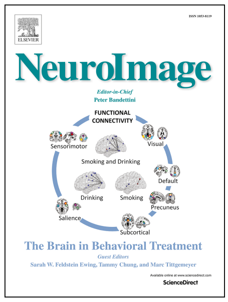
Accepted Manuscript



Making sense of objects lying around: How contextual objects shape brain activity during action observation

Nadiya El-Sourani, Moritz F. Wurm, Ima Trempler, Gereon R. Fink, Ricarda I. Schubotz

PII: S1053-8119(17)30991-6

DOI: 10.1016/j.neuroimage.2017.11.047

Reference: YNIMG 14494

To appear in: *NeuroImage*

Received Date: 26 June 2017

Revised Date: 15 November 2017

Accepted Date: 21 November 2017

Please cite this article as: El-Sourani, N., Wurm, M.F., Trempler, I., Fink, G.R., Schubotz, R.I., Making sense of objects lying around: How contextual objects shape brain activity during action observation, *NeuroImage* (2017), doi: 10.1016/j.neuroimage.2017.11.047.

Title: Making sense of objects lying around: How contextual objects shape brain activity during action observation

Authors: Nadiya El-Sourani[1,2], Moritz F. Wurm[4], Ima Trempler[1], Gereon R. Fink[2,3], Ricarda I. Schubotz[1,2]

[1]Department of Psychology, Westfälische Wilhelms-Universität, 48149 Münster, Germany

[2]Institute of Neuroscience and Medicine (INM3), Cognitive Neuroscience, Research Centre Jülich, 52425 Jülich, Germany

[3]Department of Neurology, University Hospital Cologne, 50937 Cologne, Germany

[4]Center for Mind/Brain Sciences (CIMeC), University of Trento, Rovereto, TN, Italy

Corresponding author:

Nadiya El-Sourani

Fliednerstr. 21,

48149 Münster

n_elso02@uni-muenster.de

Abstract

Action recognition involves not only the readout of body movements and involved objects but also the integration of contextual information, e.g. the environment in which an action takes place. Notably, inferring superordinate goals and generating predictions about forthcoming action steps should benefit from screening the actor's immediate environment, in particular objects located in the actor's peripersonal space and thus potentially used in following action steps. Critically, if such *contextual objects* (COs) afford actions that are semantically related to the observed action, they may trigger or facilitate the inference of goals and the prediction of following actions.

This fMRI study investigated the neural mechanisms underlying the integration of COs in semantic and spatial relation to observed actions. Specifically, we tested the hypothesis that the inferior frontal gyrus (IFG) subserves this integration. Participants observed action videos in which COs and observed actions had common overarching goals or not (*goal affinity*) and varied in their location relative to the actor.

High goal affinity increased bilateral activity in action observation network nodes, i.e. the occipitotemporal cortex and the intraparietal sulcus, but also in the precuneus and middle frontal gyri. This finding suggests that the semantic relation between COs and actions is considered during action observation and triggers (rather than facilitates) processes beyond those usually involved in action observation. Moreover, COs with high goal affinity located close to the actor's dominant hand additionally engaged bilateral IFG, corroborating the view that IFG is critically involved in the integration of action steps under a common overarching goal.

Keywords

semantic integration, object perception, action observation, fMRI, inferior frontal gyrus

1. Introduction

Ecologically, it is of utter importance that we observe and understand what people around us are doing. To properly interact with each other, the ability to derive intentions, to capture action goals and to predict future steps of an unfolding action is highly relevant. Consequently, a large body of research has focused on the question of how we capture, simply by observation, what others aim to do (Van Overwalle & Beatens, 2009; Caspers, Zilles, Laird, & Eickhoff, 2010). However, since actions are complex and carry different sorts of information it is not surprising that it remains to be investigated i) how information is exploited during the observation of actions, and ii) how this information helps to constrain the observer's expectation.

When we perceive an object-directed action we are typically confronted with *manipulation information* including hand posture and movement trajectories, and *object information* including pragmatic and semantic object properties. Both information types are relevant for action recognition and are referred to as *core information,* hereafter. For example, after watching somebody cracking (manipulation movement) an egg (object information) one could easily name the action, i.e., cracking an egg. However, it is not before detecting the pan nearby that one would expect the actor to prepare scrambled eggs rather than baking a cake. Hence, information that is not necessary for the recognition of an on-going action might be crucial for higher-level inferences of goals and the prediction of forthcoming action steps (Malcolm, Groen, & Baker, 2016). This information, which is referred to as *contextual information*, hereafter, potentially modulates action perception.

While the influence of contextual information on object recognition has been investigated (Bar, 2004; Boyce, Pollatsek, & Rayner, 1989; Hayes, Nadel, & Ryan, 2007; Zimmermann, Schnier, & Lappe, 2010), contextual influence on action perception has attracted little attention thus far. For this reason, in recent studies we concentrated on the impact of contextual information on action perception. We focused on the room and actor information during the observation of an action. As with actor information (Wurm, von Cramon, & Schubotz, 2011; Hrkać, Wurm, & Schubotz, 2014), our data revealed that participants' brains process room information spontaneously, i.e., without instruction or task requirements (Wurm & Schubotz, 2012). Participants even process room information when it is in conflict with the action itself: subjects needed longer to indicate the recognition of an action when the action took place in an incompatible as compared to a compatible or neutral room (Wurm & Schubotz, 2012). On the other hand, action-compatible room information helped when actions were difficult to recognize (Wurm & Schubotz, 2017; Wurm, Artemenko, Guiliani, & Schubotz, 2017). Moreover, a misfit between the room, the manipulated objects and the applied manipulation was associated with increased activity of brain regions associated with object and action processing, respectively, even when participants were not required to deal with this conflict (Wurm, von Cramon, & Schubotz, 2012).

The present study takes the position that *contextual objects (COs)*, which are often part of an observed action scene albeit that they are not integrated into the observed action (*Figure 1*), should also exert an impact on action observation. On the one hand, contextual objects often signify the room category (e.g., a knife block implies 'kitchen') and so inform the observer about the general probability of an action. On the other hand, contextual objects can also inform us about measures an actor has adopted before performing the observed manipulation. Thus, contextual objects are often prepared for the achievement of (and hence inform an observer about) an overarching goal (cf. Iacoboni et al., 2005), that is, a desired end state or outcome (Csibra & Gergely, 2007). This should be especially the case when contextual objects are located close and right to the actor (i.e. actor's peripersonal space), as this area yields special expectations for subsequent usage of that object. In this case, contextual objects may undergo a shift in their meaning: from contributing to the general probability of actions to becoming a potential future target of the observed actor.

While we found that manipulated objects have a significant impact on action processing due to their implicating of possible actions (Schubotz, Wurm, Wittmann, & von Cramon, 2014; Hrkać, Wurm, Kühn, & Schubotz, 2015), we know very little about the impact of contextual objects. If it is part of the information provided by an object which actions can be performed with it, contextual objects should also signal possible actions. The question hence arises: Do contextual objects also imply possible actions, and if so, how do observers deal with this additional information? We here focus on two factors that might modulate the impact of contextual object information on action perception: *goal affinity (GA)* and *location ergonomics (LE)*.

