Interaction vs. observation: distinctive modes of social cognition in human brain and behavior? A combined fMRI and eye-tracking study

Kristian Tylén[1,2]*, Micah Allen[1], Bjørk K. Hunter[1] and Andreas Roepstorff[1,3]

[1] The Interacting Minds Group, Center for Functionally Integrative Neuroscience, Aarhus University, Aarhus, Denmark
[2] Department for Aesthetics and Communication, Center for Semiotics, Aarhus University, Aarhus, Denmark
[3] Department for Culture and Society, Aarhus University, Aarhus, Denmark

Human cognition has usually been approached on the level of individual minds and brains, but social interaction is a challenging case. Is it best thought of as a self-contained individual cognitive process aiming at an "understanding of the other," or should it rather be approached as an collective, *inter*-personal process where individual cognitive components interact on a moment-to-moment basis to form coupled dynamics? In a combined fMRI and eye-tracking study we directly contrasted these models of social cognition. We found that the perception of situations affording social contingent responsiveness (e.g., someone offering or showing you an object) elicited activations in regions of the right posterior temporal sulcus and yielded greater pupil dilation corresponding to a model of coupled dynamics (joint action). In contrast, the social-cognitive perception of someone "privately" manipulating an object elicited activation in medial prefrontal cortex, the right inferior frontal gyrus and right inferior parietal lobus, regions normally associated with Theory of Mind and with the mirror neuron system. Our findings support a distinction in social cognition between *social observation* and *social interaction*, and demonstrate that simple ostensive cues may shift participants' experience, behavior, and brain activity between these modes. The identification of a distinct, interactive mode has implications for research on social cognition, both in everyday life and in clinical conditions.

Keywords: social interaction, brain imaging, theory of mind, mirror neuron system, joint action, coupled dynamics

INTRODUCTION

Recent advances in evolutionary anthropology and experimental psychology suggest that one of the keys to the unique evolutionary trajectory of the human species can be found in our advanced capacities for reciprocal social interaction (Donald, 1991, 2001; Tomasello, 1999, 2008; Tomasello et al., 2005; Csibra and Gergely, 2009, 2011). This inevitably leads to fundamental questions concerning the neurocognitive foundations of such social capacities. During the last couple of decades, an increasing number of studies have addressed the human brain mechanisms responsible for our ability to make sense of social phenomena. A number of brain networks—often referred to as "the social brain"—are found to be associated with various aspects of social cognition. For instance, the medial prefrontal and temporo-parietal cortices consistently activate in tasks involving Theory of Mind/mentalizing (e.g., Castelli et al., 2000; Gallagher et al., 2000; German et al., 2004; Walter et al., 2004), while premotor areas and inferior parietal cortices seem to be involved in mental mirroring of others' motor actions (e.g., Arbib et al., 2000; Rizzolatti et al., 2001; Stamenov and Gallese, 2002; Heiser et al., 2003; Kaplan and Iacoboni, 2006; Ocampo et al., 2011). While these studies make up an intriguing body of research on the neurobiological foundations of what we might term "social

observation" (where no contingent response is afforded), it is disputable to which degree the findings can be generalized to account for processes underlying social interaction. We argue that the distinction between 3rd person social observation and 2nd person social interaction is an important conceptual and empirical distinction that has been somewhat neglected in the neurocognitive field (Roepstorff, 2001; Tylén and Allen, 2009; Schilbach, 2010; Hasson et al., 2012).

Two prevalent conceptual frameworks have oriented the majority of studies in social neurocognition, Theory of Mind/mentalizing (hence ToM) and Simulation Theory (which is often closely associated with the Mirror-System hypothesis—hence MNS). In both cases, the overall goal is to unravel and map the neurobiological mechanisms responsible for the ability to attribute, understand, and empathize mental states of others. Although we recognize that the underlying assumptions and proposed mechanisms of ToM and MNS are indeed very different, they take the same point of departure: the individual mind. ToM and MNS models are thus mainly preoccupied with the way individuals make sense of each other from an observational point of view (Gallagher and Hutto, 2008). The fundamental processes of social cognition are described in terms of mental inference (ToM) or embodied simulation

(MNS) facilitating a "self-contained understanding" of other persons' actions. This "understanding" in turn supposedly makes it possible to choose appropriate responses, and for instance engage in *inter*actions (Frith and Frith, 2001, 2006a; Schulte-Ruther et al., 2007). In other words, individual observational processes are—more or less explicitly—given primacy as constituting the core of social cognition, while other social cognitive phenomena (e.g., social interaction) are derived from or emergent upon these fundamental processes. Hence in these frameworks, mechanisms in social interaction are extrapolated from studies of social observation and thus explained on the level of individual minds and brains. An interaction thus involves two or more individuals that recursively observe, represent and react to each other's actions based on their individual internal representational models. This has important implications for the theoretical and experimental foci of the two paradigms. Here, we will make the case that social observation and social interaction are in fact very different phenomena. While an individualistic and observational stance to social cognition may be appropriate for the study of a range of phenomena including the detection of deception, pretense, emotional expressions, etc., it is much less clear to which extent it can tackle questions related to the inherently collective and reciprocal dynamics of *social interaction*.

A growing literature within philosophy of mind and cognitive science is advancing the view that in order to adequately account for cognitive processes involved in social interaction, we need to widen the perspective beyond individual minds and brains. These approaches are largely informed by recent discussions under the headline of "extended," "enacted," and "distributed" cognition often relying on insights from complex systems theory. The main argument is that when two persons engage in joint activities their bodies, actions, and individual cognitive processes become coupled in dynamic ways. Hence rather than working in parallel as self-enclosed autonomous entities, persons involved in direct interaction get intermingled in complementary ways that enable emergent synergies (De Jaegher et al., 2010; Hasson et al., 2012). In this understanding, a sequence of joint action is better conceived of as a whole (singular, continuous) time series, rather than a synchronization of two independent processes (Black et al., 2007; Konvalinka et al., 2010; Riley et al., 2011). As an example, consider a dialogue. In conversation, interlocutors take turns in a complementary way making up the overall object of the dialog. One interlocutor's speech turn—for example, a question—is only completed by the responding speech turn of the other (cf. the concept of "adjacency pairs," Goodwin and Heritage, 1990). If we isolate an individual component, say all the speech turns of one interlocutor, we are left with a partial object that does not make any sense on its own. In other words, the dialog as a phenomenon cannot be reduced to any of the partial individual components, but can only be appropriately assessed at the collective, inter-personal level (Kello et al., 2010). We argue that turn-taking-like responsiveness is a fundamental characteristic of social interaction across a broad range of contexts from diaper-changing to tango-dancing. As a distinct phenomenon, it should not be confused with automatic mirroring or simulation. Where mirroring is assumed to be an internal representational

event, turn-taking responsiveness is rather characterized by its complementary contribution to the intersubjective scene. The ostensive act of one person (e.g., a greeting nod or an eyebrow flash) afford for the complementary response from the recipient (e.g., an "answering" nodding gesture). An offering hand gesture affords a receptive one (Newman-Norlund et al., 2007; Ferri et al., 2011; Sartori et al., 2012).

Which predictions follow from the conceptual approach to social interaction sketched above? If key dynamics of social interaction can only be found at a collective, level, how can we then study its neurocognitive underpinnings? One suggestion is that simultaneous recording from multiple agents is necessary to make claims about the dynamics of mutually coupled cognitive systems. While this may be a useful approach (see Konvalinka and Roepstorff, 2012) we here argue that recognizing the coordinative nature of social interaction allows specific predictions, even on the level of individual brains recorded in isolation. If the brain in joint action becomes a component-node in a larger interactive array, we can reframe the basic question as: What does it take for a brain to successfully engage in reciprocal coupling processes with other responsive components?

