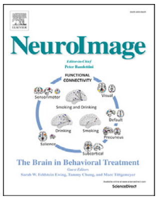
NeuroImage 165 (2018) 48–55

Contents lists available at ScienceDirect

NeuroImage

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



The role of the temporoparietal junction (TPJ) in action observation: Agent detection rather than visuospatial transformation

Moritz F. Wurm[a,b,*], Ricarda I. Schubotz[c,d]

[a] University of Trento, Center for Mind/Brain Sciences, Rovereto, TN, Italy
[b] Harvard University, Department of Psychology, Cambridge, MA, USA
[c] University Hospital of Cologne, Department of Neurology, Cologne, Germany
[d] University of Münster, Institute of Psychology, Münster, Germany

ARTICLE INFO

Keywords:
Action perception
Perspective taking
Motor simulation
Theory of mind
Temporoparietal junction

ABSTRACT

Recognizing and understanding the actions of others is usually coupled with perceiving someone else's body movements from a third person perspective (3pp) whereas we perceive our own actions from a first person perspective (1pp). From a neural viewpoint, a recent finding is that perceiving actions from a 3pp as compared to a 1pp activates the temporoparietal junction, a brain region associated with visuospatial transformation and perspective taking but also with mental state inference and Theory of Mind (ToM). The present fMRI study characterizes the response profile of TPJ to elucidate its role in action observation. Participants observed naturalistic and pixelized object-directed actions from a 3pp and 1pp. Critically, in the pixelized condition the action goal could only be inferred from the movement kinematics. Both left and right TPJ revealed an interaction: Neural activity in TPJ was enhanced for 3pp vs. 1pp actions in the naturalistic but not pixelized condition. This finding contradicts theories proposing that TPJ is generally involved in transforming the action into the observer's perspective to match perceived body movements with visuomotor representations in the observer's motor system, which would be particularly required when actions can only be inferred from movement kinematics. Instead, our results support the theory that perceptual 3pp-selective cues trigger ToM-related processes such as detection of other agents and reasoning about an action's underlying mental states.

Introduction

We typically observe actions of others from a third person perspective (3pp) whereas we see our own actions from a first person perspective (1pp). While we are aware of our intentions that drive our own actions, the intentions and other mental states of others are typically hidden and have to be inferred from the observed action and contextual cues (Jacob and Jeannerod, 2005; Kilner et al., 2007; Wurm and Schubotz, 2012). There is hence a tight relation between 3pp action observation and the attribution of mental states to others (Oosterhof et al., 2012). In a recent study (Wurm et al., 2011), we found that observing someone else's actions from a 3pp as compared to a 1pp increased the neural activity in the temporoparietal junction (TPJ) (see Ruby and Decety, 2001; Jeannerod and Anquetil, 2008 for related findings). TPJ is part of the so-called Theory-of-Mind (ToM) network, which shows enhanced activity during mental state attribution and other ToM-related tasks (Saxe and Kanwisher, 2003; Frith and Frith, 2006; Abraham et al., 2008). The finding of enhanced neural activity associated with the observation of actions observed from a 3pp relative to the 1pp suggests that certain perceptual cues specific of others' actions have the potential to induce ToM-related brain activity (ToM hypothesis). According to this interpretation, TPJ could be involved in detecting “other person” signals in observed actions to trigger the attribution of agency (Frith and Frith, 2003) and/or mental states (Van Overwalle, 2009) to observed entities. From a related viewpoint, TPJ might be involved in differentiating (visual) cues of own and other person's body parts, which in turn may provide the basis for various aspects of self processing (Blanke and Arzy, 2005) such as self-other discrimination (Jeannerod and Anquetil, 2008; Brass et al., 2009).

An alternative interpretation of increased TPJ activation for 3pp vs. 1pp actions is that TPJ is involved in transforming body posture and movements parameters of an observed action into the observer's frame of reference. This interpretation is inspired by motor theories of action recognition, which propose that action recognition relies on activation of an observer's own sensorimotor representations corresponding to the

* Corresponding author. Center for Mind/Brain Sciences (CIMeC), University of Trento, Corso Bettini 31, 38068, Rovereto, TN, Italy.
E-mail address: moritz.f.wurm@gmail.com (M.F. Wurm).

https://doi.org/10.1016/j.neuroimage.2017.09.064
Received 17 July 2017; Accepted 29 September 2017
Available online 3 October 2017

observed action (Rizzolatti and Craighero, 2004; Rizzolatti et al., 2014). One critical presumption of motor theories is that activation of sensorimotor representations builds on the perceptual similarity of the perceived action with the corresponding sensorimotor representation “as if I would perform the action myself”. However, we typically perceive the actions of others from an angle that precludes a direct overlap with the view on our own body parts and actions. Proponents of motor theories therefore postulated that action recognition requires a visuospatial transformation to match the perceived body movements with the observer's egocentric frame of reference, e.g., if the acting person is facing me the own perspective is mentally rotated by 180° into the other person's perspective (Jackson et al., 2006; Jeannerod, 2007). Visuospatial transformation and perspective taking tasks typically activate the posterior parietal cortex (Zacks, 2008) and the TPJ (Schurz et al., 2013). It was therefore proposed that TPJ is the critical neural substrate for transforming observed action parameters to enable a match with the egocentric space of the observer (Jackson et al., 2006). Taken together, following a motor account of action recognition, visuospatial transformation of actions perceived from a 3pp is necessary to enable a motor mapping and thus to decode the observed action, and this transformation takes place in the TPJ (visuospatial transformation hypothesis).

