The Context–Object–Manipulation Triad: Cross Talk during Action Perception Revealed by fMRI

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

Abstract

■ To recognize an action, an observer exploits information about the applied manipulation, the involved objects, and the context where the action occurs. Context, object, and manipulation information are hence expected to be tightly coupled in a triadic relationship (the COM triad hereafter). The current fMRI study investigated the hemodynamic signatures of reciprocal modulation in the COM triad. Participants watched short video clips of pantomime actions, that is, actions performed with inappropriate objects, taking place at compatible or incompatible contexts. The usage of pantomime actions enabled the disentanglement of the neural substrates of context–manipulation (CM) and context–object (CO) associations. There were trials in which (1) both manipulation and objects, (2) only manipulation, (3) only objects, or (4) neither manipulation nor objects were com-

patible with the context. CM compatibility effects were found in an action-related network comprising ventral premotor cortex, SMA, left anterior intraparietal sulcus, and bilateral occipito-temporal cortex. Conversely, CO compatibility effects were found bilaterally in lateral occipital complex. These effects interacted in subregions of the lateral occipital complex. An overlap of CM and CO effects was observed in the occipito-temporal cortex and the dorsal attention network, that is, superior frontal sulcus/dorsal premotor cortex and superior parietal lobe. Results indicate that contextual information is integrated into the analysis of actions. Manipulation and object information is linked by contextual associations as a function of co-occurrence in specific contexts. Activation of either CM or CO associations shifts attention to either action- or object-related relevant information. ■

INTRODUCTION

Object manipulations normally take place in particular contextual places. There is hence a tight, triadic coupling between context, objects, and manipulations (the COM triad hereafter). From a neural perspective and as a result of frequency-based learning, the COM triad is likely to be reflected by associative strengths between the three kinds of information (Turk-Browne, Scholl, Johnson, & Chun, 2010). Indeed, contextual information has been demonstrated to affect recognition of both objects (Bar, 2004) and manipulations (Wurm & Schubotz, 2012). However, it remains unclear how context–manipulation (CM) and context–object (CO) couplings interact during action recognition.

In the present fMRI study, we investigated the neural effects of CM and CO associations in a unitary experimental design. We therefore manipulated the compatibility (1) between context and manipulation information and (2) between context and object information independently. This was achieved by decoupling manipulation from their associated objects in the COM triad. To this end, participants were presented videos showing pantomime actions, that is, manipulation indicative of a particular action but actually performed with inappropriate objects (Schubotz & von Cramon, 2009). We thereby im-

plemented a 2 × 2 factorial design: First, manipulations were either compatible (comM) or incompatible (incomM) with the context. Second, pantomime actions involved objects that were either compatible (comO) or incompatible (incomO) with the context (see Figure 1). There were hence four conditions where (1) both manipulations and objects (comM/comO), (2) only manipulations (comM/incomO), (3) only objects (incomM/comO), or (4) neither manipulations nor objects (incomM/incomO) were compatible with the context.

A recent study on recognition of pantomime and real action found that the processing of manipulation information draws on the cortical motor system (left ventral premotor cortex [PMv] and left anterior intraparietal sulcus [aIPS]) and the occipito-temporal cortex (OTC), whereas the processing of object information involved the lateral occipital complex (LOC; Schubotz & von Cramon, 2009). We hypothesized that the same neural networks are differentially engaged when either manipulation or object information is compatible with the context or not. Thus, CM compatibility effects should manifest as BOLD response differences in the motor system. Likewise, CO compatibility effects should manifest as BOLD response differences in the LOC.

Moreover, because of the triadic relationship between context, objects, and manipulation, we expected interactions between CM and CO compatibility effects. Of particular

Figure 1. Experimental conditions. (A) Example frames for the employed pantomime action movies. Conditions constitute a 2 × 2 factorial design with the factors CM and CO compatibility. (B) Schema of the triadic relationship of manipulation (M), object (O), and context (C) information in the according conditions. Solid lines represent strong contextual associations, and dotted lines represent weak contextual associations.

interest was the concurrence of both CM and CO compatibility (see top left of Figure 1A and B). Here, the manipulation (opening a tin) and the objects (whisk and bowl) are associated with the same context (kitchen). This constellation is potentially conflicting because both objects and manipulations are reinforced by the context, but objects as well as object-implied invalid manipulations have to be suppressed or ignored. Hence, we expected that the influence of context on object and manipulation processing becomes particularly evident in this high-conflict situation when manipulation and objects share the same contextual affiliation. On the basis of the same findings outlined above, we expected that, in the contrast comM/comO versus incomM/comO, an increased BOLD response would be observed in the motor system due to conflicting activation/suppression of manipulation information. Likewise, in the contrast comM/comO versus comM/incomO, an increased response was expected in the LOC due to conflicting activation/suppression of object information.

The co-occurrence of objects and manipulations in particular contexts might also be reflected in strengthened associations/neural couplings within contextual categories. In other words, object and manipulation information might be represented in common networks with respect to their contextual affiliation. If this is the case, conflicting activation/suppression of manipulation and object information should also occur if a specific *visual* context is absent but manipulation and objects share the same *implied* context. We therefore additionally analyzed pantomime actions in neutral settings, that is, a neutral, white background. In these trials manipulations and objects belong to the same contextual category (sameCC) or not (diffCC). If manipulation and object information is stored in networks grouped by contextual categories, we should expect stronger con-

flict and therefore higher BOLD responses in the motor system and the LOC in the contrast sameCC versus diffCC.

METHODS

Participants

Eighteen healthy volunteers (21–29 years, mean = 24.9 years, 12 women) participated in the fMRI experiment. All participants were right-handed according to the Edinburgh Inventory Manual Preference (Oldfield, 1971), had normal or corrected-to-normal vision, and were native German speakers. No participant had a history of neurological or psychiatric disorder. Participants gave written consent before fMRI measurement. The experimental standards were approved by the local ethics committee of the Medical Faculty, University of Cologne, Germany. Data were handled anonymously.

Stimuli and Task

Stimuli consisted of 30 context-specific pantomime actions (e.g., pantomiming to open a tin with a rasp and a piece of wood instead of a tin and a tin opener) filmed in compatible and incompatible contexts. In addition, the same actions were also filmed in a neutral setting. Actions were specific for the contexts "office," "kitchen," and "workshop." Contexts were specified by the background, the working surface, and three to five context-specific stationary objects (e.g., computer screen, coffee machine, grinding machine) placed on the working surface. The neutral context consisted of a white sheet stretched on a table forming a uniform surface without any corners. Actions were filmed from an allocentric perspective (60° to the left

of the actress) ensuring a convenient view of the action and the background (Figure 1).