For a component to successfully work in tight concert with other external components it has to continuously integrate, adapt and respond to incoming stimuli at a multiplicity of temporal levels and modalities (Konvalinka et al., 2010). This suggests that rapid adaptation and coordination are crucial factors in real-time interaction. These properties are fundamentally different from those involved in "social observation." Where an observational understanding of a social phenomenon may be internally realized in terms of simulation or inference, a socially interactive practice calls for moment-to-moment reciprocity with one or more co-operative partners in the "external" social environment.

These fundamental differences between social observation and social interaction predict the involvement of distinct anatomical structures in the two processes. Regions of the temporal lopes (in particular STS, pSTS) have been consistently associated with the fine-grained continuous temporal integration of dynamic stimuli (Hasson et al., 2008; Stephens et al., 2010; Lerner et al., 2011). These structures, particularly in the right hemisphere, have indeed been found in a number of recent studies addressing the neurocognitive underpinnings of joint action and joint attention. In a fMRI study conducted by Newman-Norlund et al. (2008), activity was enhanced in right pSTS when participants performed a joint task with another person in the control room affording complementary (non-isomorphic) actions. In a study by Redcay et al. (2010), participants underwent fMRI scanning while solving a cooperative joint attention task with another person through a bidirectional video link. Again the main findings related to right pSTS/TPJ. Likewise, a fMRI study applying a dual player virtual communication game (Noordzij et al., 2009) also found the right pSTS to be modulated by social interaction in contrast to solo conditions, and finally a study by Iacoboni et al. (2001) found that the right pSTS was more active when participants imitated displayed hand movement than when they produced them from memory. We notice that the rpSTS has both been argued to belong to the ToM network (Frith and Frith, 2006b) and to the MNS

(Van Overwalle and Baetens, 2009). However, while pSTS may co-activate with both of these networks in task specific ways, no consistent pattern has so far been established, and no stable connectivity has been established between the pSTS and regions associated with ToM and MNS (Ethofer et al., 2011). We thus argue that pSTS is not a constitutive part of the ToM or the MNS network.

The findings cited above indicate the right pSTS as an area particularly sensitive to the continuous fine-grained temporal navigation and integration of stimuli necessary for immediate contingent responsiveness in social interaction. Thereby, it seems a good anatomical candidate for our hypothesized distinctive mode of social engagement. We thus predict that social interaction will recruit the pSTS, while social observation primarily will rely on networks related to ToM and MNS. How can we test such hypotheses?

This requires an experimental paradigm that directly compares interactive and observational social cognition. Here, we report an fMRI experiment that contrasts video stimuli, which either evoked an observational or interactive responsive attitude in the participant toward an actor performing simple object-related gestures. This contrast was established by modulating the ostensive character of the performed action. In the interactive conditions, the actor made interaction initiation cues (eye contact, eyebrow flashes and nods) before performing a *placing-object-for* or *showing-object-to* action (Clark, 2005). In contrast, in the non-interactive "private" condition the same actions were performed without ostensive cues. Moreover, the directionality of the action was modulated so that in some conditions the actor would face the participant while in others she/he was presented from a slightly averted perspective as if facing someone outside the perspective of the camera.

The theoretical analysis above generated specific anatomical hypotheses relating to three clusters of brain areas associated with ToM (in particular MPFC and TPJ), the MNS (pre-Motor, IPL), and Joint Action/Joint Attention (pSTS). We thus restricted the study to target these particular areas using a ROI approach (see section "Materials and Methods" for details). We predicted that ostensive object-gestures would engage contingent responsiveness in the participants, and that this would elicit differential activation in pSTS. In contrast, observing "private" object manipulations would evoke an observational attitude in the participant and hence elicit activations in ToM and MNS regions. Beside, we hypothesized that activity in these areas would be modulated by the directionality of action, as either participant-directed or other-directed.

Since the pSTS has also been associated with perspective taking, eye-gaze and saccading behaviors (Allison et al., 2000), we included simultaneous in-scanner-eye-tracking to control for effects caused by participants' simple eye gaze-behaviors. Furthermore, we used pupillometrics (pupil size measurements) to assess pupil dilation and constrictions in response to the experimental conditions (Kampe et al., 2003; Granholm and Steinhauer, 2004). We predicted that interactively engaging stimuli would be more emotionally arousing resulting in greater pupil dilation than stimuli affording a more observational attitude in the participant.

MATERIALS AND METHODS
SUBJECTS
Twenty-two healthy, right-handed adult volunteers (12 females/10 males, mean age 25 ± 4.8 STD) who had all given their written consent in correspondence with the requirements of the local ethical committee participated in the experiment. The participants were mainly recruited among students at Aarhus University, and were naïve with respect to the purpose of the study.

STIMULI AND EXPERIMENTAL DESIGN
Stimuli consisted of 32 video clips of 5 s duration, showing an actor sitting at a table in front of an object (see **Figure 1**). The videos differed on three variables: (1) actor gender (m/f), (2) object (cup or fruit) and—for the action condition—(3) action type (*placing-object-for* or *showing-object-to*) (cf. Clark, 2005). The experiment was divided into two sessions of 64 trials (i.e., all videos were shown four times).

We used two-by-two-by-two factorial design (making up in all eight conditions) with the main factors Ostention (ostensive/non-ostensive), Direction (direct/diverted perspective), and Action (action/no action). In ostensive conditions, the actor would look up and make an interaction-initiating cue by establishing eye contact (either to the participant or to an inferred other outside the scope of the camera) and making an eyebrow lift and a nod before performing one of the two object directed gestures. In non-ostensive conditions the action was performed "privately" without any addressing cues or eye contact. In direct conditions, the ostensive cues and gestures were performed directly to the
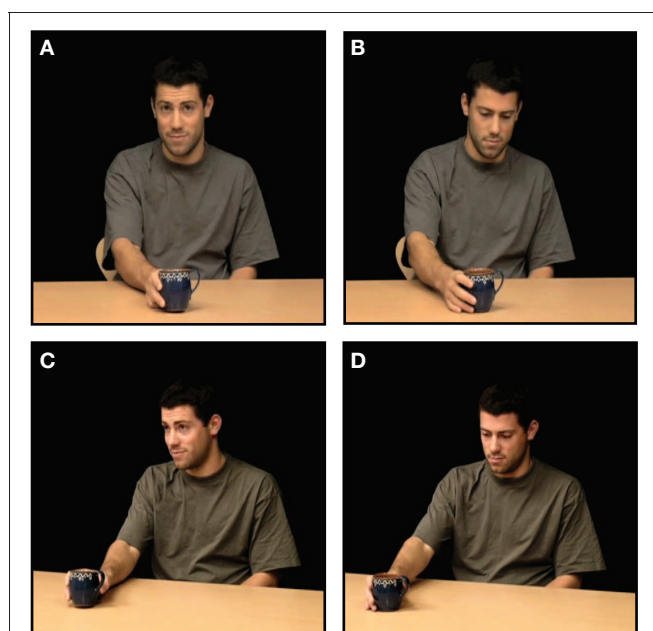

FIGURE 1 | Example of stimuli. In 5 s video clips, an actor performed simple object gestures ("*placing an object for*" or "*showing an object to*" someone) in four conditions: **(A)** ostensive and direct, **(B)** non-ostensive and direct, **(C)** ostensive and averted, **(D)** non-ostensive and averted. Besides, all four conditions were replicated without the object gesture.

participant (i.e., the camera), while in the diverted condition the actor was oriented at approx. 20° of the camera in the direction of an inferred other (see **Figure 1**). In the no action conditions, the four conditions above were replicated, but without the object gesture.