In summary, activation of TPJ for 3pp action observation seems to be compatible with both ToM and visuospatial transformation hypotheses. The present fMRI study aimed at characterizing the role of the TPJ in action observation with the particular goal to test the ToM and visuospatial transformation hypotheses against each other. Participants had to recognize object-directed actions that were shown from either a 3pp or 1pp and in an either naturalistic or pixelized fashion. Critically, pixelized actions could hardly be recognized based on object information and therefore should rely to a greater extent on the analysis of coarse movement kinematics, which were largely preserved in the videos. By contrast, recognition of naturalistic actions did not have to rely on the analysis of movement kinematics, as the involved objects were strongly indicative of the actions (e.g., orange + orange squeezer = squeezing orange). Critically, this manipulation of stimulus type emphasized different aspects of the actions: the pixelized actions emphasized the kinematics of the action whereas the naturalistic actions emphasized perceptual cues indicative of another person. Thereby, the design allowed formulating opposing predictions of the visuospatial transformation and ToM hypotheses, respectively: Following the visuospatial transformation hypothesis, a visuospatial transformation of actions from 3 PP to 1 PP should particularly support action recognition in the pixelized condition in which movement kinematics were the most critical source of information for action recognition. Naturalistic actions should less strictly require a spatial transformation as action recognition was supported by object information. Hence, if TPJ serves the visuospatial transformation of actions, then the 3pp vs. 1pp effect in TPJ should be stronger for pixelized compared to naturalistic actions. Following the ToM hypothesis, on the other hand, ToM-related processes like the detection of „other person“ signals should be particularly triggered by the perception of 3pp action cues in the naturalistic condition, which provides more perceptual cues that convey the difference between 3pp and 1pp than the perceptually impoverished pixelized actions. Hence, if TPJ serves the detection of perceptual cues that are indicative of other persons' actions, then the 3pp vs. 1pp effect in TPJ should be stronger for naturalistic compared to pixelized actions. Higher-level ToM functions such as the inference about someone else's mental states might less depend on the naturalness of the 3pp cues. In that case one would expect the 3pp vs. 1pp effect in TPJ to be equally high for naturalistic and pixelized actions.

Methods

Participants

Eighteen healthy adults (10 females, 22–28 years, mean age = 25 years) volunteered to participate in the experiment. All participants were right-handed according to the Edinburgh Inventory Manual Preference (Oldfield, 1971), had normal or corrected-to-normal vision, and had no history of neurological or psychiatric disease. One participant was excluded because of low behavioral performance (error rate of 40% in the action recognition task, which exceded the group mean by more than two standard deviations). Participants gave written informed consent prior to participation in the study. The study was approved by the Ethics Committee of the Medical Faculty the University of Cologne, Germany.

Stimuli

Stimuli consisted of 3s long videos of 40 bimanual object manipulations (e.g., cutting bread with a knife, opening a tin with a tin opener, etc.). We used a 2 × 2 factorial design (Fig. 1): The actions were shown from a 1pp or 3pp (factor PERSPECTIVE) in a naturalistic or pixelized fashion (factor STIMULUS TYPE). The actions were filmed from a top view perpendicular to the table at which the actions took place. This perspective was chosen to create both 1pp and 3pp with the same video material. We thereby avoided perceptual differences between the 1pp and 3pp conditions resulting from differences in visibility of the objects that were otherwise differentially occluded by the actresses' hands. The videos were filmed from the 1pp; the 3pp conditions were created by rotating the 1pp videos by 180° (Fig. 1). The pixelized conditions were created by averaging the grey values of pixels in 24*24 grid squares for each frame of the videos. The pixelization resolution was chosen so that the objects were not identifiable in any of the static frames of the video (as broadly tested in an object naming experiment using the first frame of the videos), i.e., in the absence of movement information (Wurm and Schubotz, 2017). Thus, after pixelization, the objects themselves were hard or impossible to identify whereas the action could still be inferred from movement kinematics (e.g. wrist transformations, horizontal and vertical arm trajectories, etc.). Videos had a presentation rate of 25 frames per second and a display width and height of 720*576 pixels. The number of actions was chosen to match the number of actions per condition in Wurm et al. (2011), resulting in 40 * 4 = 160 action trials in total (as compared to 80 trials analyzed in the contrast 3pp vs. 1pp in Wurm et al., 2011).

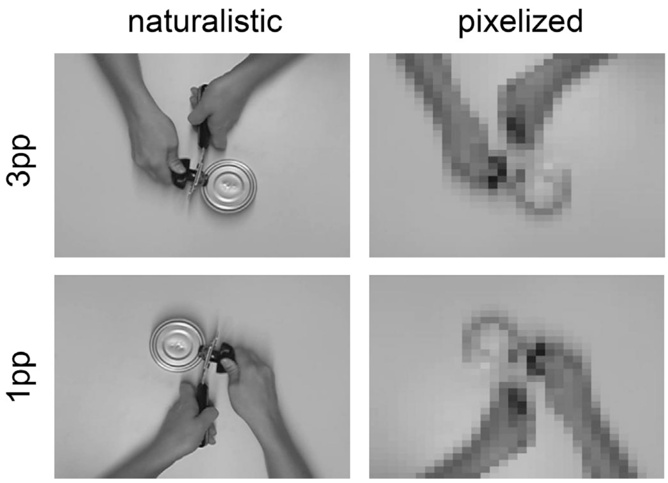


Fig. 1. (1-column fitting image). Experimental design. Participants observed 3-s long videos of object-directed actions shown from a first- or third-person perspective (1pp and 3pp, respectively; factor PERSPECTIVE). The actions were presented naturalistically or pixelized (factor STIMULUS TYPE). In the naturalistic conditions, object information supported action recognition whereas in the pixelized conditions, action recognition relied on coarse movement kinematics that remained largely preserved after pixelization.

Task

Participants were instructed to recognize the presented actions. They were informed that some of the videos were followed by a short verbal action description (e.g. "opening bottle"; question trials hereafter) that matched the preceding action (50%) or not (50%). The participants responded to question trials by button press with the right index finger (confirm) or middle finger (reject).

fMRI design

Video trials (40 per condition) were presented in an event-related design intermixed with 20 null events (6 s fixation cross) and 40 question trials. Video and question trials started with the presentation of the action video or question (3000 ms) followed by a fixation cross (2250 ms + a variable jitter of 0, 500, 1000, or 1500 ms; average trials length = 6 s). The jitter order was randomized. The order of the four action conditions, question trials, and null events was first order counterbalanced. Because each action was shown four times in the experiment (once per condition), we sought to eliminate systematic habituation effects by balancing the order of conditions per action across participants. The experiment started with 6 s fixation period and ended with an 8 s fixation period and had a total length of 22 min and 16 s.

