Quadflieg, S., Gentile, F., & Rossion, B. (2015). The neural basis of perceiving person interactions. Cortex, 70, 5-20. 10.1016/j.cortex.2014.12.020

The Neural Basis of Perceiving Person Interactions

Susanne Quadflieg[1,2], Francesco Gentile[3,4], & Bruno Rossion[3]

Authors' Affiliations:

[1]School of Experimental Psychology, University of Bristol, UK.

[2]Division of Psychology, New York University Abu Dhabi, UAE.

[3]Psychological Sciences Research Institute and Institute of Neuroscience, University of Louvain, Louvain-la-Neuve, Belgium.

[4]Department of Psychology and Neuroscience, Maastricht University, Maastricht, The Netherlands.

Corresponding author:

Susanne Quadflieg

School of Experimental Psychology

University of Bristol

12A Priory Road

Bristol BS8 1TU

United Kingdom

email: s.quadflieg@bristol.ac.uk

Abstract

This study examined whether the grouping of people into meaningful social scenes (e.g., *two people having a chat*) impacts the basic perceptual analysis of each partaking individual. To explore this issue, we measured neural activity using functional magnetic resonance imaging (fMRI) while participants sex-categorized congruent as well as incongruent person dyads (i.e., two people interacting in a plausible or implausible manner). Incongruent person dyads elicited enhanced neural processing in several high-level visual areas dedicated to face and body encoding and in the posterior middle temporal gyrus compared to congruent person dyads. Incongruent and congruent person scenes were also successfully differentiated by a linear multivariate pattern classifier in the right fusiform body area and the left extrastriate body area. Finally, increases in the person scenes' meaningfulness as judged by independent observers was accompanied by enhanced activity in the bilateral posterior insula. These findings demonstrate that the processing of person scenes goes beyond a mere stimulus-bound encoding of their partaking agents, suggesting that changes in relations between agents affect their representation in category-selective regions of the visual cortex and beyond.

Keywords: social dyad, social interaction, social relation, person perception

1. Introduction

Whether people shake hands, have a chat, a fight, or a dance, go out for a drink, or wave goodbye, they are frequently seen in each other's company. Witnessing such person interactions and making sense of them, an activity sometimes referred to as people-watching, not only allows for an entertaining everyday distraction but also poses an impressive social-cognitive feat. Initial social-psychological work indicates, for instance, that observers of person dyads portrayed in brief video clips easily decipher whether interacting people are friends, romantic partners, or work colleagues (Costanzo & Archer, 1989). Similarly, a brief look at static photographs of person dyads provides sufficient information to determine whether two people are teasing or fighting each other (Sinke, Sorger, Goebel, & de Gelder, 2010) or whether they interact for instrumental or socio-emotional reasons (Proverbio et al., 2011). The neurofunctional stage at which sensitivity to meaningful person interactions arises in the person construal process, however, remains largely unknown.

Contemporary work on person perception and person inferences focuses mainly on the processing of single individuals (see Ames, Fiske, & Todorov, 2011; Leising & Borkenau, 2010; Macrae & Quadflieg, 2010; Overwalle & Baetens, 2009; Zaki, 2013). As a result, theories on how we encode and integrate visual information involving several people remain poorly developed. In the field of object perception, however, numerous studies suggest that the encoding of visual information comprising several distinct entities depends crucially on how they relate to one another. For instance, when objects are portrayed in a meaningful rather than a meaningless interaction (e.g., a pitcher positioned to be pouring into a glass versus away from the glass), neuropsychological patients (Riddoch, Humphreys, Edwards, & Willson, 2003; Riddoch et al., 2011; Wulff & Humphreys, 2013), as well as healthy adults (Green & Hummel, 2006; Mudrik, Breska, Lamy, & Deouell, 2011; Roberts & Humphreys, 2011) display facilitated object recognition.

A likely neural substrate to underlie this facilitation effect is the lateral occipital cortex (LOC; Malach et al., 1995), a brain region that has previously been linked to object shape processing (e.g., Grill-Spector et al., 1999; Kourtzi & Kanwisher, 2001). Not only is there evidence that average activity in the LOC differentiates between meaningful and meaningless object interactions (Kim & Biederman, 2011; Roberts & Humphreys, 2010), but also that neural patterns in this region as examined using multi-voxel pattern analysis (MVPA) capture the different types of object scenes (Baeck, Wagemans, & Op de Beeck, 2013). Additional data demonstrate that disrupting LOC's normal functioning by transcranial magnetic stimulation (TMS) abolishes the interaction-based facilitation effect (Kim, Biederman, & Juan, 2011). In summary, these findings suggest that the grouping of objects into a conceptual unit modulates even basic mechanisms of object perception.

If the same mechanism holds true in the realm of person perception, the grouping of several people into one scene may also impact perceptual encoding of human faces and bodies (Hirai & Kakigi, 2009; Neri, Luu, & Levi, 2006). Such encoding is thought to occur in the so-called *core neural network of person perception* (Gobbini & Haxby, 2007; Rossion, Hanseeuw, & Dricot, 2012; Weiner & Grill-Spector, 2010), a system that comprises several brain regions, including the occipital face area (OFA), fusiform face area (FFA), posterior superior temporal sulcus (pSTS), extrastriate body area (EBA) and fusiform body area (FBA). All five regions act in concert to extract the structural and dynamic representation of an individual's facial and bodily appearance (Grosbras, Beaton, & Eickhoff, 2012; Gobbini & Haxby, 2007; Pavlova, 2012; Peelen & Downing, 2007; Weiner & Grill-Spector, 2010). When their interplay gets disturbed – through brain damage, repetitive TMS, or intracerebral electrical stimulation – face and body recognition skills decline, indicating the network's necessity for adequate person perception (e.g., Barton, Press, Keenan, & O'Connor, 2002; Jonas et al., 2012; Grossman, Battelli, & Pascual-Leone, 2005; Pitcher, Garrido, Walsh, &

Duchaine, 2009; Sorger, Goebel, Schiltz, & Rossion, 2007; Urgesi, Candidi, Ionta, & Aglioti, 2007).

While initial neuroimaging work suggests that perceiving multiple people recruits brain areas dedicated towards person perception as well as mentalizing and action understanding (Hooker, Verosky, Germine, Knight, & D'Esposito, 2010; Iacoboni et al., 2004; Sinke et al., 2010; Wagner, Kelly, & Heatherton, 2011; Walter et al., 2004), it remains uncertain at what stage in the neural processing cascade a differentiation of meaningful from meaningless person interaction occurs. Only few studies have compared the neural responses elicited by meaningful person interactions to person dyads in which both agents acted independently from each other (Centelles, Assaiante, Nazarian, Anton, & Schmitz, 2011) or systematically faced away from one another (Kujala, Carlson, & Hari, 2011). The neural differences observed across these different types of person dyads are hard to interpret, however, due to the presence of additional low level visual confounds. For instance, agents shown in these studies held different bodily postures during interactions and control scenarios, were in closer spatial proximity to each other during the former than the latter, and/or displayed mutual eye gaze and/or touch during real person interactions only.

Taking into account these limitations, the current functional magnetic resonance imaging (fMRI) study addressed the question whether the type of social relations between people can modulate the encoding of visual person information, ensuring that low-level visual differences are strictly controlled for. Specifically, participants were asked to view a series of person dyads comprising *congruent person interactions* (i.e., two agents acting in a related manner and facing each other), *incongruent person interactions* (i.e., two agents acting in an unrelated manner and facing each other), and so-called *non-interactions* (i.e., two agents acting in a related manner but facing away from each other). To ensure that perceivers held identical processing goals across all three types of dyads, a basic categorization task – judging whether the depicted agents matched with regard to sex – was

used throughout the experiment. Finally, to identify critical components of the core person perception network for each participant (i.e., FFA, FBA, OFA, EBA, pSTS), a well-controlled functional localizer was administered (Quadflieg et al., 2011). In line with previous reports of perceptual incongruency in person perception, we anticipated that activity in the core network would be modulated by dyad type, such that perceiving incongruent person interactions and non-interactions would elicit enhanced neural processing compared to congruent interactions (cf. Egner, Monti, & Summerfield, 2010; Quadflieg et al., 2011; Zaki, 2013).

2. Methods

2.1. Participants

Twelve Caucasian volunteers (5 males), aged between 21 and 31 years (mean: 26.0 years), participated in the study. All volunteers reported normal or corrected-to-normal vision. Eleven participants were right-handed, one participant left-handed as determined by the Edinburgh Handedness Inventory (Oldfield, 1971). None of the participants had a history of neurological or neuropsychiatric disorders or was currently taking psychoactive medications. Informed consent was obtained from all individuals.

2.2. Stimuli

During the person perception localizer (taken from Quadflieg et al., 2011), stimuli comprised six different types of targets including faces (42 different identities, 21 females), bodies (42 different identities, 21 females), and cars (42 different models), as well as their phase-scrambled controls. Faces were shown in frontal view with a neutral expression and direct gaze and cropped so that no hair was present. Bodies were cropped in a manner that they did not contain any head or neck information. All stimuli were presented in color, embedded in the same uniform gray background, standardized to a common size of 184 (width) x 210

(height) pixels (visual angle: 4.2° x 4.8°), and matched on mean luminosity as well as mean contrast. Image scrambling was realized by using Fourier phase randomization (Sadr & Sinha, 2004).

During the main experiment, stimuli comprised a series of color photographs of person dyads. Stimulus creation began by downloading 40 images of *congruent person interactions* from shutterstock® Photos (www.shutterstock.com). Half of these interactions were mixed-sex dyads, the other half depicted same-sex interactions (half of which took place between women, the other half between men). In order to be selected for the current study, all stimuli had to meet the following criteria: The behavior of both agents had to be related to each other in a meaningful manner but agents were not allowed to display actual bodily contact/touch or overlap with regard to their bodies and/or heads (e.g., *approaching for a handshake*, *having an argument*, *going for a run*, *waving at each other* etc., see Figure 1A). All interactions were embedded in a uniform white background and standardized to a common size of 350 x 350 pixels (visual angle: 8° x 8°).