The stimulus videos were presented in blocks of two clips from the same condition, and the order of blocks was randomized between participants. After the two clips participants were asked one of two yes/no questions: "was it the same person" or "was it the same object?" This probed whether the same or different actors or objects had appeared in the two movie clips. The questions were randomized so that the participant could not anticipate if she/he would be asked about the actor or the object. In order to solve the task, the participant thus had to pay close attention to both actors and objects during the stimulus presentation. Participants would respond by pressing one of two buttons with their right hand index and middle finger. The left/right position of the affirmative response was randomized across trials.

SCANNING PARAMETERS
We used a 3T General Electrics MR system (Waukesha, WI, USA) with an eight channel head coil to acquire the T^2-weighted gradient, echo-planar images (EPI) with Blood Oxygenation Level-Dependent (BOLD) contrast using the following parameters: echo time (TE): 30 ms, repetition time (TR): 3000 ms, and a flip angle of 90°. Whole-brain images were obtained over 39 sequential, interleaved 3.5 mm axial slices with a 128 × 128 pixel resolution matrix and a field of view of 240 × 240 mm.

EYE-TRACKING PARAMETERS
Participants' eye movements and pupil size were recorded simultaneously with the MR acquisition using a SMI/Avotec IViewX eye-tracking system in a Silent Vision 7021 MR-insert binocular visual system. Data were recorded from the right eye with a sample frequency of 50 Hz. Prior to each of the two scanning sessions, the eye-tracker was calibrated using the IViewX nine-point automated calibration procedure, which was repeated until the calibration was satisfactory. The eye-tracker was linked and synchronized with the MR stimulus computer and continuously recorded time stamps for the initiation of stimulus videos.

ADDITIONAL BEHAVIORAL TESTING
After the fMRI scanning, participants went through an extensive debriefing where they evaluated their experience on various parameters. Moreover, participants watched the stimulus videos again on a computer screen and rated how "socially engaging" they found them on a 5 point scale where 1 = not engaging, and 5 = very socially engaging.

ANALYSIS
BEHAVIORAL ANALYSES
Task performance (response accuracy) from the in-scanner task was summarized and averaged for each participant and tested against chance performance using paired t-tests. Likewise, the post-scanning ratings of the socially engaging nature of the stimuli were summarized and averaged for each participant and

each condition, and condition-related differences were tested using a within-subject, repeated measures, three-way analysis of variance (Howell, 2002), with the factors *Ostension* (+/−), *Direction* (direct/diverted), and *Action* (+/−). The analysis was thresholded at $p < 0.05$. Due to technical problems, we failed to obtain rating data from two participants, so only data from the remaining twenty participants entered this analysis. All statistical tests were performed in MATLAB 2011b.

EYE-TRACKING ANALYSIS
Since the experiment was mainly optimized for fMRI acquisition, only full eye-tracking data sets from eleven participants entered the analysis. The remaining data were lost or corrupted due to technical problems and calibration difficulties. Task related eye-tracking data (x/y coordinates and x/y pupil diameter in pixels at a 50 Hz sampling for each 5 s stimulus video) were preprocessed by removing eye blinks and outliers (deviating in distance with more that $3 \times SD$ of the mean). The data were high-pass filtered at a 100 s cut off to counter calibration drift. Saccade velocity was then calculated for each participant based on point-to-point Euclidean distance (Salvucci and Goldberg, 2000). Similarly, pupil diameter was calculated as an average of the pupil x and y diameter direction (although these were strongly correlated we used this procedure to get a more stable index of pupil size). Velocity and pupil size data were averaged for each stimulus trial before entering further analysis.

The preprocessed data were used in two ways: first, to test for condition related differences in participants' gaze behaviors (hence the "stand-alone eye-tracking analysis"). For this purpose eye-movement velocity and pupil size data from each condition entered a within-subject, repeated measures, three-way analysis of variance (Howell, 2002), with the factors *Ostension* (+/−), *Direction* (direct/diverted), and *Action* (+/−). The analysis was thresholded at $p < 0.05$. Second, for each participant, velocity data were averaged for each stimulus event to be included as a first-level parametric modulation in the fMRI analysis (hence "the combined eye-tracking/fMRI analysis").

fMRI ANALYSIS
All fMRI data analysis was conducted using SPM8 (Statistical Parametric Mapping, Wellcome Department of Imaging Neuroscience, London) implemented in MATLAB 2011b (Mathworks Inc. Sherborn, MA) using default settings unless otherwise specified. Images were spatially realigned, normalized to the MNI template and smoothed with an isotropic 8 mm FWHM Gaussian kernel.

Statistical analysis was conducted following a two-level general linear model approach (Penny and Holmes, 2007). On the first-level, task related BOLD responses were modeled for each subject by convolving condition onsets and durations with the standard hemodynamic response function and contrasting factorial main and interaction effects. Two independent first-level analyses were carried out. The first, which was carried out for all participants, included a regressor (parametric modulation) for each of the variables of the stimulus videos (gender, object and action type) as well as the six standard SPM8 motion parameters. The second first-level analysis was only carried out on data

from the 11 participants from who we recorded a full eye-tracking data set. In addition to the stimulus and motion regressors used in the analysis above, this analysis included a parametric modulation regressing out relative differences in participants' eye-movements (saccade activity). For both first-level analyses, images were high-pass filtered at a 128 s cut off.

Second level RFX analyses

Two group RFX analyses were conducted—one for each of the first-level analyses—using a Three-Way repeated measures whole brain ANOVA (corrected for non-sphericity) in SPM8. The directionality of effects was explored using one-sample t-tests. In both cases, individual subject effects were modeled using the covariate function to adjust the statistics and degrees of freedom during inference. We did not assume independence or equal variance (Christensen and Wallentin, 2011). For both analyses, the significance threshold was set to $p < 0.05$, FWE corrected for multiple comparisons. Functional images were overlaid with the standard SPM8 single subject high resolution T1 image.

To constrain the analyses to specific, predefined anatomical sites (see section "Introduction" above) for which we had hypotheses, we used a *region of interest* (ROI) approach. The analyses were carried out as small volume corrections by masking particular brain structures consistently found in neurocognitive studies on social cognition. Masks were generated in the Wake Forest University PickAtlas extension for SPM (Tzourio-Mazoyer et al., 2002; Maldjian et al., 2003, 2004) as 10 mm spheres centered in target peak voxels. These were reported as main findings in recent studies on closely related topics all using stimuli very compatible to ours (dynamic video stimuli displaying an actor performing different types of actions). We recognize that a number of other areas have previously been reported as associated to ToM and MNS, but we chose to restrict our self to a few canonical areas of the right hemisphere which are among the most consistently reported in the literature and that have been associated with tasks resembling ours. The following masks were employed: to test for neural activity related to ToM/mentalizing we masked regions in mPFC [MNI (13, 37, 2)] and rTPJ [MNI (49, −63, 29)] based on coordinates from Wurm et al. (2011). To test for neural activity related to the mirror neuron system we masked the right IFG [MNI (52, 32, 24)] and IPL [MNI (46, −48, 44)] based on coordinates from Ocampo et al. (2011). Finally, to test for neural activity related to joint action/attention, we masked right pSTS [MNI (48, −40 6,)], based on coordinates from Redcay et al. (2010).

RESULTS

BEHAVIORAL RESULTS

Participants were generally able to solve the in-scanner recollection task ("same actor/same object?"), and scored an average response accuracy of 72% ($SD = 5.33$). One participant did not perform significantly above chance due to a high number of missed responses. However, since the participant did not self-report concentration/sleepiness problems etc., and the exclusion of the data did not affect the analysis substantially, the participant was not excluded from the fMRI analysis.