MRI data acquisition

Imaging was carried out on a 3-T Siemens Magnetom Trio system (Siemens, Erlangen, Germany) equipped with a Tx/Rx birdcage head coil. Participants were placed on the scanner bed in a supine position with their right index and middle fingers positioned on the appropriate response buttons of a response box. Form-fitting cushions were utilized to prevent head, arm, and hand movements. Participants were provided earplugs in order to attenuate scanner noise. Stimulus presentation and response recoding were performed using Presentation 12.0 (Neurobehavioral Systems, San Francisco, CA, USA). Stimuli were back projected onto a screen (Optostim, Medres, Cologne, Germany; 32-inch, 1280 × 800 pixels) at the end of the scanner bore. Participants saw the screen on a mirror mounted to the head coil and adjusted individually to allow for comfortable view of the screen. For each volume, twenty-six axial slices (192 mm field of view (FOV); 64 × 64 pixel matrix; 4 mm thickness; 1 mm spacing; in-plane resolution of 3 × 3 mm) positioned parallel to the bicomissural plane (AC-PC) covering the whole brain were acquired using a single-shot gradient echo planar imaging (EPI) sequence (2000 ms repetition time (TR); 30 ms echo time (TE); 90° flip angle (FA); 116 kHz acquisition bandwidth) sensitive to BOLD contrast. The experiment was preceded by an unrelated experiment with a duration of 21 min and 4 s. In total, 1300 volumes (627 + 668 + 5 volumes between the two experiments) + 3 dummy volumes were acquired in a single run (duration of acquisition = 43 min and 20 s). Prior to functional imaging, 26 anatomical T1-weighted 2D-FLASH MDEFT images (Norris, 2000; Ugurbil et al., 1993) were acquired in the same spatial orientation as the functional slices (192 mm FOV, 128 × 128 pixel matrix, 4 mm thickness, 1 mm spacing, in-plane resolution 3 × 3 mm). 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 3-D-segmented MDEFT sequence (2250 ms TR, 3.93 TE, 256 mm FOV, FA 9°, thickness 1 mm, 0.5 mm spacing, in-plane resolution 1 × 1 mm, 160 sagittal slices).

MRI data analysis

After motion correction using rigid-body registration to the central volume (Siemens motion protocol PACE), fMRI data were processed using the software package LIPSIA 1.5.0 (Lohmann et al., 2001). Functional data slices were aligned with a 3-D stereotactic coordinate reference system. To this end, matching parameters (six degrees of freedom; three rotational, three translational) of the T1-weighted 2D-FLASH images onto the high-resolution 3D MDEFT reference set were computed. The resulting transformation matrix for rigid spatial co-registration was standardized to the Talairach stereotactic space (Talairach and Tournoux, 1988) by linear scaling. This normalized transformation matrix was then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system. The generated output had a spatial resolution of 3 × 3 × 3 mm (27 mm^3). To correct for the temporal offset between the slices acquired in one image, a cubic-spline interpolation was employed. Low-frequency signal changes and baseline drifts were removed using a temporal high-pass filter with a cut-off frequency of 1/90 Hz. Spatial smoothing was performed with a Gaussian filter of 8 mm FWHM. 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 contained predictors of the four action conditions, question trials, and null events. For each predictor, box-car functions were generated and convolved with a hemodynamic response function (Glover, 1999). Action trials were modeled as epochs lasting from video onset to offset (3s), question trials were modeled as epochs lasting from question onset to time point of button press (i.e., epoch duration = reaction time, max. 3s), and null events were modeled as epochs lasting from fixation onset to offset (6s). The resulting reference time courses were used to fit the signal time courses of each voxel. 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). For each participant, maps of beta value estimates of experimental conditions and of contrasts between specified conditions were generated and entered into second-level random effects analyses.

Condition-specific beta estimates were used for a ROI analysis. The ROI analysis was conducted to test the specific hypotheses on the role of the TPJ regions identified in Wurm et al. (2011). In addition, the display of each condition's beta allowed a better assessment of the nature of putative interactions and their directions. ROIs were based on the peak coordinates of the clusters found by Wurm et al. (2011) for the contrast for 3pp vs. 1pp (Talairach coordinates of left TPJ: 47 -60 27, right TPJ: 49–60 30). For each ROI and participant, beta values were extracted from the four experimental conditions using spherical ROIs (6 mm radius) centered on the peak voxel. For each ROI and participant, beta value estimates were averaged and resulting mean beta estimates were entered into statistical analyses (repeated measures ANOVA and paired samples t tests).

Whole brain analyses using the contrasts between conditions were used to identify effects in other brain regions not specified in the hypotheses. In addition, the whole brain analysis allowed supporting and refining the results of the ROI analysis by providing additional information about the extent and peak location of putative TPJ clusters. Contrasts were computed between (1) 3pp vs. 1pp for both pixelized and natural conditions separately and (2) pixelized vs. natural for both 3pp vs. 1pp conditions separately. (3) A directed interaction contrast was computed using the contrast vector [1 -1 -1 1] (3pp_nat, 1pp_nat, 3pp_pix, 1pp_pix). Single-subject contrast images were entered into a second-level random effects analysis for each of the contrasts. One-sample t tests were employed for the group analyses across the whole-brain 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 the whole-brain statistical maps for false-positive results, in a first step, an initial voxelwise 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 1000 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 (Lohmann et al., 2008).

Conjunctions were calculated by extracting the minimum Z value of the two input contrasts for each voxel (Nichols et al., 2005).

Statistical maps were projected on a cortical surface for visualization using BrainVoyager QX 2.6 (BrainInnovation).

Results

Behavioral results

Mean error rates and reaction times on responses to question trials are shown in Table 1. A repeated measures ANOVA on error rates with the factors STIMULUS TYPE and PERSPECTIVE revealed a main effect of STIMULUS TYPE ($F(1,16) = 9.9$; $p = 0.006$). This shows that participants made significantly more errors on question trials following pixelized compared to naturalistic actions. This was expected, as the pixelized actions were more difficult to recognize than the naturalistic actions. The error rates were below chance performance (0.5) demonstrating that subjects paid attention to the actions and that in all four conditions the actions were recognizable. No main effect of PERSPECTIVE and no interaction were observed (both $p > 0.2$). A repeated measures ANOVA on reaction times revealed no significant effects (all $p > 0.1$).