Using Adobe Photoshop CS5, the stimuli were then altered to create *incongruent person interactions*. To do so, two agents from different congruent interactions were paired in a pseudorandom manner so that they would face each other but without engaging in a meaningful social encounter. The pseudorandom pairing was done so that agents were paired with new partners of the same sex as their original interaction partner (see Figure 1B). In addition, a non-interaction condition was added in which the position of the agents in the original person interactions were simply swopped, resulting in a set where agents were shown with their backs turned towards each other (see Figure 1C). While this manipulation allowed us to present the exact same agents as in the interaction condition, the two agents shown were no longer involved in a meaningful exchange. In sum, each individual appeared three times during the experiment: Once shown in a congruent person interaction, once shown in an incongruent person interaction, and once shown in a non-interaction.

Care was taken that non-interactions and incongruent interactions resembled the original congruent interactions on a number of perceptual dimensions. In particular, we aimed for the three conditions to be similar with regard to dyad width, inter-agent distance, and inter-agent center of mass (CoM) distance. The image processing toolbox in MATLAB (VersionR2012b, ©The MathWorks, Inc.) was used to compute all three indicators for each image (see Table 1 and Figure 2). To determine image width, for both agents the point that was closest to the image's outer frame on the left or on the right side of the image (for the left and the right agent, respectively) was determined. Then, the difference in the image's x dimension between these two points was computed in pixels (see Figure 2A). Submitting these to a repeated measures Analysis of Variance (ANOVA) with the three-level factor dyad type (congruent interactions, incongruent interactions, non-interactions) revealed no significant differences [$F(2, 78) = .22$, $p = .80$].

In a next step, inter-agent distance was quantified. Thus, for both agents the point that was closest to the other agent was determined (see Figure 2B). Again, the difference between these two points in the image's x dimension was calculated. Note, this value can be negative when these points (e.g., due to outstretched limbs) cross in the image's x-dimension. Again, the above ANOVA revealed that inter-agent distance was equivalent across dyad types [$F(2, 78) = .38$, $p = .69$]. Finally, the CoM for each agent in an image was determined to compute the distance (in pixel) between these two sets of coordinates for each dyad (see Figure 2C). This time significant differences emerged [$F(2, 78) = 15.20$, $p < .01$]. As a result of keeping overall dyad width and inter-agent distance constant, swopping the order of agents led to a reduction in CoM distance for non-interaction images. In other words, these images showed a significantly smaller CoM distance than congruent interactions [$t(39) = 4.16$, $p < .01$] and incongruent interactions [$t(39) = 4.90$, $p < .01$]. Most importantly, however, the latter two main conditions of the study did not differ from each other [$t(39) = 1.03$, $p = .31$].

In an additional preparatory step, it was examined that the created images had the intended semantic effects. Therefore, twenty volunteers (10 males, mean age = 25.4 years) were asked to view all images (one at a time, in a randomized manner) on a computer screen to judge each picture for its content via a button press. Specifically, volunteers were asked to rate to what extent the two people in an image were interacting in a meaningful manner (1= "*not at all*" to 9= "*entirely*"). Though participants were under no time pressure to respond, they were encouraged to rely on their first impressions. Submitting these semantic ratings (see Table 1) to the above ANOVA revealed the intended main effect [$F(2, 78) = 244.76$, $p < .01$]. That is, congruent person interactions were rated more meaningful than incongruent interactions [$t(39) = 20.44$, $p < .01$], which in turn were rated more meaningful than non-interactions [$t(39) = 2.46$, $p = .01$]. This perceptually standardized and semantically pre-tested set of images was then used in the current fMRI study.

2.3. FMRI Tasks and Procedure

During the person perception localizer participants viewed blocks of subsequently presented images and performed a 1-back repetition detection task, requiring them to press a button for any immediate repetition of the same image. The functional localizer comprised three separate runs, each lasting about 11 min. In each run, participants encountered 4 blocks of each of the 6 types of visual stimuli (faces, scrambled faces, bodies, scrambled bodies, cars, scrambled cars) resulting in a total of 24 alternating blocks per run (see Quadflieg et al., 2011). Each block consisted of 18 stimuli from the same visual category and lasted 18 sec. Within each block each stimulus was presented for 750 ms followed by a blank screen for 250 ms. One or two out of the 18 stimuli per block were repeated resulting in at least four repetitions for each visual category per run and in exactly 14 repetitions per visual category across runs. On each trial, image presentation on the screen's uniform white background was centralized but varied slightly in location (X = ~5%; Y = ~5%) to prevent participants from adapting to low level cues and to avoid repetition decisions based on the inspection of

only small sectors of the images. Between blocks a white fixation cross was shown at the center of the screen for 9 sec. For each localizer run a fixed, pseudo-randomized block order was created so that the same visual category was never run back-to-back and every visual category followed any other category at least twice but no more than three times. In addition, the three localizer runs were presented in counterbalanced order across participants. Responses were given by pressing a button on a button box with the index finger of the right hand.

The main experiment was set up as a slow event-related fMRI design during which images of person dyads were shown. For each image participants were asked to report via button press whether the two people shown were of the same or of a different sex. The main experiment comprised two separate runs, each lasting about 13 min. In each run, participants encountered 20 trials of each image type resulting in 60 total trials per run. Trials were presented in a new random order for each run and participant. For each trial, participants saw a target image presented centrally on a uniform white background. After 2250 ms (= 1TR) the image was replaced by a black fixation cross that varied in duration between 8 and 12 sec (with an average of 10 sec). Responses were given by pressing one of two buttons on a button box with the index or middle finger of the right hand. All participants classified same-sex images with their index finger and different-sex images with their middle finger. In each experimental condition, exactly half of the images showed agents that matched with regard to sex, whereas the other half differed. In consequence, equal numbers of index and middle finger presses were required across experimental conditions.

Overall, the order of the five experimental runs was fixed such that all participants first encountered two person dyad runs that were then followed by three localizer runs. During all runs, stimuli were back projected onto a screen visible via a mirror mounted on the MRI head coil. Stimulus presentation and recording of participants' responses and associated latencies was done using Presentation® software (version 9.13, Neurobehavioral Systems, inc.,

Albany, California). To familiarize participants with both tasks, they completed 2 practice blocks of the localizer task (1 face block, 1 scrambled bodies block) and 12 person dyad trials on a MacBook Pro portable computer equipped with a 15 inch screen outside the scanner. None of the practice images were included in the actual experiment.

2.4. Image Acquisition

Image acquisition was undertaken at the University of Maastricht (The Netherlands) on a 3 Tesla head scanner (Siemens Allegra, Erlangen, Germany) with an 8 channels phased array head coil. For registration purposes T1-weighted anatomical images were acquired for each participant using an ADNI sequence (192 sagittal slices, TE = 2.6 ms, TR = 2250 ms, flip angle = 9°, voxel size = 1 x 1 x 1 mm, acquisition matrix: 256 x 256). Functional images were collected using a repeated single-shot echo-planar imaging sequence sensitive to BOLD contrast (TR=2250 ms, TE 30 ms, flip angle = 90°, 3.5 x 3.5 in-plane resolution; field of view 224 mm; acquisition matrix 64 x 64). For each volume 36 axial slices; 3.5 mm slice thickness, 0 mm skip between slices was acquired. In total, 298 volumes were collected for each of the three runs of the functional localizer and 338 volumes for each run of the main experiment. To account for T1 saturation effects the first 5 volumes of each run were discarded.

2.5. Data Analysis

Behavioral data were analyzed using SPSS for Windows. For image analyses we used the BrainVoyager software package (Version 2.4.1, Brain Innovation, Maastricht, The Netherlands). For the functional localizer, image pre-processing included slice scan time correction with sinc interpolation, 3-dimensional motion correction with trilinear interpolation, spatial smoothing with 6 mm full-width at half-max Gaussian filter, and temporal filtering using a high-pass filter of 3 cycles over the run’s length for linear trend removal. For the main

experiment, the same pre-processing procedures were applied but a temporal high-pass filter of 8 rather than 3 cycles was used to account for the slow event-related design. During pre-processing the direction and magnitude of motion for each participant over the course of each functional run was examined. Estimated translation and rotation parameters were inspected and never exceeded 3 mm or 3 degrees. All functional images were coregistered to each individual subject's intensity-inhomogeneity corrected anatomical scan. Anatomical scans were transformed into Talairach coordinates (Talairach & Tournoux 1988), on which all the statistical analyses were performed.

For each run of each participant's localizer and main experiment, a BrainVoyager protocol file (PRT) was derived representing the onset and duration of the events for the different conditions. From the created protocols, design matrices were defined using an appropriate boxcar waveform with a doublegamma hemodynamic response function to create regressors of interest for each experimental condition (Friston et al., 1998). We then used individual-level GLM analyses to identify regions of interest (ROI) for each participant based on the localizer runs (see Quadflieg et al., 2011; Rossion et al., 2012): To determine face-selective ROIs (i.e., FFA, OFA, and pSTS), a conjunction of the contrasts faces versus cars and faces versus scrambled faces [(faces – car) & (faces – scrambled faces)] was computed to control for low-level visual differences between faces and objects as well as high-level representations that are common for faces and objects. Similarly, to determine body ROIs (i.e., EBA, FBA), a conjunction of the contrasts bodies versus cars and bodies versus scrambled bodies [(bodies – car) & (bodies – scrambled bodies)] was computed. Based on these conjunctions, for each ROI in each participant, the most significantly activated voxel was identified within a restricted part of cortex based on previously reported anatomical locations and mean Talairach coordinates (Downing, Chan, Peelen, Dodds, & Kanwisher, 2006; Rossion et al., 2012). ROIs were defined as the set of at least 5 contiguous voxels that were significantly activated (all $p < .05$, uncorrected) within a 9 x 9 x 9

mm cube of this region-specific peak voxel. This procedure was chosen to ensure that ROIs could be segregated from nearby activations and to roughly equate the number of voxels included across different ROIs.