The post-scanning rating of the socially engaging nature of the stimulus videos showed a number of significant between-condition differences and interactions. The main effect of ostension yielded an F ratio of $F_{(1, 19)} = 203$, $p < 0.000$, indicating that the ostensive behavior of the actor made the scenes overall more socially engaging ($M = 3.05$, $SD = 0.77$) than non-ostensive scenes ($M = 1.28$, $SD = 0.35$). The main effect of direction was also significant, $F_{(1, 19)} = 10.5$, $p < 0.01$, likewise indicating that direct perspective was found more socially engaging ($M = 2.69$, $SD = 0.44$) than diverted perspective ($M = 1.63$, $SD = 0.68$). Finally, the main effect of action was also found significant: $F_{(1, 19)} = 23.4$, $p < 0.000$, indicating that the more dynamic scenes including object manipulations (placing for/showing) were more socially engaging ($M = 2.32$, $SD = 0.58$), than non-dynamic scenes ($M = 1.99$, $SD = 0.53$). Besides, all interactions were significant. Ostension interacted thus significantly with direction: $F_{(1, 19)} = 54.2$, $p < 0.000$, indicating that scenes were found more socially engaging when ostensive cues were performed directly to the participant. Ostension interacted significantly with action: $F_{(1, 19)} = 56.6$, $p < 0.000$, suggesting that the scenes were found more socially engaging when ostension and action accompanied each other (to form communicative gestures). Direction and action also interacted significantly, although to a somewhat lesser extent: $F_{(1, 19)} = 7.2$, $p < 0.05$, and, finally, the factors showed a significant three-way interaction: $F_{(1, 19)} = 10.1$, $p < 0.005$ (see **Figure 2**).

STAND-ALONE EYE-TRACKING RESULTS

The analysis of participants' eye-movements (saccade velocity) related to the conditions showed some significant effects. The main effect of direction yielded an F ratio of $F_{(1, 10)} = 8$,
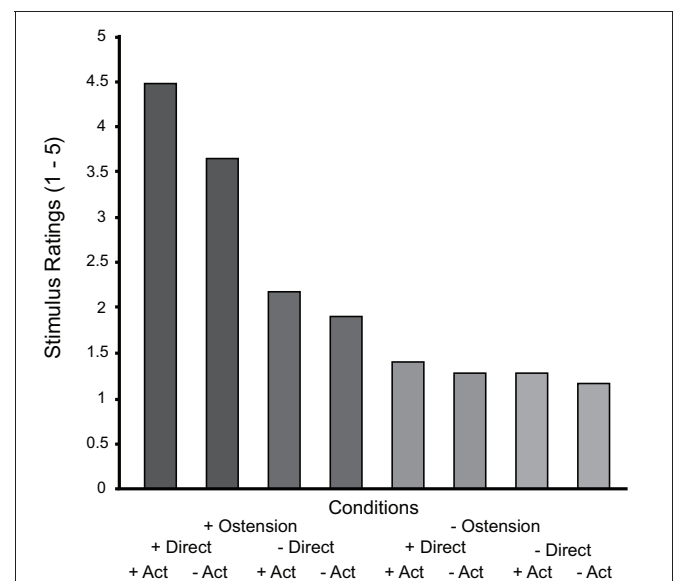

FIGURE 2 | Results from the post-scan stimulus ratings sorted by condition. Notice that the graph summarizes the "within-subject" results across all subjects and therefore we have not included error bars (between subject variance would not reflect the actual analysis).

$p < 0.05$, suggesting that participants generally displayed more saccading behaviors in the diverted conditions ($M = 0.31$, $SD = 0.09$) than in the direct ($M = 0.29$, $SD = 0.08$). The main effect of action was also found significant: $F_{(1, 10)} = 19.2$, $p < 0.005$, indicating that participants made more saccades in the action ($M = 0.32$, $SD = 0.09$) than the no-action conditions ($M = 0.29$, $SD = 0.08$). The main effect of ostension and all interaction effects were non-significant.

The analysis of pupil diameter changes also showed significant effects. The main effect of ostension yielded an F ratio of $F_{(1, 10)} = 5.2$, $p < 0.05$, indicating pupil dilation (measured in pixels) in response to ostensive cues ($M = 78$, $SD = 10.7$) relative to non-ostensive scenes ($M = 77.7$, $SD = 10.6$). Likewise, the main effect of direction was found significant: $F_{(1, 10)} = 18.4$, $p < 0.005$, suggesting dilation in response to direct perspective ($M = 78.4$, $SD = 10.8$) relative to diverted perspective ($M = 77.3$, $SD = 10.5$). The main effect of action had no effect on pupil size and all interaction effects were non-significant (see **Figure 3**).

fMRI RESULTS

As predicted, the positive main effect of ostension significantly modulated activity in regions associated with Joint Action/Attention, i.e., the ROI in right pSTS [peak voxel: MNI (48, −38, 0)]. However, no above threshold activations were found in ROIs associated with ToM and MNS (i.e., mPFC, rTPJ, rIFG, and rIPL) (see **Figure 4A** and **Table 1**). In contrast, the negative main effect of ostension was found significant in a number of ROIs related both to ToM and MNS: mPFC [peak voxel: MNI (6 42 0)], rIPL [MNI (54, −48, 38)], and rIFG [MNI (44, 36, 28)]

(see **Figure 4B** and **Table 1**). No significant effects were found in the rTPJ and pSTS for this contrast.

The main effect of direction (both positive and negative) did not modulate activity in any of the predefined ROIs. However, explorative whole-brain analysis revealed activity in early visual areas (V1) possibly related to participants' increased eye-movements in this condition (see section "Stand-Alone Eye-Tracking Results" above). These results will thus not be considered any further.

The positive main effect of action elicited significant activity in a number of ROIs relating to the MNS and Joint Action: right pSTS [peak voxel: MNI (52, −48, 2)], rIPL [MNI (42, −44, 52)], and rIFG [MNI (58, 30, 30)] (see **Table 1**). However, ROIs associated with ToM (mPFC and rTJP) did not give significant results. The negative main effect of action did not show any effects. Likewise, none of the interaction effects showed significant results.