ROI analysis in left and right TPJ

To test whether the neural activity in TPJ in response to observed actions is modulated by stimulus type and perspective we performed a ROI analysis in left and right TPJ. A repeated measures 2 × 2 × 2 ANOVA with the factors STIMULUS TYPE (pixelized, naturalistic), PERSPECTIVE (3pp, 1pp), and HEMISPHERE (left, right TPJ) revealed an interaction of STIMULUS TYPE and PERSPECTIVE ($F(1,16) = 10.77$; $p = 0.005$): TPJ activation was higher for 3pp than for 1pp actions when actions were naturalistically presented but not when they were pixelized (Fig. 2A). In addition, we found a weak main effect of STIMULUS TYPE ($F(1,16) = 4.56$; $p = 0.049$) indicating that naturalistic actions activated TPJ to a stronger degree than pixelized actions. Finally, the analysis revealed a three-way interaction of STIMULUS TYPE, PERSPECTIVE, and HEMISPHERE ($F(1,16) = 7.329$; $p = 0.016$) indicating that the interaction of STIMULUS TYPE and PERSPECTIVE was more pronounced in the left than in the right hemisphere. Notably, the pattern of the interaction in left and right TPJ was qualitatively different from the pattern of behavioral effects (ER, RT), which suggests that the interaction effects in TPJ are unlikely to be due to differences in task difficulty.

Paired samples two-tailed t-tests revealed significant activation differences between 3pp and 1pp actions for the naturalistic conditions (left: $t(16) = 2.72$, $p = 0.015$; right: $t(16) = 2.37$, $p = 0.031$) but not for the pixelized conditions (left: $t(16) = -2.05$, $p = 0.057$; right: $t(16) = -1.49$, $p = 0.157$). Likewise, the activation differences between naturalistic and pixelized actions were significant only for 3pp (left: $t(16) = 4.07$, $p = 0.001$; right: $t(16) = 2.80$, $p = 0.013$) but not 1pp conditions (left: $t(16) = -1.69$, $p = 0.111$; right: $t(16) = -1.33$, $p = 0.201$).

Whole-brain analysis

A subsequent whole-brain analysis aimed at further specifying and supporting the results of the ROI analysis and to identify potential other regions showing an interaction of stimulus type and perspective. The interaction map revealed, in line with the results of the ROI analysis, significant clusters in left and right TPJ (Fig. 2B. Table 2). The left cluster comprised the ventral anterior part of the angular gyrus, the posterior part of the supramarginal gyrus (SMG), and the posterior superior temporal sulcus (pSTS). The right cluster was slightly more posterior and comprised most parts of the angular gyrus and pSTS. In addition, we found an interaction effect in the medial prefrontal cortex (at the section between BA 9 and 10), which, together with TPJ, is considered to be a part of the ToM network. No interactions in the opposite direction (increased activation of 1pp vs. 3pp in naturalistic but not pixelized actions or increased activation for pixelized vs. naturalistic in 3pp but not 1pp actions) were observed.

Table 1
Behavioral responses to question trials.

	Naturalistic, 1pp	Naturalistic, 3pp	Pixelized, 1pp	Pixelized, 3pp
ER	0.076 (0.020)	0.076 (0.026)	0.176 (0.043)	0.129 (0.033)
RT (in ms)	993 (46)	963 (49)	1 021 (44)	979 (45)

Mean error rates (ER) and reaction times (RT) of the responses to question trials. SEM in parentheses.

Furthermore, we investigated the effects of stimulus type and perspective independent of each other. To identify specific effects of stimulus type irrespective of perspective we computed the conjunction of pixelized vs. naturalistic actions seen from the 3pp and pixelized vs. naturalistic actions seen from the 1pp ((pixelized vs. naturalistic 3pp) ∩ (pixelized vs. naturalistic 1pp)). We found stronger neural responses for pixelized vs. naturalistic actions in the left (and with a more liberal threshold for multiple comparison correction also right) inferior frontal gyrus (IFG) extending posteriorly into the ventral premotor cortex (PMv) and ventrally into the anterior operculum (Fig. 3, Table 2). This activation might reflect increased efforts in action recognition, possibly selectively a particular sub-process of action recognition as only the IFG/PMv, but not other parts of the so-called action observation network such as inferior parietal and occipitotemporal cortex (Caspers et al., 2010), showed increased neural activation. A likely candidate process important for action recognition in our study is the retrieval and selection of semantic action and object information (Badre and Wagner, 2007; Binder and Desai, 2011) as this information was depleted in the pixelized actions.

For the reverse contrast, i.e., naturalistic vs. pixelized actions, we found stronger neural responses in bilateral lateral occipital gyrus (LOG) and bilateral ventral postcentral gyrus (PoCG) extending posteriorly into the anterior part of the SMG. Activation of LOG likely reflects enhanced mid-level visual processing, e.g. of contour and shape of hands and objects (Kourtzi and Kanwisher, 2000, 2001), as clear-cut object information was perceptually accessible in the naturalistic but not in the pixelized actions. A possible interpretation of the enhanced activation of PoCG is that this region is involved in processing specific and fine-grained perceptual details of the action that were present in the naturalistic but not in the pixelized actions. Note, however, that decoding studies suggest PoCG to also code perceptually invariant action information that generalizes across various perceptual features, such as kinematics (Oosterhof et al., 2010; Wurm and Lingnau, 2015) and involved objects (Wurm and Lingnau, 2015; Wurm et al., 2016), and even across concrete action subtypes (Leshinskaya and Caramazza, 2015).

To identify specific effects of perspective irrespective of stimulus type we computed the conjunction of naturalistic actions seen from 3pp vs. 1pp and pixelized actions seen from the 3pp vs. 1pp ((naturalistic 3pp vs. 1pp) ∩ (pixelized 3pp vs. 1pp)). We found stronger neural responses in the ventral occipital pole corresponding to early visual cortex (EVC) for 3pp vs. 1pp actions and stronger neural responses in dorsal EVC for the reverse contrast (Fig. 3, Table 2). These effects can be explained by visual differences in the upper and lower parts of the stimuli of the 3pp and 1pp conditions: In the 3pp videos, the arms were more visible in the upper half of the visual field, which therefore contained more visual information to be processed by the ventral EVC (DeYoe et al., 1996). By contrast, in 3pp videos the arms were visible in the lower half of the visual field, which increased activation of the dorsal EVC.