For the main experiment, group-level GLM analyses treating participants as a random effect were computed. To minimize false-positive results, effects were considered statistically significant using a voxelwise threshold of $p < .001$ and a cluster-based threshold of $p < .05$ (implemented via the BrainVoyager QX cluster threshold estimator plugin, see Goebel, Esposito, & Formisano, 2006). In addition, the mean parameter estimates from the main experiment were extracted for each participant and ROI as identified in the functional localizer. These estimates were then submitted to a one factor (dyad type: *congruent interaction*, *incongruent interaction*, *non-interaction*) repeated measures ANOVA. Moreover, we conducted a pattern classification analysis for the localizer-based ROIs (see Formisano, De Martino, & Valente, 2008; Etzel, Gazzola, & Keysers, 2009) using a support vector machine (SVM) as implemented in BrainVoyager with a linear kernel and the cost parameter fixed at 1. For each class (congruent interactions, incongruent interactions, and non-interactions), beta parameters were extracted fitting a 2 gamma HRF (parameter range for fit: -.60 to .60) for each individual trial across a time window of 1 volume pre-stimulus onset to 4 volumes post-stimulus onset. The resulting single-trial responses across relevant voxels (e.g., ROI voxels with signal values ≥ 100) then formed the feature vectors used to train the classifier. For each participant and for each of the two pairs of conditions of relevance (i.e., congruent interactions versus incongruent interactions as well as congruent interactions versus non-interactions), a linear SVM was trained using a *leave-five-out* procedure. Thus, a classifier that best separated the data from the two conditions of relevance was constructed for each participant and then applied to 10 test items (5 per condition) to determine its decoding accuracy. The procedure was repeated 20 times for each perceiver and pair of conditions. In addition, labels were assigned randomly before training the classifier to

establish that the classifier's mere chance-based generalization performance (number of permutations: 1000) was at chance level (i.e., 50%) for each participant.

Finally, we also aimed to identify brain regions in which activation during the presentation of person dyads (regardless of its type) was a linear function of the semantic ratings for each scenario as obtained by independent participants. Along similar lines, we examined activation during person dyad viewing that was a linear function of the visual properties (i.e., image width, inter-agent distance, inter-agent CoM distance) of each scenario over all experimental conditions per participant. For this parametric analysis, all trials were treated as belonging to a single condition, and the four variables of interest were included as parametric modulators one after the other (i.e., four separate models were run). The obtained *t*-statistics were analyzed on the group level by means of a conjunction [(stimulus-baseline) & (parametric modulation-baseline)] to identify regions that a) responded towards the scenarios and b) did so in a parametric manner. Again, participants were treated as a random effect (i.e., Brain Voyager's 'conjunction of RFX option' was used). A voxelwise threshold of $p < .001$ with a cluster-based threshold of $p < .05$ was applied.

3. Results

3.1. Behavioral Analysis

Accuracy rates in the person dyad task (see Table 1) revealed no significant main effect of dyad type [$F(2,22) = 0.31$, $p = .74$], indicating that accuracy did not differ across experimental conditions (*congruent interactions* versus *incongruent interactions* versus *non-interactions*). Error trials were excluded from further statistical analyses before median response times were calculated (see Table 1). Sex categorization times on correct trials did not differ across the three types of dyads [$F(2,22) = .24$, $p = .79$].

3.2. Localizer-Based FMRI Analysis

Table 2 lists the average peak Talairach coordinates of all ROIs (i.e., FFA, FBA, OFA, EBA, pSTS) across participants, including the number of individuals for which each ROI was identified. The regions observed are in agreement with previous fMRI studies (e.g., Peelen et al., 2007; Rossion et al., 2012; Weiner & Grill-Spector, 2010) and replicate reports according to which face-selective activity is right-lateralized and elicited more consistently in the FFA than the OFA (e.g., Andrews, Davies-Thompson, Kingstone, & Young, 2010; Engell & McCarthy, 2013; Rossion et al., 2003). Mean parameter estimates in all five ROIs were extracted from the main experiment for each participant (see Figure 3). For the OFA, no significant effects emerged in either of the two hemispheres (both Fs < 1.89, *ns*). For the right and left FFA, a significant main effect of image type emerged [right: $F(2,20) = 7.69$; left $F(2,20) = 10.75$, ps < .05]. The same was found for the pSTS [right: $F(2,20) = 22.02$; left $F(2,12) = 11.16$, ps < .05]. For the body perception network significant main effects emerged for both the EBA [right: $F(2,22) = 12.86$; left $F(2,22) = 13.45$, ps < .05] and the FBA [right: $F(2,22) = 8.45$; left $F(2,20) = 7.42$, ps < .05].

Follow-up t-tests conducted for ROIs with a significant main effect revealed that activity was enhanced throughout these regions during the perception of incongruent interactions compared to congruent person interactions (all ts > 2.55, ps < .05) as well as compared to non-interactions (all ts > 3.83, ps < .05). When comparing non-interactions and congruent person interactions, only a marginally significant effect in the right EBA emerged [signaling a larger response for non-interactions see Figure 4; $t(11) = 1.91$, p =.08; all other ts < 1.10, *ns*]. Moreover, SVM accuracy was significantly above chance in the right FBA [M = 53.58; $t(11) = 2.21$, p = .03 one-sided] and the left EBA [M = 55.25; $t(11) = 2.34$, p = .02 one-sided] when discriminating between congruent and incongruent interactions (see Figure 5). In contrast, classification accuracy for congruent interactions versus non-interactions was at chance level in the entire person perception network [all ts < 1.12, ps ≥ .10 one-sided].

3.3. Whole-Brain FMRI Analysis

Exploratory univariate whole-brain analyses were undertaken to examine the effects of person dyad type beyond those observed in the core person perception network (see Table 4). First, congruent and incongruent interactions were compared. The contrast congruent interactions > incongruent interactions yielded no significant effects. The reverse contrast revealed an enhanced response bilaterally in the posterior middle temporal gyrus (BA 37, see Figure 6). Similarly, the contrast non-interactions > incongruent interactions also failed to yield suprathreshold activation, whereas the reverse contrast revealed enhanced activity bilaterally in the posterior middle temporal gyrus (BA 37), in the right parahippocampal gyrus (BA 19), and the left pSTS (BA 21), as displayed in Figure 7. Finally, comparing congruent interactions and non-interactions revealed no significant effects.

3.4. Parametric FMRI Analysis

At a standard level of thresholding no parametric effects emerged for any of the variables under investigation (i.e., image width, inter-agent CoM difference, inter-agent distance, semantic ratings). For exploratory reasons, we repeated all parametric analyses adopting more lenient thresholding criteria (i.e., a voxelwise threshold of $p < .05$ and a minimal cluster-size of 20 voxel). Despite this adjustment, no suprathreshold effects emerged for image width. In contrast, an enlargement in absolute inter-agent distance was accompanied by enhanced activity in the lingual gyrus bilaterally (RH: BA 19; peak voxel x = 8, y = -56, z = -6; $t = 3.32$, 22 voxel; LH: BA 18; peak voxel x = -7, y = -68, z = 0; $t = 3.06$; 87 voxel). Similarly, an enlargement in CoM distance [note that this indicator was unrelated to absolute inter-agent distance: $r(118) = .08$, $p > .10$] was accompanied by increased brain activity in three brain regions, specifically in the bilateral lingual gyrus, (BA 17; peak voxel x = -10, y = -89, z = -3; $t = 4.72$, 170 voxel; cluster extending to the right hemisphere), in the left parahippocampal gyrus (BA 30; peak voxel x = -14, y = -35, z = -6; $t = 3.16$, 36 voxel), and

in the left posterior middle temporal gyrus (BA 37; peak voxel x = -52, y = -61, z = 0; *t* = 3.80, 51 voxel). Finally, an increase in the scenes' semantic meaningfulness was accompanied by enhanced activity in the bilateral posterior insula [BA 13; RH: peak voxel x = 32, y = -2, z = 15, *t* = 3.62, 25 voxel; LH: peak voxel x = -46, y = -5, z = 9; *t* = 4.17, 24 voxel; see Figure 8]. Note that increases in the above variables were not accompanied by any systematic decreases in brain activity.

4. Discussion

Contemporary behavioral as well as neuroimaging work on person perception and person inferences focuses largely on the processing of single individuals (cf. Ames et al., 2011; Leising & Borkenau, 2010; Macrae & Quadflieg, 2010; Overwalle & Baetens, 2009; Zaki, 2013). As a result, theories on how observers encode and integrate visual information involving several people are poorly developed. The lack of scientific inquiry regarding the topic seems surprising given that initial studies suggest that humans are highly skilled at making sense of scenes involving multiple individuals (Costanzo & Archer, 1989; Neri et al., 2006; Sinke et al., 2010; Proverbio et al., 2011). Recent fMRI studies additionally suggest that witnessing meaningful social interactions recruits brain regions dedicated towards person perception and social-cognitive reasoning (Centelles et al., 2011; Kujala et al., 2011).

Building on this fascinating work, the current study examined the extent to which different types of person dyads modulate the visual encoding of their partaking agents. Note that this question deserves particular attention as it also addresses the broader topic of how face and body encoding is accomplished in the human brain (cf. Quadflieg & Rossion, 2011). According to Downing and Peelen (2011), the core person perception network provides a "*cognitively unelaborated* perceptual framework that allows *other* cortical systems to exploit the rich, socially relevant information" (p. 186). An alternative view to such stimulus-bound processing in the core network suggests that social-cognitive reasoning about the mental

states, intentions and goals of others may penetrate (i.e., cognitively elaborate) even the perceptual analyses of human targets (e.g., Freeman, Johnson, Adams, & Ambady, 2012; Gilbert & Li, 2013; Teufel, Fletcher, & Davis, 2011).