Eye-movement corrected fMRI results

When factoring in parametric modulations expressing participants' relative eye-movements, activation patterns largely resemble the results from the analysis above. This indicates that the results reported in **Table 1** are not confounded by condition-related differences in participants' eye-movement patterns. However, the overall statistical strength is considerably weaker, possibly due to the reduced number of participants entering this analysis (full data sets could only be obtained from 11 participants). The positive main effect of ostension was significant in right pSTS [peak voxel: MNI (46, −40, 6)], but not in any of the remaining ROIs. The negative main effect of ostension did not
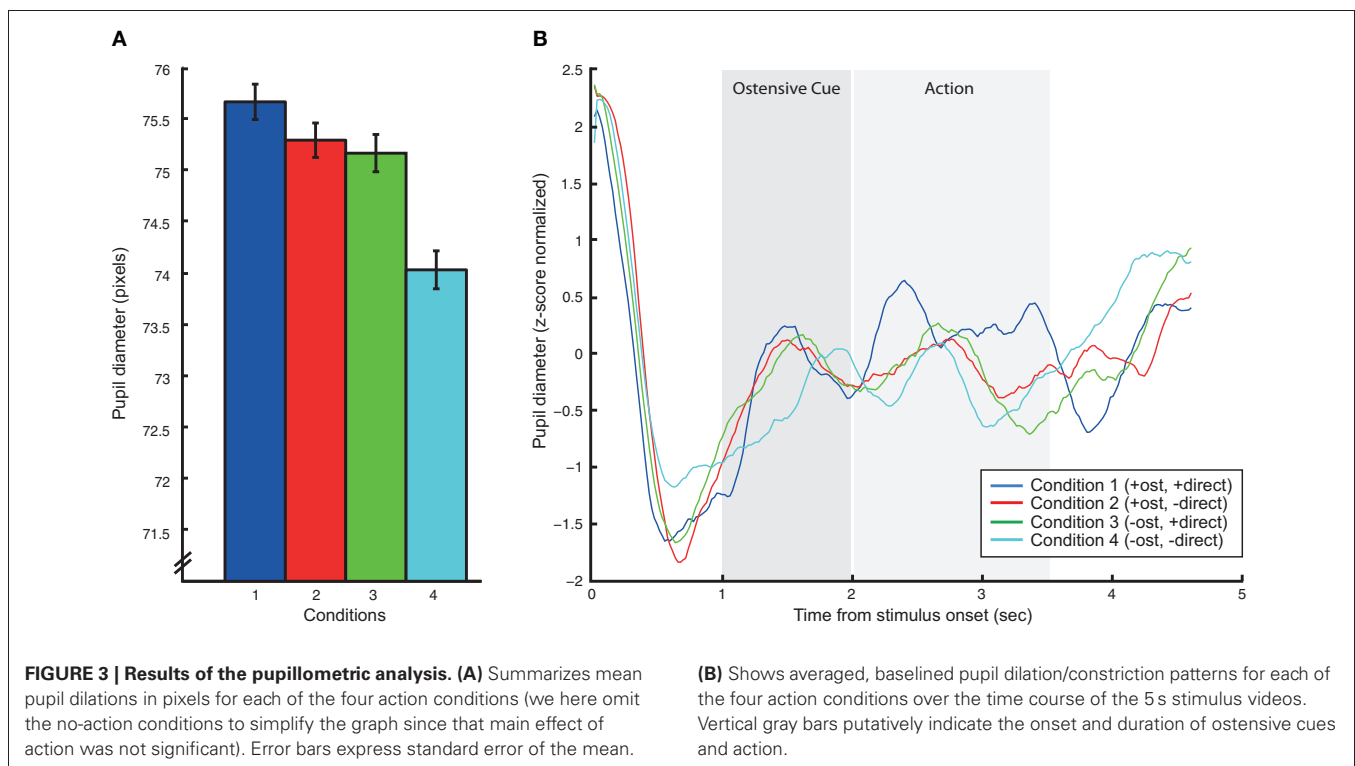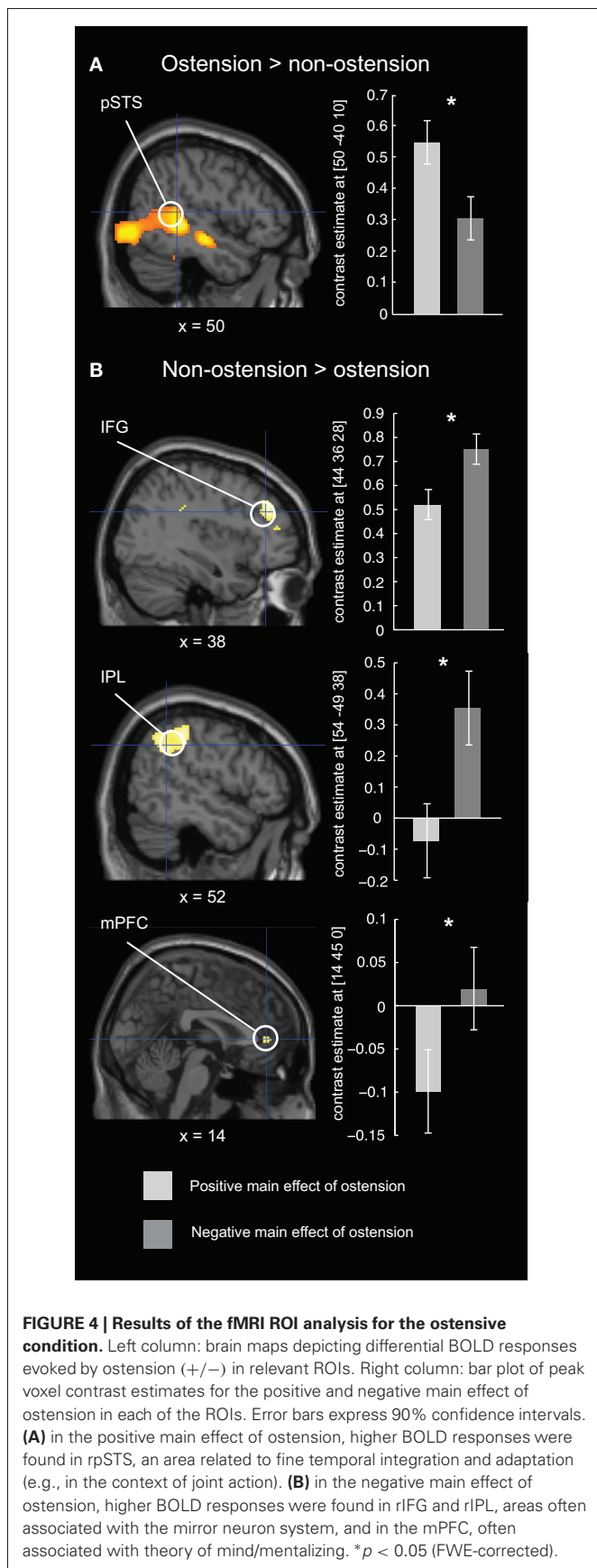

FIGURE 3 | Results of the pupillometric analysis. (A) Summarizes mean pupil dilations in pixels for each of the four action conditions (we here omit the no-action conditions to simplify the graph since that main effect of action was not significant). Error bars express standard error of the mean.

(B) Shows averaged, baselined pupil dilation/constriction patterns for each of the four action conditions over the time course of the 5 s stimulus videos. Vertical gray bars putatively indicate the onset and duration of ostensive cues and action.

FIGURE 4 | Results of the fMRI ROI analysis for the ostensive condition. Left column: brain maps depicting differential BOLD responses evoked by ostension (+/−) in relevant ROIs. Right column: bar plot of peak voxel contrast estimates for the positive and negative main effect of ostension in each of the ROIs. Error bars express 90% confidence intervals. **(A)** in the positive main effect of ostension, higher BOLD responses were found in rpSTS, an area related to fine temporal integration and adaptation (e.g., in the context of joint action). **(B)** in the negative main effect of ostension, higher BOLD responses were found in rIFG and rIPL, areas often associated with the mirror neuron system, and in the mPFC, often associated with theory of mind/mentalizing. $*p < 0.05$ (FWE-corrected).

Table 1 | fMRI results (ROI analysis).

Putative anatomical regions	Z-scores	Coordinates		
		x	y	z
OSTENSION > NON-OSTENSION				
pSTS (right)	4.61	48	−38	0
NON-OSTENSION > OSTENSION				
IPL (right)	3.86	54	−48	38
IFG/MFG (right)	3.23	44	36	28
mPFC	2.99	6	42	0
ACTION > NON-ACTION				
pSTS (right)	7.78	52	−48	2
IPL (right)	4.19	42	−44	52
IFG/MFG (right)	5.16	58	30	30

Main effects of experimental conditions (t-statistics) thresholded at $p < 0.05$ (FWE corrected for multiple comparisons).

Table 2 | Eye-movement corrected fMRI results (11 participants).