Discussion

This study aimed at characterizing the neural response profile of the temporoparietal junction (TPJ) with regard to observed actions presented from a 3pp or 1pp and in a naturalistic or pixelized fashion. Critically, the pixelized actions emphasized the action kinematics

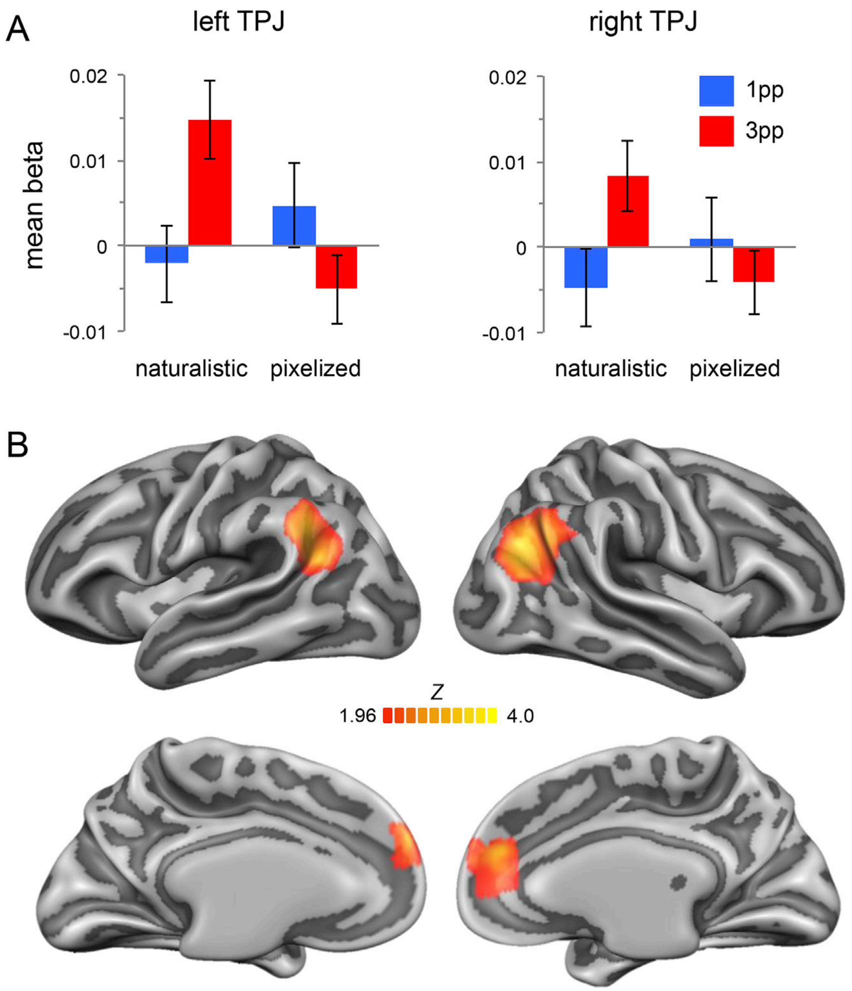


Fig. 2. (1-column fitting image). Interaction of perspective and stimulus type. (A) Average BOLD responses (beta coefficients) for each action condition in left and right temporoparietal junction (TPJ). Error bars indicate standard errors of the mean. (B) Whole-brain interaction map (thresholded at $z = 2.576$, corrected cluster threshold $p = 0.05$).

Table 2
Brain areas showing significant effects of stimulus type (pixelized, naturalistic), perspective (3pp, pp), and the interaction of stimulus type and perspective.

	cluster		local maxima			
	mean Z	mm3	Z	x	y	z
pixelized vs. naturalistic						
right IFG (BA 47)	2.91	3996	3.69	28	28	6
right IFG (BA 45)			2.77	40	21	24
right PMv			3.12	34	6	30
naturalistic vs. pixelized						
right PoCG	3.11	4320	3.95	58	−17	30
left PoCG	2.86	1836	3.43	−59	−17	33
right LOG	3.22	1890	4.5	31	−86	0
left LOG	3.17	2214	4.42	−35	−86	−3
3pp vs. 1pp						
right lingual gyrus	2.93	1485	3.46	1	−86	−3
1pp vs. 3pp						
left cuneus	3.38	7668	4.91	−11	−95	15
right cuneus			4.66	13	−93	2
interaction						
right TPJ	3.19	3699	3.51	52	−47	33
left TPJ	3.17	7128	3.55	−50	−56	36
left mPFC	3.15	3186	3.38	−5	43	21

Peak and cluster Z-values, cluster size, and peak Talairach coordinates of identified clusters in the whole-brain analysis ($p = 0.05$ corrected at the cluster level, initial voxelwise threshold $p = 0.005$). Abbreviations: IFG, inferior frontal gyrus; LOG, lateral occipital gyrus; mPFC, medial prefrontal cortex; PMv, ventral premotor cortex; PoCG, postcentral gyrus; TPJ, temporoparietal junction.

whereas the naturalistic actions more strongly conveyed perceptual person cues. Both ROI and whole-brain analyses revealed a significant interaction: Activity in TPJ correlated more strongly with actions seen from a 3pp but only when they were presented naturalistically. When actions were pixelized, no significant difference between 3pp and 1pp was observed.

This result is incompatible with the visuospatial transformation hypothesis, which claims that TPJ's role in action recognition is to take the perspective of the acting person to allow a matching of the observed movement kinematics with the observer's visuomotor representations of the action. Following the logic of this hypothesis, recruitment of this function would be particularly needed when the object is pixelized and hence barely informative for action recognition. In this case action recognition would primarily rely on the spatial transformation of the observed movement kinematics to enable a mapping with the observer's own action repertoire.