The following action conditions were employed in the experiment: (1) context-compatible manipulations performed with context-compatible objects (comM/comO), (2) context-incompatible manipulations performed with context-compatible objects (incomM/comO), (3) context-compatible manipulations performed with context-incompatible objects (comM/incomO), (4) context-incompatible manipulations performed with context-incompatible objects (incomM/incomO), and (5) the same set of pantomime actions in neutral contexts. To control for low-level differences between stimuli, we balanced the occurrence of contexts, objects, and manipulations across conditions: Each context occurred 10 times (except for the neutral condition, where no visual context was provided), and each object pair and each manipulation occurred one time in each condition. Each video had a length of 3 sec, a presentation rate of 25 frames per second and a display width and height of 720×576 pixels. Care was taken that movement trajectories of each pantomime action were performed in the same manner for each of the five conditions. 30 trials per condition were presented in a event-related design, intermixed with three conditions employing normal actions (Wurm & Schubotz, 2012).

Participants were instructed to infer the presented actions to ensure that participants were able to recognize and paid attention to the actions. They were informed that some of trials (20%) would be followed by verbal action descriptions (question trials). Participants had to indicate via button press whether a description matched or did not match the action of the preceding trial ($n - 1$). The descriptions matched the action in 50% of the question trials. Every second question trial was followed by an empty trial in which participants were instructed to fixate a small dot in the center of the screen, providing a resting baseline.

All trials had a length of 6 sec, starting with a variable jitter (0, 500, 1000, or 1500 msec) to enhance the temporal resolution of the BOLD response. Videos and verbal action descriptions were followed by a fixation dot until the next trial started.

The trial order was pseudorandomized to equate transition probabilities between settings and transition probabilities between conditions. Furthermore, the occurrence of question trials and baseline trials were balanced so that in each condition there were the same number of trials preceded and followed by question and baseline trials. Finally, to rule out habituation effects with respect to the actions, the succession of conditions was balanced across participants for each action. Care was taken that orthogonality is maintained.

MRI Data Acquisition

After instruction and a short practice of the experiment outside the scanner, participants were placed supine on the scanner bed, with the right index and middle finger positioned over the two buttons of a response box. To attenuate scanner noise, participants were provided with earplugs and headphones. Form-fitting cushions were used to prevent arm and head motion.

Imaging was performed using a 3-T Siemens Trio system (Erlangen, Germany) using a standard birdcage head coil. 26 axial slices (192 mm field of view; 64×64 pixel matrix; 4 mm thickness; 1 mm spacing; in-plane resolution of 3×3 mm), covering the whole brain, were acquired using a single-shot gradient EPI sequence (repetition time = 2000 msec; echo time = 30 msec; flip angle = 90°; acquisition bandwidth = 116 kHz) sensitive to BOLD contrast. Functional imaging (90 images per condition) was followed by the acquisition of 26 anatomical T1-weighted MDEFT images (Norris, 2000). In a separate session, high-resolution whole-brain images were acquired from each participant using a T1-weighted 3-D-segmented MDEFT sequence.

MRI Data Analysis

fMRI data were processed using the software package LIPSIA 1.5.0 (Lohmann et al., 2001). An off-line motion correction was performed with the Siemens motion correction protocol PACE (Siemens, Erlangen, Germany). A cubic spline interpolation was employed to correct for the temporal offset between slices acquired in one image. A temporal high-pass filter with a cut-off frequency of 1/80 Hz was used. Spatial smoothing was performed with a Gaussian filter of 8 mm FWHM. To align the functional data slices with a 3-D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (three rotational, three translational) was performed. An optimal match between the 3-D individual data set and the EPI-T1 slices was achieved by using the rotational and translational parameters acquired on the basis of the MDEFT and the EPI-T1 slices. The MDEFT volume data set with 160 slices and 1 mm slice thickness was standardized to the Talairach stereotactic space (Talairach & Tournoux, 1988). Rotational and translational parameters were subsequently transformed by linear scaling to a standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system, thus generating output data with a spatial resolution of $3 \times 3 \times 3$ mm (27 mm^3). The statistical evaluation was based on a least-squares estimation using the general linear model for serially autocorrelated observations (Friston et al., 1995; Worsley & Friston, 1995). For the main experiment, the design matrix was generated with a gamma function, convolved with the hemodynamic response function. Brain activations were analyzed time-locked to onset of the movies, and the analyzed epoch comprised the full duration (3 sec) of the presented movies, the duration of the null events (6 sec), and the RT in question trials (maximum of 3 sec), respectively. Analyzed epochs of the localizer

experiment comprised the duration of blocks (17 sec) and were convolved with a Gaussian function. The model equation, including the observation data, the design matrix, and the error term, was convolved with a Gaussian kernel of dispersion of 4 sec FWHM to account for the temporal autocorrelation (Worsley & Friston, 1995). In the following, contrast images, that is, beta value estimates of the raw score differences between specified conditions, were generated for each participant. As all individual functional data sets were aligned to the same stereotactic reference space, the single-participant contrast images were entered into a second-level random effects analysis for each of the contrasts.

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

To correct for false-positive results, in a first step, an initial voxelwise z threshold was set to $z = 2.576$ ($p = .005$). In a second step, the results were corrected for multiple comparisons using cluster size and cluster value thresholds obtained by Monte Carlo simulations at a significance level of $p = .05$, that is, the reported activations are significantly activated at $p < .05$, corrected for multiple comparisons at the cluster level.

Conjunctions were calculated by outputting the minimum z value of the two input contrasts for each voxel (Nichols, Brett, Andersson, Wager, & Poline, 2005).

RESULTS

Behavioral Results

Behavioral performance was monitored during the experiment to ensure that participants were paying attention to the presented actions. Performance was assessed by analyzing the error rate to question trials: 88.5% (±1.1%, *SEM*) of question trials were answered correctly, indicating adequate performance. The differences in performance between conditions were not analyzed because of low statistical power (20% of trials, i.e., 6 of 30 trials per condition, were followed by question trials).

fMRI Results

CM Compatibility

Effects of CM compatibility were expected to manifest in (1) the contrast comM/incomO versus incomM/incomO and (2) comM/comO versus incomM/comO.