In the context of the current study, we showed participants a series of person dyads comprising congruent interactions, incongruent interactions, and non-interactions, before they completed a face and body perception localizer task. Challenging a strictly stimulus-bound view, we found that activity within the core person perception network – with the exception of the OFA – was enhanced when observers viewed the same targets involved in socially incongruent interactions compared to congruent person interactions. A multivariate analysis additionally revealed that based on neural activity in the right FBA as well as in the left EBA, a linear classifier successfully distinguished between congruent and incongruent interactions. These observations indicate that high-level visual areas of the core network do not merely process face and body stimuli as single, unrelated entities.

Enhanced neural processing within the FFA, pSTS, FBA, and EBA for incongruent interactions is likely to reflect increased processing demands to form a coherent person percept when witnessing ambiguous person dyads. Along similar lines, enhanced activity in the person perception network has also been reported for targets that displayed incongruent information in their facial and bodily appearance, either due to expressing conflicting emotional states (e.g., Meeren, van Heijnsbergen, & de Gelder, 2005) or due to exhibiting contradictory gender cues (Quadflieg et al., 2011). Furthermore, incongruence of emotional valence between facial expressions and learned associations with background colors has been found to increase activity in the mid-lateral fusiform gyrus and the inferior occipital gyrus (i.e., in brain regions that correspond to the FFA and OFA; Frühholz, Fehr, & Herrmann, 2009). However, existing studies on incongruency in person perception have focused exclusively on the processing of single individuals. In contrast, the current study demonstrates that enhanced processing in the core person perception network can also be

elicited when incongruency arises from the relation between two separate individuals. Importantly, the neural increase for incongruent relative to congruent social interactions was found in the current study during an explicit processing task that did not induce a concomitant congruency-related behavioral effect and for stimuli well matched on low-level visual properties. In other words, the current study further supports the notion that visual processing is compromised for incongruent relative to congruent dyads of entities, regardless whether people or objects are concerned (cf. Riddoch et al., 2003; 2011; Wulff & Humphreys, 2013; Green & Hummel, 2006; Roberts & Humphreys, 2011).

This observation raises several questions that deserve future examination. The first question concerns the exact source of the effect of incongruency as observed in the current study. Incongruent and congruent dyads differed not only with regard to their relatedness of both agents' body postures (i.e., their actions), but also regarding the relatedness of their facial expressions and clothing style. In consequence, combining individuals from different original interactions resulted in incongruent dyads that were inherently less frequent and more unusual than their congruent counterparts. Neural effects of incongruency may therefore reflect the combination of any of these elements in an unlikely manner. While we aimed to account for this issue by also including non-interactions, these control items differed from congruent and incongruent dyads in several other important aspects (e.g., a systematically reduced CoM distance between agents, a lack of mutual eye gaze). Future work must therefore carefully examine which specific aspects of incongruency elicit systematically enhanced activity in the person perception network.

Another pivotal question is whether the effect of incongruency takes place during early visual encoding or results from late reentrant processes. Both a mismatch between perceivers' pre-existing perceptual representations and the actual visual input or an increase in difficulty of information integration may account for the neural increases observed towards incongruent interactions in the present study (cf. Bar, 2004; Biederman, Mezzanotte, &

Rabinowitz, 1982; Mudrik et al., 2011; Palmer, 1975). Thus, on the one hand, different types of person dyads might modulate the early perceptual encoding of both targets, so that their visual representation occurs in a context-dependent manner. On the other hand, multiple targets may be encoded independently from one another with incongruency emerging only at a post-perceptual processing stage but ultimately feeding back into person perception areas, driving the observed activity enhancement. Both mechanisms are widely debated in the domain of visual scene processing with some authors favoring the idea of contextual processing (e.g., Bar, 2004; Biederman et al., 1982; Boyce, Pollatsek, & Rayner, 1989; Kosslyn, 1994; Neri, 2014), while others argue for a post-perceptual explanation (e.g., De Graef, 1992; Hamm, Johnson, & Kirk, 2002; Hollingworth & Henderson, 1998).

In the current study at least a univariate whole brain analysis failed to detect dyad-dependent neural differences in post-perceptual (i.e., higher cognitive) brain regions, therefore tentatively challenging a reentrant account. Given that such reasoning builds on a null finding, however, and considering the low temporal resolution of fMRI, electromagnetic recordings during congruent and incongruent person dyad processing would provide more relevant data to address this issue. So far, event-related potentials (ERPs) evoked by scenes that depict a person performing an action using either a congruent or an incongruent object (e.g., a man shaving with a razor or with a fork) report congruity effects starting at approximately 300 ms (Mudrik, Lamy, &, Deouell, 2010; Mudrik, Shalgi, Lamy, & Deouell, 2014). Future studies should examine whether incongruency in social interactions may elicit effects at even earlier latencies related to the perceptual encoding of human faces and/or bodies (i.e., around 170 ms after stimulus onset; Bentin et al., 1996; Pourtois, Peelen, Spinelli, Seeck, & Vuilleumier, 2007; Rossion & Jacques, 2011; Soria Bauser & Suchan, 2013).

Interestingly, beyond the observed localizer-based results, multivariate differentiation of congruent and incongruent interactions was observed in the right FBA and in the left EBA.

Recent work on the LOC suggests that the region's response to object pairs reflects a weighted average of the response patterns elicited by its constituent objects, with the maximum single-object response getting weighted more than the minimum response (Baeck et al., 2013). Most importantly, the difference in weights between the two responses has been found to be systematically larger for interacting than non-interacting object pairs, a circumstance allowing for the successful classification of different types of object pairs. A similar mechanism might account for the current observation. Importantly, our findings indicate that the successful classification of congruent and incongruent interactions occurs only in body-specific ROIs, suggesting that the representation of individuals' bodily postures is particularly relevant during the perception of incongruent interactions. Future research will be required, however, to understand why the observed effects occurred in a lateralized manner as the functional significance of lateralization effects in body perception regions remains poorly understood (cf. Downing & Peelen, 2011; Romaiguère, Nazarian, Roth, Anton, & Felician, 2014).

Beyond the effects observed in the core person perception network, incongruent interactions also increased activation in the posterior middle temporal gyrus (pMTG). The bilateral activity was located slightly posterior to the face-selective pSTS as well as superior-anterior to the body-selective EBA (see Figure 9), in a region typically referred to as the middle temporal area (MT/V5; cf. Dumoulin et al., 2000; Hampson, Olson, Leung, Skudlarski, & Gore, 2004). While this region has generally been linked to the perception of visual motion (Hampson et al., 2004; Kolster, Peeters, & Orban, 2010), it has been shown to display particular sensitivity to the perception of (implied) human movement (Ferri, Kolster, Jastorff, & Orban, 2013; Grosbras et al., 2012). Therefore, our results suggest that the extraction of motion cues as pivotal signals of agents' intentions and goals gains in relevance when perceivers process ambiguous social scenes (cf. Hirai & Kakigi, 2009, Neri et al., 2006).

In contrast to our findings on incongruent interactions, the current study largely failed to uncover differences in neural processing when comparing non-interactions with congruent interactions. Only a marginally enhanced mean activity difference in the right EBA towards non-interactions relative to person interactions was observed. This finding may seem surprising given that non-interactions were considered the least meaningful dyads based on the collected pilot ratings. It must be kept in mind, however, that non-interactions closely resembled congruent interactions in various aspects of relevance. For instance, both types of dyads showed the same semantically related agents (e.g., a direction giver and a direction receiver) as well as their corresponding bodily postures, facial expressions, and clothing styles. This far-reaching overlap may have resulted in a similar visual analysis for both non-interactions and congruent interactions. Yet, effects of non-interactions (or the lack thereof) must be interpreted with caution because these dyads differed from both congruent and incongruent interactions on several low-level visual properties (e.g., the agents' did not look at each other, their CoM distance was systematically reduced).

It must also be noted that the observed lack of neural differentiation between congruent interactions and non-interactions in the current study conflicts with previous reports in the literature (cf. Kujala et al., 2011). Several factors might explain this discrepancy. First and foremost, the current study carefully standardized low-level visual properties of the different image types while such confounds may have partially driven earlier results. Our parametric analyses indicate, for instance, that increases in inter-agent distance as well as in CoM distance across people scenes suffice to elicit enhanced activity in several brain regions including the bilateral lingual gyrus, the left parahippocampal gyrus and/or the left posterior middle temporal gyrus. It seems likely that additional factors such as systematic differences in the presence of mutual touch (cf. Bolognini, Rossetti, Convento, & Vallar, 2013) or body postures (Wiggett & Downing, 2011) across conditions – as present in previous work – could also impact neural processing (as discussed by Kujala et al., 2011).

Beyond low-level visual confounds, however, the role of observers' processing goals during scene perception might also be of pivotal importance to explain the diverging results. Specifically, in the object perception literature, it has been reported that neural differences for meaningful and meaningless object pairs emerge more strongly when participants are required to explicitly evaluate object relations rather than when performing a mere 1-back detection task (Baeck et al., 2013). In the current study, participants adopted a processing goal (i.e., sex categorization) that did not require explicitly attending to the dyads' specific relations. While this instruction ensured that participants were not artificially forced to process person dyads as a unit (as we were interested in their spontaneous response), it failed to encourage them to process the scenes' social content. As a result, participants' inclination to mentalize about the agents' actions or intentions was likely to be reduced compared to that of perceivers involved in previous studies who were encouraged to attend to the agents' attitudes towards each other (cf. Kujala et al., 2011) or to judge whether two agents acted in a coordinated manner (Centelles et al., 2011). Future research will need to examine the specific modulatory influence of processing goals on person dyad processing (cf. with work on single targets: Overwalle & Baetens, 2009; Spunt & Lieberman, 2012; Spunt, Satpute, & Lieberman, 2011).