Instead, the enhanced activation difference between 3pp and 1pp for naturalistic actions supports the ToM hypothesis, which claims that the perception of 3pp information functions as trigger for ToM-related activity in TPJ. Following this hypothesis, perceiving actions from a 3pp should trigger ToM-related processes to a stronger degree when cues that are indicative of a 3pp are fully visible but not, or less so, when they are degraded by pixelizing the actions. The ToM hypothesis is further corroborated by a similar interaction effect in mPFC, which is associated with ToM-related functions as well. Notably, our findings do not preclude the possibility that in certain situations taking the perspective of another person is beneficial for cognitive processes related to action

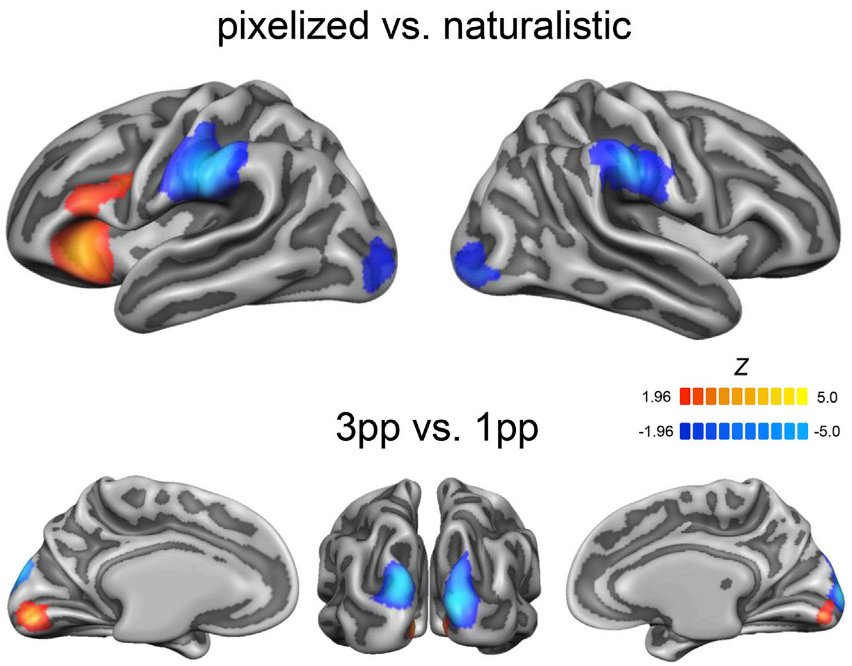


Fig. 3. (1-column fitting image). Effects of stimulus type (pixelized vs. naturalistic) and perspective (3pp vs. 1pp). For isolating the effect of stimulus type independent of perspective we computed the conjunction (3pp pixelized vs. naturalistic) ∩ (1pp pixelized vs. naturalistic). For isolating the effect of perspective independent of stimulus type we computed the conjunction (pixelized 3pp vs. 1pp) ∩ (naturalistic 3pp vs. 1pp). Maps are thresholded at $z = 2.576$, corrected cluster threshold $p = 0.05$.

understanding, e.g., to estimate which objects are in the other person's visual field and/or reach of grasp, which may help predicting forthcoming action steps.

Two questions arise from our results: First, what are the critical cues of naturalistic 3pp actions that trigger activity in TPJ? Second, what exactly is TPJ doing with these cues?

TPJ modulation by perspective: is action information needed?

What exact information in the 3pp action videos is constitutive for TPJ activation? As the 3pp actions were horizontally flipped versions of the 1pp action videos, 3pp and 1pp actions contained the same amount of information in terms of low-level visibility of hands and objects. Therefore, the critical information selective for the 3pp videos seem to be the “not me” view on body parts (hands shown upside down). Crucially, this information was present – in a degraded way – in the pixelized action videos too. The significantly lower TPJ activation in the pixelized conditions suggests that a critical receptive feature for the perspective-dependent TPJ sensitivity is that the 3pp cues are naturalistic and particularized.

Does TPJ activation depend on a 3pp view on body parts per se? Or is further information required, e.g., body part movements or meaningful, possibly object-involving actions? Studies using static images of hand postures perceived from 3pp vs. 1pp found stronger neural responses for 3pp vs. 1pp body views in the right extrastriate body area (EBA), which is in the wider vicinity of TPJ (Chan et al., 2004; Saxe et al., 2006). However, no significant TPJ (and EBA) modulation was reported in whole brain analyses of these studies suggesting that 3pp body part information alone does not substantially activate TPJ. Thus, body information alone may not provide hints about possible mental states, and at least one other trigger in addition to a 3pp may be needed to activate TPJ. Actions likely function as such trigger as they are strong indicators of underlying intentions. However, it is noteworthy that some of the fMRI studies contrasting actions shown from 3pp vs. 1pp did not find significant TPJ activation for 3pp actions (Jackson et al., 2006; Shmuelof and Zohary, 2008; Hesse et al., 2009). Although the absence of effects is generally difficult to interpret, one might speculate that differences in stimulus design and action complexity are potential reasons for the difference between these studies on the one hand and the present study and Wurm et al. (2011) on the other. For example, it is possible that observation of real life object manipulations triggers the inference of underlying intentions to a stronger degree than observation of meaningless intransitive body part movements (Jackson et al., 2006) or reaching toward abstract objects (Hesse et al., 2009). Likewise, observers might engage in intention inference to a stronger degree when observing many different actions compared to repetitive observation of few similar actions. However, without directly studying the interaction between perspective and action complexity, this argument will remain speculative. In addition to action stimuli, other sources of information, e.g. facial/emotional expressions, may provide equivalent information about another person's mental states such as feelings and desires (Wurm et al., 2011).

It is unclear to which degree the information about perspective has to be a perceptual cue of another person's body to modulate TPJ activation. Our findings suggest that concrete body part information is needed in the context of action observation but it should be pointed out that also conjugated action verbs presented in 3pp vs. 1pp (scrive = s/he writes vs. scrivo = I write) increases activation in pSTS, which is located in vicinity to TPJ (Papeo and Lingnau, 2015). This suggests that not only perceptual 3pp body information in combination with action information but also other types of perspective information, such as action verbs conjugated in 3pp, are likely to be effective triggers of TPJ activity.