Putative anatomical regions		Z-scores	Coordinates		
			x	y	z
OSTENSION > NON-OSTENSION					
rpSTS	$p < 0.000$, corrected	4.36	46	−40	6
NON-OSTENSION > OSTENSION					
rIPL	$p = 0.01$, uncorrected	2.34	54	−48	48
mPFC	$p = 0.005$, uncorrected	2.57	14	28	−2
ACTION > NON-ACTION					
rpSTS	$p < 0.000$, corrected	4.37	46	−42	6
rIPL	$p = 0.01$ corrected	3.52	38	−44	48
rTPJ	$p = 0.01$ corrected	3.53	50	−64	20

Main effects of experimental conditions (t-statistics).

give any above threshold activity. However, an exploratory lowering of significance thresholds revealed strong trends in mPFC [MNI (14, 28, −2), $p = 0.005$, uncorrected] and rIPL [MNI (54, −48, 48), $p = 0.01$, uncorrected] (see **Table 2**). No results were found in pSTS, rTPJ and rIFG.

The main effect of action was found to significantly modulate activity in right pSTS [MNI (46, −42, 6)], in rIPL [MNI (38, −44, 48)], and the rTPJ [MNI (50, −64, 20)] (see **Table 2**). No effects were found in the remaining ROIs and for the negative main effect of action. Likewise, none of the interaction effects reached above threshold significance.

DISCUSSION

Which brain structures facilitate contingent complementary coordination between interacting individuals? This study attempts to make an experimental contribution to current disputes concerning the foundations of social interaction. Based on recent directions in philosophy of mind and complex systems approaches we argue that social interaction may be conceptualized as a collective, interpersonal phenomenon constituted by multi-modal intersubjective coordination processes.

This approach departs from ToM and MNS based frameworks, where interaction is founded on or extrapolated from individual processes of social observation, and it allows for specific predictions regarding individual brain activity during social interaction.

Participants were presented with dynamic situations that afforded different styles of social perception. In some situations, an actor "privately" manipulated objects in a non-ostensive context, while in others object gestures were accompanied with interaction-initiating, ostensive cues. Our results demonstrate that the ostensive contextualization of action radically altered the perceptual attitude of participants. While the non-ostensive scenes called for an observational attitude concerned with "understanding" the actions and intentions of the actor, the ostensive act of *placing an object for* or *showing an object to* someone strongly affords complementary completion by the recipient. The non-ostensive and ostensive scenes thus engage the participants in fundamentally different ways as "observational bystanders" or as "potential interactive recipients." While the first type of situation (*social observation*) can be fully described on the level of individual cognition (mental inference of simulation), the second (*social interaction*) is more appropriately approached as a continuous adaptive coupling between minds (Tylén and Allen, 2009; Hasson et al., 2012). We thus predicted quite different behavioral and neurocognitive results for the two conditions.

Participants' ratings of the socially engaging character of stimulus scenes confirm such predictions. Overall, the scores suggest that although the video stimuli are inherently unresponsive (compared to "live" interaction), they successfully evoked feelings of social contingency in the participants. By far the strongest result is obtained for the positive main effect of ostension, followed by action. Curiously, and contrary to our expectations, the effect of direction is substantially weaker, indicating that the recipient design ("facing you" vs. "facing someone else") is less important for the participants' experience of social engagement with the displayed actor. However, there are strong interaction effects indicating that direct perspective matters for the ostensive conditions while the effect is substantially weaker for the non-ostensive conditions (see **Figure 2**).

Analogous results are found for the fMRI brain imaging data. Among the pre-defined regions of interest, the rpSTS was most strongly activated by scenes affording social responsiveness. In these scenes, an actor looked up and made interaction-initiating ostensive cues (eye contact, eyebrow flashes and nods). The rpSTS area has been repeatedly associated with eye-gaze (Allison et al., 2000; Pelphrey et al., 2004). In a related study, Redcay et al. (2010) suggested that uncontrolled condition related differences in participants' eye-movement patterns could potentially confound their findings. However, we employed in-scanner eye-tracking to test for eye-movement related effects. Analyses of saccade velocities did not show significant differences for ostensive/non-ostensive conditions. Beside, when participants' eye-movements were factored into the fMRI analysis, results related to rpSTS were not influenced. Our findings relating interactive conditions to activity in the pSTS can therefore not be explained simply by differences in participants' eye movements.

Other studies have indicated that rpSTS may be particularly sensitive to gaze direction, such as the distinction between direct and averted gaze (Kuzmanovic et al., 2009; Ethofer et al., 2011). In our study, we modulated the body orientation of the actor, so that she/he was either facing the participant or presented from a slightly averted perspective. Contrary to our expectations, we did not find any modulations of the rpSTS related to this contrast (we had expected direction to interact with ostension). In fact, similar to the behavioral stimulus ratings, the body orientation of the actor showed relatively weak effects. In sum, the rpSTS effects found in this study cannot be reduced to stimulus-induced differences in participants' eye-movement patterns or effects related to the actors' body and gaze directions. Instead, they suggest that rpSTS activity may be related to a socially interactive contextualization of the scene as a whole whether or not the participant was addressed as the intended recipient of the act.

Interestingly, pupillometric analyses showed a strong main effect of direction. When the actor was oriented toward the experimental participant, we recorded stronger pupil dilations relative to diverted orientations. Similar (although slightly weaker) effects were found for ostension. Pupil dilation has been reported as a reliable marker of low-level emotional arousal related to the sympathetic nervous system (Laeng et al., 2012) and has likewise been shown to provide a sensitive index of subtle and complex cognitive and affective processes (Partala and Surakka, 2003; Granholm and Steinhauer, 2004). The pupillometric findings in this study are thus taken in support of our predictions: actors' ostensive cues and direct body orientation induce participants with increased levels of attentional alertness due to affordances for complementary responsive action. It should be noted that although the activation of rpSTS does not follow the same pattern as the pupil dilations (rpSTS seems insensitive to direction), we cannot fully exclude the possibility that arousal rather than complementary interactive dynamics drives some of the brain activation patterns in this study.

Together, the findings can inform discussion between different models of social cognition. While many "observational" social cognition tasks rely solely on participants to internally represent other agents' behaviors, intentions and beliefs, social interaction is more appropriately depicted as a continuous contingent coupling between two or more individuals (Hasson et al., 2012). Right pSTS has been reported in a number of studies contrasting situations where participants solve tasks relying on continuous coordination with external social stimuli with situations where they solve tasks purely based on internal reasoning processes (Iacoboni et al., 2001; Newman-Norlund et al., 2008; Noordzij et al., 2009; Wyk et al., 2009; Redcay et al., 2010; Carter et al., 2011). A subset of these studies (Newman-Norlund et al., 2008; Noordzij et al., 2009; Redcay et al., 2010) even facilitated live contingent interaction between experimental participants lying in the scanner and cooperative partners in the control room in cooperative tasks. It should be noted that—based solely on our data—we cannot exclude the possibility that the rpSTS effect found in our study reflects a mentalizing strategy. Following the work of e.g., Gergely and colleagues (Csibra and Gergely, 2009; Southgate et al., 2009), ostensive cues can act to direct and enhance attention to a subsequent behavior and thereby

facilitate "understanding of the social goal" of the agent. However, considering the growing literature associating the rpSTS with contingent social interaction, we favor the interpretation that the effect relates to the socially engaging affordances of the ostensive stimulus scenes evoking a strong inclination to respond in complementary ways (Sartori et al., 2012).