With respect to (1), no significant increases or decreases in BOLD response were found. With respect to (2), increases in BOLD response were found in the motor system, specifically, bilateral superior PMv extending into inferior frontal gyrus in the left hemisphere, bilateral SMA, left supramarginal gyrus (SMG), left aIPS, and bilateral OTC. Furthermore, activation was found in bilateral dorsal premotor cortex (PMd)/superior frontal sulcus (SFS) and left superior parietal lobe (SPL; Figure 2A, Table 1).

CO Compatibility

Effects of CO compatibility were expected to manifest in the contrasts (3) incomM/comO versus incomM/incomO and (4) comM/comO versus comM/incomO.

With respect to (3), no significant increased or decreased activation was found. With respect to (4), increased activation was found in both subdivisions of the LOC bilaterally (Grill-Spector, Kourtzi, & Kanwisher, 2001), that is,

Figure 2. CM and CO compatibility effects. (A) CM compatibility (comM/comO vs. incomM/comO) enhanced neural responses particularly in the motor system (bilateral superior PMv, left SMG extending into the IPS, SMA, and bilateral OTC). (B) CO compatibility (comM/comO vs. comM/incomO) enhanced activity particularly in the LOC (bilateral lateral occipital cortex [LO] and posterior fusiform and parahippocampal gyrus [pFs]). Contrasts are z-thresholded at 2.576; corrected cluster threshold $p < .05$.
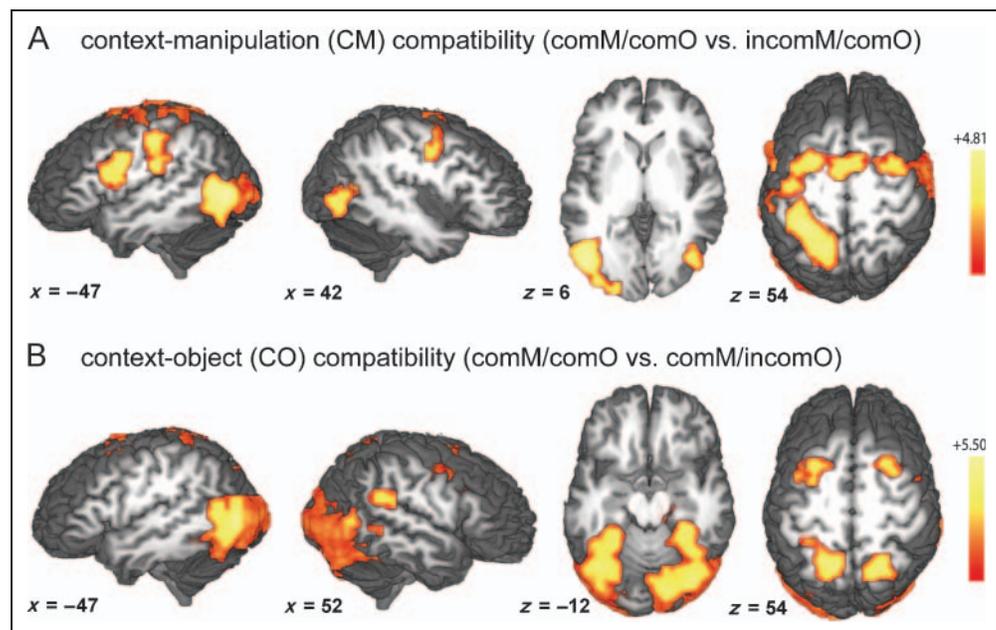
Table 1. Activations for CM and CO Compatibility and Same versus Different Contextual Categories

Area		BA	x	y	z	Z
CM Compatibility (comM/comO vs. incomM/comO)						
L	pMTG/ITG (OTC)	18/19/37	−44	−72	6	4.79
L			34	−69	6	3.46
R	Superior PMv	44/6	−47	0	24	3.75
R	aIPS/SMG	40	−29	−42	48	4.95
L	SPL	7	−20	−54	63	4.59
R	PMd/SFS	6/8	−20	3	63	4.01
L	PMv	6	52	−6	35	4.06
R			25	−3	60	3.81
R	SMA	6	−2	0	54	3.34
CO Compatibility (comM/comO vs. comM/incomO)						
L	pMTG/ITG (OTC)	18/19/37	−38	−75	3	5.63
L	Fusiform gyrus	37	−29	−48	−3	4.11
R	pMTG/ITG	18/19/37	19	−87	6	5.67
R	Fusiform gyrus	37	28	−36	−9	4.12
L	PMd/SFS	6/8	−32	0	36	4.21
R			22	15	45	4.30
L	SPL	7	−17	−54	51	4.47
R			16	−60	48	4.25
R	SMG	40	58	−36	21	4.25
Actions in Neutral Context with Objects from Same versus Different Contextual Category (sameCC vs. diffCC)						
L	pMTG/ITG (OTC)	18/19	−44	−81	12	3.61
R			37	−69	22	3.66
L	SPL	7	−23	−51	60	4.64
L	IPS	40	−56	−24	39	3.91
L	PMd	6	−23	−3	51	3.54
R			34	−24	42	5.14
R	SPL	7	10	−41	69	3.84
	Paracentral gyrus	4	4	−33	63	4.21
R			28	−9	51	3.58
	SMA	6	4	−3	51	3.45

Hemisphere (left, right), macroanatomical specification, Brodmann's area, Talairach coordinates (*x*, *y*, *z*), maximal *Z* scores (*Z*), and cluster volumes (mm^3); *z*-thresholded at 2.576, corrected cluster threshold *p* < .05. Abbreviations: L = left; R = right; BA = Brodmann's area; ITG = inferior temporal gyrus; pMTG = posterior middle temporal gyrus.