Goal affinity is defined as an object's semantic relation to the observed action, quantified by the probability with which the object becomes employed in the same action. Accordingly, the higher the goal affinity of a contextual object, the more an observer will expect the contextual object to be soon integrated into the observed action. Functional MRI studies have shown that the mere perception of objects implicates manipulation and action (Buxbaum, Kyle, Tang, & Detre, 2006; Johnson-Frey, 2004; Schubotz et al., 2014). Correspondingly, we expected the goal affinity of contextual objects to increase activity in brain areas that are related to object-related action representation, particularly the occipito-temporal cortex (OTC; Wigget & Downing, 2011) and the inferior parietal lobule (IPL; Buxbaum & Kalénine, 2010; Schubotz et al., 2014). Moreover, we expected that the goal affinity of contextual objects correlates with activity in the inferior frontal gyrus (IFG), as we found this area to be modulated by other types of contextual information, specifically room and actor information, during action observation (Wurm & Schubotz, 2012; Hrkać et al., 2014; Hrkać et al., 2015). In action observation, the IFG is suggested to be engaged in the retrieval and integration of action-relevant semantic information (see also Caspers et al., 2010; Kilner, 2011).

On the other hand, *location ergonomics*, the spatial relation to an observed action, are meant to quantify the probability with which an object can be reached by an actor. For example, right-handers have a general bias for objects on their right-hand-side (Toney & Thomas, 2006; Rezaee, Shojaee, Ghasemi, Moghaddam & Momeni, 2010; Bryden, Pryde & Roy, 2000). This could be confirmed by a pilot study, in which we asked participants to place an object in the most convenient location they could think of for subsequent use. In order to build expectations about the observed action, the location of a contextual object in relation to the actor is meaningful in such a way that objects that are needed for an action have to be easily reached and grasped for subsequent efficient use. So far, object reachability has been studied only with regard to its effect on action execution, but not regarding the effect on goal inferences. It was found that grip facilitation depends on the object's availability to manipulation, pointing to an interaction between the object's pragmatic properties and the action goal implied by the contextual setting (Kalénine, Shapiro, Flumini, Borghi, & Buxbaum, 2013). Furthermore, current research on the effects of hand proximity in relation to objects showed that placing the hands near an object was found to initiate the visual system for the processing of visually guided actions (Gozli, West, & Pratt, 2012). Against this backdrop, we expected that goal affinity of contextual objects should have a specific effect when located close and right to the actor. Our own previous findings implied that objects with a low goal affinity would then lead to increased IFG activity as they would challenge attempts to be integrated into the currently observed action (Wurm & Schubotz, 2012; Hrkać et al, 2014). However, in principle, also the opposite effect is conceivable: IFG activity increases for contextual objects with both a strong semantic and spatial relation to the observed action, because when contextual objects score high on both dimensions, they might trigger expectations of a concrete action sequence. These opposite effects would speak in favor of alternative processing modes: in the former, the brain selects more information from the scene than is minimally needed to interpret the currently observed object manipulation; then, it strives to resolve potentially conflicting information. In contrast, the latter effect would be expected when the brain focuses only on core information (manipulated object, manipulating hands) and selects from the rest of the scene only compatible information to strengthen the current interpretation of the object manipulation with regard to upcoming action steps.

2. Materials and Methods

2.1 Participants

Twenty-one right-handed subjects (12 females; 24.00 ± 3.25 years old; range, 20-30 years) with normal or corrected-to-normal vision participated in the study. None of them reported a history of medical, neurological or psychiatric disorders, or substance abuse. The study protocol was conducted in accordance with ethical standards of the Declaration of Helsinki and approved by the local Ethics Committee of the University of Münster. Each participant submitted a signed informed consent and received reimbursement or course credits for their participation afterwards.

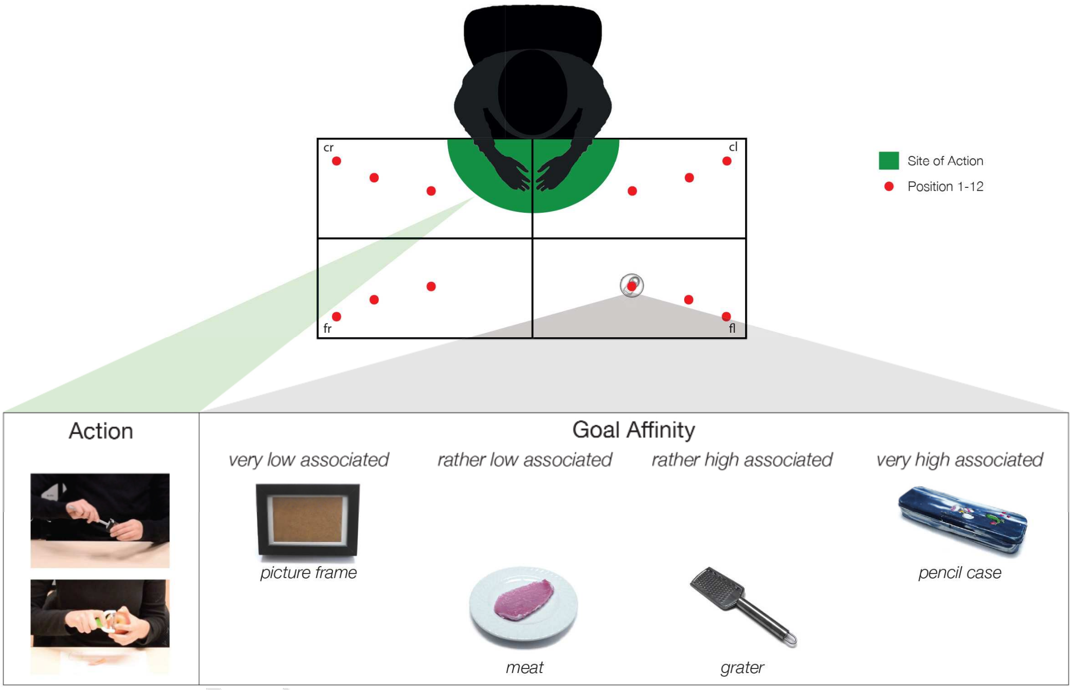


Figure 1. Exemplary demonstration of experimental factors: Red dots refer to location ergonomics and include positions 1-12. Further positions are subdivided into quadrants from the actor's perspective: close-right (cr), close-left (cl), far-right (fr) and far-left (fl). Goal affinity levels of contextual objects for distinct actions are depicted in the box.

2.2 Stimuli

During the scanning session participants were presented with 360 video clips showing actions (action trials) and with 72 written action descriptions referring to these actions (question trials) (presentation software: Presentation Version 13.1, Neurobehavioral Systems). Each trial (6 s) started with a video clip or a question (3 s), which was followed by a fixation phase (3 s). To enhance the temporal resolution of the BOLD response, a variable jitter (500, 1000, 1500 ms) was included after the fixation phase. In 5% (N=18) of the trials a null event was implemented (6 s), which required participants to fixate a fixation cross.

