# Title: Predictive impact of contextual objects during action observation: evidence from fMRI

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

[1]Department of Psychology, Westfälische Wilhelms-Universität, 48149 Münster, Germany
[2]Department of Neurology, University Hospital Cologne and University of Cologne, 50937 Cologne, Germany
[3]Center for Mind/Brain Sciences (CIMeC), University of Trento, Rovereto, TN, Italy
[4]Institute of Neuroscience and Medicine (INM3), Cognitive Neuroscience, Research Centre Jülich, 52425 Jülich, Germany

Corresponding author:
Nadiya El-Sourani
Fliednerstr. 21,
48149 Münster
n_elso02@uni-muenster.de

## Abstract

The processing of congruent stimuli, such as an object or action in its typical location, is usually associated with reduced neural activity, probably due to facilitated recognition. However, in some situations, congruency increases neural activity, e.g. when objects next to observed actions are likely vs. unlikely to be involved in forthcoming action steps. Here, we investigated whether the processing of contextual cues during action perception is driven by their (in-)congruency and, thus, informative value to make sense of an observed scene. Specifically, we tested whether both highly congruent contextual objects (COs), which strongly indicate a forthcoming action step, and highly incongruent COs, which require updating predictions about possible forthcoming action steps, thus provide more anticipatory information about the action course than moderately congruent COs. In line with our hypothesis that especially the inferior frontal gyrus (IFG) subserves the integration of the additional information into the predictive model of the action, we found highly congruent and incongruent COs to increase bilateral activity in action observation nodes, i.e. the IFG, the occipitotemporal cortex and the intraparietal sulcus. Intriguingly, Brodmann Area 47 was significantly stronger engaged for incongruent COs reflecting updating of prediction in response to conflicting information. Our findings imply that the IFG reflects the informative impact of COs on observed actions, by using contextual information to supply and update the currently operating predictive model. In case of incongruent CO, this model has to be reconsidered and extended towards a new overarching action goal.

**Keywords**
Action observation, action prediction, congruence, context, IFG

## 1. Introduction

In daily life, it is essential to understand what people around us are doing, that is, to predict the goal they currently aim to achieve (cf. van Overwalle & Beatens, 2009; Caspers, Zilles, Laird, & Eickhoff, 2010). To this end, action observers can exploit various sources of information, including not only moving body parts (i.e. manipulation movements) and manipulated objects, but also various contextual factors, such as the room (Wurm & Schubotz, 2012), the actor (Hrkac et al., 2013), additional objects in a scene (*contextual objects*, COs; El-Sourani et al., 2018), and spatial relations between objects and agents (El-Sourani et al. in prep). While the influence of contextual information on object recognition has been intensively investigated (Bar, 2004; Boyce, Pollatsek, & Rayner, 1989; Hayes, Nadel, & Ryan, 2007; Zimmermann, Schnier, & Lappe, 2010; Barenholtz, 2013), its impact on action understanding has so far been addressed by only a few studies. These studies found that participants process contextual information spontaneously, that is, without task requirements: While participants need longer to recognize an action when it takes place in an incompatible vs. a compatible or a neutral room (Wurm & Schubotz, 2012), action-compatible room information can help when actions are difficult to recognize, leading to increased recognition accuracy (Wurm & Schubotz, 2017). Moreover, brain activation during action recognition suggested interference effects of action-incompatible contexts rather than facilitation effects of action-compatible contexts (Hrkac et al., 2014; Wurm & Schubotz, 2012). For example, when the manipulated object does not fit to room and manipulation, or when the manipulation does not fit to object and room, activity increased in brain regions associated with object and manipulation processing, respectively (Wurm, von Cramon, & Schubotz, 2012). In particular, the inferior frontal gyrus (IFG) has been frequently reported for action-incompatible information processing, for instance when actions took place in incompatible rooms (Wurm & Schubotz, 2012). Also, an increase in IFG activity was found when participants observed an actor performing actions that did not match their current goal, supposedly reflecting attempts to integrate incoherent action steps into a common goal (Hrkać, Wurm, & Schubotz, 2014). So far, the involvement of the IFG underlines its central role in integration of contextual information during action perception *(*Badre & Wagner, 2007; Kilner, 2011), and, from a broader perspective, its role in effortful contextual integration in different cognitive domains, including language (Poldrack et al., 1998; Smirnov et al., 2014; van Schie et al., 2006). These findings indicate that different types of contextual information impact on the processing of observed actions.

In a recent fMRI study (El-Sourani et al., 2018), we focused on *contextual objects (COs),* which are part of an observed action scene, yet not part of the action itself. By modulating the semantic relation (*Goal Affinity*) as well as the spatial relation (*Location Ergonomics*) of the CO to the observed action, we investigated under which conditions such task-irrelevant objects modulate an action observer's brain activity. We argued that such effects can be taken to reflect attempts to incorporate these COs into an internal model of the observed action in order to anticipate an overarching action goal. Functional MRI results confirmed that COs are processed during action observation, even though participants' attention was tied to the observed action and considering the COs was not necessary to identify the action at hand. Contrary to the previously observed interference effects for action-incompatible information, for instance an action-incompatible room (Wurm & Schubotz, 2012), we found significant engagement of brain areas associated with object-related action representation when COs were *highly compatible* with the observed action, e.g. a frying pan next to cracking an egg. Specifically, Brodmann Area (BA) 44 and BA 45 of the IFG showed increased activation when the location of the CO and its semantic relation to the observed action strongly implied its use in immediately upcoming action steps (El-Sourani et al., 2018). This apparent discrepancy may be explained by the different operationalization of context-action incompatibility in the different studies, and point towards a more specific interpretation of the IFG's function in action observation. For example, Wurm and Schubotz (2012) investigated the effect of compatibility and incompatibility of room information on action perception (e.g. squeezing lemons in the bathroom). By contrast, in El-Sourani et al. (2018) the observed action (e.g. cracking an egg) was generally compatible with the room information implied by the CO but was more or less associated with the CO itself (frying pan vs. wine opener). Hence, no strong *conflict* was induced by low congruent COs, suggesting that this object category was processed as part of the room, or room category, rather than with regard to a potential usage by the actor. As we are used to be surrounded by room-compatible objects of a low congruence to our currently performed action, such low-congruence COs can usually be ignored, whereas COs with a high congruence (e.g. a frying pan) are probably perceived as comparatively *highly informative* (and thus relevant) for action observers in such a way that a specific overarching action goal (e.g. preparing scrambled eggs) can be inferred. Importantly, a similar degree of informativeness and thus relevance conceivably also applies to highly incongruent contextual information, as mismatches between the observed action and contextual information signal the need to reconsider the action’s anticipated outcome (cf. Wurm & Schubotz, 2012; Hrkac et al., 2015). In this case, the current predictive model of the observed action should be revised (cf. Kilner, 2007; Kilner et al., 2011).

