Predictive Impact of Contextual Objects during Action Observation: Evidence from Functional Magnetic Resonance Imaging

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

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—for example, when objects next to observed actions are likely versus unlikely to be involved in forthcoming action steps. Here, we investigated using fMRI 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 future action step, and highly incongruent COs, which require updating predictions about possible forthcoming action steps, 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, that is, the IFG, the occipitotemporal cortex, and the intraparietal sulcus. Intriguingly, BA 47 was significantly stronger engaged for incongruent COs reflecting the 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 the case of an incongruent CO, this model has to be reconsidered and extended toward a new overarching action goal. ■

INTRODUCTION

In daily life, it is essential to understand what people around us are doing—that is, to predict the goal they currently pursue (cf. Caspers, Zilles, Laird, & Eickhoff, 2010; van Overwalle & Beatens, 2009). 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 (Hrkać, Wurm, & Schubotz, 2014), additional objects in a scene (contextual objects [COs]; El-Sourani, Wurm, Trempler, Fink, & Schubotz, 2018), and spatial relations between objects and agents (Brozzoli, Gentile, Bergouignan, & Ehrsson, 2013; Costantini, Committeri, & Sinigaglia, 2011). Although the influence of contextual information on object recognition has been intensively investigated (Barenholtz, 2014; Zimmermann, Schnier, & Lappe, 2010; Hayes, Nadel, & Ryan, 2007; Bar, 2004; Boyce, Pollatsek,

& Rayner, 1989), its impact on action understanding has so far been addressed by only a few studies. The results of these studies suggest that participants process contextual information spontaneously, that is, without task requirements: Although participants need longer time to recognize an action when it takes place in an incompatible versus 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 (Hrkać et al., 2014; Wurm & Schubotz, 2012). For example, when the manipulated object does not fit room and manipulation or when the manipulation does not fit object and room, neural 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 linked to 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

[1]Westfälische Wilhelms-Universität, Münster, [2]University Hospital Cologne and University of Cologne, [3]University of Trento, [4]Institute of Neuroscience and Medicine (INM3: Cognitive Neuroscience), Research Centre Jülich

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 et al., 2014). So far, the involvement of the IFG underlines its central role in the integration of contextual information during action perception (Kilner, 2011; Badre & Wagner, 2007) and, from a broader perspective, its role in effortful contextual integration in different cognitive domains, including language (Smirnov et al., 2014; van Schie, Toni, & Bekkering, 2006; Poldrack et al., 1999). These findings indicate that different types of contextual information may impact the processing of observed actions.

In a recent fMRI study (El-Sourani et al., 2018), we focused on COs (e.g., a tomato) that are part of an observed action scene (e.g., cutting a cucumber in a kitchen setting), yet not part of the action itself (e.g., cutting a cucumber). 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 may reflect attempts to incorporate these COs into an internal model of the observed action to anticipate an overarching action goal. fMRI results confirmed that COs are processed during action observation, even though the 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, for example, a frying pan next to cracking an egg. Specifically, Brodmann's 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 of brain activation in response to manipulating the congruency of contextual information during action observation may be explained by the different operationalization approaches of context–action incompatibility in these studies and point toward a more specific interpretation of the IFG's function in action observation. Whereas Wurm and Schubotz (2012) investigated the effect of compatibility and incompatibility of room information on action perception (e.g., squeezing lemons in the bathroom vs. kitchen), 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 lowly congruent COs, suggesting that this object category was processed as part of the room or room category, rather than concerning a potential usage

by the actor. As we are used to be surrounded by room-compatible objects with low congruency to our currently performed action, such lowly congruent COs can usually be ignored. In contrast, 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. Hrkać, Wurm, Kühn, & Schubotz, 2015; Wurm & Schubotz, 2012). In this case, the current predictive model of the observed action should be revised (cf. Kilner, 2011; Kilner, Friston, & Frith, 2007).

Extending upon these findings, the present fMRI study aimed at investigating if COs that neither match the current action nor the belonging room category exert a substantial impact on action observation. More specifically, when COs 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 contextual background (i.e., highly congruent CO), only the contextual background but not the action (i.e., lowly congruent CO), or neither the contextual background nor the action (i.e., incongruent CO). To replicate 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 al., 2018).

We hypothesized that particularly high compatibility between an observed action and its context (highly congruent COs) as well as 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 with lowly 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, Wurm, Wittmann, & von Cramon, 2014), as the perception of an object can already imply its manipulation and action (Schubotz et al., 2014; Buxbaum, Kyle, Tang, & Detre, 2006; Johnson-Frey, 2004). This is also referred to as the concept of "affordance," according to which the environment (including objects) implies information about possible actions (Fagg & Arbib, 1998; Jeannerod, 1994; Gibson, 1977, 1979). In

line with the concept of affordance, Tucker and Elis (1998, 2001) argued that objects can potentiate actions even when the goal of a task is not to directly interact with the object. Here, it is worth noting that the IPL and the IFG have been linked to the forward planning of potential object use (cf. Randerath, Goldenberg, Spijkers, Li, & Hermsdörfer, 2010).