Despite perceivers not being explicitly required to process the social relations embedded in our people images, a parametric analysis revealed sensitivity to the social context in which individuals were shown. Note that even though all three types of person dyads differed systematically on social meaningfulness on average, within each type a range of evaluations was recorded ($MIN_{congruent}$ = 5.10 to $MAX_{congruent}$ = 8.60; $MIN_{incongruent}$ = 1.60 to $MAX_{incongruent}$ = 7.75; $MIN_{non\text{-}interaction}$ = 1.60 to $MAX_{non\text{-}interaction}$ = 6.65), indicating that occasionally even incongruent interactions and non-interactions were perceived as meaningful by naïve perceivers. The more a social scene (regardless of dyad type) had been rated as socially meaningful by perceivers of an unrelated pilot study, the more posterior

insula activity was observed when this scene was seen by our fMRI participants. Intriguingly, the posterior insula has recently been linked to bodily self-awareness, body-related self-other discrimination, and the observation of movement/action imitation (Heydrich & Blanke, 2013; Kühn et al., 2010; Tsakiris, 2010; Watanabe et al., 2011). Thus, the obtained results might signal that easily decipherable social scenes invite observers' to link the perceived actions to their own bodily experience. In the present inquiry, however, the insula finding was returned by an exploratory analysis at rather lenient thresholding. While our observation of a bilateral activation makes it likely that the result is reliable, the finding must be considered with caution and requires future examination.

The observation also highlights a potential limitation of the current work. In line with previous fMRI investigations examining the processing of social interactions (cf. Centelles et al., 2011: n = 14; Hooker et al., 2010: n = 15; Iacobini et al., 2004: n = 13; Sinke et al., 2010: n = 14; Walter et al., 2004: n = 13), a relatively small number of participants was recruited for the present study (n = 12). This shortcoming may explain why some of our findings have failed to emerge at standard levels of significance/thresholding. Despite this limitation, the current study demonstrated that the perception of socially incongruent person interactions compared to congruent interactions and non-interactions increased the level and altered the patterns of neural activity in the core person perception network. It also showed that these socially ambiguous person scenes recruited processing resources beyond the core network dedicated toward the analysis of human movement, such as the pMTG. Finally, the findings suggest that linear increases in the scenes' semantic meaningfulness are accompanied by enhanced posterior insula activity. Taken together, these data begin to elucidate how observers encode and integrate visual information involving several people, extending recent efforts to study the comprehension of social scenes from a third-person perspective.

5. Acknowledgments

The authors thank Goedele Van Belle for processing all stimuli using the Matlab Image Processing Toolbox. This work was supported by an ARC grant 13/18-053, and a collective grant from the Louvain Academy Foundation. Bruno Rossion is supported by the Belgian National Fund for Scientific Research (FNRS).

6. References

Ames, D. L., Fiske, S. T., & Todorov, A. (2011). Impression formation: A focus on others' intents (pp. 419-433). In J. Decety & J. Cacioppo (Eds.), *The Oxford Handbook of Social Neuroscience*. Oxford University Press.

Andrews, T.J., Davies-Thompson, J., Kingstone, A., & Young, A.W. (2010). Internal and external features of the face are represented holistically in face-selective regions of visual cortex. *Journal of Neuroscience*, *30*, 3544–3552.

Baeck, A., Wagemans, J., & Op de Beeck, H. P. (2013). The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: The weighted average as a general rule. *NeuroImage*, *70*, 37-47.

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

Barton, J. J., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, *58*, 71-78.

Bentin S., Allison T., Puce A., Perez E., & McCarthy G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*, 551-565.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143-177.

Bolognini, N., Rossetti, A., Convento, S., & Vallar, G. (2013). Understanding others' feelings: The role of the right primary somatosensory cortex in encoding the affective valence of others' touch. *The Journal of Neuroscience*, *33*, 4201-4205.

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

Centelles, L., Assaiante, C., Nazarian, B., Anton, J.-L., & Schmitz, C. (2011). Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: A neuroimaging study. *PLoS ONE*, *6*, e15749.

Costanzo, M., & Archer, D. (1989). Interpreting the expressive behavior of others: The interpersonal perception task. *Journal of Nonverbal Behavior*, *13*, 225-244.

De Graef, P. (1992). Scene-context effects and models of real-world perception. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 243-259). New York: Springer.

Dumoulin, S. O., Bittar, R. G., Kabanai, N. J., Baker Jr. C. L., Le Goualher, G., Pike, G. B., & Evans, A. C. (2000). A new anatomical landmark for reliable identification of human area V5/MT: A quantitative analysis of sulcul patterning. *Cerebral Cortex*, *10*, 454-463.

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

Downing, P. E., & Peelen, M. V. (2011). The role of occipitotemporal body-selective regions in person perception. *Cognitive Neuroscience*, *2*, 186-226.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *The Journal of Neuroscience*, *30*, 16601-16608.

Engell A.D., & McCarthy G. (2013). fMRI activation by face and biological motion perception: Comparison of response maps and creation of probabilistic atlases. *NeuroImage*, *74*, 140-151.

Etzel, J. A., Gazzola, V., & Keysers, C. (2009). An introduction to anatomical ROI-based fMRI classification analysis. *Brain Research*, *1282*, 114-125.

Ferri, S., Kolster, H., Jastorff, J., & Orban, G. A. (2013). The overlap of the EBA and the MT/V5 cluster. *NeuroImage*, *66*, 412-425.

Formisano, E., De Martino, F., & Valente, G. (2008). Multivariate analysis of fMRI time series: Classification and regression of brain responses using machine learning. *Magnetic Resonance Imaging*, *26*, 921-934.

Freeman, J. B., Johnson, K. L., Adams, R. B. Jr.,& Ambady, N. (2012). The social-sensory interface: Category interactions in person perception. *Frontiers in Integrative Neuroscience*, doi: 10.3389/fnint.2012.00081

Friston, K. J., Fletcher, P., Josephs, O., Holmes, A. P., Rugg, M. D., & Turner, R. (1998). Event-related fMRI: Characterizing differential responses. *NeuroImage*, *7*, 30-40.

Frühholz, S., Fehr, T., & Herrmann, M. (2009). Interference control during recognition of facial affect enhances the processing of expression specific properties – an event-related fMRI study. *Brain Research*, *1269*, 143-157.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*, 350-363.

Gobbini, M. I., & Haxby, J. V. (2007). Neural systems for recognition of familiar faces. *Neuropsychologia*, *45*, 32-41.

Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of FIAC data with BrainVoyager QX. *Human Brain Mapping*, *27*, 392-401.

Green, C., & Hummel, J. E. (2006). Familiar interacting object pairs are perceptually grouped. *Journal of Experimental Psychology: Human Perception and Performance*, *32*, 1107-1119.

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.

Grosbras, M.-H., Beaton, S., & Eickhoff, S. B. (2012). Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Human Brain Mapping*, *33*, 431-454.

Grossman, E. D., Battelli, L., & Pascual-Leone, A. (2005). Repetitive TMS over STSp disrupts perception of biological motion. *Vision Research*, *45*, 2847-2853.

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, *113*, 1339-1350.

Hampson, M., Olson, I. R., Leung, H.-C., Skudlarski, P., & Gore, J. C. (2004). Changes in functional connectivity of human MT/V5 with visual motion input. *NeuroReport*, *15*, 1315-1319.

Heydrich, L., & Blanke, O. (2013). Distinct illusory own-body perceptions caused by damage to posterior insula and Extrastriate cortex. *Brain*, *136*, 790-803.

Hirai, M., & Kakigi, R. (2009). Differential orientation effect in the neural response to interacting biological motion of two agents. *BMC Neuroscience*, *10*, doi: 10.1186/1471-2202-10-39

Hollingworth, A., & Henderson, J. M. (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General*, *127*, 398-415.

Hooker, C. I., Verosky, S. C., Germine, L. T., Knight, R. T., & D'Esposito, M. (2010). Neural activity during social signal perception correlates with self-reported empathy. *Brain Research*, *1308*, 100-113.

Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., & Fiske, A. P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *NeuroImage*, *21*, 1167-1173.

Jonas, J., Descoins, M., Koessler, L., Colnat-Coulbois, S., Sauvée, M., Guye, M., et al. (2012). Focal electrical intracerebral stimulation of a face-sensitive area causes transient prosopagnosia. *Neuroscience*, 222, 281-288.

Kim, J. B., & Biederman, I. (2011). Where do objects become scenes? *Cerebral Cortex*, *21*, 1738-1746.

Kim, J. G., Biederman, I., & Juan, C.-H. (2011). The benefit of object interactions arises in the lateral occipital cortex independent of attentional modulation from the intraparietal sulcus: A transcranial magnetic stimulation study. *The Journal of Neuroscience*, *31*, 8320-8324.

Kolster, H., Peeters, R., & Orban, G. A. (2010). The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. *The Journal of Neuroscience*, *30*, 9801-9820.

Kosslyn, S. M. (1994). *Image and Brain*. Cambridge, MA: MIT Press.

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

Kühn, S., Müller, B. C. N., van Baaren, R. B., Wietzker, A., Dijksterhuis, A., & Brass, M. (2010). Why do I like you when you behave like me? Neural mechanisms mediating positive consequences of observing someone being imitated. *Social Neuroscience*, *5*, 384-392.

Kujala, M. V., Carlson, S., & Hari, R. (2011). Engagement of the amygdala in third-person view of face-to-face interaction. *Human Brain Mapping*, doi: 10.1002/hbm.21317.

Leising, D., & Borkenau, P. (2010). Person perception, disopositional inferences, and social judgment. In L. M. Horowitz & S. Strack (Eds.), *The Handbook of Interpersonal Psychology: Theory, Research, Assessment, and Therapeutic Interventions* (1st ed., pp. 157-170). New York: Wiley.

Macrae, C. N., & Quadflieg, S. (2010). Perceiving people. In Fiske, S., Gilbert, D. T., & Lindzey, G. (Eds.), *The Handbook of Social Psychology* (5th ed., pp. 428-463). New York: McGraw-Hill.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jlang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences USA, 92,* 8135–8139.

Meeren, H. K. M., van Heijnsbergen, C. C. R., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences USA, 102,* 16518-16523.

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: Expanding the limits of unconscious processing. *Psychological Science*, *22*, 764-770.

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*, *48*, 507-517.

Mudrik, L., Shalgi, S., Lamy, D., & Deouell, L. Y. (2014). Synchronous contextual irregularities affect early scene processing: replication and extension. *Neuropsychologia*, *56*, 447-458.