The puzzling role of TPJ: “someone else” or “not me”?

The perspective-dependent TPJ modulation in the naturalistic but not in the pixelized condition argues against a role of TPJ in visuospatial transformation of action parameters but rather suggests a ToM-related function. What specific function could TPJ have in 3pp action observation?

One possibility is that TPJ is involved in higher-level inference of action-associated mental states such as intentions, desires, preferences, and emotions of the acting agent. As we typically infer the mental states of others by others' behavior from a 3pp, the perception of “someone else” cues might automatically trigger such inferential processes. Indeed, TPJ is activated during inference of intentions (Van Overwalle, 2009; Van Overwalle and Baetens, 2009) and other mental states like preferences (David et al., 2008). An argument against this interpretation in the present study is that one would expect that pixelized 3pp actions trigger mental state inference at least to some extent. In that case TPJ activity related to pixelized 3pp actions should be weaker than activity in the natural 3pp actions but still higher than in the two 1pp conditions. Instead, we found that TPJ activity for pixelized 3pp actions was at the same level as for the two 1pp conditions. However, it may be that TPJ does not encode mental states per se but rather provides associated functions that are required either for attributing mental states to other persons (Van Overwalle, 2009) or to discriminate the mental states of others from own mental states.

Regarding the latter option, there is robust evidence that TPJ is involved in self-other discrimination (Blanke and Arzy, 2005; Jeannerod, 2007; Brass et al., 2009). Following this line, 3pp-specific cues might help to differentiate one's own action intentions from intentions triggered by observed actions. As only the naturalistic actions provided clear-cut perceptual cues of another person's body a self-other distinction would only be required for the naturalistic but not the pixelized actions. On the other hand, one would actually predict the opposing effect: because the 1pp action looks “like me” there is increased effort in self-other discrimination, possibly in addition to a mismatch between perceived and own movement signals, which should result in increased neural activity for 1pp vs. 3pp. This is not what we observed. Another related explanation refers to integratory capacities of TPJ/pSTS in visual perception in general (Huberle and Karnath, 2012; Pollmann et al., 2014) and action recognition in particular (Giese and Poggio, 2003). Thus, TPJ might be involved in the integration of different visual features of observed actions, such as the acting agent's body parts and movements, involved objects, and their locations in space. Such a „features of others“ integration could then subserve higher level ToM functions. However, it is not entirely clear why this integration would be selectively recruited for action cues perceived from a 3pp and absent for cues that resemble the view on the own body and movements.

A more parsimonious interpretation might therefore be that TPJ is involved in attributing mental states or, more generally, intentionality to others (Van Overwalle, 2009) or in detecting cues indicative of other intentional agents in a stimulus (Gao et al., 2012; Lee et al., 2014). From this perspective, TPJ neurons tuned to the detection of other agents, which is strongly correlated with the naturalistic 3pp view of acting persons, could provide a behavioral advantage to prepare for quick reactions. Agent detection might have become a more general function that applies also to non-visual contexts (Abraham et al., 2008; Papeo and Lingnau, 2015). Detecting agents and ascribing intentionality to other entities likely involves the discrimination from and relation to one's own mental states. This interpretation fits with the co-modulation of mPFC, which is activated when one's own motor plans conflict with observed behavior of others (Brass et al., 2009). Naturalistic 3pp action cues hence may trigger higher-level ToM functions in mPFC such as self-referential and/or relational processing that are critical for counterfactual reasoning (e.g., “she, not me” or “there, not here”) (Schubotz, 2011).

Conclusion

TPJ activity during action observation is enhanced for actions perceived from a 3pp and when detailed cues of the acting person's body are present. This finding narrows down possible functional properties of action-sensitive neurons in TPJ: In the context of action observation, TPJ neurons respond to fine-grained information of acting persons, but not to coarse action kinematics, perceived from a 3pp but not from a 1pp. This finding contradicts the view that TPJ is involved in a spatial transformation of action movement parameters in general but supports the hypothesis that TPJ is involved in ToM-related processes such as detecting cues that are specific of another agent and his or her intentional activities. Even if in some situations increased sensitivity and automaticity of agency detection might result in false alarms, detection of such cues could be an evolutionary adaption that potentially guarantees survival.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgements

We would like to thank Liuba Papeo for helpful comments on the manuscript and Ilona Schneider for assistance in stimulus preparation.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.neuroimage.2017.09.064.

References

Abraham, A., Werning, M., Rakoczy, H., von Cramon, D.Y., Schubotz, R.I., 2008. Minds, persons, and space: an fMRI investigation into the relational complexity of higher-order intentionality. Conscious Cogn. 17, 438–450.

Badre, D., Wagner, A.D., 2007. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia 45, 2883–2901.

Binder, J.R., Desai, R.H., 2011. The neurobiology of semantic memory. Trends Cogn. Sci. 15, 527–536.

Blanke, O., Arzy, S., 2005. The out-of-body experience: disturbed self-processing at the temporo-parietal junction. Neuroscientist 11, 16–24.

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.

Caspers, S., Zilles, K., Laird, A.R., Eickhoff, S.B., 2010. ALE meta-analysis of action observation and imitation in the human brain. Neuroimage 50, 1148–1167.

Chan, A.W., Peelen, M.V., Downing, P.E., 2004. The effect of viewpoint on body representation in the extrastriate body area. Neuroreport 15, 2407–2410.

David, N., Aumann, C., Santos, N.S., Bewernick, B.H., Eickhoff, S.B., Newen, A., Shah, N.J., Fink, G.R., Vogeley, K., 2008. Differential involvement of the posterior temporal cortex in mentalizing but not perspective taking. Soc. Cogn. Affect Neurosci. 3, 279–289.

DeYoe, E.A., Carman, G.J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., Miller, D., Neitz, J., 1996. Mapping striate and extrastriate visual areas in human cerebral cortex. Proc. Natl. Acad. Sci. U. S. A 93, 2382–2386.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210.

Frith, C., Frith, U., 2006. The neural basis of mentalizing. Neuron 50, 531–534.

Frith, U., Frith, C.D., 2003. Development and neurophysiology of mentalizing. Philos. Trans. R. Soc. Lond B Biol. Sci. 358, 459–473.