In this study, we particularly focused on the IFG because of its role in the retrieval and integration of action-relevant semantic information (Caspers et al., 2010; Badre & Wagner, 2007). As outlined above, we argue that previous findings of IFG activation patterns can be reconciled if the IFG not simply reflects integration attempts but rather signals how informative a CO is concerning an observed action's anticipated outcome. If so, IFG activity should be low for COs with a low congruence to the observed action, but high 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 & Schubotz, 2006; Grafman, 2002), thereby potentially supporting the anticipation of upcoming action steps during action perception (Friston, Mattout, & Kilner, 2011; Schubotz &

von Cramon, 2009; Csibra, 2007; Fagg & Arbib, 1998). Although BA 45 activation supports the selection among competitively activated semantic representations (Gold et al., 2006; Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; 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 BA 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 with highly congruent COs.

METHODS

Participants

Thirty-five right-handed individuals (20 women; 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 because of either poor performance or strong head motion (more than 3 mm between two scans). None of the remaining 32 participants reported a history of medical, neurological/psychiatric disorders or substance abuse. The study
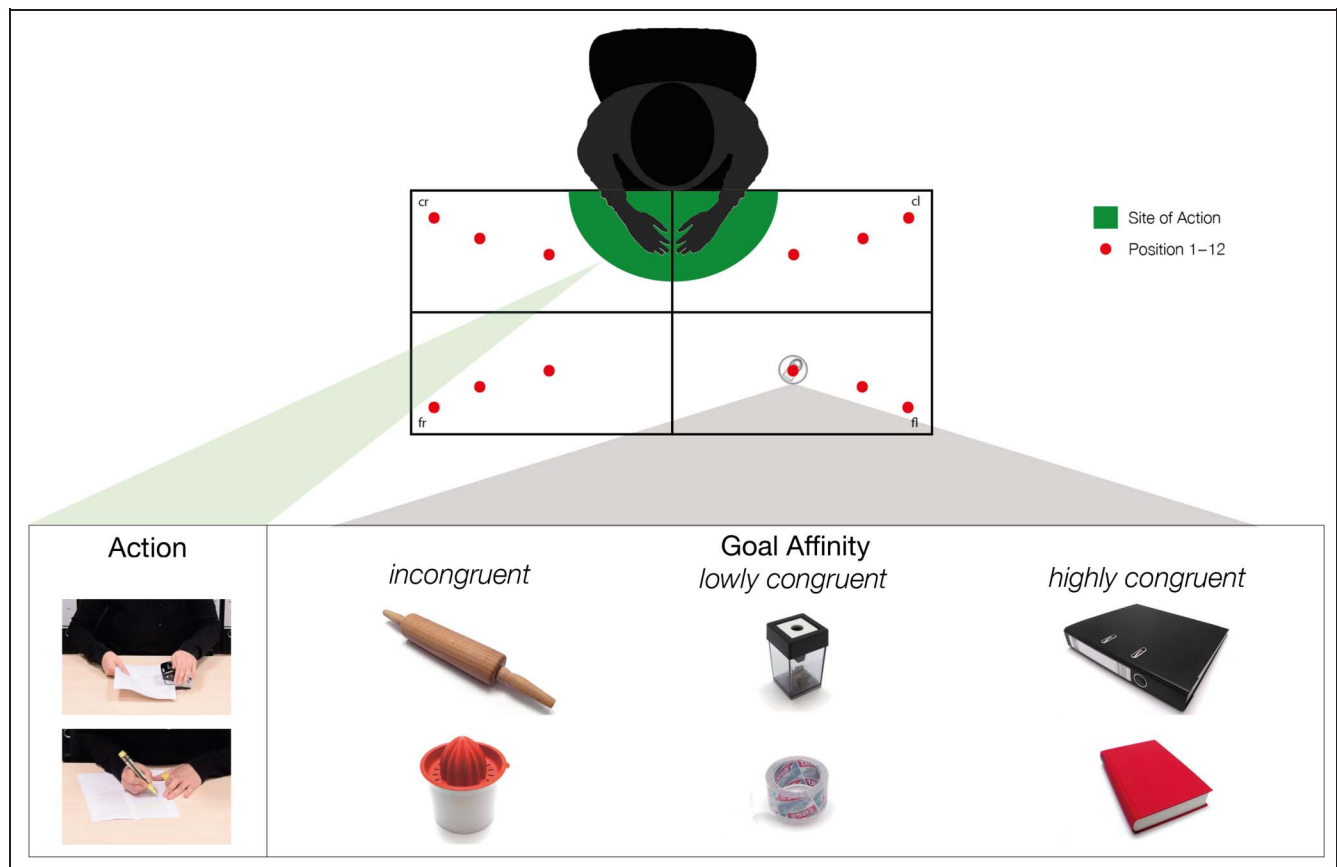
Figure 1. Example stimuli for implementing the three factor levels for goal affinity of COs, depicted for two different actions (punching and writing). Red dots refer to possible CO locations on the table (location ergonomics).
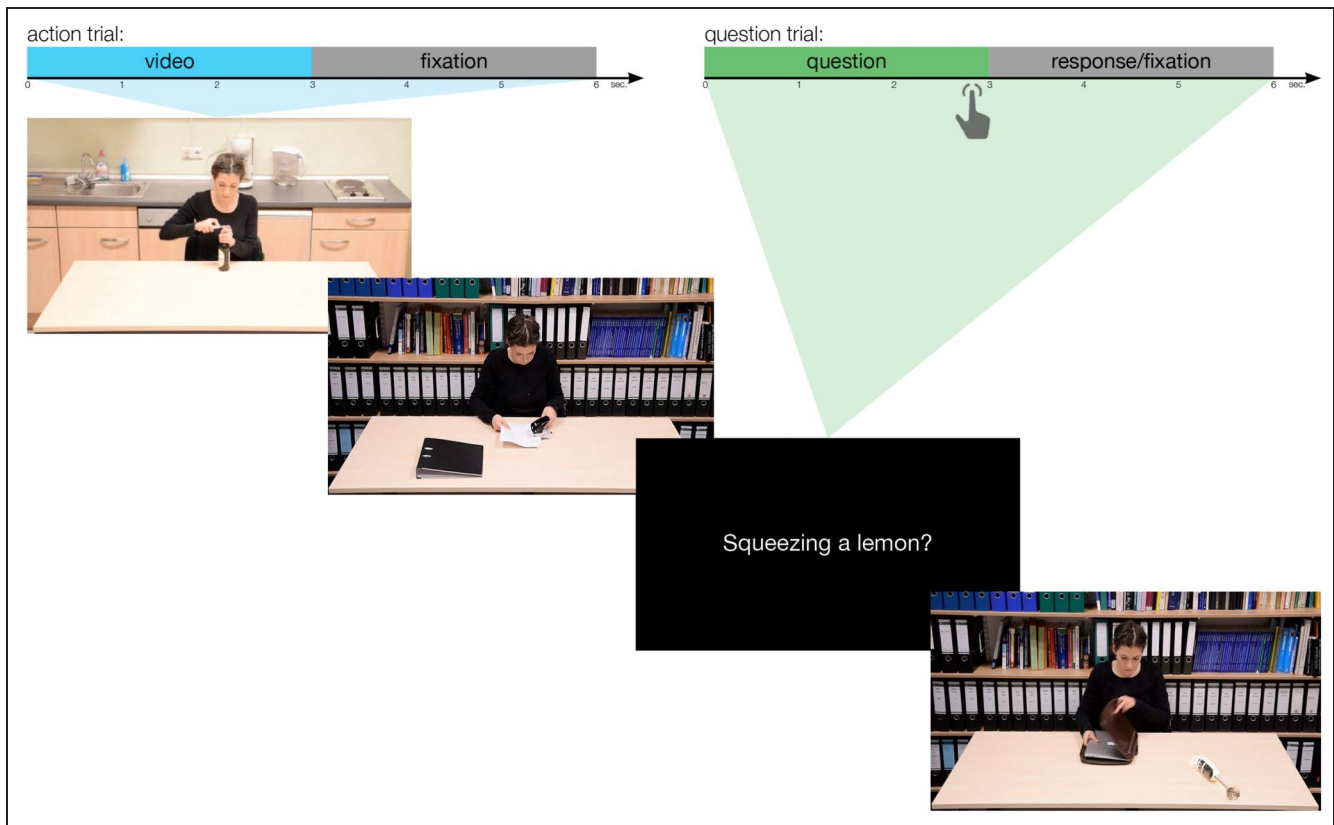
Figure 2. Schematic diagram of the task. Action trials consisted of an action video (3 sec) and a fixation phase (3 sec). Question trials consisted of a question regarding the previous video trial ($n - 1$), followed by a response and fixation phase. Retrieved from El-Sourani et al. (2018) and partly modified.

protocol was conducted following the 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).