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, *9*, 1186-1192.

Neri, P. (2014). Semantic control of feature extraction from natural scenes. *Journal of Neuroscience*, *34*, 2374-2388.

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

Overwalle, van F. & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*, 564-584.

Palmer, S. E. (1975). Effects of contextual scenes on identification of objects. *Memory and Cognition*, *3*, 519-526.

Pavlova, M. A. (2012). Biological motion processing as a hallmark of social cognition. *Cerebral Cortex*, *22*, 981-995.

Peelen, M. V., Atkinson, A. P., Andersson, F., & Vuilleumier, P. (2007). Emotional modulation of body-selective visual areas. *Social Cognitive and Affective Neuroscience*, *2*, 274-283.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience*, *8*, 636–648.

Pitcher, D., Garrido, L., Walsh, V., & Duchaine, B. C. (2009). Triple dissociation of faces, bodies, and objects in extrastriate cortex. *Current Biology*, *19*, 319-324.

Pourtois, G., Peelen, M., Spinelli, L., Seeck, M., & Vuilleumier, P. (2007). Direct intracranial recording of body-selective responses in human extrastriate visual cortex. *Neuropsychologia*, *45*, 2621-2625.

Proverbio, A. M., Riva, F., Paganelli, L., Cappa, S. F., Canessa, N., Perani, D., & Zani, A. (2011). Neural coding of cooperative vs. affective human interactions. *PLoS One*, *6*, e22026.

Quadflieg, S., Flannigan, N., Waiter, G. D., Rossion, B., Wig, G. S., Turk, D. J., & Macrae, C. N. (2011). Stereotype-based modulation of person perception. *NeuroImage*, *57*, 549-557.

Quadflieg, S. & Rossion, B. (2011). When perception and attention collide: Neural processing in EBA and FBA. *Cognitive Neuroscience*, *2*, 209-210.

Riddoch, M. J., Humphreys, G. W., Edwards, T., & Willson, K. (2003). Seeing the action: neuropsychological evidence for action-based effects on object selection. *Nature Neuroscience*, *6*, 82-89.

Riddoch, M. J., Pippard, B., Booth, L., Rickell, J., Summers, J., Brownson, A., & Humphreys, G. W. (2011). Effects of action relations on the configural coding between objects. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 580-587.

Roberts, K. L., & Humphreys, G. W. (2010). Action relationships concatenate representations of separate objects in the ventral visual system. *NeuroImage*, *52*, 1541-1548.

Roberts, K. L., & Humphreys, G. W. (2011). Action relations facilitate the identification of briefly-presented objects. *Attention, Perception, and Psychophysics*, *73*, 597-612.

Romaiguère, P., Nazarian, B., Roth, M., Anton, J.-L., & Felician, O. (2014). Lateral occipitotemporal cortex and action representation. *Neuropsychologia*, http://dx.doi.org/10.1016/j.neuropsychologia.2014.01.006

Rossion, B., Caldara, R., Seghier, M., Schuller, A.M., Lazeyras, F., & Mayer, E., (2003). A network of occipito-temporal face-sensitive areas beside the right middle fusiform gyrus is necessary for normal face processing. *Brain*, *126*, 2381-2395.

Rossion, B., Hanseeuw, B., & Dricot, L. (2012). Defining face perception areas in the human brain: A large-scale factorial fMRI face localizer analysis. *Brain & Cognition*, *79*, 138-157.

Rossion B., & Jacques C. (2011). The N170: Understanding the time-course of face perception in the human brain. In S. Luck, E. Kappenman (eds). *The Oxford Handbook of ERP Components* (pp. 115-142). Oxford: University Press.

Rossion, B., Schiltz, C., & Crommelinck, M. (2003). The functionally defined 'face areas' are sensitive to long-term visual familiarity. *NeuroImage*, *19*, 877-883.

Sinke, C. B. A., Sorger, B., Goebel, R., & de Gelder, B. (2010). Tease or threat? Judging social interactions from bodily expressions. *NeuroImage*, *49*, 1717-1727.

Sorger, B., Goebel, R., Schiltz, C., & Rossion, B. (2007). Understanding the functional neuroanatomy of prosopagnosia. *NeuroImage*, *35*, 836-852.

Soria Bauser, D. & Suchan, B. (2013). Behavioral and electrophysiological correlates of intact and scrambled body perception. *Clinical Neurophysiology*, *124*, 686-696.

Spunt, R. P., & Lieberman, M. D. (2012). Dissociating modality-specific and supramodel neural systems for action understanding. *The Journal of Neuroscience*, *32*, 3575-3583.

Spunt, R. P., Satpute, A. B., & Lieberman (2011). Identifying the what, why, and how of an observed action: An fMRI study of mentalizing and mechanizing during action observation. *Journal of Cognitive Neuroscience*, *23*, 63-74.

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

Teufel, C., Fletcher, P. C., & Davis, G. (2011). Seeing other minds: Attributed mental states influence perception. *Trends in Cognitive Sciences*, *14*, 376-382.

Tsakiris, M. (2010). My body in the brain: A neurocognitive model of body-ownership. *Neuropsychologia*, *48*, 703-712.

Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. M. (2007). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature Neuroscience*, *10*, 30-31.

Wagner, D. D., Kelley, W. M., & Heatherton, T. F. (2011). Individual differences in the spontaneous recruitment of brain regions supporting mental state understanding when viewing natural social scenes. *Cerebral Cortex*, *21*, 2788-2796.

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., & Bara, B. G. (2004). Understanding intentions in social interaction: The role of the anterior paracingulate cortex. *Journal of Cognitive Neuroscience*, *16*, 1854-1863.

Watanabe, R., Watanabe, S., Kuruma, H., Murakami, Y., Seno, A., & Matsuda, T. (2011). Neural activation during imitation of movements presented from four different perspectives: A functional magnetic resonance imaging study. *Neuroscience Letters*, *503*, 100-104.

Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *NeuroImage*, *52*, 1559-1573.

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

Wulff, M., & Humphreys, G. (2013). Visual responses to action between unfamiliar object pairs modulate extinction. *Neuropsychologia*, *51*, 622-632.

Zaki, J. (2013). Cue integration: A common framework for social cognition and physical perception. *Perspectives on Psychological Science*, *8*, 296-312.

7. Tables

Table 1.

Average image properties, semantic ratings and response times across experimental conditions (standard deviations in brackets).

Person Dyad Type	Congruent Interactions	Incongruent Interactions	Non-Interactions
Image width in pixels	269 (49)	272 (42)	271 (52)
Inter-agent distance in pixels	-5 (28)	-2 (25)	-3 (19)
Center of mass distance in pixels	144 (34)	149 (30)	126 (24)
Semantic rating on a 9-point rating scale	7.87 (0.83)	3.67 (1.29)	2.96 (1.18)
Accuracy rates for the sex categorization task in %	96 (3)	96 (4)	97 (3)
Median response times for the sex categorization task in ms	955 (110)	963 (101)	970 (112)

Table 2.

Average peak Talairach coordinates of person perception regions as determined based on the localizer task (coordinates in Talairach space).

Region	Hemisphere	n	x	y	z
Fusiform Face Area (FFA)	R	11	37	-45	-18
	L	11	-41	-47	-19
Fusiform Body Area (FBA)	R	12	38	-38	-19
	L	11	-40	-41	-18
Extrastriate Body Area (EBA)	R	12	45	-71	-2
	L	12	-45	-71	-1
Occipital Face Area (OFA)	R	10	29	-84	-11
	L	9	-34	-79	-12
Posterior Superior Temporal Sulcus (pSTS)	R	11	48	-46	6
	L	7	-50	-46	7

Table 3.

Peak voxel and number of voxels for brain regions as identified by whole brain analyses at a voxelwise threshold of $p<.001$ and a cluster-size threshold of $p<.05$ (coordinates in Talairach space).

Region	Hemisphere	Voxels	x	y	z	*t*-value
Congruent Interactions > Incongruent Interactions						
no suprathreshold activation						
Incongruent Interactions > Congruent Interactions						
Posterior Middle Temporal Gyrus (BA 37)	R	33	38	-65	6	*6.44*
	L	9	-40	-62	6	*6.32*
Congruent Interactions > Non-Interactions						
no suprathreshold activation						
Non-Interactions > Congruent Interactions						
no suprathreshold activation						
Incongruent Interactions > Non-Interactions						
Posterior Middle Temporal Gyrus (BA 37)	R	71	50	-56	3	*7.22*
	L	41	-43	-62	3	*10.34*
Parahippocampal Gyrus (BA 19)	R	15	17	-44	-3	*6.37*
Posterior Superior Temporal Gyrus (BA 21)	L	18	-52	-44	9	*8.31*
Non-Interactions > Incongruent Interactions						
no suprathreshold activation						

8. Figure Captions

Figure 1.

Example images as used in the person dyad task across experimental conditions. Participants viewed 40 person dyads per condition.

Figure 2.

Visual properties as determined for all person dyads.