Each video showed a single object-directed action performed by one and the same actress sitting at a table and was filmed from a frontal third person perspective (3pp). Each of the actions was performed in a typical setting (e.g., cracking an egg in the kitchen, writing a letter in the office), resulting in compatibility between action and context-background. Settings were either a kitchen background or an office background. In total, 39 action videos in the kitchen and 33 in the office were selected.

Each video involved two target objects (e.g., a lemon and a squeezer in „squeezing a lemon“). The actress's face was inclined downwards in order to minimize cognitive effects of person perception (*Figure 2)*. In a pilot study (C) all videos were investigated regarding their recognizability. Only videos were selected in which 100% of the participants recognized the action correctly.

Out of the 360 action videos presented to each participant, 288 included an additional contextual object, which was always positioned in front of the actress on a table (*Figure 2*). In order to investigate whether those objects are processed in terms of action probabilities we modulated them due to two experimental factors: *location ergonomics* and *goal affinity*.

Location ergonomics was quantified by a pilot study A (N=24) in which the reachability of an object at 60 varying positions on a table had to be rated. The mean standard deviation of this rating was 0.579. Based on these results the table was subdivided into four quadrants: close-right (cr), close-left (cl), far-right (fr), and far-left (fl). Furthermore, for each quadrant three positions were selected, which differed with their regards to reachability (high, middle, low) and, at the same time, mirrored the positions of the other quadrants. This resulted in a total number of 12 positions for the placement of contextual objects during the observation of action videos (*Figure 1*). In a second task we asked participant to place a "power-grip" object in the most convenient location they can think of for subsequent use. Mean values of x and y coordinates on the table show that participants' preferred location for subsequent usage was close and right to the action site.

Goal affinity was quantified on the basis of subjective ratings of a large sample (N = 500) of students (pilot study B). The mean standard deviation of this rating was 0.645. Based on pilot data B, 144 different objects were assigned to four different levels of goal affinity ranging from "very low associated" to "very high associated". Each of the 72 actions was paired with two objects of two different goal affinity levels. However, these pairs were selected in such a way that the factor goal affinity was evenly distributed across all levels (*Figure* 1). Videos were arranged and selected for the fMRI session so that each action was seen five times in total, once without and four times with one contextual object. Furthermore, those action videos containing a contextual object were arranged in a way that all goal affinity levels at all 12 positions occurred in an evenly distributed number (12 positions x 4 goal affinity levels x 6 occurrences = 288 action videos with a contextual object). Videos were presented in a pseudo-randomized fashion in order to avoid direct repetition of action, goal affinity, and location of the contextual object. All factor levels were presented in an evenly distributed manner.

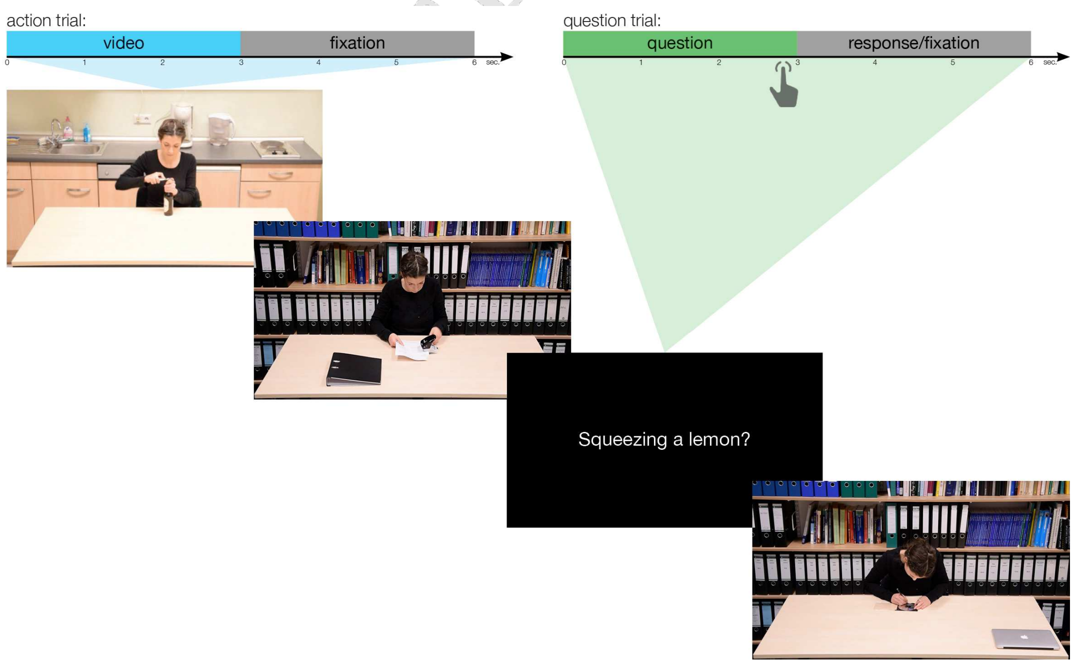


Figure 2. Schematic diagram of the task. Action trials consisted of an action video (3s) and a fixation phase (3s). Question trials consisted of a question regarding the preceding video trial (n-1), followed by a response and fixation phase.

2.3 Task

Participants were instructed to watch the video clips attentively. They were told that after some of the video clips an action description would appear that referred to the content of the preceding video, and that had either to be accepted or rejected by the participants. Participants were naïve with regard to the ratio of videos followed by a question trial. The ratio of action and catch trials was 1:5 (20% catch trials); 50% of trials were to be affirmed and 50% to be rejected. Action descriptions (e.g., squeezing a lemon) did not amount to an overarching action goal (e.g., preparing a meal) but rather to each single action itself and hence a short-term goal. Action descriptions were presented in a pseudo-randomized fashion mixed with the experimental trials. Responses were given on a two-button response box, using the index finger to accept and the middle finger to reject the action description at hand. This method was used for the purpose of keeping the participants' attention on track while watching the videos (Wurm & Schubotz, 2012; Hrkać et al, 2015). Error rates were analyzed to assess participants' behavioral performance. Finally, a training phase of five minutes was included before the fMRI session in order to familiarize the participants with the task.

2.4 fMRI Image Acquisition

Imaging was performed on a 3 Tesla Siemens Magnetom Prisma MR tomograph using a 20 channel head coil. Participants were located in a supine position on the scanner bed with their right index and middle fingers positioned on the appropriate response buttons of a response box. To minimize head and arm motions, head and arms were tightly fixated with form-fitting cushions. Furthermore, participants were provided with earplugs in order to attenuate the scanner noise. Whole-brain functional images were acquired using a gradient T2*-weighted single-shot echo-planar imaging (EPI) sequence sensitive to blood oxygenation level dependent (BOLD) contrast (64 x 64 data acquisition matrix, 192 mm field of view, 90° flip angle, TR = 2000 ms, TE = 30 ms). Each volume consisted of thirty adjacent axial slices with a slice thickness of 4 mm and a gap of 1 mm, which resulted in a voxel size of 3 x 3 x 5 mm. Images were acquired in ascending order along the AC-PC plane to provide a whole-brain coverage. After functional imaging, structural data were acquired for each participant using a standard Siemens 3D T1-weighted MPRAGE sequence for detailed reconstruction of anatomy with isotropic voxels (1 x 1 x 1 mm) in a 256 mm field of view (256 x 256 matrix, 192 slices, TR = 2130, TE = 2.28).