Extending upon these findings, the present fMRI study aimed at investigating if COs that neither match the current action nor the according room category have a strong impact on action observation. More specifically, as they point to actions associated with a different room category, they should generate a real conflict for, and hence complicate, goal inference.

To test this assumption, participants watched action videos containing COs that varied with regard to three levels of the factor goal affinity: they either matched the currently observed action and the context background (highly congruent CO), only the context background but not the action (low congruent CO) or neither the context background nor the action (incongruent CO). To test for replication of previous findings, we also implemented the factor LOCATION ERGONOMICS, with varying positions of the COs on the table on which the action was performed (cf. El-Sourani et el., 2018).

We hypothesized that a particularly high compatibility between an observed action and its context (highly congruent COs) as well as a particularly high action-context incompatibility (incongruent COs) both provide rich information regarding potentially upcoming next action steps. Therefore, we expected brain activity to initially reflect an in-depth processing of these two object categories, demonstrated by an assumed overlap of neural activity elicited by highly congruent and incongruent COs, as compared to low congruent COs. Based on previous findings, this effect was expected to be reflected in brain areas linked to object-related action representations, especially the occipito-temporal cortex (OTC; Wigget & Downing, 2011) and the inferior parietal lobule (IPL; Buxbaum & Kalénine, 2010; Schubotz et al., 2014), as the perception of objects can already imply manipulation and action (Buxbaum et al., 2006; Johnson-Frey, 2004; Schubotz et al., 2014). Most importantly, we particularly focused on the IFG due to its role in the retrieval and integration of action-relevant semantic information (Casper et al., 2010; Badre & Wagner, 2007). As outlined above, we argue that previous findings can be reconciled if the IFG not simply reflects integration attempts but rather signals how informative a contextual object is with regard to an observed action's anticipated outcome. If so, IFG activity should be low for COs with a low congruence to the observed action, but strongly engaged for action scenes with (a) COs that are highly congruent to the observed action and (b) COs that neither match the observed action nor the room category (incongruent COs) in which the action is observed. More specifically, within the IFG, BA 44 is suggested to be involved in structuring sequences to realize particular outcomes (Fiebach and Schubotz, 2006; Grafman, 2002) thereby potentially supporting the anticipation of upcoming action steps during action perception (Csibra, 2007; Fagg and Arbib, 1998; Friston et al., 2011; Schubotz & von Cramon, 2009). While BA 45 activation supports the selection among competitively activated semantic representations (Badre et al., 2005; Gold et al., 2006; Moss et al., 2005), BA 47 is suggested to be involved in top-down semantic retrieval of goal-relevant knowledge (e.g., when participants are asked to think of unusual functions of an object, Kröger et al., 2012). Hence, we expected BA 44 and 45 to be engaged to a similar degree for both highly congruent COs and incongruent COs, whereas BA 47 might be stronger engaged by incongruent COs as compared to highly congruent COs.

## 2. Materials and Methods

### *2.1 Participants*

Thirty-five right-handed subjects (20 females; 24,6±3.1 years old; range, 19-30 years) with normal or corrected-to-normal vision participated in the study. Three of these participants were excluded due to either poor performance or strong head motion (more than 3mm between two scans). None of the remaining thirty-two participants reported a history of medical, neurological/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 before they participated in the study. Afterwards, participants either received course credits or reimbursement.