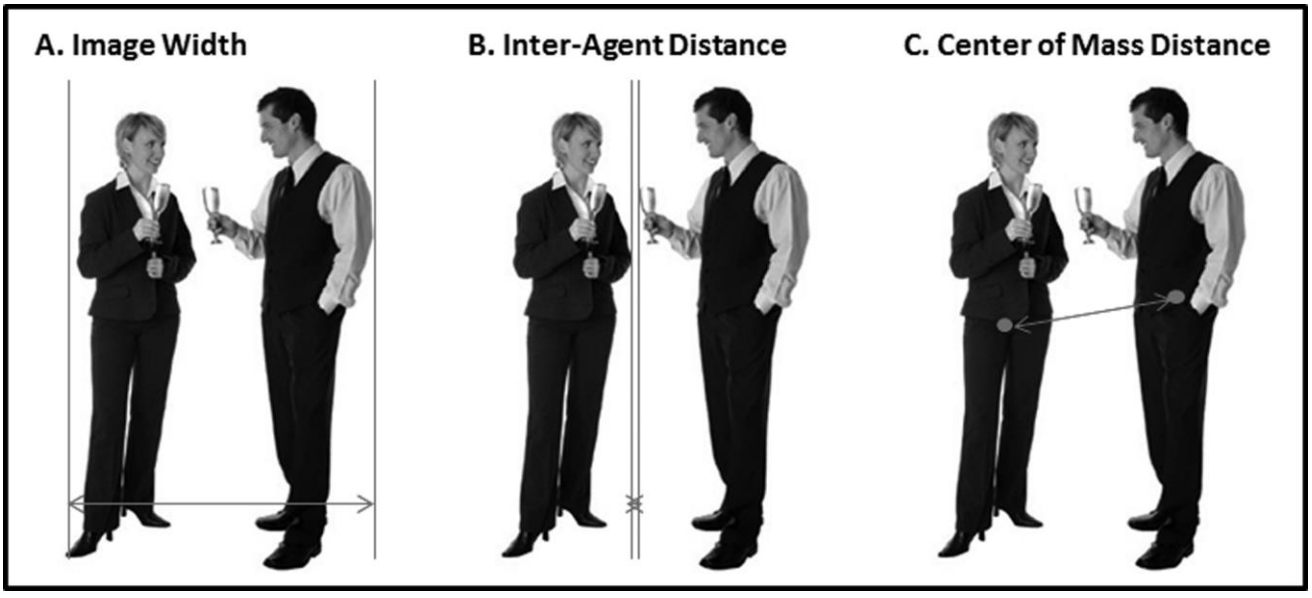

Figure 3.

Panel A displays the average peak Talairach coordinates of all face- and body-selective regions of interest (ROI) based on the localizer task. Panel B shows the localizer-based mean parameter estimates across experimental conditions for each ROI. Error bars represent the standard error of the mean. A difference in activity from the congruent person interaction condition is signaled by * ($p < .05$) or ° ($p = .08$).

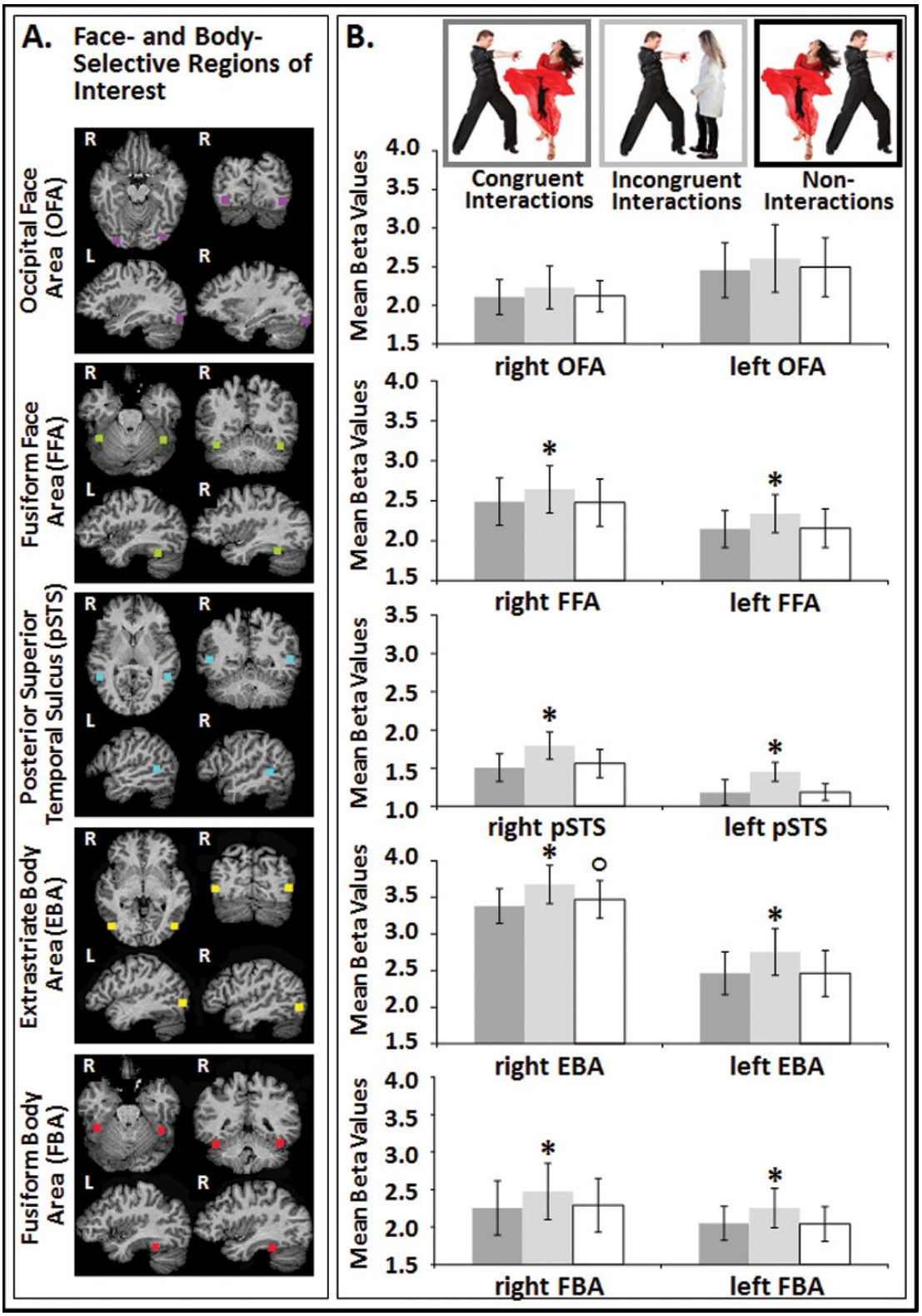

Figure 4.

Event-related averaging of BOLD signal change for illustrative purpose across image type in the right extrastriate body area (A) and the left extrastriate body area (B). BOLD signal change was extracted from a 9 mm cube around the regions' average peak coordinates based on the localizer task. Error bars represent the standard error of the mean.

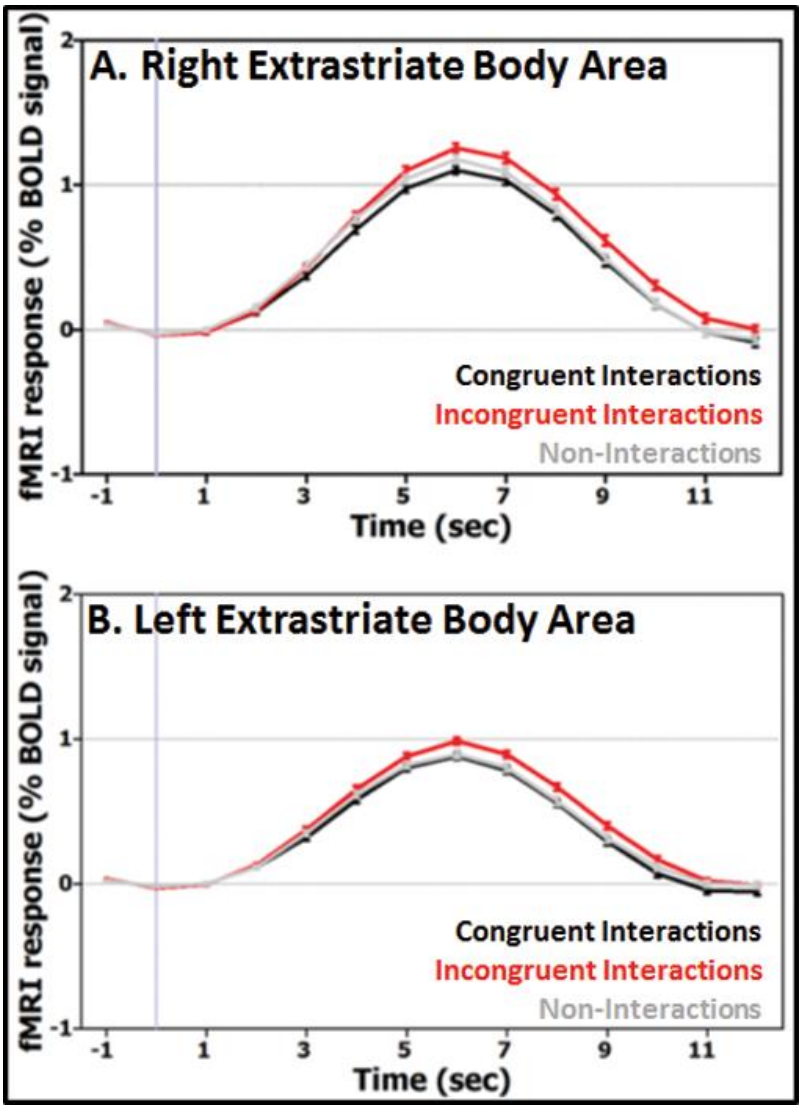

Figure 5.

Support vector machine based pattern classification results: Mean classification accuracies in the person perception network when discriminating between different person dyads. Error bars represent the standard error of the mean (*p < .05 one-sided).