In order to present the stimuli during the scanner session, a 45° mirror was fixated on the top of the head coil. A video-projector projected the experiment on a screen that was positioned behind the subject's head, so that participants could see the stimuli via the mirror. The mirror was adjusted for each participant to provide a perfect view (center of the field of vision). In a pilot study C we controlled for recognizability of actions and contextual objects in a similar experimental setting. Only action videos in which the action and the contextual object could be identified by at least 95% of the participants were employed in the present study.

2.5 fMRI Data Analysis

2.5.1 fMRI data preprocessing

Brain image preprocessing and basic statistical analyses were conducted using LIPSIA software package, version 3.0 (Lohmann et al., 2001). Initially, spikes in time series were corrected by interpolating them with adjacent time points. To correct for temporal offsets between the slices acquired in one scan, a cubic-spline interpolation was used. Furthermore, individual functional MR (EPI) images were motion-corrected with the first time-step as reference and six degrees of freedom of which three are rotational and three translational. Then, the average across all time points of this corrected data was used as reference scan for a second pass of motion-correction. Motion correction estimates were inspected visually. Coregistration was done using statistical parametric mapping package (SPM12, Wellcome Department of Imaging Neuroscience, London, UK). The images were coregistered and transformed into a standard stereotactic space using the intercommissural line as the reference plane for transformation (Ashburner & Friston, 1997). Anatomical datasets were normalized to the ICBM/MNI space by linear scaling. The resulting parameters were then used to transform all functional slices employing a trilinear interpolation. Resulting data had a spatial resolution of 3 x 3 x 3 mm (27 mm^3). Normalized functional images were spatially smoothed with a Gaussian kernel of 6 mm full width at half-maximum

(FWHM). Finally, a temporal high-pass filter of 1/125 HZ was applied to the data in order to remove low-frequency noise such as scanner drift.

2.5.2 Design specification

The statistical evaluation was based on a least-squares estimation using the general linear model (GLM) for serially autocorrelated observations (Friston et al., 1995; Worsley & Friston, 1995). The design matrix was generated with delta functions and convolved with a canonical hemodynamic response function. Activations were analyzed time-locked to the onset of the videos and the analyzed epoch comprised the full duration (3s) of the presented videos, the duration of the null events (6s), and the reaction time in question trials (max. 3s). The GLM contained 12 regressors: eight predictors for the experimental conditions, one predictor for videos without contextual objects (noCO), one including all the null events (6s fixation phase), one predictor for question trials, and finally one parametric regressor for the iteration of action, which was included as a regressor of nuisance to control for effects of action repetition. The eight predictors for the experimental conditions were assigned with regard to location ergonomics in terms of the four quadrants. Furthermore, goal affinity levels were merged into high (level 3 + 4) and low level (level 1 + 2), which resulted in the following predictors: 1. High goal affinity at close-right (cr) positions; 2. Low goal affinity at close-right (cr) positions; 3. High goal affinity at close-left (cl) positions; 4. Low goal affinity at close-left (cl) positions; 5. High goal affinity at far-right (fr) positions; 6. Low goal affinity at far-right (fr); 7. High goal affinity at far-left (fl) positions; 8. Low goal affinity at far-left (fl) positions.

Influence of contextual objects on action perception. In order to investigate whether the brain automatically seeks to integrate contextual objects with the observed action, we contrasted action videos containing a contextual object with those, which did not (CO > noCO).

Influence of contextual objects goal affinity. For the main effect of goal affinity, all eight predictors regarding goal affinity were used in order to contrast high goal affinity with low goal affinity ($GA_{high} > GA_{low}$) on a first level GLM.

For the interaction analysis of goal affinity and location ergonomics only close-right and close-left positions were included in the analysis reflecting high and low location ergonomics, respectively. This was done as we hypothesized close-right positions to have a specific effect on the processing of contextual objects, due to the preferences of actors to deposit objects that they plan to use in upcoming action steps at such locations (Pilot Study A).

2.5.3 Group analysis

To obtain group statistics, the resulting contrast images of all participants were entered into a second level random-effects analysis using a one-sample *t*-test across participants to test for significant deviation from zero. We then corrected for multiple comparisons across all voxels using the threshold-free cluster enhancement (TFCE) method (Smith & Nichols, 2009). The significance level for whole-brain activations was set to $p < 0.05$ TFCE-corrected. Default TFCE parameters H = 2 and E = 0.5 were used (Smith & Nichols, 2009).

SVC Analysis. To specifically test the hypothesis that the IFG is modulated as function of the interaction of the goal affinity and the location of the contextual object, a small volume correction (SVC) was performed on the interaction contrast goal affinity (high, low) x location ergonomics (close-right (cr), close-left (cl)) at $p < 0.05$ TFCE-SVC-corrected. Anatomical masks of left and right Brodmann area (BA) 44 and 45 (Amunts et al., 1999) were defined based upon the Anatomical Toolbox (Eickhoff et al., 2005).

3. Results

3.1 Behavioral Results

Performance during the fMRI session was assessed by error rates and reaction times on correctly answered trials. The average response time was 1353 ± 65ms and the average error rate was low (1.23 ± 2.16%) indicating that participants attentively observed and recognized the actions.

Moreover, in a post-fMRI interview 4 out of the 21 subjects reported not having noticed the CO during the course of the experiment, i.e. they did not consciously process these objects. It would be great, however, to find out whether implicit and explicit processing of these contextual objects would lead to differential activation patterns. Unfortunately the sample size was too small in order to conduct a valid statistical comparison.

3.2 *fMRI Results*

3.2.1Main effect of contextual objects

To investigate brain activity triggered by the mere presence of contextual objects, we contrasted action trials containing contextual objects compared with action trials, which did not (CO > noCO). After applying a TFCE-correction at $p < 0.05$, the inclusion of a contextual object revealed activation of the left intraparietal sulcus (IPS; descending branch, middle occipital gyrus), peristriate regions of the occipital cortex (BA 18, 19) and the fusiform gyrus (BA 19, 37) bilaterally, with stronger activation in the left than the right hemisphere (*Figure 3*; *Table 1*).