lateral occipital cortex (LO) and bilateral posterior para-hippocampal and fusiform gyrus (pFs). Activation (left LO: −38, −75, 3; right LO: 19, −87, 6; left pFs: −29, −48, −3; right pFs: 28, −37, −9) overlapped with activation obtained in a localizer scan for objects versus scrambled objects (left LO: −50, −69, 12; right LO: 37, −72, 9; left pFs: −32, −39, −9; right pFs: 25, −45, −3; Wurm & Schubotz, 2012). Additional activation was found in bilateral PMd/SFS and bilateral SPL, right superior temporal gyrus/SMG, and right superior PMv (Figure 2B, Table 1).
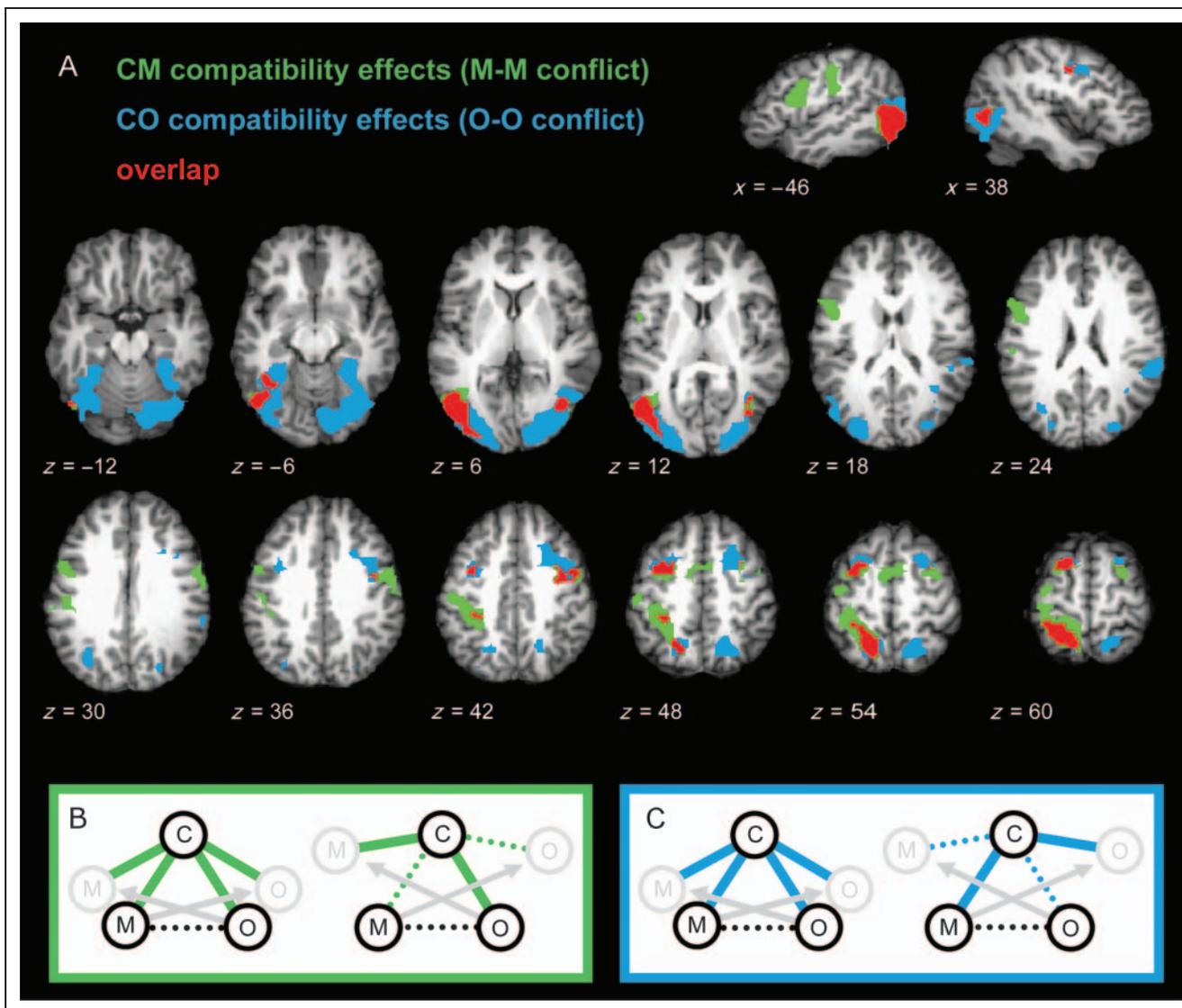
Figure 3. Influence of context on attentional focus in the COM triad. (A) Overlap of CM and CO compatibility effects (comM/comO vs. comM/incomO and comM/comO vs. incomM/comO). (B) M-M conflict arises when object-implied, invalid manipulation information is contextually related to the pantomimed action. (C) O-O conflict arises when perceived, invalid objects are contextually related to objects that are implied by the manipulation. Context information compatible with either manipulation or objects fixes contextual association on either the manipulation- or object-side of the triad. Solid lines represent strong contextual associations, and dotted lines represent weak contextual associations. Black circles represent present/cued context, object, or manipulation information, and gray circles represent object- or manipulation-implied (gray arrows) manipulation or object information.

Dissociation of CM and CO Compatibility Effects

CM- and CO-specific and common activations were identified by computing an overlap of the corrected contrasts comM/comO versus incomM/comO (CM compatibility) and comM/comO versus comM/incomO (CO compatibility).

Activations specific for the contrast comM/comO versus incomM/comO (CM compatibility) were found in the left superior PMv, the left SMG, and the left IPS. Activation specific for the contrast comM/comO versus comM/incomO (CO compatibility) was found in bilateral parahippocampal and fusiform gyrus and right SMG. An overlap was estab-lished in bilateral OTC, bilateral PMd/SFS, and left SPL (Figure 3A). Activation of PMd/SFS and SPL resembled activation found in imaging studies on selective attention (Cristescu & Nobre, 2008; Wager, Jonides, & Reading, 2004; Yantis & Serences, 2003; Nobre, 2001; Fink, Dolan, Halligan, Marshall, & Frith, 1997).

ROI Analysis

To dissociate CM and CO compatibility effects in the motor system (left PMv and left aIPS) and the LOC (bilateral LO and bilateral pFs), an ROI analysis was performed. ROIs

were determined using coordinates reported in Wurm and Schubotz (2012): LO and pFs were identified in a functional localizer scan for objects versus scrambled objects (left LO: −50, −69, 12; right LO: 37, −72, 9; left pFs: −32, −39, −9; right pFs: 25, −45, −3); left PMv and aIPS were identified in the contrast action recognition versus rest (PMv: −50, 0, 36; aIPS: −53, −24, 42; Wurm & Schubotz, 2012). Mean beta values were extracted from the peak voxel plus six adjacent voxels for the contrasts comM/comO versus rest, incomM./comO versus rest, comM./incomO versus rest, and incomM/incomO versus rest. Mean beta values are shown in Figure 4. Repeated-measures (2 × 2 factorial) ANOVAs revealed significant effect of CM compatibility in PMv, aIPS, left and right LO, and left pFs. Significant main effects of CO Compatibility were found in left and right LO and left and right pFs. Interaction effects were found in left LO and right pFs (Table 2).