Interestingly, when participants were confronted with non-ostensive scenes featuring an actor "privately" manipulating objects, we found increased activation of areas normally associated with ToM (mPFC) and MNS (rIPL and rIFG). We notice that although the frontal component of our MNS mask was centered in the IFG (see section "Materials and Methods" above), the activation peak found in our study is slightly more anterior and thus rather resembles findings from Weissman et al. (2008) relating social observation to the DLPFC. In contrast to the rIPL activation, we will thus not make any strong claims about this frontal component in relation to the MNS. However, our findings suggest that the effects found in ToM and MNS related areas could be explained by reference to the quite different affordances of the control stimuli. The non-ostensive character of these scenes frames the participant as an observing bystander *making sense* of the scenes rather than *responding* to them. This form of "social observation" does not to the same extent depend on fine temporal coupling and coordination with the external social environment. Rather, it can be characterized as a decoupled process relying on inferential reasoning (mentalizing) and mental action simulation.

It has been argued that the MNS is indeed sensitive to socially complementary action affordances (Newman-Norlund et al., 2007). While an interesting TMS study could be interpreted in favor of this account (Newman-Norlund et al., 2010), other evidence is more mixed. We thus notice that in a study from the same lab, the strongest effect of complementary actions was seemingly found in the rpSTS (Newman-Norlund et al., 2008). Furthermore, other researchers have not been able to replicate the MNS findings for complementary actions (Kokal et al., 2009; Ocampo et al., 2011).

The differential activation and deactivation patterns found for interaction vs. observation conditions seem to resonate with findings on intrinsic variability of macroscopic networks associated with attention and social-cognitive action control. Indeed, evidence suggests that the neural apparatus supporting social observation (in particular mPFC and IPL) are directly inhibited by tasks requiring high cognitive demand and focused attention (Raichle et al., 2001; McKiernan et al., 2003; Spreng et al., 2009; Allen and Williams, 2011). Similarly, the continuous tracking and contingent responding required of social interaction may necessitate going "online" to the extent of actually de-activating networks associated with ToM and self-related cognition (Fox et al., 2005; Schilbach et al., 2008; Andrews-Hanna et al., 2010).

We also found a number of our regions of interest to be modulated by the positive main effect of action. In particular, significant activation was found in rIFG, rIPL and the rpSTS, while no effect was found in rTPJ and mPFC. While the activation of MNS related regions (rIFG and rIPL) is possibly related to participants' mirroring of displayed actions, we speculate that the ostensive properties of the object gestures themselves observed in the behavioral study (stimulus ratings) might account for the rpSTS component finding.

CONCLUSION

The current study contrasted brain activity elicited by short videos, which evoked in the participants *observational* social cognition and *interactive* social cognition. The difference was triggered by the presence of ostensive cues, which open a channel of communication and interaction. Observational social cognition differentially evoked activity in regions hitherto associated with Theory of Mind (mPFC) and the Mirror Neuron Systems (IPL, IFG). Interactive social cognition differentially evoked activity in right posterior STS, a region known to be involved in continuous fine-grained temporal navigation and integration of stimuli. Brain imaging findings are supported by behavioral tests showing that participants found interactive conditions more socially engaging and pupillometric analyses indicating higher levels of arousal (pupil dilation) for interactive than observational conditions.

Our findings demonstrate that very simple cues may shift both the experience of participants and the brain activity associated with social cognition between an *observational mode* and an *interactive mode*. The identification of a distinct interactive mode of social cognition opens a new avenue for research on social cognition, both under normal conditions and in clinical disorders, such as autism and schizophrenia, characterized by disturbances in social cognition.

REFERENCES

Allen, M., and Williams, G. (2011). Consciousness, plasticity, and connectomics: the role of intersubjectivity in human cognition. *Front. Psychol.* 2:20. doi: 10.3389/fpsyg.2011.00020

Allison, T., Puce, A., and McCarthy, G. (2000). Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4, 267–278.

Andrews-Hanna, J. R., Reidler, J. S., Huang, C., and Buckner, R. L. (2010). Evidence for the default network's role in spontaneous cognition. *J. Neurophysiol.* 104, 322–335.

Arbib, M. A., Billard, A., Iacoboni, M., and Oztop, E. (2000). Synthetic brain imaging: grasping, mirror neurons and imitation. *Neural Netw.* 13, 975–997.

Black, D. P., Riley, M. A., and McCord, C. K. (2007). Synergies in intra- and interpersonal interlimb rhythmic coordination. *Motor Control* 11, 348–373.

Carter, E. J., Hodgins, J. K., and Rakison, D. H. (2011). Exploring the neural correlates of goal-directed action and intention understanding. *Neuroimage* 54, 1634–1642.

Castelli, F., Happe, F., Frith, U., and Frith, C. (2000). Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage* 12, 314–325.

Christensen, K. R., and Wallentin, M. (2011). The locative alternation: distinguishing linguistic processing cost from error signals in Broca's region. *Neuroimage* 56, 1622–1631.

Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse Stud.* 7, 507–525.

Csibra, G., and Gergely, G. (2009). Natural pedagogy. *Trends Cogn. Sci.* 13, 148–153.

Csibra, G., and Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 366, 1149–1157.

De Jaegher, H., Di Paolo, E., and Gallagher, S. (2010). Can social interaction constitute social cognition? *Trends Cogn. Sci.* 14, 441–447.

Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition.* Cambridge, MA: Harvard University Press.

Donald, M. (2001). *A Mind So Rare: The Evolution of Human Consciousness.* New York, NY: Norton.

Ethofer, T., Gschwind, M., and Vuilleumier, P. (2011). Processing social aspects of human gaze: a combined fMRI-DTI study. *Neuroimage* 55, 411–419.

Ferri, F., Campione, G. C., Dalla Volta, R., Gianelli, C., and Gentilucci, M. (2011). Social requests and social affordances: how they affect the kinematics of motor sequences during interactions between conspecifics. *PLoS ONE* 6:e15855. doi: 10.1371/journal.pone.0015855

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9673–9678.

Frith, C. D., and Frith, U. (2006a). How we predict what other people are going to do. *Brain Res.* 1079, 36–46.

Frith, C. D., and Frith, U. (2006b). The neural basis of mentalizing. *Neuron* 50, 531–534.

Frith, U., and Frith, C. D. (2001). The biological basis of social interaction. *Curr. Dir. Psychol. Sci.* 10, 151–155.

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., and Frith, C. D. (2000). Reading the mind in cartoons and stories: an fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia* 38, 11–21.

Gallagher, S., and Hutto, D. (2008). "Understanding others through primary interaction and narrative practice," in *The Shared Mind: Perspectives on Intersubjectivity,* eds J. Zlatev, T. P. Racine, C. Sinha, and E. Itkonen (Amsterdam; Philadelphia: John Benjamins Publishing Company), 17–38.

German, T. P., Niehaus, J. L., Roarty, M. P., Giesbrecht, B., and Miller, M. B. (2004). Neural correlates of detecting pretense: automatic engagement of the intentional stance under covert conditions. *J. Cogn. Neurosci.* 16, 1805–1817.

Goodwin, C., and Heritage, J. (1990). Conversation analysis. *Annu. Rev. Anthropol.* 19, 283–307.

Granholm, E., and Steinhauer, S. R. (2004). Pupillometric measures of cognitive and emotional processes. *Int. J. Psychophysiol.* 52, 1–6.

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* 16, 114–121.

Hasson, U., Yang, E., Vallines, I., Heeger, D. J., and Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *J. Neurosci.* 28, 2539–2550.

Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., and Mazziotta, J. C. (2003). The essential role of Broca's area in imitation. *Eur. J. Neurosci.* 17, 1123–1128.

Howell, D. C. (2002). *Statistical Methods for Psychology.* Pacific Grove, CA: Duxbury Wadsworth Group.

Iacoboni, M., Koski, L. M., Brass, M., Bekkering, H., Woods, R. P., Dubeau, M. C., et al. (2001). Reafferent copies of imitated actions in the right superior temporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 98, 13995–13999.