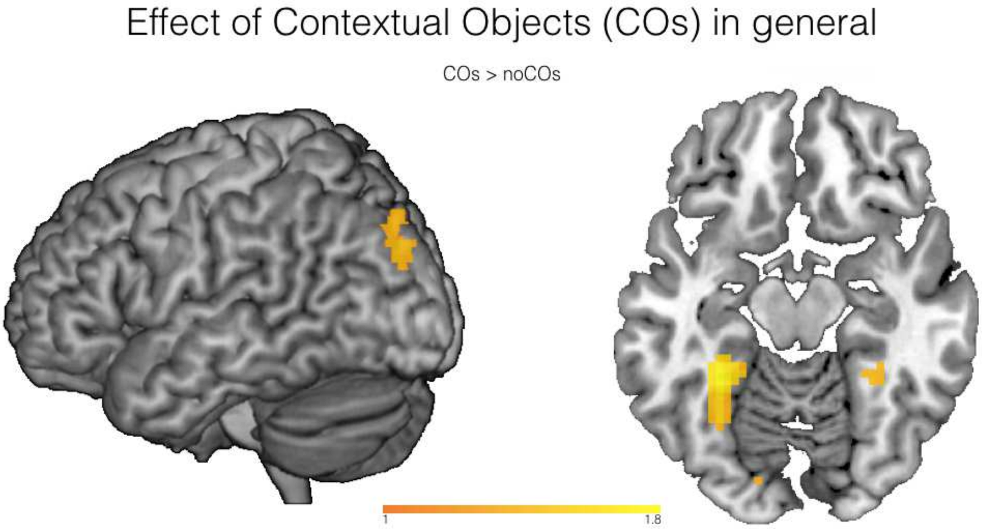


Figure 3. Effects of contextual objects, at $p < 0.05$, TFCE-corrected. P-values are scaled to $-\log10$ (p).

Table 1. fMRI activations for the presence of contextual objects (CO > noCO)

			MNI Coordinates			
Region[*]	*Side*	*BA*	*x*	*y*	*z*	*p values*[**] */ local maxima*
Fusiform gyrus[***]	L	37/19	-24	-76	-7	1.9
	R	37/19	30	-46	-10	1.5
Middle occipital gyrus	L	-	-27	-85	32	1.4
	R	19	36	-85	26	1.4
Cuneus	R	18	18	-91	8	1.3

R, Right; L, Left; x, y, z, MNI coordinates of peak voxel activation; [*]according to the Anatomy Toolbox (Amunts et al., 1999); [**]TFCE-corrected for multiple comparison: p-values are scaled to –log10 (p); [***]extending into left BA 18 + 19

3.2.1 Main effect of goal affinity

For the main effect of goal affinity, we tested for differences between high vs. low goal affinity trials (GAhigh > GAlow). We hypothesized that goal affinity of the contextual objects modulates activity of the OTC and the IPL. TFCE corrected at $p < 0.05$, we found the OTC (local maxima in posterior middle temporal gyrus and inferior temporal gyrus) and the IPL (IPS) to be activated bilaterally, for high compared to low goal affinity trials. In addition, the dorsolateral prefrontal cortex (dlPFC) was activated bilaterally, more specifically in BA 8, 9 and 10. Finally, we found activity in the precuneus and the cerebellum. The reverse contrast (GAlow > GAhigh) revealed no significant activation.

To test whether the activations found for high compared to low goal affinity were due to an increase in activation for contextual objects with a high goal affinity rather than a decrease for contextual objects with low goal affinity, we calculated a conjunction of the TFCE-corrected contrast GAhigh > GAlow and the TFCE-corrected contrast GAhigh > Rest (Null-Events), i.e. (GAhigh > GAlow) ∩ (GAhigh > Rest). When corrected at $p < 0.05$, the conjunction analysis revealed bilateral activation of the IPS (including aIPS), the dlPFC, the OTC, the precuneus, and the cerebellum (*Figure* 4;Table 2). The results indicate that the effect of $GA_{high} > GA_{low}$ did in fact become apparent due to an increase in activation for contextual objects with a high goal affinity. The conjunction analysis for low GA $((GA_{low} > GA_{high}) \cap (GA_{low} > Rest))$ revealed no significant activations.

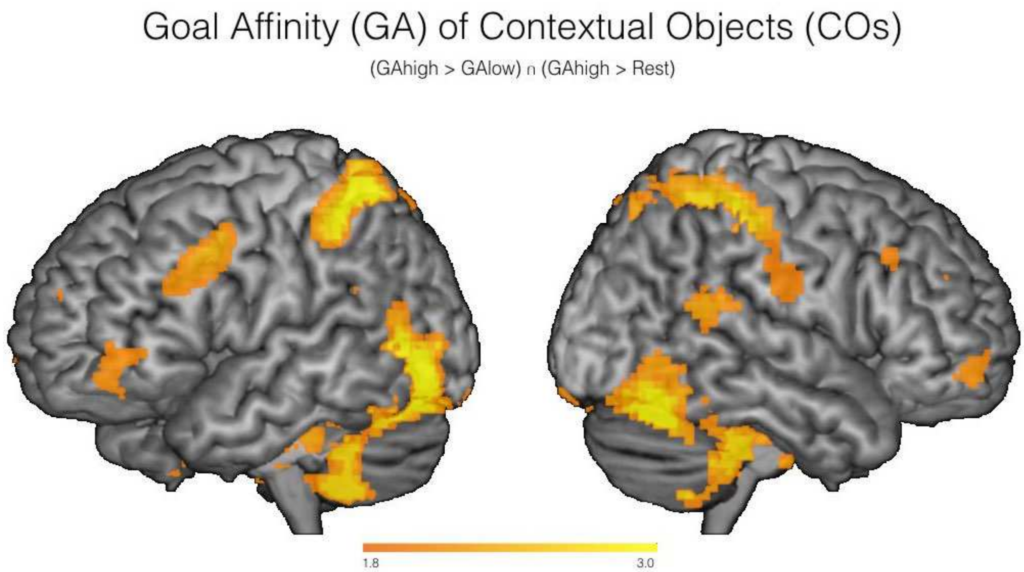


Figure 4. Effects of goal affinity at p < 0.05, TFCE-corrected. P-values are scaled to –log10 (p).

Table 2. fMRI activations for Goal Affinity (GAhigh > GAlow) ∩ (GAhigh > Rest)

Region[*]	*Side*	*BA*	*MNI Coordinates* *x*	*y*	*z*	*p values*[**] */ local maxima*
IPL	L	-	-54	-34	44	3
			-42	-52	59	3
	R	-	45	-40	53	3
			36	-46	59	2.7
MTG/ITG	L	21/22/37	-57	-61	-1	3
	R	21/22/37	52	-66	-6	3
Fusiform gyrus	L	20/36	-27	-7	-37	2.15
	R	20/36	27	3	-44	1.8
Precuneus	L	7	-12	-73	56	2.4
	R	7	12	-67	59	2.3
Middle orbital gyrus	L	10	-45	56	-5	1.9
	R	10	45	59	-7	2.0
MFG	L	8/9	-51	14	38	2.4
	R	8/9	51	23	35	1.86
Cerebellum (Crus 1)	L		-14	-79	-22	3
Cerebellum (V1)	R		15	-73	-22	3

R, Right; L, Left; IPL, inferior parietal lobule; MTG, middle temporal gyrus; ITG, inferior temporal gyrus; MFG, middle frontal gyrus; x, y, z, MNI coordinates of peak voxel activation; [*]according to the Anatomy Toolbox (Amunts et al., 1999); [**]TFCE corrected for multiple comparison: p-values are scaled to –log10 (p).