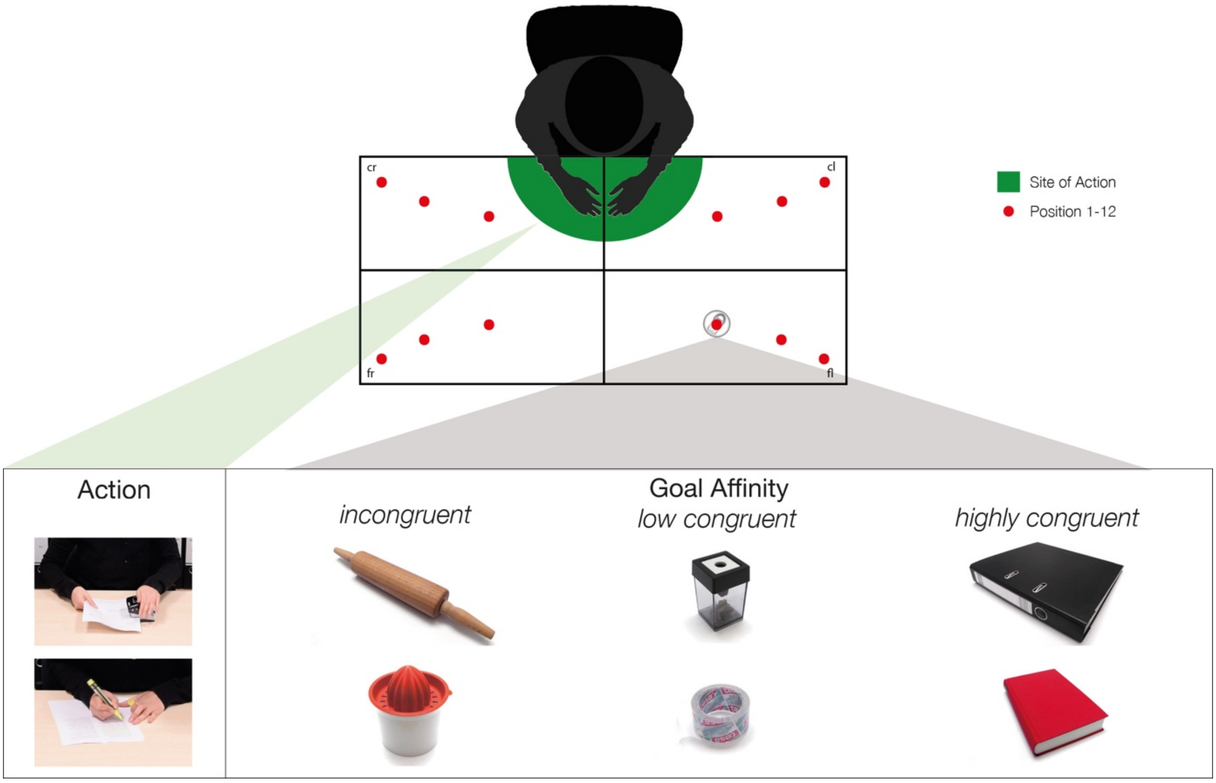


*Figure 1.* Example stimuli for implementing the three factor levels for goal affinity of COs, depicted for two different actions (punching and writing)**.** Red dots refer to possible CO positions on the table (*Location Ergonomics)*.

### *2.2 Stimuli*

Stimuli were presented using Presentation 18.1 (Neurobehavioral Systems, San Francisco, CA, USA). In total, participants were presented with 360 action videos (action trials). Action trials were intermixed with 72 questions trials (20%), i.e., written action descriptions that referred to these actions (see 2.3 Task). Action and question trials had a duration of 6s and consisted of either an action video (3s) or a question (3s), followed by a fixation phase (3s). A variable jitter

(500, 1000, 1500 ms) was included after the fixation phase in order to enhance temporal resolution of the BOLD response. Finally, in 5% of the trials a null event (fixation cross) was implemented (6s).

All action videos were performed by the same actress throughout the experiment and were filmed from a third person perspective (3pp). 72 actions were used. Each of the actions was performed in its typical setting that was either the kitchen (39 actions) or the office (33 actions), i.e., action and context background were always compatible (cf. El-Sourani et al. 2018). Each action video depicted a single object-directed action with two target objects. Out of the 360 action videos, 288 contained an additional contextual object (CO) that was positioned in front of the actress on the table (Figure 2). In a pilot study (N=24), action videos with and without a CO were tested with regard to the recognizability of the action and the CO, and only those actions and COs were employed that were recognized by all participants.

COs varied according to two experimental factors: GOAL AFFINITY*,* the semantic relation of the CO to the observed action, and LOCATION ERGONOMICS*,* the spatial relation of the CO to the observed action. Note that the latter factor was not relevant for testing the hypotheses of the present study, but was only employed to replicate previous findings from the precursor study (El-Sourani et al., 2018). The factor GOAL AFFINITY had three levels:

1. *Highly congruent CO* ($GA_{high}$), depicting COs that are compatible with the context background and the action
2. *Low congruent CO* ($GA_{low}$)*,* depicting COs that are compatible with the context background but not the action
3. *Incongruent CO* ($GA_{no}$)*,* depicting COs that are neither compatible with the context background nor with the action

*Goal affinity* was initially quantified on the basis of subjective ratings of a large sample (N = 500) of students, in which participants had to rate the associative strength of objects (N = 144) and actions (N = 72). Based on this pilot data, objects were assigned to four different levels of goal affinity ranging from “very low associated” to “very high associated”. Subsequently, COs of level one (“very low associated”) and level two (“rather low associated”) were merged to form the level *low congruent CO* in the present study, whereas level three (“rather high associated”) and level four (“very high associated”) were merged to form the level *highly congruent CO ,* corresponding to the categories of our previous study (El-Sourani et al., 2018) To determine *incongruent COs,* we conducted a further pilot study, where COs belonging to the one context background (kitchen) were tested for their probability of occurrence in the other context background (office), and vice versa. 24 right-handed participants rated on a 6-Point

Likert scale how strongly the presented CO (e.g. rolling pin) fits into the alternate room category (e.g. office). Objects creating the biggest mismatch according to these pilot data were chosen for the level *incongruent CO*.

As in our previous study (El-Sourani et al., 2018), the factor LOCATION ERGONOMICS was implemented by varying locations on the table corresponding to close-right, close-left, far-right and far-left with regard to the action site (*Figure 1*).

Subsequently, each of the 72 actions was paired (using Adobe Premiere Pro CS, Adobe Photoshop and/or Matlab) with two COs of two different goal affinity levels ensuring a balanced distribution of the GOAL AFFINITY levels, that is, 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 3 goal affinity levels x 8 occurrences = 288 action videos with a contextual object).

Resulting videos were presented in a pseudo-randomized fashion by avoiding direct repetition of the presented action, and the goal affinity and location of the CO. Levels of both factors were presented in an evenly distributed manner.