Kampe, K. K., Frith, C. D., and Frith, U. (2003). "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *J. Neurosci.* 23, 5258–5263.

Kaplan, J. T., and Iacoboni, M. (2006). Getting a grip on other minds: mirror neurons, intention understanding, and cognitive empathy. *Soc. Neurosci.* 1, 175–183.

Kello, C. T., Brown, G. D., Ferrer, I. C. R., Holden, J. G., Linkenkaer-Hansen, K., Rhodes, T., et al. (2010). Scaling laws in cognitive sciences. *Trends Cogn. Sci.* 14, 223–232.

Kokal, I., Gazzola, V., and Keysers, C. (2009). Acting together in and beyond the mirror neuron system. *Neuroimage* 47, 2046–2056.

Konvalinka, I., and Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Front. Hum. Neurosci.* 6:215. doi: 10.3389/fnhum.2012.00215

Konvalinka, I., Vuust, P., Roepstorff, A., and Frith, C. D. (2010). Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Q. J. Exp. Psychol.* 63, 220–230.

Kuzmanovic, B., Georgescu, A. L., Eickhoff, S. B., Shah, N. J., Bente, G., Fink, G. R., et al. (2009). Duration matters: dissociating neural correlates of detection and evaluation of social gaze. *Neuroimage* 46, 1154–1163.

Laeng, B., Sirois, S., and Gredebäck, G. (2012). Pupillometry. *Perspect. Psychol. Sci.* 7, 18–27.

Lerner, Y., Honey, C. J., Silbert, L. J., and Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *J. Neurosci.* 31, 2906–2915.

Maldjian, J. A., Laurienti, P. J., and Burdette, J. H. (2004). Precentral gyrus discrepancy in electronic versions of the Talairach atlas. *Neuroimage* 21, 450–455.

Maldjian, J. A., Laurienti, P. J., Kraft, R. A., and Burdette, J. H. (2003). An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. *Neuroimage* 19, 1233–1239.

McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., and Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *J. Cogn. Neurosci.* 15, 394–408.

Newman-Norlund, R. D., Bosga, J., Meulenbroek, R. G., and Bekkering, H. (2008). Anatomical substrates of cooperative joint-action in a continuous motor task: virtual lifting and balancing. *Neuroimage* 41, 169–177.

Newman-Norlund, R. D., Ondobaka, S., Van Schie, H. T., Van Elswijk, G., and Bekkering, H. (2010). Virtual lesions of the IFG abolish response facilitation for biological and non-biological cues. *Front. Behav. Neurosci.* 4:5. doi: 10.3389/neuro.08.005.2010

Newman-Norlund, R. D., Van Schie, H. T., Van Zuijlen, A. M., and Bekkering, H. (2007). The mirror neuron system is more active during complementary compared with imitative action. *Nat. Neurosci.* 10, 817–818.

Noordzij, M. L., Newman-Norlund, S. E., De Ruiter, J. P., Hagoort, P., Levinson, S. C., and Toni, I. (2009). Brain mechanisms underlying human communication. *Front. Hum. Neurosci.* 3:14. doi: 10.3389/neuro.09.014.2009

Ocampo, B., Kritikos, A., and Cunnington, R. (2011). How

frontoparietal brain regions mediate imitative and complementary actions: an fMRI study. *PLoS ONE* 6:e26945. doi: 10.1371/journal.pone.0026945

Partala, T., and Surakka, V. (2003). Pupil size variation as an indication of affective processing. *Int. J. Hum. Comput. Stud.* 59, 185–198.

Pelphrey, K. A., Viola, R. J., and McCarthy, G. (2004). When strangers pass: processing of mutual and averted social gaze in the superior temporal sulcus. *Psychol. Sci.* 15, 598–603.

Penny, W., and Holmes, A. P. (2007). "Random effects analysis," in *Statistical Parametric Mapping: The Analysis of Functional Brain Images,* eds K. J. Friston, J. Ashburner, S. Kiebel, T. Nichols, and W. Penny (London: Academic Press), 156–165.

Raichle, M. E., Macleod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., and Shulman, G. L. (2001). A default mode of brain function. *Proc. Natl. Acad. Sci. U.S.A.* 98, 676.

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D., et al. (2010). Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *Neuroimage* 50, 1639–1647.

Riley, M. A., Richardson, M. J., Shockley, K., and Ramenzoni, V. C. (2011). Interpersonal synergies. *Front. Psychol.* 2:38. doi: 10.3389/fpsyg.2011.00038

Rizzolatti, G., Fogassi, L., and Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev. Neurosci.* 2, 661–670.

Roepstorff, A. (2001). Brains in scanners: an Umwelt of cognitive neuroscience. *Semiotica* 134, 747–765.

Salvucci, D. D., and Goldberg, J. H. (2000). "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the 2000 Symposium on Eye Tracking Research and Applications,* ed A. T. Duchowski (Palm Beach Gardens, FL: ACM), 71–78.

Sartori, L., Cavallo, A., Bucchioni, G., and Castiello, U. (2012). From simulation to reciprocity: the case of complementary actions. *Soc. Neurosci.* 7, 146–158.

Schilbach, L. (2010). A second-person approach to other minds. *Nat. Rev. Neurosci.* 11:449. doi: 10.1038/nrn2805-c1

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., and Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of

cognizing and its putative relationship to the "default system" of the brain. *Conscious. Cogn.* 17, 457–467.

Schulte-Ruther, M., Markowitsch, H. J., Fink, G. R., and Piefke, M. (2007). Mirror neuron and theory of mind mechanisms involved in face-to-face interactions: a functional magnetic resonance imaging approach to empathy. *J. Cogn. Neurosci.* 19, 1354–1372.

Southgate, V., Chevallier, C., and Csibra, G. (2009). Sensitivity to communicative relevance tells young children what to imitate. *Dev. Sci.* 12, 1013–1019.

Spreng, R. N., Mar, R. A., and Kim, A. S. N. (2009). The common neural basis of autobiographical memory, prospection, navigation, theory of mind, and the default mode: a quantitative meta-analysis. *J. Cogn. Neurosci.* 21, 489–510.

Stamenov, M., and Gallese, V. (2002). *Mirror Neurons and the Evolution of Brain and Language.* Amsterdam, Philadelphia: John Benjamins Pub.

Stephens, G. J., Silbert, L. J., and Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14425–14430.

Tomasello, M. (1999). *The Cultural Origins of Human Cognition.* Cambridge, MA: Harvard University Press.

Tomasello, M. (2008). *Origins of Human Communication.* Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H., (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691. discussion: 691–735.

Tylén, K., and Allen, M. (2009). "Interactive sense-making in the brain," in *Enacting Intersubjectivity: Paving the Way for a Dialogue Between Cognitive Science, Social Cognition and Neuroscience,* eds A. Carassa, F. Morganti, and G. Riva (Lugano, Switzerland: Lariopring), 224–241.

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

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

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., and Bara, B. G. (2004). Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *J. Cogn. Neurosci.* 16, 1854–1863.

Weissman, D. H., Perkins, A. S., and Woldorff, M. G. (2008). Cognitive control in social situations: a role for the dorsolateral prefrontal cortex. *Neuroimage* 40, 955–962.

Wurm, M. F., Von Cramon, D. Y., and Schubotz, R. I. (2011). Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. *Hum. Brain Mapp.* 32, 2141–2150.

Wyk, B. C., Hudac, C. M., Carter, E. J., Sobel, D. M., and Pelphrey, K. A. (2009). Action understanding in the superior temporal sulcus region. *Psychol. Sci.* 20, 771–777.