3.2.2 Interaction of goal affinity by actor side

In order to investigate whether goal affinity effects differed depending on the contextual objects location, we conducted a SVC-based analysis of the left and right IFG for the interaction contrast goal affinity (GAh, GAl) x location ergonomics (close-right (cr), close-left (cl)). The analysis revealed significant engagement of the left and right IFG (*Figure 5*), when TFCE-corrected at $p < 0.05$. Mean beta values of the significantly activated clusters can be found in *Figure 5*.

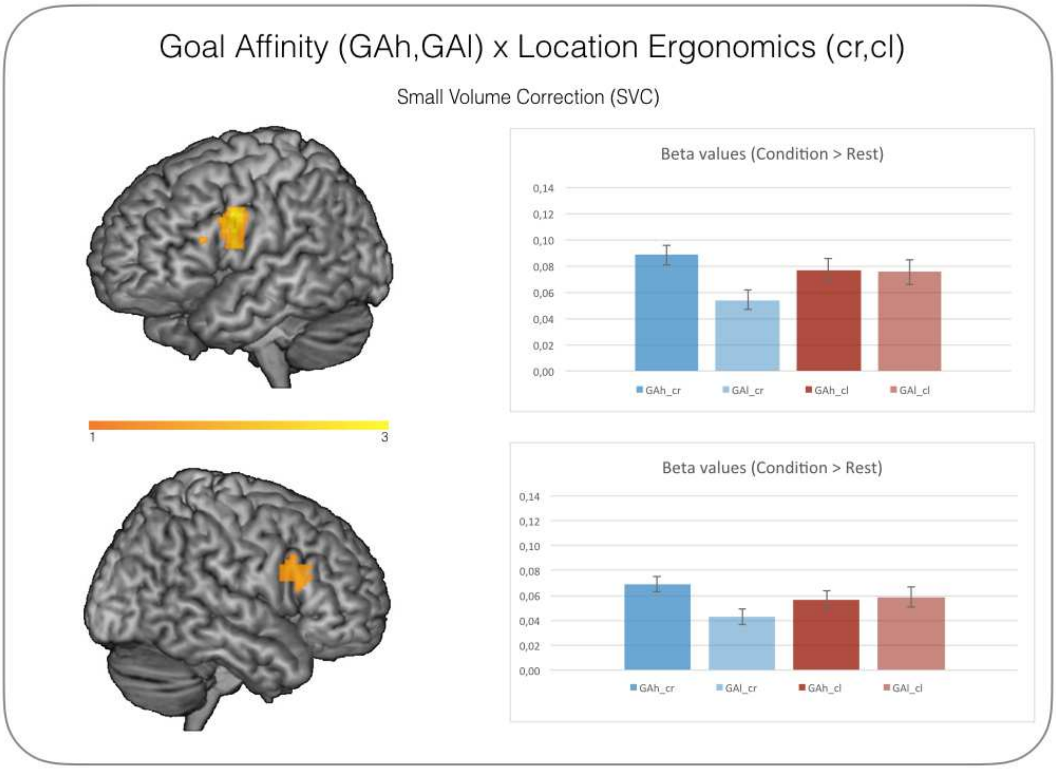


Figure 5. Effects of IFG SVC, TFCE-corrected at $p < 0.05$. P-values are scaled to $-\log10$ (p). The bar graphs depict beta values of the significantly activated clusters, in order to provide a better understanding of the interaction effect. The IFG was enhanced by contextual objects with high goal affinity when presented close to the actor's right (cr) vs. left (cl).

4. Discussion

The present study aimed at gaining insight into the question if and how the brain processes nearby objects (contextual objects, COs) that are not involved in an observed action. We manipulated the contextual objects' semantic relation (goal affinity, GA) as well as their spatial relation (location ergonomics, LE) to the observed action. We expected both factors to influence action observation such that, first, the higher the contextual object's goal affinity the stronger the observing brain's readiness to integrate it into the observed action. Second, we expected that goal affinity would have an additional effect for contextual objects presented close and right to the actor, i.e., those contextual objects that were potentially easy to reach for the actor (Pilot study A).

We found that contextual objects are processed by brain areas that transcend simple visual processing of man-made objects. In particular, activity increased at several sites when the contextual objects' semantic relationship to the observed action was strong. Notably, when such contextual objects with a high goal affinity were presented close and to the right side of the actor in contrast to close left, we observed additional effects. We had hypothesized modulation of the inferior frontal gyrus (IFG), which is known to be involved in the retrieval and integration of action-relevant semantic information. In the following, we will outline the view that the impact of the actor's location on the contextual object's goal affinity suggests that objects lying close to the actor's dominant hand are processed as being potential future targets of the observed actor.

Our findings show that the brain processes objects in the vicinity of an observed action, and it does so in a telling manner. We here have to distinguish between the effects induced by the mere presence of a contextual object on the table (vs. no contextual object), and the effects induced by the goal affinity this contextual object has with regard to the observed object manipulation. Regarding the presence of contextual objects, in our experiment contextual objects were from trial to trial located at different sites on a table including locations that were not easily accessible to the actor. Independent of this accessibility, the mere presence of an object anywhere on the table increased neural activation in areas known for visual, haptic, and action-related object processing, including the left IPS (Creem-Regehr, 2009; Grefkes & Fink, 2005; Singh-Curry & Husain, 2009; Ramsey, Cross, & Hamilton, 2011), the fusiform gyrus (cf. Bar et al., 2001; Tyler, Stamatakis, Bright & Acres, 2004), and visual association areas, i.e., BA 18 and 19 (Chawla, Rees, & Friston, 1999). Activity in the visual association areas, especially BA 19, corresponds to the occipital place area (OPA), which is suggested to represent local elements of an action scene (Kamps, Julian, Kubilius, Kanwisher, & Dilks, 2016).

Remarkably, an additional set of brain areas was engaged in processing *the goal affinity* of the contextual object in relation to the currently observed action, indicating an in-depth processing for this particular object property. This set of areas included the posterior components of the action observation network (AON), i.e., the right IPS (in addition to the left IPS already active for contextual objects vs. no objects) and the OTC. We additionally found activation of the precuneus and the dlPFC, which we had not a priori hypothesized and only shortly speculate upon in the following. While AON activity is already triggered by the observation of short and simple object manipulations, dlPFC and precuneus are often seen to join this network for longer and multi-step actions (e.g., Wurm, Hrkać, Morikawa & Schubotz, 2014). The posterior dorsal precuneus, which we found in our contrast, is strongly connected to the co-activated middle frontal gyrus (dlPFC), and both are engaged in visuomotor imagery and action planning (Zhang & Chiang-shan, 2012). Balser and co-workers (2014a) reported enhanced precuneus and cerebellar activity for experts vs. novices during the anticipation of sports movements. In a further study, these areas' activity increased as a function of sport experts' anticipatory performance (Balser et al., 2014b). In a similar vein, Calvo-Merino and co-workers (2005) reported an almost identical network as ours to increase its activity when expert dancers watched previously learned versus novel movements. These activations were interpreted as a recall of acquired action representations, enabling greater skills in anticipating these actions' outcomes (cf. Aglioti, Cesari, Romani, & Urgesi, 2008). In sum, we take the network for high vs. low goal affinity to reflect enhanced action anticipation in the observers' brains, and thus to corroborate the notion that contextual objects are processed with regard to their potential upcoming usage in an observed action.