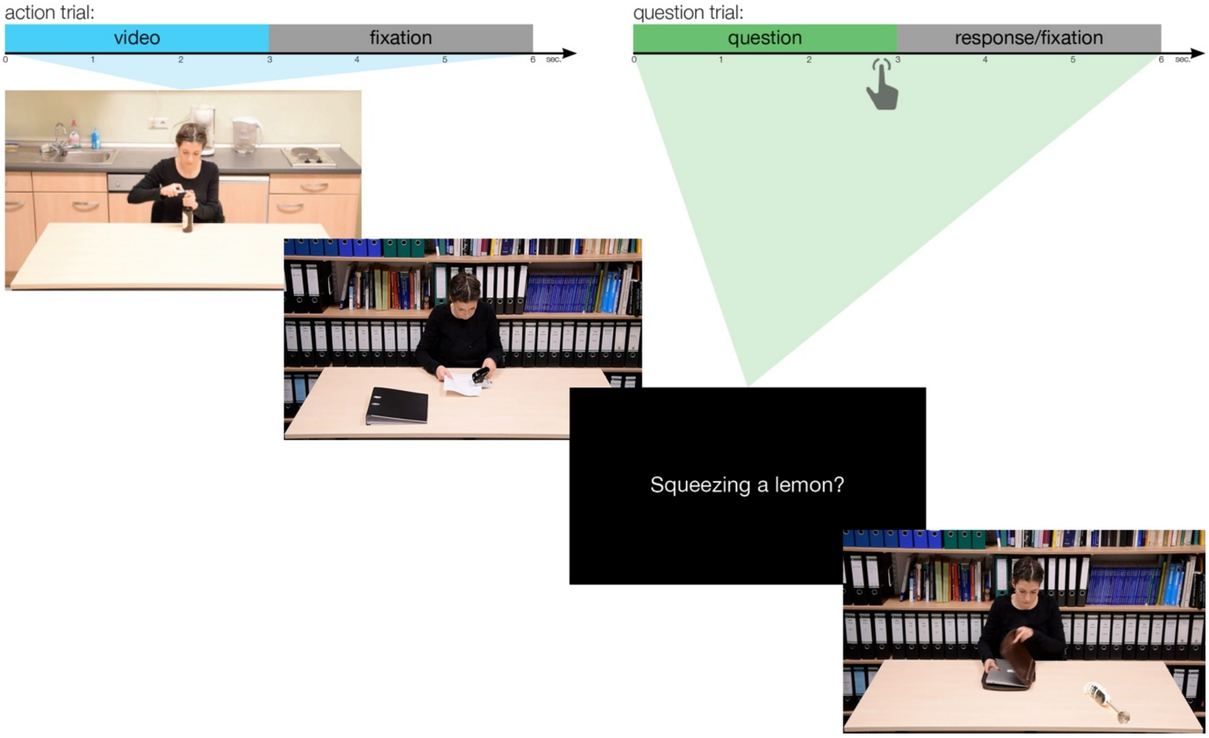


*Figure 2.* Schematic diagram of the task. Action trials consisted of an action video (3s) and a fixation phase (3s). Question trials consisted of a question regarding the preceding video trial (n-1), followed by a response and fixation phase. Retrieved from El-Sourani et al. (2018) and partly modified.

### *2.3 Task*

To keep the participants' attention on track while watching the videos, we asked them to watch the video clips attentively and to respond to the action description (20%) that either referred to the content of the preceding video (50%) or not (50%). Participants had to either accept or reject the action description on a two-button response box.

### *2.4 fMRI Image Acquisition*

Imaging was performed on a 3 Tesla Siemens Magnetom Prisma MR tomograph using a 20channel 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 and headphones in order to attenuate the disturbance by 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 33 adjacent axial slices with a slice thickness of 3 mm and a gap of 1 mm, which resulted in a voxel size of 3 x 3 x 4 mm. Images were acquired in interleaved 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).

For stimuli presentation, 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 participant's head, so that they 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, we controlled for recognizability of actions and contextual objects using the final video selection. 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 SPM12 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm/software/spm12/). Functional images were slice-timed to the middle slice to correct for differences in slice acquisition time. To correct for three-dimensional motion, individual functional MR (EPI) images were realigned to the mean epi image and further motion correction estimates were inspected visually. The anatomical scan was co-registered (rigid body transformation) to the mean functional image. Each subject's co-registered anatomical scan was segmented into native space tissue components. The parameters obtained were applied to normalize the subject's functional scans to the template brain MNI space. Finally, the normalized images were spatially smoothed using a Gaussian kernel of 8mm3 full width at half-maximum (FWHM). A 128 s temporal high-pass filter was applied to the data to remove low-frequency noise.

### *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. The subject-specific six rigid-body transformations obtained from residual motion correction were included as covariates of no interest. Activations were analyzed time-locked to the onset of the videos and the analyzed epoch comprised the full duration (3s) of the presented videos and the reaction time in question trials (max. 3s). In order to make results as comparable as possible between the current study and our previous study (El-Sourani et al., 2018), we aimed at having a similar design regarding our regressors. Therefore, our GLM contained 15 regressors in total: 12 predictors for the experimental conditions, one predictor for videos without contextual objects (noCO), one including all the null events (6s fixation phase) and one predictor for question trials. The 12 experimental regressors were assigned to the level combination of the factor LOCATION ERGONOMICS (close-right, close-left, far-right and far-left) as well as the factor GOAL AFFINITY (high, low, conflict inducing). To test for the effects of the factor GOAL AFFINITY, low congruent COs served as control condition. Thus, to test for the effect of incongruent COs, all predictors containing incongruent COs were contrasted with all predictors containing low congruent COs ($GA_{no} > GA_{low}$) on a first level GLM. To replicate the main effect of GOAL AFFINITY as found in El-Sourani et al. (2018), high goal affinity regressors were contrasted with low goal affinity regressors ($GA_{high} > GA_{low}$).

### *2.5.3 Group analysis*

To obtain group statistics, the resulting contrast images of all participants for our contrasts of interests ($GA_{high} > GA_{low}$; $GA_{no} > GA_{low}$; $GA_{no} > GA_{high}$) were entered a second level random-effects analysis using a one-sample *t*-test across participants to test for significant deviations from zero. Subsequently we corrected for multiple comparisons using the false discovery rate (FDR) method with $p < 0.01$. Significant activation maps were superimposed on a ch2better.nii.gz atlas using MRIcron software (https://www.nitrc.org/projects/mricron).

### *2.5.4 ROI analysis*

To specifically test whether different areas of the IFG are differentially modulated by the compatibility of the presented contextual information, we performed a ROI analysis the contrasts of interest ($GA_{high} > GA_{low}$; $GA_{no} > GA_{low}$). Thus, low congruent COs served as baseline. Anatomical masks of left and right IFG were defined according to the automated anatomical labeling (AAL) atlas, implemented in SPM12 (Mazoyer & Joliot, 2002). To this end, we extracted mean beta scores per ROI and entered them into two-sided one-sample t-tests. Note that we aggregated beta values across hemispheres, as we did not hypothesize differential activation patterns regarding left and right IFG. To specifically test for a difference between incongruent and highly congruent COs regarding the different IFG areas (pars opercularis (BA 44), pars triangularis (BA 45), pars orbitalis (BA 47)), we performed one-sided paired-sample *t*-tests.

## 3. Results

### *3.1. Behavioral results*

Performance was assessed by error rates and reaction times (on correctly answered trials). The average response time was 1257.69 ± 39ms and the average error rate was low (2.47 ± 1.14%) indicating that participants attentively observed and recognized the actions.