Stimuli

Stimuli were presented using Presentation 18.1 (Neurobehavioral Systems). In total, participants were shown 360 action videos (action trials). Action trials were intermixed with 72 questions trials (20%), that is, written action descriptions that referred to these actions (see the Task section). Action and question trials had a duration of 6 sec and consisted of either an action video (3 sec) or a question (3 sec), followed by a fixation phase (3 sec). A variable jitter (500, 1000, and 1500 msec) was included after the fixation phase to enhance the temporal resolution of the BOLD response. Finally, in 5% of the trials, a null event (fixation cross) was implemented (6 sec).

All action videos were performed by the same actress throughout the experiment and were filmed from a third-person perspective. Seventy-two actions were used. Each of the actions was performed in its typical setting that was either a kitchen (39 actions) or an office (33 actions), that is, action and contextual background were always compatible (cf. El-Sourani et al., 2018). Each action video depicted a single object-directed action with two target objects. Of the 360 action videos, 288 contained an additional 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 for their recognizability, and only those actions and COs that were employed were recognized by all participants.

COs varied according to two experimental factors: goal affinity, that is, the semantic relation of the CO to the observed action, and location ergonomics, i.e., the spatial relation of the CO to the observed action. Note that the latter factor was not relevant for testing the hypotheses of this study but was employed to replicate findings from the precursor study (El-Sourani et al., 2018). Location ergonomics will be considered a follow-up study in depth.

The factor goal affinity had three levels:

1. highly congruent CO (GA_{high}), depicting COs that are compatible with the contextual background and the action;
2. lowly congruent CO (GA_{low}), depicting COs that are compatible with the contextual background but not the action; and

3. incongruent CO (GA$_{no}$), depicting COs that are neither compatible with the contextual background nor with the action.

Goal affinity was initially quantified based on 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 1 ("very low associated") and Level 2 ("rather low associated") were merged to form the level "lowly congruent CO" in this study, whereas Level 3 ("rather high associated") and Level 4 ("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 contextual background (kitchen) were tested for their probability of occurrence in the other contextual background (office), and vice versa. Twenty-four right-handed participants rated on a 6-point Likert scale how strongly the presented CO (e.g., rolling pin) fitted into the other 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 the locations of the CO on the table, corresponding to close-right (cr), close-left (cl), far-right (fr), and far-left (fl) concerning the action site (Figure 1).

Subsequently, each of the 72 actions was paired (using Adobe Premiere Pro CS, Adobe Photoshop, and/or MATLAB) with only two COs of two different goal affinity levels to have more variability within the stimulus material. This ensured a balanced distribution of the goal affinity levels, that is, videos containing a CO were arranged in a way that all goal affinity levels at all 12 positions occurred in an evenly distributed number (12 positions × 3 goal affinity levels × 8 occurrences = 288 action videos with a CO). In addition, each action was once seen without a CO, resulting in a total of 360 action videos.

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

Task

To keep the participants' attention focused on the videos, we implemented a cover task, in which participants were asked 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 using a two-button response box.

fMRI Image Acquisition

Imaging was performed on a 3-T 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 the response box. To minimize head and arm movements, head and arms were tightly fixated with form-fitting cushions. Furthermore, participants were provided with earplugs and headphones to attenuate the scanner noise. Whole-brain functional images were acquired using a gradient T2*-weighted single-shot EPI sequence sensitive to BOLD contrast (64 × 64 data acquisition matrix, 192 mm field of view, 90° flip angle, repetition time = 2000 msec, echo time = 30 msec). 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 × 3 × 4 mm. Images were acquired in interleaved order along the AC–PC plane to provide whole-brain coverage. After functional imaging, structural data were acquired for each participant using a standard Siemens 3-D T1-weighted MPRAGE sequence for detailed reconstruction of anatomy with isotropic voxels (1 × 1 × 1 mm) in a 256-mm field of view (256 × 256 matrix, 192 slices, repetition time = 2130, echo time = 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 visual field). In a pilot study, we controlled for recognizability of actions and COs using the final video selection. Only action videos in which the action and the CO could be identified by at least 95% of the participants were employed.

fMRI Data Analysis

fMRI Data Preprocessing

Brain image preprocessing and basic statistical analyses were conducted using SPM12 (Wellcome Department of Imaging Neuroscience; 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 3-D 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 coregistered (rigid body transformation) to the mean functional image. Each subject's coregistered 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 8 mm³ FWHM. A 128-sec temporal high-pass filter was applied to the data to remove low-frequency noise.