To further explore the impact of contextual objects on brain activity during action observation, we tested the hypothesis that goal affinity would have a specific effect when contextual objects were located close and right to the actor. This idea derived from the fact that actors have a preference to deposit objects that they plan to use in upcoming action steps at this location (Pilot study A). We reasoned that the observers' brains should be tuned to integrate particularly those contextual objects in the currently observed action that are endorsed by their ergonomic location; moreover, if these contextual objects score particularly high on goal affinity to the ongoing action, semantic integration should be especially facilitated. Based on former studies, we expected IFG to be the area where such integration effects could surface (Wurm & Schubotz, 2012; Hrkać et al, 2014; see also Kilner, 2011). Indeed, the comparison between close right and close left contextual objects revealed an enhanced goal affinity effect in both right and left IFG: while high and low goal affinity objects induced comparable activity in this brain region when contextual objects were presented to the actor's left hand side, they differed significantly when presented to the actor's right (*Figure 5*). The engagement of the IFG in close-left positions of the contextual objects could be interpreted as baseline activation. Hence, when presented close and right to the actor, contextual objects with high goal affinity received a boost in IFG engagement, while low goal affinity contextual objects led to a suppression of IFG activation. This finding supports the view that contextual objects were processed in their relation to the actor and the action: not only with regard to their goal-relatedness, but also with regard to their spatial reachability. The influence of the actor's location on the goal affinity effect implies that contextual objects are perceived as being possible upcoming targets of the actor: the stronger the spatial and semantic relation of the contextual object to the observed manipulation, the more information regarding a specific upcoming action sequence is provided, and possibly, the more a specific action goal can be anticipated. This result is in good accordance with the presumed core function of the IFG in action observation, i.e., the retrieval and integration of action-relevant semantic information (Caspers et al., 2010; Kilner, 2011). Beyond its role in action observation, the IFG has also been linked to response inhibition (e.g. Menon et al., 2001, Rubia et al., 2003, Aron et al., 2004) and to the detection of important cues (Hampshire et al. 2010). We cannot exclude that, and therefore it remains to be elucidated in future studies, whether such processes might have taken part to enhance focus on the currently observed action, as required by our task.

We suggest that, as long as contextual objects are not presented in the close right area of the actor, they are processed with regard to their goal affinity (relative to the observed action), whereas when entering the preferred action zone, this information is used as current impact on upcoming overarching goals.

In previous studies of our own (Hrkać et al, 2014; Wurm & Schubotz 2012) we found interference rather than enhancement effects for contextual information in action processing. This apparent discrepancy can be resolved by the fact that in Wurm and Schubotz (2012) we investigated the effect of compatibility and incompatibility of room information on action perception, whereas the present study emphasized the effect of compatibility, ranging from unspecific compatibility (low goal affinity) to very specific compatibility (high goal affinity). Thus, in the present study no strong conflict or mismatch emerged by the presence of contextual objects with low goal affinity. Going into more detail, in Wurm and Schubotz (2012) we created compatible room information by employing a set of non-specifically related contextual objects in order to form a room category (e.g., kitchen: stove, kettle, bottle of milk). Hence, contextual objects were processed as part of the 'room information'. As we are used to be surrounded by room-compatible objects not directly fitting our action goals, compatible objects can be neglected by an observer, whereas objects revealing incompatible room information are given more weight. In contrast, in the present study contextual objects were presented in isolation and in front of the actor, and action site while additional room information provided behind the actor was kept constant. Therefore, contextual objects received a special emphasis with regard to their implications for probable goals or – at least – subsequent action steps. Due to their potential to soon become core information, an interference effect of low goal affinity contextual objects was still conceivable, especially when positioned close and right to the actor. The fact that we did not find an effect of low goal affinity on IFG activation suggests that – in line with Wurm and Schubotz (2012) - contextual objects with a non-specific compatibility to the room category are in fact treated as non-informative regarding a specific action goal. Hence, we assume that low goal affinity contextual objects are perceived as part of the room category comparable to the compatibility condition in Wurm and Schubotz (2012). Indeed, in everyday life most of the objects that surround us are room-compatible but have a low goal affinity to the object manipulation we are about to perform. On the contrary, contextual objects with a high goal affinity are perceived as highly informative in such a way that a specific overarching action goal and hence a concrete action sequence can be generated, especially when presented close and right to the actor.

4.2 Conclusion

Our study shows that objects provide a meaningful context for observed actions, even when they are neither involved in the action nor relevant to the observer's task. Brain activation elicited by a contextual object varied as a function of the currently observed action, i.e., the contextual object's goal affinity. Moreover, when highly goal-related objects were presented close to the actor's preferred action space, additional brain activity signified efforts to integrate the contextual object into potentially upcoming action steps.

Acknowledgments

We would like to thank Monika Mertens for her great help in the data acquisition. Further we thank Sarah Fromme, Haris Orucevic, Daniel Westen and Daniel Kluger for assistance with the creation of the stimulus material. Finally, we want to thank Christiane Ahlheim and the workgroup AE Schubotz for fruitful discussions.

References

Aglioti, S.M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience*, *11*(9), 1109–1116.

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H.B.M., & Zilles, K. (1999). Broca's region revisited: Cytoarchitexture & intersubject variability. Journal of comparative Neurology 412, 319-341.

Ashburner, J., & Friston, K. J. (1997). Multimodal image coregistration and partitioning - a unified framework. *Neuroimage* 6, 209–217. doi: 10.1006/nimg.1997.0290

Aron A.R., Robbins T.W., Poldrack R.A. (2008). Inhibition and the right inferior frontal cortex. Trends Cogn., 8, 170–177.

Balser, N., Lorey, B., Pilgramm, S., Naumann, T., Kindermann, S., Stark, R., … & Munzert, J. (2014a). The influence of expertise on brain activation of the action observation network during anticipation of tennis and volleyball serves. *Frontiers in Human Neuroscience*. *8*, 35-47. doi: 10.3389/fnhum.2014.00568

Balser, N., Lorey, B., Pilgramm, S., Stark, R., Bischoff, M., Zentgraf, K., … & Munzert, J. (2014b). Prediction of human actions: expertise and task-related effects on neural activation of the action observation network. *Human Brain Mapping*, *35*(8), 4016–4034. doi: 10.1002/hbm.22455

Bar, M., Tootell, R. B. H., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., Rosen, B. R., Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*. 29(2): 529–35.

Bar, M. 2004. Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617-629.

Boyce, S.J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception Performance*, *15*(3), 556-566.

Buxbaum, L. J., Kyle, K. M., Tang, K., & Detre, J. A. (2006). Neural substrates of knowledge of hand postures for object grasping and functional object use: evidence from fMRI. *Brain Research*, *1117*(1), 175-185.