### *3.2. fMRI results*

#### *3.2.1 Highly congruent COs ($GA_{high} > GA_{low}$)*

To test for the effect of high highly congruent COs, we contrasted action videos containing highly congruent COs with low congruent COs, irrespective of their location on the table. Largely replicating previous findings (El-Sourani et al. 2018), highly congruent COs increased activity in the posterior parietal cortex (PPC) and the OTC (posterior temporal gyrus (pMTG), fusiform gyrus and the lateral occipital complex (LOC)) bilaterally. In addition, right middle and superior frontal gyrus were significantly activated (*Figure 3*). Interestingly, in contrast to our previous study, where we found an increase of IFG activity for highly congruent objects only when they were positioned close right to the actress, the IFG here now became significantly activated independent of the position of the CO on the table (cf. *Figure 3*). The reverse contrast did not reveal any significant activation patterns after FDR-correction

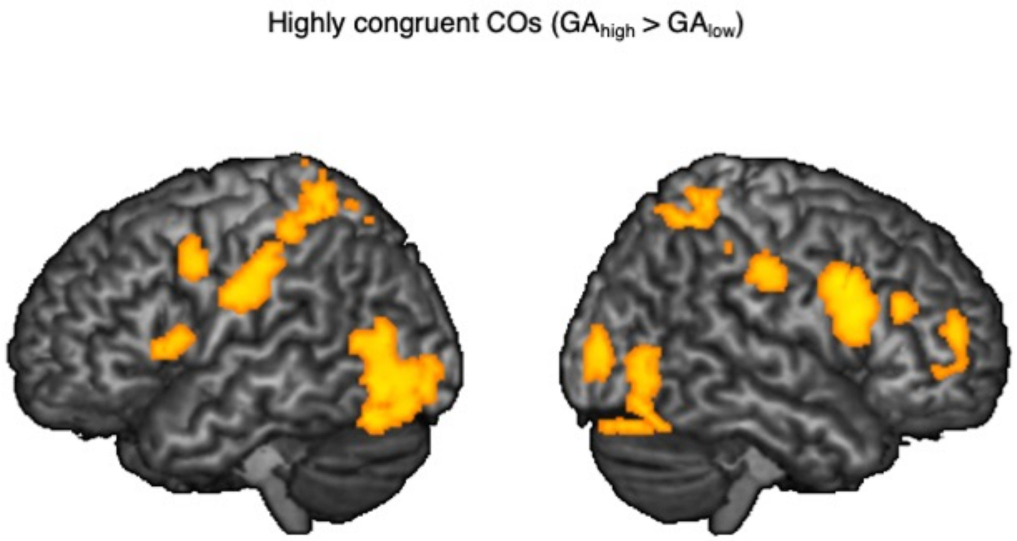


*Figure 3*. Brain activations for highly congruent vs. low congruent COs ($GA_{high} > GA_{low}$), FDR corrected at $p < 0.01$.

*Table 1.* fMRI activations for highly congruent COs ($GA_{high} > GA_{low}$)

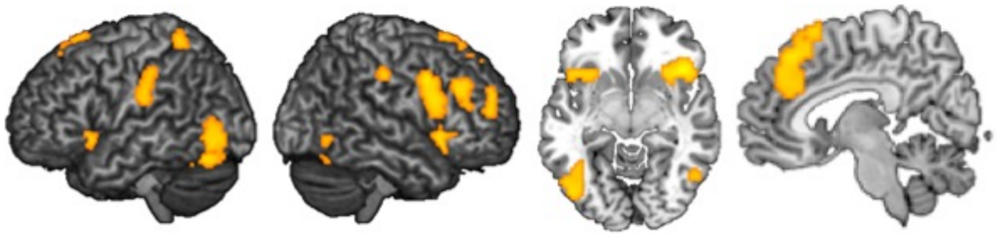
| | | | *MNI Coordinates* | | | |
|---|---|---|---|---|---|---|
| *Region* | *Hemisphere* | *BA* | *x* | *y* | *z* | *t-scores* |
| **Inferior/middle occipital gyrus** | l | 18/19 | -39 | -70 | -4 | 6.25 |
| | r | | 33 | -82 | 5 | 5.04 |
| **Fusiform gyrus** | l | 37 | -33 | -58 | -18 | 5.01 |
| | r | | 36 | -61 | -16 | 5.22 |
| **MTG /ITG** | l | 19 | -51 | -64 | 5 | 4.38 |
| | r | | 41 | -64 | 5 | 4.71 |
| **PMv** | l | 6 | -54 | 5 | 41 | 4.61 |
| | r | | 60 | 11 | 35 | 5.91 |
| **IFG (pars opercularis)** | l | 44 | -60 | 11 | 9 | 4.13 |
| | r | | 60 | 17 | 20 | 5.58 |
| **IPL /IPS** | l | 7 | -30 | -46 | 53 | 4.35 |
| | r | | 37 | -49 | 56 | 4.64 |
| **SPL/IPS** | l | 7 | -27 | -49 | 68 | 4.59 |
| | r | | 33 | -49 | 60 | 5.38 |
| **Supramarginal gyrus** | l | 40 | -60 | -22 | 41 | 4.99 |
| | r | | 54 | -24 | 39 | 4.68 |
| **Postcentral gyrus** | l | | -36 | -40 | 65 | 4.63 |
| **Precuneus** | l | 7 | -9 | -52 | 74 | 4.13 |
| **Superior medial frontal gyrus** | l | 9 | -3 | 56 | 29 | 4.97 |
| **Superior frontal gyrus** | r | 10 | 32 | 54 | 11 | 4.29 |
| **Middle frontal gyrus** | r | 9 | 45 | 32 | 23 | 4.43 |
| **Cerebellum** | r | - | 18 | -55 | -43 | 4.94 |

R, right; L, left; x, y, z, MNI coordinates of peak voxel activation; MTG, middle temporal gyrus, ITG, inferior temporal gyrus; PMv, ventral premotor cortex; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; SPL, superior parietal lobule; $p < 0.01$ FDR-corrected for multiple comparisons. *ranging into left IFG (BA 44).