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 (3 sec) of the presented videos and the RT in question trials (max 3 sec). 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 COs, one including all the null events (6-sec 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 (highly congruent, lowly congruent, incongruent). To test for the effects of the factor goal affinity, lowly congruent COs served as a control condition. Thus, to test for the effect of incongruent COs, all predictors containing incongruent COs were contrasted with all predictors containing lowly congruent COs ($GA_{no} > GA_{low}$) on a first-level GLM. We contrasted highly congruent CO regressors with lowly congruent CO regressors ($GA_{high} > GA_{low}$) to replicate the main effect of goal affinity as found in El-Sourani et al. (2018).

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 into 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 $p < .01$. Significant activation maps were superimposed on a ch2better.nii.gz atlas using MRIcron software (https://www.nitrc.org/projects/mricron).

ROI Analysis

To test whether different areas of the IFG are differentially modulated by the compatibility of the contextual information, we performed an ROI analysis of the contrasts of interest ($GA_{high} > GA_{low}$; $GA_{no} > GA_{low}$). Thus, lowly congruent COs served as baseline. Anatomical masks of left and right IFG were defined according to the automated anatomical labeling atlas, implemented in

SPM12 (Tzourio-Mazoyer et al., 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.

RESULTS

Behavioral Results

The performance was assessed by error rates and RTs (of correctly answered trials). The average RT was 1257.69 ± 39 msec, and the average error rate was low (2.47% ± 1.14%), indicating that participants attentively observed and recognized the actions.

fMRI Results

Highly Congruent COs ($GA_{high} > GA_{low}$)

To test for the effect of highly congruent COs, we contrasted action videos containing highly congruent COs with lowly 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, fusiform gyrus, and the lateral occipital complex) bilaterally. Also, the right middle and superior frontal gyrus were significantly activated (Figure 3). In contrast to our previous study, where we found an increase of IFG activity or highly congruent objects only when they were positioned close right to the actress, the IFG here now became significantly activated independently 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 (Table 1).
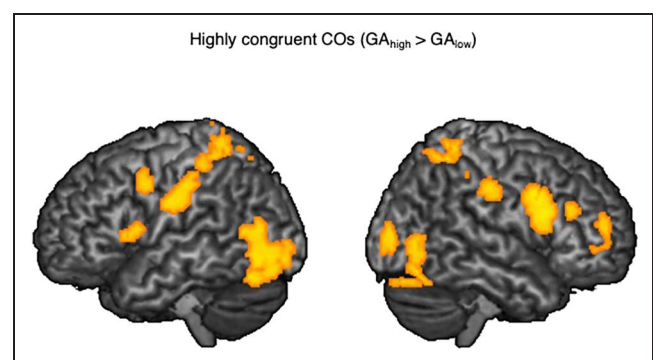
Figure 3. Brain activations for highly congruent versus lowly congruent COs ($GA_{high} > GA_{low}$), FDR-corrected at $p < .01$.

Table 1. fMRI Activations for Highly Congruent COs (GA$_{high}$ > GA$_{low}$)

Region	Hemisphere	BA	MNI Coordinates			t Scores
			x	y	z	
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; SPL = superior parietal lobule. $p < .01$, FDR-corrected for multiple comparisons. *Extending into left IFG (BA 44).

Incongruent COs (GA$_{no}$ > GA$_{low}$)

Incongruent versus lowly congruent COs yielded significantly increased activation in the bilateral anterior dorsal insula, ACC, SMA, and IFG (BA 47). 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 (Table 2).

ROI Analysis of CO Congruency Effects in the IFG

To assess a putative differential contribution of different subregions of the IFG to the processing of incompatible versus highly compatible contextual information, we performed ROI analyses of the pars opercularis, pars triangularis, and pars orbitalis roughly corresponding to BA 44, BA 45, and BA 47, respectively (Figure 5). We extracted beta values for highly congruent COs (GA$_{high}$ > GA$_{low}$) and incongruent COs (GA$_{no}$ > GA$_{low}$). As we did not hypothesize differences in activation between left and right IFG, we aggregated beta values of both hemispheres. One-sample t tests revealed significant activations for all conditions of interest (GA$_{high}$, GA$_{no}$) as compared with baseline (GA$_{low}$; Figure 5). Finally, one-sided paired-sample t tests revealed a significant difference between incongruent and highly congruent COs in BA 47, $t(63) = 2.479$, $p < .01$.