Actions in Neutral Contexts

We hypothesized that object and manipulation information is grouped in category-specific networks with stronger associations within than between contextual categories. Therefore, highest BOLD responses for the condition comM/comO might have been due to the contextual relatedness of to-be-suppressed object/manipulation
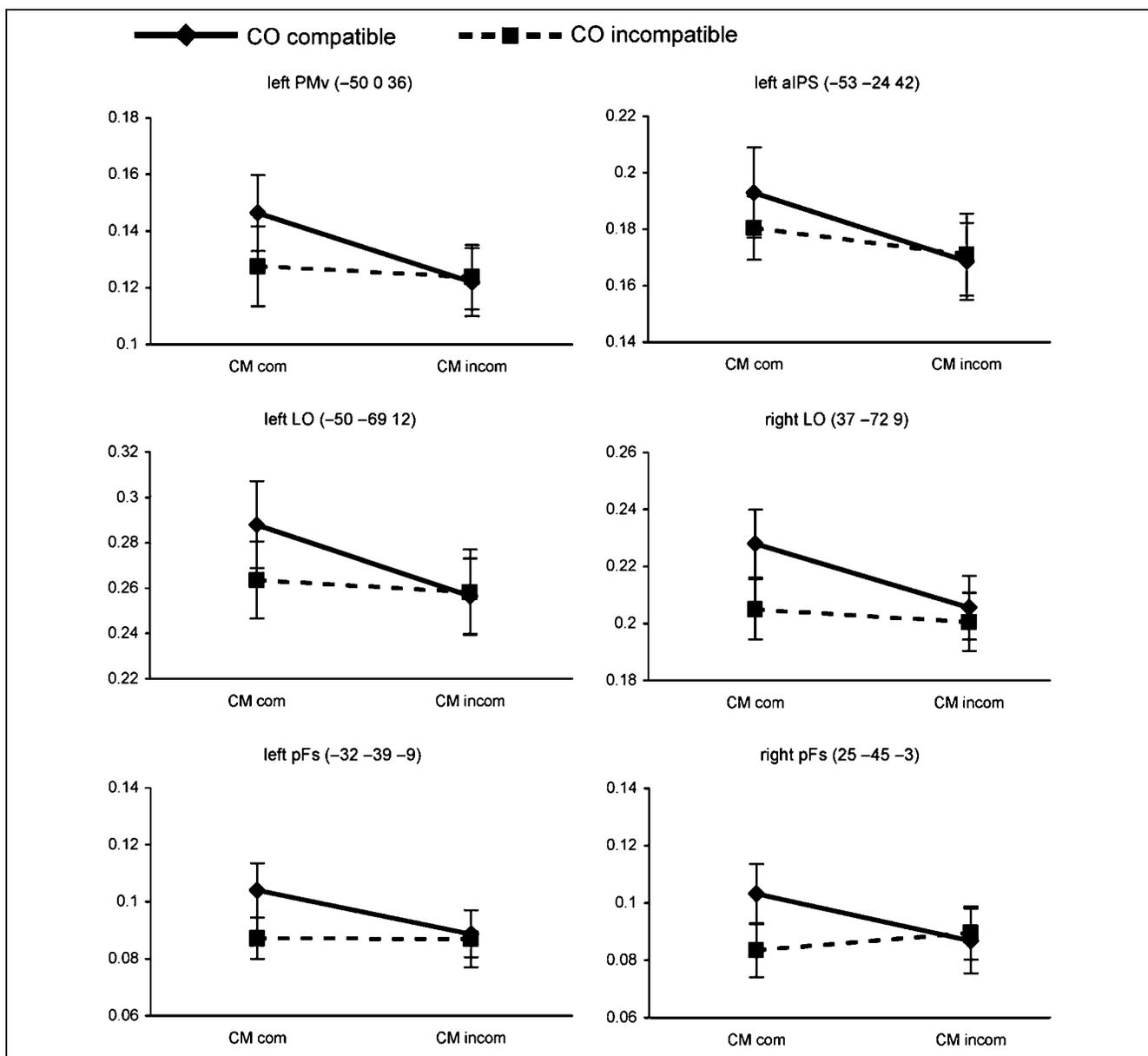

Figure 4. CM and CO compatibility effects in regions associated with action (left PMv, left aIPS) and object perception (left and right LO, left and right pFs) ROI coordinates from Wurm and Schubotz (2012). Mean beta values were extracted from the peak coordinate voxel plus six adjacent voxels of the baseline contrasts experimental condition versus rest. Error bars indicate standard error of means.
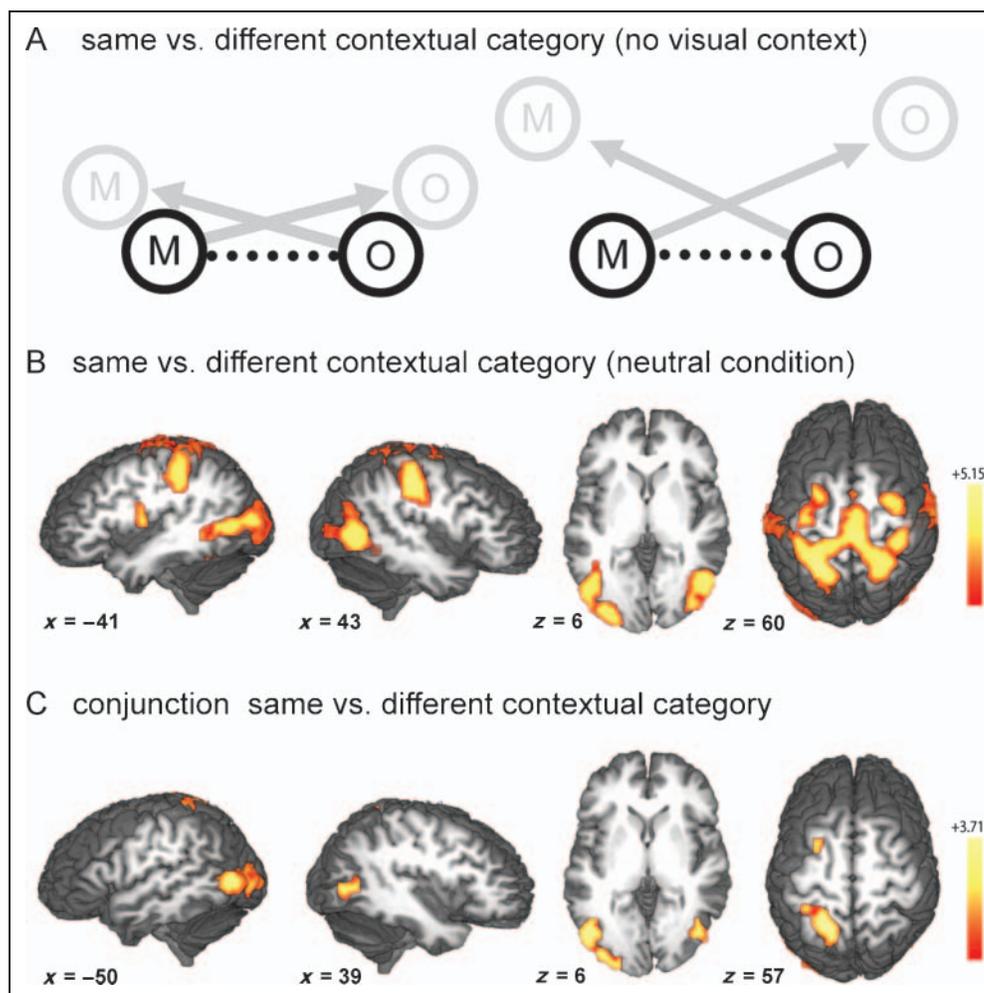
Table 2. Results of the ROI Analysis in the Motor System (Left PMv and aIPS) and LOC (Left and Right LO, Left and Right pFs)