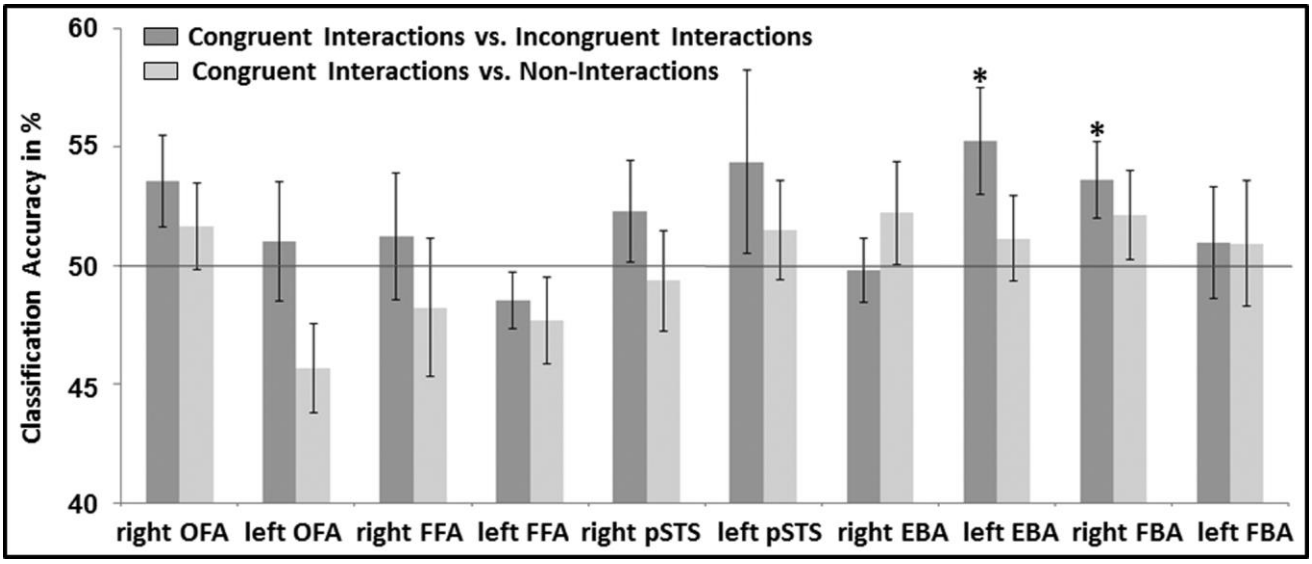

Figure 6.

Exploratory whole brain analysis: Enhanced activity during incongruent interactions compared to congruent interactions (voxelwise p < .001; cluster-based p < .05).

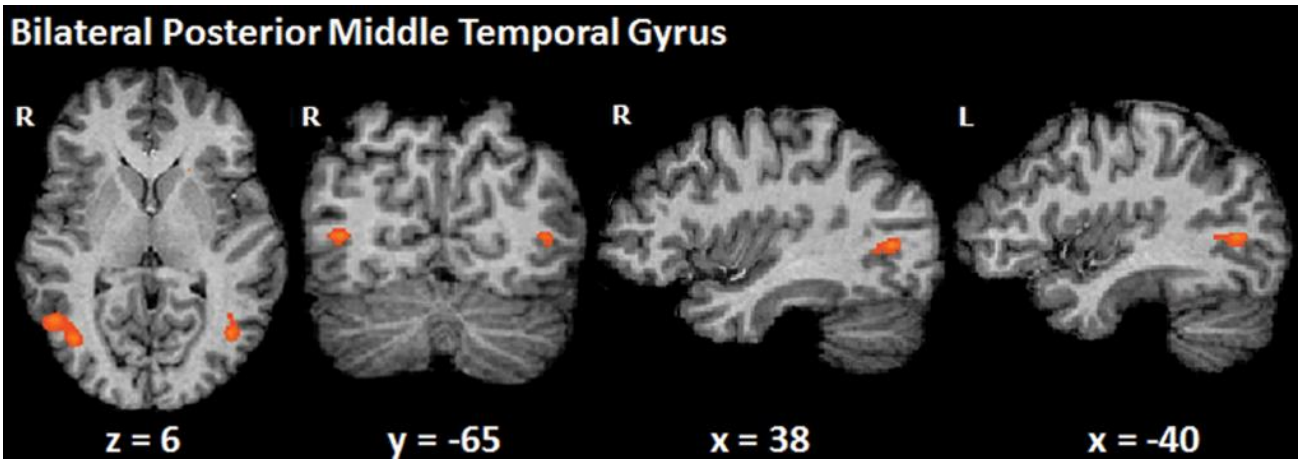

Figure 7.

Exploratory whole brain analysis: Enhanced activity during incongruent interactions compared to non-interactions (voxelwise $p < .001$; cluster-based $p < .05$).

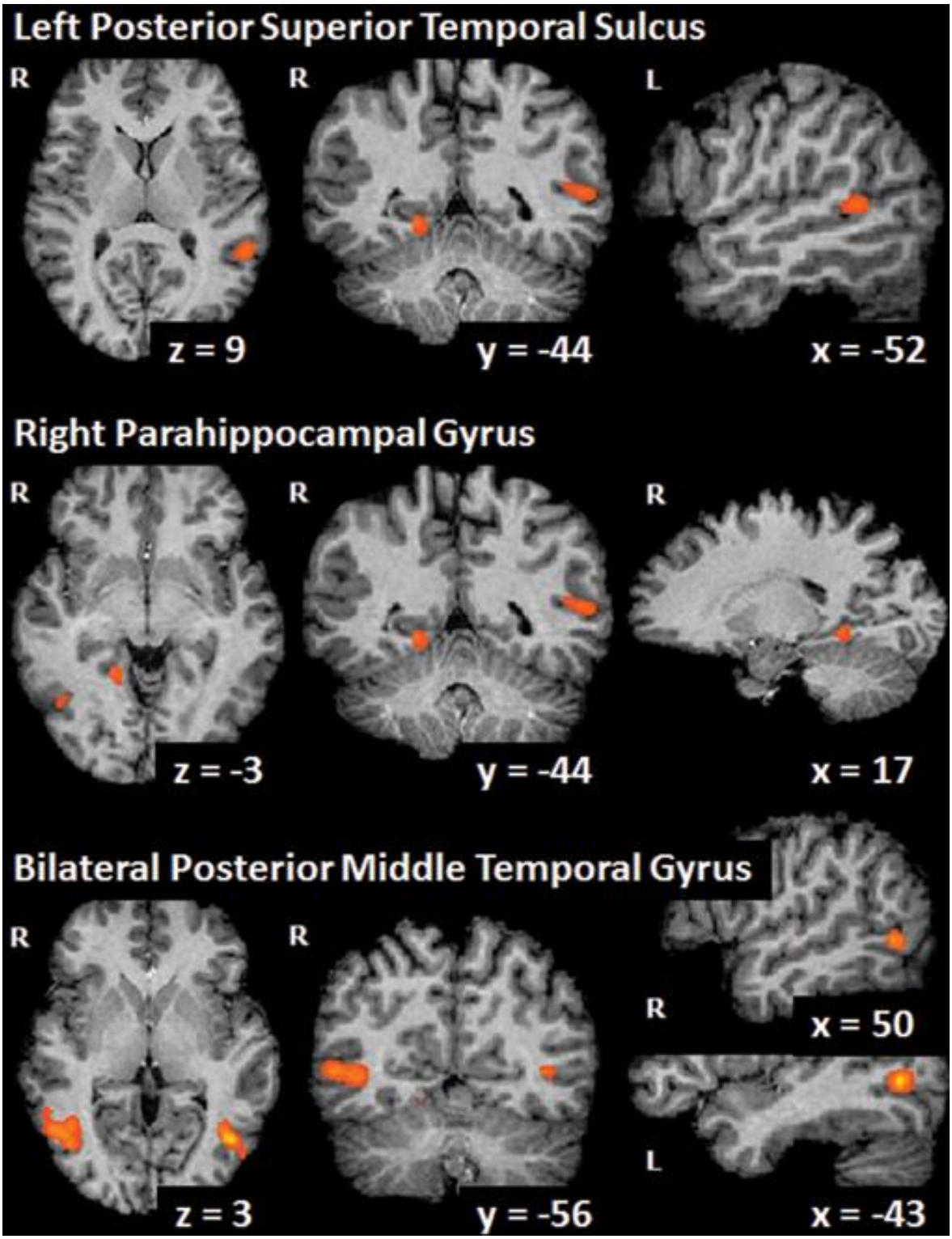

Figure 8.

Posterior insula activity was found to increase parametrically based on a scene's perceived meaningfulness (as rated in a pilot study), regardless of image type (voxelwise $p < .05$; minimum cluster size: 20 voxels).

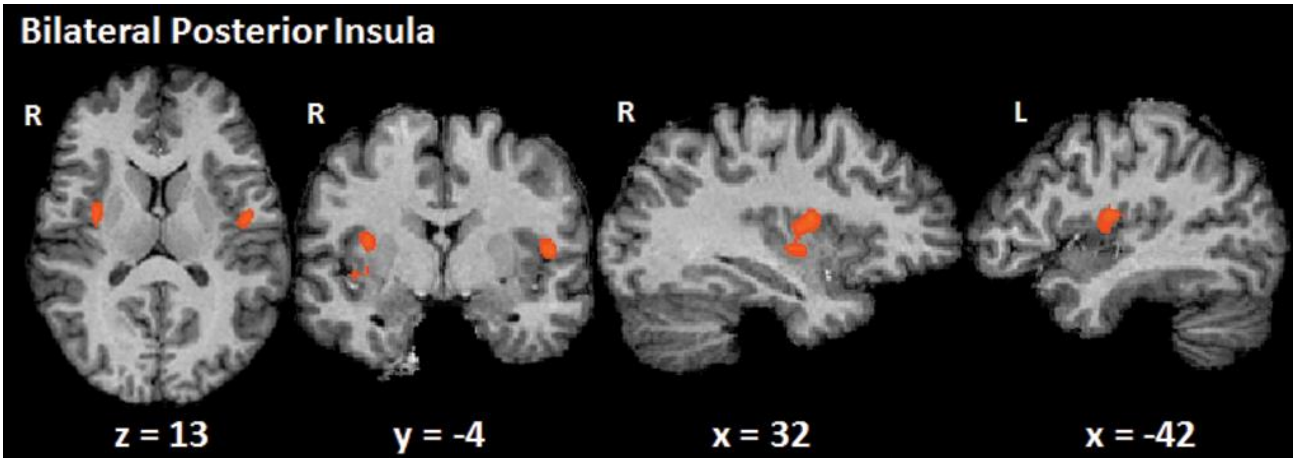

Figure 9.

Posterior middle temporal gyrus activity as detected for incongruent interactions during whole brain analyses displayed in relation to the average peak Talairach coordinates for the face-selective posterior superior temporal sulcus and the body-selective extrastriate body area as derived from the localizer task.