DISCUSSION

When we observe others' actions, specific brain regions are involved in integrating this action with contextual
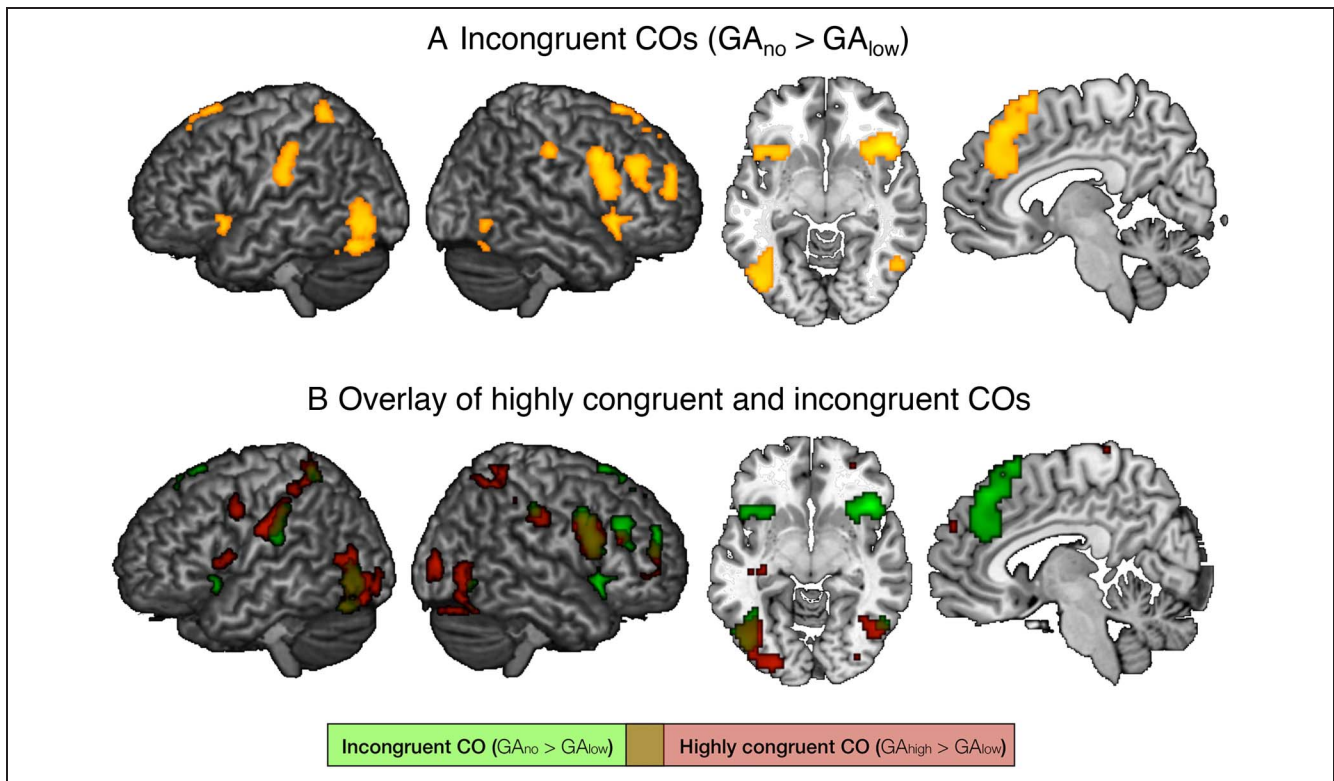
Figure 4. (A) Brain activation patterns for incongruent COs, FDR-corrected at $p < .01$. (B) Overlay of activations for highly congruent (red) and incongruent COs (green). Activations of both object categories overlap in the IPL and OTC as well as in the right ventral premotor cortex and right IFG.

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 COs is not driven by the COs congruency or incongruency to the observed action, but rather to the CO's potential to inform expectations toward 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 sites, among these, as hypothesized, the OTC and the PPC (especially the intraparietal sulcus [IPS]). The same effect was found for the IFG, the area we focused on because of its well-established role in the processing of semantic information. Interestingly, BA 47 of the IFG was especially engaged in the processing of 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 action observation network areas associating objects with actions, including the OTC (Wigget & Downing, 2011; Grill-Spector, Kourtzi, & Kanwisher, 2001) and the IPS (Ramsey, Cross, & Hamilton, 2011; Creem-Regehr, 2009; Singh-Curry & Husain, 2009; Grefkes & Fink, 2005), were significantly more active for both highly congruent and incongruent COs as compared with lowly congruent COs. Moreover, processing incongruent COs engaged a set of brain areas related to conflict and error processing: the ACC (Botvinick, Cohen, & Carter, 2004) and the anterior insula (Klein, Ullspenger, & Danielmeier, 2013). More precisely, these areas have been suggested to operate as a (response) inhibition network (Kana, Keller, Minshew, & Just, 2007; see also Hoffstaedter et al., 2014; Botvinick, Braver, Barch, Carter, & Cohen, 2001), 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. Supporting this line of interpretation, it has been suggested that the

Table 2. fMRI Activations for Incongruent COs ($GA_{no} > GA_{low}$)

