. *Cereb Cortex* 19:13–23.
- Leibenluft E, Gobbini MI, Harrison T, Haxby JV (2004): Mothers' neural activation in response to pictures of their children and other children. *Biol Psychiatry* 56:225–232.
- Lohmann G, Muller K, Bosch V, Mentzel H, Hessler S, Chen L, Zysset S, von Cramon DY (2001): LIPSIA—a new software system for the evaluation of functional magnetic resonance images of the human brain. *Comput Med Imaging Graph* 25(6):449–457.
- Mitchell JP (2008): Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb Cortex* 18:262–271.
- Norris DG (2000): Reduced power multislice MDEFT imaging. *J Magn Reson Imaging* 11(4):445–451.
- Pasley BN, Mayes LC, Schultz RT (2004): Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron* 42:163–172.
- Perner J, Aichhorn M, Kronbichler M, Staffen W, Ladurner G (2006): Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Soc Neurosci* 1:245–258.
- Puce A, Perrett D (2003): Electrophysiology and brain imaging of biological motion. *Philos Trans R Soc Lond B Biol Sci* 358:435–445.
- Rizzolatti G, Craighero L (2004): The mirror-neuron system. *Annu Rev Neurosci* 27:169–192.
- Ruby P, Decety J (2001): Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nat Neurosci* 4:546–550.
- Ruby P, Decety J (2004): How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *J Cogn Neurosci* 16:988–999.
- Saxe R, Jamal N, Powell L (2006a): My body or yours? The effect of visual perspective on cortical body representations. *Cereb Cortex* 16:178–182.
- 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.
- Saxe R, Powell LJ (2006): It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychol Sci* 17:692–699.
- Saxe R, Schulz LE, Jiang YV (2006b): Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Soc Neurosci* 1:284–298.
- Schubotz RI, von Cramon DY (2009): The case of pretense: Observing actions and inferring goals. *J Cogn Neurosci* 21: 642–653
- Shmuelof L, Zohary E (2008): Mirror-image representation of action in the anterior parietal cortex. *Nat Neurosci* 11:1267–1269.
- Smith CD, Lori NF, Akbudak E, Sorar E, Gultepe E, Shimony JS, McKinstry RC, Conturo TE (2009): MRI diffusion tensor tracking of a new amygdalo-fusiform and hippocampo-fusiform pathway system in humans. *J Magn Reson Imaging* 29: 1248–1261.
- Snodgrass JG, Corwin J (1988): Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *J Exp Psychol Gen* 117:34–50.
- Sommer M, Dohnel K, Sodian B, Meinhardt J, Thoermer C, Hajak G (2007): Neural correlates of true and false belief reasoning. *Neuroimage* 35:1378–1384.
- Spiridon M, Fischl B, Kanwisher N (2006): Location and spatial profile of category-specific regions in human extrastriate cortex. *Hum Brain Mapp* 27:77–89.
- Talairach J, Tournoux P (1988): Co-planar Stereotaxic Atlas of the Human Brain. New York: Thieme.
- Taylor JC, Wiggett AJ, Downing PE (2007): Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *J Neurophysiol* 98:1626–1633.
- Todorov A, Gobbini MI, Evans KK, Haxby JV (2007): Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia* 45(1):163–173.
- Trinkler I, King JA, Doeller CF, Rugg MD, Burgess N (2009): Neural bases of autobiographical support for episodic recollection of faces. *Hippocampus* 19:718–730.
- Van Overwalle F (2009): Social cognition and the brain: A meta-analysis. *Hum Brain Mapp* 30:829–858.
- Van Overwalle F, Baetens K (2009): Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *Neuroimage* 48:564–584.
- Vogeley K, Bussfeld P, Newen A, Herrmann S, Happe F, Falkai P, Maier W, Shah NJ, Fink GR, Zilles K (2001): Mind reading: Neural mechanisms of theory of mind and self-perspective. *Neuroimage* 14(1 Pt 1):170–181.
- Vogeley K, May M, Ritzl A, Falkai P, Zilles K, Fink GR (2004): Neural correlates of first-person perspective as one constituent of human self-consciousness. *J Cogn Neurosci* 16:817–827.
- Ugurbil K, Garwood M, Ellermann J, Hendrich K, Hinke R, Hu X, Kim SG, Menon R, Merkle H, Ogawa S and others (1993): Imaging at high magnetic fields: initial experiences at 4 T. *Magn Reson Q* 9(4):259–277.
- Wakusawa K, Sugiura M, Sassa Y, Jeong H, Horie K, Sato S, Yokoyama H, Tsuchiya S, Inuma K, Kawashima R (2007): Comprehension of implicit meanings in social situations involving irony: A functional MRI study. *Neuroimage* 37:1417–1426.
- Wohlschlager A, Gattis M, Bekkering H (2003): Action generation and action perception in imitation: An instance of the ideomotor principle. *Philos Trans R Soc Lond B Biol Sci* 358:501–515.
- Worsley KJ, Friston KJ (1995): Analysis of fMRI time-series revisited—again. *Neuroimage* 2(3):173–181.