### *3.2.2 Incongruent COs ($GA_{no} > GA_{low}$)*

Incongruent versus low congruent COs yielded significantly increased activation in the bilateral anterior dorsal insula, ACC, SMA and IFG (BA47). Moreover, we found a significant engagement of the left PPC, OTC as well as the right middle and superior frontal gyri (*Figure 4A*). As expected, brain areas partly overlapped with those engaged for highly congruent COs (*Figure 4B*). Again, the reverse contrast did not reveal any significant activation patterns after applying FDR-correction. Finally, directly contrasting incongruent COs with highly congruent COs did not reveal any significant whole-brain effects after FDR-correction.

a) Incongruent COs ($GA_{no} > GA_{low}$)



b) Overlay of highly congruent and incongruent COs

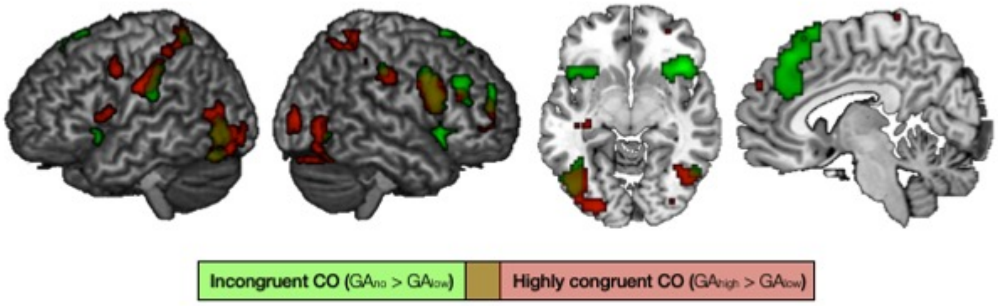


*Figure 4.* a) depicts brain activation patterns for conflict inducing COs, FDR corrected at $p < 0.01$. b) shows an overlay of activations for high goal affinity (red) and conflict inducing COs (green). Activations of both object categories overlap in the IPL and OTC as well as in the right PMv and right IFG.

*Table 2*. fMRI activations for incongruent COs ($GA_{no} > GA_{low}$)

| | | | *MNI Coordinates* | | | |
|---|---|---|---|---|---|---|
| *Region* | *Hemisphere* | *BA* | *X* | *y* | *z* | *t-scores* |
| **Inferior/middle occipital gyrus** | l | 19 | -45 | -73 | -10 | 5.65 |
| **Fusiform gyrus** | l | 37 | -30 | -52 | -19 | 5.78 |
| | r | | 33 | -58 | -19 | 4.5 |
| **MTG/ITG** | l | 37 | -51 | -61 | -4 | 4.51 |
| | r | | 52 | -58 | -5 | 4.71 |
| **IPL/ IPS** | l | 40 | -54 | -25 | 38 | 4.91 |
| | | 7 | -33 | -52 | 55 | 4.35 |
| **SPL/ IPS** | l | 7 | -28 | -49 | 68 | 4.67 |
| **Supramarginal gyrus /Postcentral gyrus** | l | 40/1 | -51 | -25 | 38 | 5.17 |
| | r | | 54 | 24 | 41 | 5.27 |
| **Insula** | l | 13 | -30 | 17 | -7 | 4.99 |
| | r | | 42 | 23 | -7 | 7.44 |
| **SMA** | l | 6 | -3 | 20 | 66 | 4.71 |
| | r | | 3 | 20 | 65 | 5.46 |
| **ACC/MCC** | l | 24/32 | -3 | 11 | 23 | 4.16 |
| | r | | 5 | 40 | 31 | 4.87 |
| **PMv** | r | 6 | 57 | 11 | 35 | 6.59 |
| **IFG (pars opercularis)** | r | 44 | 60 | 14 | 26 | 5.85 |
| **IFG (pars orbitalis)** | l | 47 | -54 | -25 | 38 | 5.09 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | r | | 47 | 23 | -4 | 5.62 |
| **Superior medial frontal gyrus** | l | 8 | -2 | 34 | 51 | 5.27 |
| | r | 8 | 3 | 35 | 50 | 6.81 |
| **Superior frontal gyrus** | r | 10 | 30 | 47 | 9 | 4.79 |
| **Middle frontal gyrus** | r | 46 | 27 | 56 | 23 | 6.10 |
| * | r | 9 | 42 | 35 | 30 | 6.00 |
| **Cerebellum** | l | | -38 | -72 | -22 | 4.73 |

R, right; L, left; x, y, z, MNI coordinates of peak voxel activation; MTG, middle temporal gyrus, ITG, inferior temporal gyrus, PMv, ventral premotor cortex; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; SPL, superior parietal lobule; SMA, supplementary motor area; IPS, intraparietal sulcus; ACC, Anterior cingulate cortex; MCC, middle cingulate cortex; $p < 0.01$ FDR-corrected for multiple comparisons; * ranging into IFG (pars triangularis; BA45)

### *3.2.3 ROI analysis of CO congruency effects in the IFG*

To assess a putative differential contribution of different sub-regions of the IFG to the processing of incompatible vs. compatible contextual information, we performed ROI analyses of the pars opercularis, pars triangularis and pars orbitalis roughly corresponding to BA 44, 45 and 47, respectively (*Figure 5*). We extracted beta values for highly congruent COs ($GA_{high} > GA_{low}$) and incongruent COs ($GA_{no} > GA_{low}$). One sample *t*-tests revealed significant activations for all conditions of interest ($GA_{high}$, $GA_{no}$) as compared to baseline ($GA_{low}$) (*Figure 5*). As we did not hypothesize differences in activation between left and right IFG we aggregated beta values of both hemispheres. Finally, one-sided paired sample *t*-tests revealed a significant difference between incongruent and highly congruent COs in BA 47 ($t_{(63)} = 2.479$; $p < 0.01$).