Region	Hemisphere	BA	MNI Coordinates 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; SPL = superior parietal lobule; MCC = middle cingulate cortex. $p < .01$, FDR-corrected for multiple comparisons. *Extending into IFG (pars triangularis; BA 45).

processing of affordances is dependent on the availability of cognitive resources of an individual as well as the context and cues of the specific situation. If the context grants affordances importance with regard to an action goal and cognitive resources are available, these affordances may impact action interpretation and anticipation (cf. Randerath, Martin, & Frey, 2013). This relevance seems to apply for highly congruent and highly incongruent COs, but not for lowly congruent COs. Moreover, recent studies on stroke patients with an impaired perceptual sensitivity relate action selection processes to a frontoparietal and a claustrum-cingulate system, which roughly correspond to findings of this study (cf. Randerath et al., 2018). To anticipate an overarching action goal, one needs to select an action based on the action opportunities the environment has to offer (i.e., affordances), which has been depicted by the hierarchical affordance competition hypothesis (Pezzulo & Cisek, 2016). According to this hypothesis, the brain is suggested to be continuously engaged in generating predictions (e.g., about future action opportunities) rather than just reacting to already available affordances. Thereby, the brain is able to link different levels of abstraction and to bias the selection of immediate actions (e.g., cutting a tomato) according to the predicted long-term opportunities they make possible (e.g., making a salad).
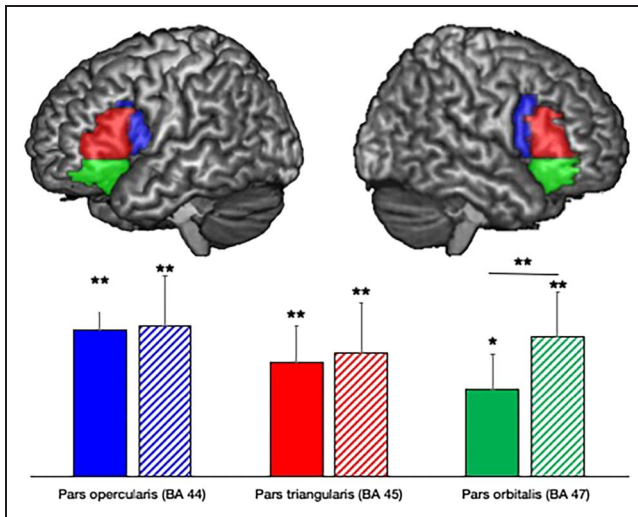
Figure 5. ROI analysis of IFG subregions according to the automated anatomical labeling atlas. Applied masks are illustrated in blue, red, and green for BA 44, BA 45, and BA 47, respectively. Corresponding beta values for highly incongruent versus lowly congruent COs are depicted in full colored bars, whereas beta values for incongruent COs versus lowly 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 subregions of the IFG. Regarding BA 47 (pars orbitalis), paired-sample t tests revealed a significant increase of activation for incongruent as compared with highly congruent COs. None of the other pairwise comparisons revealed a significant difference. $*p < .05$, $**p < .01$.

The current study focused particularly on BOLD effects in the IFG, which is known for its role in retrieval and integration of semantic information (Kilner, 2011; Caspers et al., 2010). We here aimed to extend this role and hypothesized that the IFG is sensitive to the informative impact of COs concerning 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 with lowly congruent COs. The three subregions of the IFG have been associated with different functions across different domains, including language (cf. Liakakis, Nickel, & Seitz, 2011; Bookheimer, 2002), emotion processing (Seitz et al., 2008; Carr, Iacoboni, Dubeau, Mazzioatta, & Lenzi, 2003), and creativity (Kröger et al., 2012). According to different accounts regarding the IFG function (Uddén & Bahlmann, 2012; Badre & Wagner, 2007; Koechlin & Summerfield, 2007), the IFG is hierarchically organized in a functional stepwise gradient along the rostrocaudal axis, where top–down control is exerted from anterior to posterior regions (Koechlin, Ody, & Kouneiher, 2003). In specific reference to action observation, the more posterior the area, the more it is suggested to contribute to constraining the immediate action requirements or options. More anterior sites, in turn, are more content-independent and associated with high-level goals (Badre