Buxbaum, L. J., & Kalénine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. *Annals of the New York Academy of Sciences*, *1191*(1), 201-218.

Bryden, P. J., Pryde, K. M., & Roy, E. A. (2000). A performance measure of the degree of hand preference. *Brain Cognition* 44, 402-414.

Calvo-Merino, B., Glaser, D. E., Grezes, J., Passingham, R.E., & Haggard, P. (2005). Action observation and acquired motor skills: an fMRI study with expert dancers. *Cerebral Cortex*, *15*(8), 1243–1249.

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*(3), 1148-1167.

Chawla, D., Rees, K., & Friston, J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *Nature Neuroscience*, *2*(7), 671 – 676. doi:10.1038/10230

Creem-Regehr S. H. (2009). Sensory-motor and cognitive functions of the human posterior parietal cortex involved in manual actions. *Neurobiology of Learning and Memory*, 91(2), 166–171 10.1016/j.nlm.2008.10.004

Csibra, G., & Gergerly, G. (2007). 'Obsessed with goals': functions and mechanisms of teleological interpretations of actions in humans. *Acta Psychologica*, 124(1), 60-78.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*(4), 1325-1335.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 189–210.

Gozli, D. G., West, G. L., & Pratt, J. (2012). Hand position alters vision by biasing processing through different visual pathways. *Cognition*, 124(2), 244-250.

Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, 207(1), 3–17

10.1111/j.1469-7580.2005.00426.x

Hayes, S. M., Nadel, L., & Ryan, L. (2007). The effect of scene context on episodic object recognition: parahippocampal cortex mediates memory encoding and retrieval success. *Hippocampus*, *17*(9), 873-889.

Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., & Owen, A. M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage*, *50*(3-3), 1313–1319. http://doi.org/10.1016/j.neuroimage.2009.12.109

Hrkać, M., Wurm, M. F., & Schubotz, R. I. (2014). Action observers implicitly expect actors to act goal- coherently, even if they do not: An fMRI study. *Human Brain Mapping*, *35*(5), 2178-2190. doi: 10.1002/hbm.22319.

Hrkać, M., Wurm, M. F., Kühn, A. B., & Schubotz, R. I. (2015). Objects mediate goal integration in ventrolateral prefrontal cortex during action observation. *PloS one*, *10*(7), e0134316.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G., (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biol*, *3*(3), e79.

Johnson-Frey, S. H. (2004). The neural bases of complex tool use in humans. *Trends in Cognitive Sciences*, *8*(2), 71-78.

Kalénine, S., Shapiro, A. D., Flumini, A., Borghi, A. M., & Buxbaum, L. J. (2013). Visual context modulates potentiation of grasp types during semantic object categorization. *Psychonomic Bulletin & Review*, 21(3), 645-651. doi: 10.3758/s13423-013-0536-7

Kamps F.S., Julian J.B., Kubilius J., Kanwisher N. & Dilks D. D. (2016). The occipital place area represents the local elements of scenes, *NeuroImage*, 132, 417-424. doi:10.1016/j.neuroimage.2016.02.062

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, *15*(8), 352-357.

Lohmann, G., Müller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., … & von Cramon, D. Y. (2001). LIPSIA—A new software system for the evaluation of functional magnetic resonance images of the human brain. *Computerized Medical Imaging and Graphics*, *25*(6), 449–457.

Malcolm, G. L, Groen, I. I., & Baker, C. I. (2016). Making Sense of Real-World Scenes. *Trends in Cognitive Sciences*, *20*(11), 843-856.

Menon V., Adleman N.E., White C.D., Glover G.H., Reiss A.L. (2001). Error-related brain activation during a Go/NoGo response inhibition task. *Human Brain Mapping*, *12*, 131–143.

Ramsey R., Cross, E. S., & Hamilton, A. F. D. C. (2011). Eye Can See What You Want: Posterior Intraparietal Sulcus Encodes the Object of an Actor's Gaze. *Journal of Cognitive Neuroscience*, *23*(11), 3400-3409.

Rezaee, M., Shojaee, M., Ghasemi, A., Moghaddam, A., & Momeni, M. (2010). Which Factors Affect Hand Selection in Adults? Combined Effects of Ocular Dominance, Task Demand and Object Location. *Journal of Human Kinetics, 26*, 39-44.

Rubia K., Smith A.B., Brammer M.J., Taylor E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *Neuroimage, 20*, 351–358.

Schubotz, R. I., Wurm, M. F., Wittmann, M. K., & von Cramon, D. Y. (2014). Objects tell us what action we can expect: Dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Frontiers in Psychology*, *5*, 636. doi: 10.3389/fpsyg.2014.00636

Singh-Curry, V., & Husain M. (2009). The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. *Neuropsychologia*, *47*(6):1434-1448.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, *44*(1), 83–98.

Toney, A., & Thomas, B. H. (2006). Considering Reach in Tangible and Table Top Design. *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*.

Tyler, L. K., Stamatakis, E. A., Bright, P., & Acres, K., Abdallah, S., Rodd, J. M., & Moss, H. E. (2004). Processing objects at different levels of specificity. Journal of Cognitive Neuroscience, 16, 351–362.

Van Overwalle, F., & Beatens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, 48 564-584.

Wigget, A. J., & Downing, P. E. (2011). Representation of Action in Occipito-temporal Cortex. *Journal of Cognitive Neuroscience*, *23*(7), 1765-1780. doi: 10.1162/jocn.2010.21552.

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI time-series revisited—Again. *Neuroimage*, 2, 173–181.

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. *Human Brain Mapping*, *32*(12), 2141-2150. doi: 10.1002/hbm.21176.

Wurm, M. F., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: contextual information modulates action recognition. *Neuroimage*, *59*(2), 1551-1559. doi: 10.1016/j.neuroimage.- 2011.08.038

Wurm, M. F., von Cramon, D. Y., & Schubotz, R. I. (2012). The context-object-manipulation triad: cross talk during action perception revealed by fMRI. *Journal of Cognitive Neuroscience*, *24*(7), 1548-1559.

Wurm, M. F., Hrkać, M., Morikawa, Y., & Schubotz, R. I. (2014). Prediciting goals in action episodes attenuates BOLD response in inferior frontal and occipitotemporal cortex. *Behavioral Brain Research, 274,* 108-117. doi: 10.1016/j.bbr.2014.07.053

Wurm, M. F., & Schubotz, R. I. (2017). What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychonomic Bulletin & Review*, *24*(2), 503-509. Advance online publication. http://dx.doi.org/10.3758/s13423-016-1108-4

Wurm, M. F., Artemenko, C., Giuliani, D., & Schubotz, R. I. (2017). Action at its place: Contextual settings enhance action recognition in 4-8 year old children. *Developmental Psychology*, *53*(4), 662-671.

Zhang, S., & Chiang-shan, R. L. (2012). Functional connectivity mapping of the human precuneus by resting state fMRI. *Neuroimage*, *59*(4), 3548–3562.

Zimmermann, E., Schnier, F., Lappe, M., (2010). The contribution of scene context on change detection performance. *Vision Research*, *50*(20), 2062-2068.