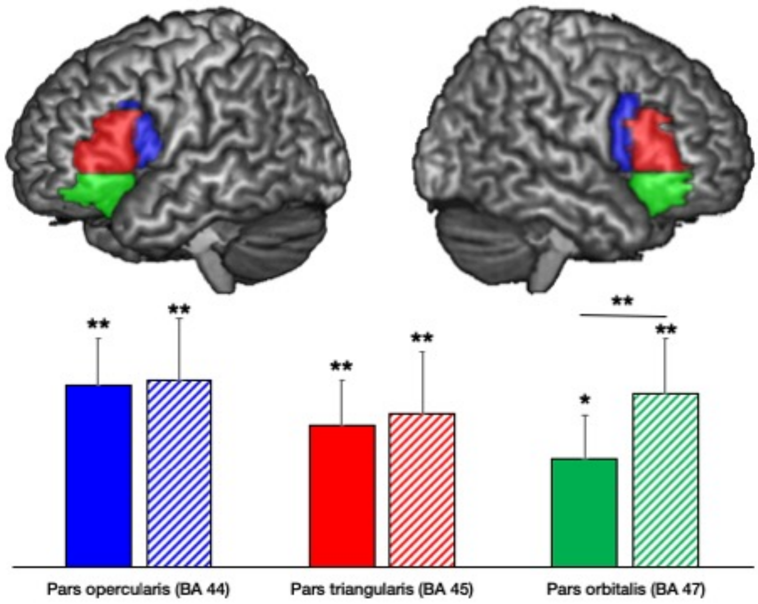


*Figure 5.* ROI analysis of IFG sub-regions according to the AAL atlas. Applied masks are illustrated in blue, red and green for BA 44, 45 and 47, respectively. Corresponding beta values for highly incongruent versus low congruent COs are depicted in full colored bars, whereas beta values for incongruent COs versus low congruent COs are depicted in striped colored bars. Beta values for right and left were aggregated. The two conditions significantly differed from zero in all sub-regions of the IFG. Regarding BA 47 (pars orbitalis), paired sample *t*-tests revealed a significant increase of activation for incongruent as compared to highly congruent COs. None of the other pairwise comparisons revealed a significant difference. * corresponds to a significance level of $p < 0.05$ and ** to a significance level of $p < 0.01$.

## 4. Discussion

When we observe others' actions, specific brain regions are involved in integrating this action with contextual information to enable the inference of action goals. This contextual information includes not only the environment and the actor, but as recently found also unused objects nearby (El-Sourani et al., 2018). In the current study, we aimed to better understand the processes underlying the latter effect. Specifically, we tested the assumption that the brain's engagement in processing contextual objects (CO) is not driven by the COs congruency or incongruency to the observed action, but rather to the CO's potential to inform expectations towards upcoming action steps.

Relative to COs with a low congruency to the observed action, both highly congruent and entirely incongruent COs were accompanied by increased brain activation at several action observation network (AON) sites, among of those, as hypothesized, the OTC and the PPC (especially the IPS). The same effect was found for the IFG, the area we had a special emphasis on due to its role in the processing of semantic information. Interestingly, BA 47 of the IFG was especially engaged for incongruent COs.

These findings support the view that when observing an action, the brain is particularly tuned to highly informative context. Contextual information may exert its impact via probabilistic associative knowledge about rooms in which certain classes of actions are frequently observed (Wurm & Schubotz, 2012) or about objects that are frequently used in the same sequence of actions (El-Sourani et al., 2018). In support of the latter assumption, posterior AON areas associating objects with actions, including the OTC (Grill-Spector et al., 2001; Wigget & Downing, 2011) and the IPS (Creem-Regehr, 2009; Grefkes & Fink, 2005; Singh-Curry & Husain, 2009; Ramsey, Cross, & Hamilton, 2011), were significantly more active for both highly congruent and incongruent COs as compared to low congruent COs. Moreover, processing incongruent COs engaged a set of brain areas related to conflict processing: the ACC and the anterior dorsal insula. These areas have been suggested to operate as a (response) inhibition network (Kana et al., 2007, see also Botvnick et al., 2001; Hoffstadter et al., 2014)), indicating that the processing of this object category also entailed processing its conflict to the observed action. Put into broader context, the efficient processing of an observed action scene includes the selection of sensory input that is crucial for informing the expectation of potential outcomes of the observed action (cf. Csibra, 2007; Kilner et al., 2007), irrespective of whether the information fits or contradicts the action.

The current study focused particularly on BOLD effects in the IFG, which is known for its role in retrieval and integration of semantic information (Caspers, Zilles, Laird, & Eickhoff, 2010; Kilner, 2011). We here aimed to extend this picture and hypothesized that the IFG is sensitive to the informative impact of COs with regard to a potential refinement of expectations of action outcomes. As expected, ROI analyses revealed significant engagement of all IFG compartments - BA 44, BA 45 and BA 47 - for both incongruent and highly congruent COs as compared to low congruent COs. The three sub-regions of IFG have been associated with different functions across different domains, including language (cf. Bookheimer, 2002; Liakakis et al., 2011), emotion processing (Car et al., 2003; Seitz et al., 2008) and creativity (Kröger et al., 2012). According to different accounts on the IFG function, (Koechlin & Summerfield, 2007; Udden & Bahlmann, 2012; Badre & Wagner, 2007), the IFG is hierarchically organized to a functional stepwise gradient along the rostro-caudal axis, where top-down control is exerted from anterior to posterior regions (Koechlin et al., 2003). In specific reference to action observation, the more posterior the area the more it is suggested to contribute to constrain the immediate action requirements or options. More anterior sites, in turn, are more content-independent and associated with high-level goals (Badre, Poldrack, Pare- Blagoev, Insler, & Wagner, 2005; Badre & D'Esposito, 2009; cf. Buckner, 2003). More specifically, BA 44 (together with premotor regions) supports structured sequence processing (cf. Udden & Bahlman, 2012) in order to realize a particular outcome (Fiebach and Schubotz, 2006; Grafman, 2002). Regarding BA 45 and 47, Badre and Wagner (2007) specified two sub-functions exerting cognitive control: strategic memory retrieval in anterior sites (BA 47) and post-retrieval selection among competing memory in BA 45. While semantic retrieval is necessary when bottom-up cues are not sufficient to activate goal-relevant information, post-retrieval selection is necessary to resolve the competition between simultaneously activated memory representations (e.g. grasping to clean vs. grasping to drink).