Gao, T., Scholl, B.J., McCarthy, G., 2012. Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. J. Neurosci. 32, 14276–14280.

Giese, M.A., Poggio, T., 2003. Neural mechanisms for the recognition of biological movements. Nat. Rev. Neurosci. 4, 179–192.

Glover, G.H., 1999. Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9, 416–429.

Hesse, M.D., Sparing, R., Fink, G.R., 2009. End or means–the “what” and “how” of observed intentional actions. J. Cogn. Neurosci. 21, 776–790.

Huberle, E., Karnath, H.O., 2012. The role of temporo-parietal junction (TPJ) in global Gestalt perception. Brain Struct. Funct. 217, 735–746.

Jackson, P.L., Meltzoff, A.N., Decety, J., 2006. Neural circuits involved in imitation and perspective-taking. Neuroimage 31, 429–439.

Jacob, P., Jeannerod, M., 2005. The motor theory of social cognition: a critique. Trends Cogn. Sci. 9, 21–25.

Jeannerod, M., 2007. Being oneself. J. Physiol. Paris 101, 161–168.

Jeannerod, M., Anquetil, T., 2008. Putting oneself in the perspective of the other: a framework for self-other differentiation. Soc. Neurosci. 3, 356–367.

Kilner, J.M., Friston, K.J., Frith, C.D., 2007. Predictive coding: an account of the mirror neuron system. Cogn. Process 8, 159–166.

Kourtzi, Z., Kanwisher, N., 2000. Cortical regions involved in perceiving object shape. J. Neurosci. 20, 3310–3318.

Kourtzi, Z., Kanwisher, N., 2001. Representation of perceived object shape by the human lateral occipital complex. Science 293, 1506–1509.

Lee, S.M., Gao, T., McCarthy, G., 2014. Attributing intentions to random motion engages the posterior superior temporal sulcus. Soc. Cogn. Affect Neurosci. 9, 81–87.

Leshinskaya, A., Caramazza, A., 2015. Abstract categories of functions in anterior parietal lobe. Neuropsychologia 76, 27–40.

Lohmann, G, Neumann, J, Müller, K, Lepsien, J, Turner, R., 2008. The multiple comparison problem in fmria new method based on anatomical priors. In: Proceedings of the First Workshop on Analysis of Functional Medical Images, pp 1–8.

Lohmann, G., Muller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., Zysset, S., von Cramon, D.Y., 2001. LIPSIA–a new software system for the evaluation of functional magnetic resonance images of the human brain. Comput. Med. Imaging Graph 25, 449–457.

Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B., 2005. Valid conjunction inference with the minimum statistic. Neuroimage 25, 653–660.

Norris, D.G., 2000. Reduced power multislice MDEFT imaging. J. Magn. Reson. Imaging 11, 445–451.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113.

Oosterhof, N.N., Tipper, S.P., Downing, P.E., 2012. Viewpoint (in)dependence of action representations: an MVPA study. J. Cogn. Neurosci. 24, 975–989.

Oosterhof, N.N., Wiggett, A.J., Diedrichsen, J., Tipper, S.P., Downing, P.E., 2010. Surface-based information mapping reveals crossmodal vision-action representations in human parietal and occipitotemporal cortex. J. Neurophysiol. 104, 1077–1089.

Papeo, L., Lingnau, A., 2015. First-person and third-person verbs in visual motion-perception regions. Brain Lang. 141, 135–141.

Pollmann, S., Zinke, W., Baumgartner, F., Geringswald, F., Hanke, M., 2014. The right temporo-parietal junction contributes to visual feature binding. Neuroimage 101, 289–297.

Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. Annu. Rev. Neurosci. 27, 169–192.

Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., Rozzi, S., 2014. Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. Physiol. Rev. 94, 655–706.

Ruby, P., Decety, J., 2001. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. Nat. Neurosci. 4, 546–550.

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., Jamal, N., Powell, L., 2006. My body or yours? The effect of visual perspective on cortical body representations. Cereb. Cortex 16, 178–182.

Schubotz, R.I., 2011. In: Welsch, W., Singer, W.J., Wunder, A. (Eds.), Long-Term Planning and Prediction: Visiting a Construction Site in the Human Brain. Interdisciplinary Anthropology. Continuing Evolution of Man. Springer, Berlin/Heidelberg, pp. 79–104.

Schurz, M., Aichhorn, M., Martin, A., Perner, J., 2013. Common brain areas engaged in false belief reasoning and visual perspective taking: a meta-analysis of functional brain imaging studies. Front. Hum. Neurosci. 7, 712.

Shmuelof, L., Zohary, E., 2008. Mirror-image representation of action in the anterior parietal cortex. Nat. Neurosci. 11, 1267–1269.

Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain. Thieme, New York.

Ugurbil, K., Garwood, M., Ellermann, J., Hendrich, K., Hinke, R., Hu, X., Kim, S.G., Menon, R., Merkle, H., Ogawa, S., et al., 1993. Imaging at high magnetic fields: initial experiences at 4 T. Magn. Reson. Q 9, 259–277.

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.

Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited–again. Neuroimage 2, 173–181.

Wurm, M.F., Schubotz, R.I., 2012. Squeezing lemons in the bathroom: contextual information modulates action recognition. Neuroimage 59, 1551–1559.

Wurm, M.F., Lingnau, A., 2015. Decoding actions at different levels of abstraction. J. Neurosci. 35, 7727–7735.

Wurm, M.F., Schubotz, R.I., 2017. What's she doing in the kitchen? Context helps when actions are hard to recognize. Psychonomic Bull. Rev. 24, 503–509.

Wurm, M.F., von Cramon, D.Y., Schubotz, R.I., 2011. Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. Hum. Brain Mapp. 32, 2141–2150.

Wurm, M.F., Ariani, G., Greenlee, M.W., Lingnau, A., 2016. Decoding concrete and abstract action representations during explicit and implicit conceptual processing. Cereb. Cortex 26, 3390–3401.

Zacks, J.M., 2008. Neuroimaging studies of mental rotation: a meta-analysis and review. J. Cogn. Neurosci. 20, 1–19.