& D'Esposito, 2009; Badre et al., 2005; cf. Buckner, 2003). More specifically, BA 44 (together with premotor regions) supports structured sequence processing (cf. Uddén & Bahlmann, 2012) to realize a particular outcome (Fiebach & Schubotz, 2006; Grafman, 2002). Regarding BA 45 and BA 47, Badre and Wagner (2007) associated strategic memory retrieval with BA 47 and postretrieval selection among competing memories with BA 45. Although semantic retrieval is necessary when bottom–up cues are not sufficient to activate goal-relevant information, postretrieval 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 the activation pattern observed in our current study: incorporating relevant CO information into an observed action scene to anticipate its outcome draws on all IFG subregions. However, although BA 44 and BA 45 were recruited to an equivalent 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, that is, 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 both incongruent COs (i.e., a highlighted scientific paper) and the observed action (washing a salad) under an overarching goal requires a much higher level of abstraction, evoked by the strength of the CO's association with an incompatible room category (office as compared with a kitchen) and hence with actions associated with this room category (e.g., learning for a test). Importantly, the observed pattern of activation in the IFG does not simply reflect demands on integrating more or less compatible contextual information (here: COs) in the observed action. In that case, one would see a parametric increase of IFG activation with increasing incongruence, that is, 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 considers the informative value of COs when observing an action. More specifically, our results suggest that the neural activity of the IFG reflects the informational impact of COs on an 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 toward a particular outcome, or when the contextual information reveals a strong conflict with the observed manipulation, in which case the

currently operating predictive model has to be reconsidered and possibly extended toward a new overarching action goal.

REFERENCES

Badre, D., & D'esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience*, *10*, 659–669.

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*, 2883–2901.

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

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

Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*, 151–188.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 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*, 539–546.

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

Brozzoli, C., Gentile, G., Bergouignan, L., & Ehrsson, H. H. (2013). A shared representation of the space near oneself and others in the human premotor cortex. *Current Biology*, *23*, 1764–1768.

Buckner, R. L. (2003). Functional–anatomic correlates of control processes in memory. *Journal of Neuroscience*, *23*, 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*, 201–218.

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*, 175–185.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences, U.S.A.*, *100*, 5497–5502.

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

Costantini, M., Committeri, G., & Sinigaglia, C. (2011). Ready both to your and to my hands: Mapping the action space of others. *PLoS One*, *6*, e17923.

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*, 166–171.

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, United Kingdom: 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.

Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal–premotor interactions in primate control of grasping. *Neural Networks*, *11*, 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*, *42*, 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*, 189–210.

Friston, K. J., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological Cybernetics*, *104*, 137–160.

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing*. Hillsdale, NJ: Erlbaum.

Gibson, J. J. (1979). The theory of affordances. In *The ecological approach to visual perception*. Boston, MA: Houghton Mifflin.

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*, 6523–6532.

Grafman, J. (2002). The structured event complex and the human prefrontal Cortex. In D. T. Stuss & R. T. Knight (Eds.), *Principles of frontal lobe function*. New York: Oxford University Press.

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

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, *41*, 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*, 873–889.

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

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*, e0134316.

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*, 2178–2190.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*, 187–202.

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

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*, 198–206.

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

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

Klein, T. A., Ullsperger, M., & Danielmeier, C. (2013). Error awareness and the insula: Links to neurological and psychiatric diseases. *Frontiers in Human Neuroscience*, *7*, 14.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*, 1181–1185.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in Cognitive Sciences*, *11*, 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.

Liakakis, G., Nickel, J., & Seitz, R. J. (2011). Diversity of the inferior frontal gyrus—A meta-analysis of neuroimaging studies. *Behavioural Brain Research*, *225*, 341–347.

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

Pezzulo, G., & Cisek, P. (2016). Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. *Trends in Cognitive Sciences*, *20*, 414–424.

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. *Neuroimage*, *10*, 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*, 3400–3409.

Randerath, J., Finkel, L., Shigaki, C., Burris, J., Nanda, A., Hwang, P., et al. (2018). Does it fit?–Impaired affordance perception after stroke. *Neuropsychologia*, *108*, 92–102.

Randerath, J., Goldenberg, G., Spijkers, W., Li, Y., & Hermsdörfer, J. (2010). Different left brain regions are essential for grasping a tool compared with its subsequent use. *Neuroimage*, *53*, 171–180.

Randerath, J., Martin, K. R., & Frey, S. H. (2013). Are tool properties always processed automatically? The role of tool use context and task complexity. *Cortex*, *49*, 1679–1693.

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

Seitz, R. J., Schäfer, R., Scherfeld, D., Friederichs, S., Popp, K., Wittsack, H.-J., et al. (2008). Valuating other people's emotional face expression: A combined functional magnetic resonance imaging and electroencephalography study. *Neuroscience*, *152*, 713–722.

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

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

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 830.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition*, *8*, 769–800.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*, 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: Series B: Biological Sciences*, *367*, 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*, 495–498.

Wigget, A. J., & Downing, P. E. (2011). Representation of action in occipito-temporal cortex. *Journal of Cognitive Neuroscience*, *23*, 1765–1780.

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

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

Wurm, M. F., & Schubotz, R. I. (2017). What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychonomic Bulletin & Review*, *24*, 503–509.

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*, 1548–1559.

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