The assumed functions of the IFG compartments concur with our observed activation pattern: incorporating relevant CO information into an observed action scene in order to anticipate its outcome draws on all IFG sub-regions. However, while BA 44 and 45 have been recruited to a comparable degree by highly congruent and incongruent COs, BA 47 was significantly stronger engaged by incongruent COs. This underlines the ascribed function of the BA 47 in the controlled retrieval, i.e. a top-down process activating goal-relevant knowledge especially in the face of contradicting representations. Thus, increased BA 47 in response to incongruent versus highly congruent COs can be explained by the increased demand to retrieve an action outcome when confronted with conflicting action-related information. More specifically, incorporating incongruent COs and the observed action under an overarching goal requires a much higher level of abstraction, evoked by the associative strength of the CO to an incompatible room category and hence to actions associated with this room-category. Importantly, the observed pattern of activation in the IFG does not simply reflect demands on integrating more or less compatible contextual information (here: contextual objects) in the observed action. In that case, one would see a parametric increase of IFG activation with increasing incongruence, i.e. lowest IFG engagement for highly congruent COs. Instead, the IFG rather appears to respond to contextual information that specifies and/or enriches the interpretation of an observed action and ignores contextual information that is less informative for action interpretation.

Taken together, our findings imply that the brain cares for the informative value of contextual objects when observing an action. More specifically, we suggest that the IFG reflects the informational impact of COs on the observed action at several circumstances: either when the contextual information depicts a strong match so that the currently operating predictive model can be updated and specified towards a particular outcome; or when the contextual information reveals a strong conflict to the observed manipulation, in which case the currently operating predictive model has to be reconsidered and possibly extended towards a new overarching action goal.

*Acknowledgments*

We thank Monika Mertens for her great help with data acquisition. Further we thank Haris Orucevic for his assistance with the creation of the figures.

## Literature

Badre, D., Poldrack, R.A., Pare-Blagoev, E.J., Insler, R.Z., Wagner, A.D., 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral pre-frontal cortex. *Neuron 47*, 907–918

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883-2901.

Badre, D., & D'esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical?. *Nature Reviews Neuroscience, 10*(9), 659.

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

Barenholtz, E. (2014). Quantifying the role of context in visual object recognition. *Visual Cognition, 22*(1), 30-56.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D., (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624-652.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in cognitive sciences, 8*(12), 539-546.

Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cognitive, Affective, & Behavioral Neuroscience*, *7*(4), 356-366.

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.

Brandeis, R., Brandys, Y., & Yehuda, S. (1989). The use of the Morris water maze in the study of memory and learning. *International Journal of Neuroscience*, 48(1-2), 29-69

Buckner, R. L. (2003). Functional–anatomic correlates of control processes in memory. *Journal of Neuroscience*, *23*(10), 3999-4004.

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

Carter, C.S. et al. (1998) Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science 280*, 747–749

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.

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., 2007. Action mirroring and action interpretation: An alternative account. In: P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII* (pp. 435-459). Oxford: Oxford University Press

El-Sourani, N. Wurm, M. F., Trempler, I., Fink, G. R., & Schubotz, R. I. (2018). Making sense of objects lying around: How contextual objects shape brain activity during action observation. *Neuroimage*, 167, 429-437.

El-Sourani, N., Trempler, I., Wurm, M. F., Fink, G. R., & Schubotz, R. I. (in prep.). Is it mine or is it yours? Proximal unused objects invite third-person perspective action observers to create their own action space. *XX*, XX, XX-XX.

Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal–premotor interactions in primate control of grasping. *Neural Networks*, *11*(7-8), 1277-1303.

Fiebach, C. J. & Schubotz, R. I. (2006). Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains? *Cortex, Special Issue, 42*(4), 499-502

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.

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological cybernetics*, *104*(1-2), 137-160

Grafman, J., 2002. The structured event complex and the human prefrontal Cortex. In: Stuss DT. & Knight RT. Principles of frontal lobe function. Oxford University Press, New York

Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *Journal of Neuroscience*, *26*(24), 6523-6532.

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

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10-11), 1409-1422.

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.

Hoffstaedter, F., Grefkes, C., Caspers, S., Roski, C., Palomero-Gallagher, N., Laird, A. R., & Eickhoff, S. B. (2014). The role of anterior midcingulate cortex in cognitive motor control: evidence from functional connectivity analyses. *Human. Brain Mapping,* 35 (6), 2741–2753.

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.

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.

Kana, R. K., Keller, T. A., Minshew, N. J., & Just, M. A. (2007). Inhibitory control in high-functioning autism: decreased activation and underconnectivity in inhibition networks. *Biological psychiatry*, 62(3), 198-206.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, *8*(3), 159-166.

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

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. Trends in cognitive sciences, 11(6), 229-235.

Kröger, S., Rutter, B., Stark, R., Windmann, S., Hermann, C., & Abraham, A. (2012). Using a shoe as a plant pot: neural correlates of passive conceptual expansion. *Brain research, 1430*, 52-61.

Moss, H.E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., Tyler, L.K. (2005). Selecting among competing alternatives: selection and retrieval in the left inferior frontal gyrus. *Cereb. Cortex 15*, 1723–1735.

Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 781-795.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. N*euroimage*, 10(1), 15-35.

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

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., Simmons, A., Williams, S. C. R., Giampietro, V., Andrew, C. M., & Taylor, E. (2001). Mapping motor inhibition: conjunctive brain activations across different versions of go/no-go and stop tasks. *Neuroimage*, *13*(2), 250-261

Smirnov, D., Glerean, E., Lahnakoski, J. M., Salmi, J., Jääskeläinen, I. P., Sams, M., & Nummenmaa, L. (2014). Fronto-parietal network supports context-dependent speech comprehension. *Neuropsychologia, 63*, 293-303.

Schubotz, R. I., & von Cramon, D. Y. (2009). The case of pretense: Observing actions and inferring goals. *Journal of Cognitive Neuroscience*, *21*(4), 642-653.

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.

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

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15*(1), 273-289.

Uddén, J., & Bahlmann, J. (2012). A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1598), 2023-2032.

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

Van Schie, H. T., Toni, I., & Bekkering, H. (2006). Comparable mechanisms for action and language: neural systems behind intentions, goals, and means. *Cortex*, 42(4), 495-498.

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

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI Time-Series Revisited – Again. *NeuroImage*, *2*(3), 173-181.

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., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: contextual information modulates action recognition. *Neuroimage*, *59*(2), 1551-1559.

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.

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