ROI	CM Compatibility		CO Compatibility		Interaction	
	F(1, 17)	p	F(1, 17)	p	F(1, 17)	p
Left PMv	11.51	**.003**	2.27	.150	2.77	.114
Left aIPS	11.28	**.004**	0.87	.364	1.17	.293
Left LO	46.87	**< .001**	14.34	**.001**	6.15	**.024**
Right LO	15.81	**.001**	21.06	**< .001**	4.02	.061
Left pFs	5.38	**.033**	6.51	**.021**	2.73	.118
Right pFs	1.35	.261	7.10	**.016**	7.73	**.013**

Significant *p* values are in **bold** font.

information and to-be-inferred object/manipulation information (Figure 3B and C). If this assumption is correct, similar effects should be observed when object and manipulation belong to same versus different contextual categories but visual context information is absent (Figure 5A).

We tested this hypothesis by analyzing pantomime actions taking place in a neutral context. Trials were divided into two subsets: Actions of the first subset involved objects from the same contextual category as the manipulation (sameCC hereafter), actions of the other subset involved objects from a different contextual category as the manipulation (diffCC hereafter).

The contrast sameCC versus diffCC revealed activation in the bilateral dorsal attention network (bilateral PMd/SFS and bilateral SPL) and bilateral OTC. Further activation was found in the right SMA, right paracentral gyrus, bilateral IPS, and right anterior middle temporal gyrus (MTG; Figure 5B, Table 1).

Overlap of Activation due to Contextual Category Effects

To test whether the activation pattern of same versus different context categories (sameCC vs. diffCC) overlapped with the activation pattern of CM and CO compatibility (comM/comO vs. incomM/comO and comM/comO vs. comM/incomO), we computed a conjunction of the three contrasts. The conjunction revealed common activation in the left SPL, left PMd/SFS, and bilateral OTC (Figure 5C).

Figure 5. Same versus different contextual category. (A) Schematic depiction of M-M and O-O conflicts in the absence of visual context. Black circles represent present/cued object or manipulation information, gray circles represent object- or manipulation-implied (gray arrows) manipulation or object information. (B) Actions in neutral settings employing objects of same versus different contextual category (sameCC vs. diffCC). (C) Conjunction of contrasts employing objects of same versus different contextual category [(comM/comO versus comM/incomO) ∩ (comM/comO versus incomM/comO) ∩ (sameCC versus diffCC)]. Contrasts are *z*-thresholded at 2.576; corrected cluster threshold *p* < .05.

DISCUSSION

The present experiment investigated the triadic relationship of context, object, and manipulation information during action recognition. We employed pantomime action to decouple manipulations from their associated objects. Thereby, it was possible to disentangle effects of CM and CO compatibility.

An ROI analysis revealed main effects in (1) PMv, aIPS, left and right LO, and left pFs with respect to CM compatibility and (2) left and right LO and left and right pFs with respect to CO compatibility. Interactions of these effects were significant in left LO and right pFs. Furthermore, whole-brain analyses revealed activation in (1) left PMv, aIPS, and SMA for CM compatibility when objects were also compatible with the context and (2) the fusiform component of the LOC for CO compatibility when actions were also compatible with the context. Common activations for both contrasts were found in bilateral OTC and bilateral PMd/SFS-SPL loop. Crucially, this common network was also found when analyzing pantomime actions in neutral settings if objects from either the same or a different contextual category as the manipulation were used. The common activation pattern is suggested to reflect increased demands to focus attention on visible manipulation information and mentally associated object information while suppressing attention on visible interfering object information and mentally associated manipulation information, if manipulation and objects belong to the same contextual category.

Together, the findings indicate that contextual information affects object and manipulation information in an interactive way: Activation is strongest in either manipulation-associated or object-associated brain regions dependent on reinforced CM and CO associations in the COM triad. Interaction effects in the left LO and the right pFs demonstrate that visual context, although task-irrelevant, is integrated into the analysis of actions. Furthermore, context effects in the absence of visual context suggest stronger associations between contextually related versus unrelated objects and manipulations, pointing to the encoding of object and manipulation information in networks grouped by contextual category.

Contextual Modulation of Manipulation and Object Processing

We demonstrated that CM compatibility modulates neural activity in the left PMv, left aIPS, left and right LO, and left pFs (Table 2). These effects became particularly apparent when objects were also context-compatible in both contrast conditions (Figures 2A and 4). Here, activation was found in bilateral superior PMv, left SMG/aIPS, SMA, and bilateral OTC. These regions form a network that is typically active during the perception of action (Caspers, Zilles, Laird, & Eickhoff, 2010).

The extrastriate OTC comprises several functional areas, for example, the EBA, hMT, and pSTS. These regions are involved in the perceptual analysis of manipulation information (Thompson & Parasuraman, 2011; Jastorff & Orban, 2009; Kable & Chatterjee, 2006; Beauchamp, Lee, Haxby, & Martin, 2002). According to Giese and Poggio (2003), information about form, for example, postures of body parts and shape of objects, is processed by the form pathway in V1, V4, EBA, and finally pSTS, where so-called snapshot patterns are combined with optic flow patterns detected by the motion pathway, that is, V1, V2, hMT, and MST.

Predictive accounts interpret the activation of the premotor–parietal network during action perception as reflecting a generation of internal models to anticipate forthcoming events (Jacob, 2009; Csibra, 2007; Schubotz, 2007). The SMA is involved in temporal sequencing of action steps (Mita, Mushiake, Shima, Matsuzaka, & Tanji, 2009; Tanji, 1994). The SMA is connected to the PMv (Luppino, Matelli, & Rizzolatti, 1990), thereby providing signals to update the generation of the internal action models (Schubotz, 2007; Fagg & Arbib, 1998). PMv and aIPS are reciprocally connected (Geyer, Matelli, Luppino, & Zilles, 2000). PMv thus receives perceptual updates of the ongoing action and sends predictions of visual consequences of the internally generated action model (Kilner, Friston, & Frith, 2007; Keysers & Perrett, 2004).

We further demonstrated that CO compatibility modulates neural activity in left and right LO and left and right pFs (Table 2). Parallel to CM effects reported above, CO compatibility effects were particularly present when manipulations were also context-compatible in both contrast conditions (Figures 2B and 4). Increased activation was found in the LOC, specifically, the lateral occipito-temporal component (LO) and the component in posterior and mid-fusiform gyrus extending into the occipito-temporal sulcus (pFs; Grill-Spector et al., 2001). The LOC plays a central role in object recognition (Grill-Spector et al., 2001; Kourtzi & Kanwisher, 2001). Both of its subdivisions are sensitive to object shape, irrespective of low-level visual features such as color or texture (Grill-Spector et al., 1999). The LOC, particularly LO, is also involved in representation of semantic object knowledge (Binder, Desai, Graves, & Conant, 2009). In terms of manipulable objects, LO overlaps with the MTG tool area, which is sensitive to visual perception of tools (Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Chao, Haxby, & Martin, 1999) and to rigid tool motion (Beauchamp, Lee, Haxby, & Martin, 2003).

Contextual Categories and the COM Triad

To understand why the neural correlates of manipulation and object processing show an enhanced response particularly when both manipulation and objects are compatible with the context, we need to reconsider the COM triad: In the top left cells of Figure 1A and B, both the presented manipulation (opening a tin) as well as the used objects (whisk and bowl) are associated with the visible context (kitchen). Thus, although the manipulation and the object are incompatible, they belong to the same contextual

category. In the top right cells of Figure 1A and B, only the objects (pliers and cable) are associated with the context (workshop) whereas the manipulation (opening a tin) is not. Similarly, in the bottom left cells of Figure 1A and B, only the manipulation is associated with the context (kitchen) whereas the objects (hole puncher and documents) are not. In both cases, manipulation and object belong to a different contextual category. During inference of the pantomimed action, the observed manipulation interferes with manipulations mentally associated with the objects. If the observed and the object-implied manipulations are contextually related to each other, the manipulation–manipulation (M-M) conflict is stronger, increasing the load on the motor system. The same logic applies to object information within the COM triad: The observed invalid objects interfere with object information mentally associated with the manipulation. This object–object (O-O) conflict is stronger if concurrent object representations belong to the same versus a different contextual category, increasing the load in LOC. Notably, stronger M-M and O-O conflicts become apparent in both constellations for CM and CO compatibility, where same versus different contextual categories were contrasted. Overlapping effects therefore reflect the influence of implied context on action perception. However, the CM and CO constellations differ with regard to the reinforcement of either CM or CO associations by the visual context: In the case of CM compatibility, there is no manipulation bias by visual context information (hence increasing recognition demands on manipulation information within the COM triad; see Figure 3A and B), whereas in the case of CO compatibility, the objects are unbiased by visual context information (hence increasing recognition demands on object information; see Figure 3A and C). The fact that the activation patterns of CM and CO compatibility differ in the predicted regions in the predicted way and the effects interact in some of these regions indicates that visual context is integrated into the analysis of observed actions (Figure 3A and Table 2).

Contextual Modulation of Attention Orienting toward Object and Manipulation Information

For both CM and CO compatibility, neural activity increased bilaterally in PMd/SFS and SPL. These regions are reciprocally connected with each other (via SLF I; Geyer et al., 2000) and form the so-called dorsal attention network (Corbetta & Shulman, 2002, 2011). Notably, the SPL is also connected to the parahippocampal place area (Schmahmann & Pandya, 2006), a region involved in the recognition of places (Epstein & Kanwisher, 1998) and the encoding of contextual associations (Bar & Aminoff, 2003; but see Epstein & Ward, 2010). The dorsal attention network is involved in directing attention to a target and in selecting appropriate responses (e.g., limb or eye movements; Corbetta & Shulman, 2002). According to the biased competition model, a top–down signal, determined by task demands, biases among mutually

inhibitory sensory representations, so that an attended cue "wins" the competition by producing more robust cortical activity than unattended cues (Desimone & Duncan, 1995). Parietal cortices are thought to act as priority maps (Bisley & Goldberg, 2010) that integrate bottom–up visual stimuli and top–down influences provided by corresponding frontal regions. Locations or objects with high priority are represented with greater neural activity, thereby guiding attention and appropriate behavioral responses (Bisley & Goldberg, 2010). During identification of the pantomimed action, attention was directed to hand postures and movements and associated object information in semantic memory. This requires object- or feature-based (Liu, Slotnick, Serences, & Yantis, 2003; Yantis & Serences, 2003) and semantic-based attention (Cristescu, Devlin, & Nobre, 2006). With reference to behavioral findings on semantic distance effects (Vigliocco, Vinson, Damian, & Levelt, 2002; Glaser & Dungelhoff, 1984; see below), we suggest that, during pantomime recognition, orienting attention toward hand posture and movements and mentally associated object information is more demanding when these share a common contextual category with the visible but inappropriate (and thus interfering) objects and corresponding mentally associated manipulations.

The dorsal attention network may anatomically and, probably also functionally, overlap with regions involved in oculomotor processing (Corbetta, 1998). It is thus possible that activation of PMd/SFS and SPL is due to differences in fixation and eye movement behavior in the high conflict CM/CO compatible condition. Future studies should address which parts of the action scenes are fixated and how long they are fixated.

Contextual Category Effects in the Absence of Visual Context Information

The proposed mechanism of COM interaction should yield similar neural responses when no bias by visual context information is introduced (Figure 5A). We therefore analyzed pantomime actions in neutral conditions. Half of the actions involved objects from the same contextual category as the manipulation; the other half involved objects from a different contextual category. Observing pantomime actions involving objects from the same versus different contextual categories as the manipulation revealed activation in bilateral OTC, PMd/SFS, and SPL/IPS. The activation pattern closely overlapped with the dorsal attention network and OTC activation found for CM and CO compatibility. This result supports the assumption that the contextual compatibility effects reported above reflect a suppression/activation conflict, which is stronger when to-be-suppressed and to-be-activated information belongs to the same versus different contextual categories.

The observed effects for same versus different contextual categories resemble, to some degree, semantic distance effects during picture–word interference: Participants are slower in naming pictures of objects when distracter

words with related meanings are presented than if distracter words have unrelated meanings (Glaser & Dungelhoff, 1984). Similar findings were obtained when categorically related versus unrelated body, hand, and mouth actions were to be named (Vigliocco et al., 2002). These and other studies (Harris & Little, 2010; Vigliocco, Vinson, Lewis, & Garrett, 2004; Caramazza & Shelton, 1998) provide support for the idea that objects and actions are represented in categorical neighborhoods in semantic memory by sharing a greater number of conceptual features if they belong to the same contextual category than if they belong to a different contextual category. To our knowledge, this is the first study providing evidence for a fine-grained dissociability of category-specific items within the domains hand actions and tools, that is, kitchen, workshop, and office actions and tools. The present findings hence corroborate and extend previous findings on category-specificity in semantic memory.

Acknowledgments

REFERENCES

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

Bar, M., & Aminoff, E. (2003). Cortical analysis of visual context. *Neuron, 38,* 347–358.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron, 34,* 149–159.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). fMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience, 15,* 991–1001.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex, 19,* 2767–2796.

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience, 33,* 1–21.

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience, 10,* 1–34.

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.

Chao, L. L., Haxby, J. V., & Martin, A. (1999). Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nature Neuroscience, 2,* 913–919.

Corbetta, M. (1998). Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Sciences, U.S.A., 95,* 831–838.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3,* 201–215.

Corbetta, M., & Shulman, G. L. (2011). Spatial neglect and attention networks. *Annual Review of Neuroscience, 34,* 569–599.

Cristescu, T. C., Devlin, J. T., & Nobre, A. C. (2006). Orienting attention to semantic categories. *Neuroimage, 33,* 1178–1187.

Cristescu, T. C., & Nobre, A. C. (2008). Differential modulation of word recognition by semantic and spatial orienting of attention. *Journal of Cognitive Neuroscience, 20,* 787–801.

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.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18,* 193–222.

Downing, P. E., Chan, A. W., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex, 16,* 1453–1461.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598–601.

Epstein, R. A., & Ward, E. J. (2010). How reliable are visual context effects in the parahippocampal place area? *Cerebral Cortex, 20,* 294–303.

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

Fink, G. R., Dolan, R. J., Halligan, P. W., Marshall, J. C., & Frith, C. D. (1997). Space-based and object-based visual attention: Shared and specific neural domains. *Brain, 120,* 2013–2028.

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

Geyer, S., Matelli, M., Luppino, G., & Zilles, K. (2000). Functional neuroanatomy of the primate isocortical motor system. *Anatomy and Embryology (Berlin), 202,* 443–474.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience, 4,* 179–192.

Glaser, W. R., & Dungelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 640–654.

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

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron, 24,* 187–203.

Harris, I. M., & Little, M. J. (2010). Priming the semantic neighbourhood during the attentional blink. *PLoS One, 5,* e12645.

Jacob, P. (2009). The tuning-fork model of human social cognition: A critique. *Consciousness and Cognition, 18,* 229–243.

Jastorff, J., & Orban, G. A. (2009). Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *Journal of Neuroscience, 29,* 7315–7329.

Kable, J. W., & Chatterjee, A. (2006). Specificity of action representations in the lateral occipitotemporal cortex. *Journal of Cognitive Neuroscience, 18,* 1498–1517.

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: A Hebbian perspective. *Trends in Cognitive Sciences, 8,* 501–507.

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

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

Liu, T., Slotnick, S. D., Serences, J. T., & Yantis, S. (2003). Cortical mechanisms of feature-based attentional control. *Cerebral Cortex, 13,* 1334–1343.

Lohmann, G., Muller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., et al. (2001). LIPSIA-A new software system for the evaluation of functional magnetic resonance images of the human brain. *Computerized Medical Imaging and Graphics, 25,* 449–457.

Luppino, G., Matelli, M., & Rizzolatti, G. (1990). Cortico-cortical connections of two electrophysiologically identified arm representations in the mesial agranular frontal cortex. *Experimental Brain Research, 82,* 214–218.

Mita, A., Mushiake, H., Shima, K., Matsuzaka, Y., & Tanji, J. (2009). Interval time coding by neurons in the presupplementary and supplementary motor areas. *Nature Neuroscience, 12,* 502–507.

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

Nobre, A. C. (2001). The attentive homunculus: Now you see it, now you don't. *Neuroscience and Biobehavioral Reviews, 25,* 477–496.

Norris, D. G. (2000). Reduced power multislice MDEFT imaging. *Journal of Magnetic Resonance Imaging, 11,* 445–451.

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

Schmahmann, J. D., & Pandya, D. N. (2006). *Fiber pathways of the brain*. New York: Oxford University Press.

Schubotz, R. I. (2007). Prediction of external events with our motor system: Towards a new framework. *Trends in Cognitive Sciences, 11,* 211–218.

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

Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.

Tanji, J. (1994). The supplementary motor area in the cerebral cortex. *Neuroscience Research, 19,* 251–268.

Thompson, J., & Parasuraman, R. (2011). Attention, biological motion, and action recognition. *Neuroimage, 59,* 4–13.

Turk-Browne, N. B., Scholl, B. J., Johnson, M. K., & Chun, M. M. (2010). Implicit perceptual anticipation triggered by statistical learning. *Journal of Neuroscience, 30,* 11177–11187.

Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition, 85,* B61–B69.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology, 48,* 422–488.

Wager, T. D., Jonides, J., & Reading, S. (2004). Neuroimaging studies of shifting attention: A meta-analysis. *Neuroimage, 22,* 1679–1693.

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 modulation of action recognition. *Neuroimage, 59,* 1551–1559.

Yantis, S., & Serences, J. T. (2003). Cortical mechanisms of space-based and object-based attentional control. *Current Opinion in Neurobiology, 13,* 187–193.