Look at this: the neural correlates of initiating and responding to bids for joint attention

Elizabeth Redcay [1]*, Mario Kleiner [2] and Rebecca Saxe [3]

[1] Department of Psychology, University of Maryland, College Park, MD, USA
[2] Department of Human Perception, Cognition, and Action, Max Planck Institute for Biological Cybernetics, Tuebingen, Germany
[3] Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

When engaging in joint attention, one person directs another person's attention to an object (Initiating Joint Attention, IJA), and the second person's attention follows (Responding to Joint Attention, RJA). As such, joint attention must occur within the context of a social interaction. This ability is critical to language and social development; yet the neural bases for this pivotal skill remain understudied. This paucity of research is likely due to the challenge in acquiring functional MRI data during a naturalistic, contingent social interaction. To examine the neural bases of both IJA and RJA we implemented a dual-video set-up that allowed for a face-to-face interaction between subject and experimenter via video during fMRI data collection. In each trial, participants either followed the experimenter's gaze to a target (RJA) or cued the experimenter to look at the target (IJA). A control condition, solo attention (SA), was included in which the subject shifted gaze to a target while the experimenter closed her eyes. Block and event-related analyses were conducted and revealed common and distinct regions for IJA and RJA. Distinct regions included the ventromedial prefrontal cortex for RJA and intraparietal sulcus and middle frontal gyrus for IJA (as compared to SA). Conjunction analyses revealed overlap in the dorsal medial prefrontal cortex (dMPFC) and right posterior superior temporal sulcus (pSTS) for IJA and RJA (as compared to SA) for the event analyses. Functional connectivity analyses during a resting baseline suggest joint attention processes recruit distinct but interacting networks, including social-cognitive, voluntary attention orienting, and visual networks. This novel experimental set-up allowed for the identification of the neural bases of joint attention during a real-time interaction and findings suggest that whether one is the initiator or responder, the dMPFC and right pSTS, are selectively recruited during periods of joint attention.

Keywords: fMRI, superior temporal sulcus, social cognition, social interaction, face-to-face, dorsal medial prefrontal cortex

INTRODUCTION

Imagine a typical scene at a zoo: a two-year-old child points into an enclosure, while looking at her father and saying "Ba." The father looks at the child, then into the enclosure, then back at the child, and says "Yes! It's a bear!" In this scenario, the child has made a bid to *initiate* joint attention on something in the enclosure; the parent then *responds* by attending to the likely target (the bear), and then returning attention to the child to share the rewards of the interaction.

These simple, automatic, and everyday behaviors are the foundations of our abilities to communicate with and learn from others from infancy through adulthood. Joint attention skills in early infancy are predictive of later language development (Morales, 2000; Brooks and Meltzoff, 2005; Mundy et al., 2007; Brooks, 2008), social competence (Vaughan Van Hecke et al., 2007), and theory of mind abilities (Nelson et al., 2008). Joint attention behaviors are reported to be atypical in individuals with autism spectrum disorders (ASD), and are proposed to be a source of characteristic deficits in language and social interaction (Charman, 2003).

One unresolved question is the extent to which responding and initiating joint attention (IJA) behaviors rely on the same cognitive and neural systems or distinct but interacting systems (e.g., Mundy and Newell, 2007). In a dyad, one person initiates joint attention (IJA) while the other responds to a joint attention bid (RJA). In both, two people share attention on a common object. Importantly, this is distinct from coincidental shared attention where two people may happen to attend to the same thing. True joint attention requires the intention to share attention, or shared intentionality. If the core of both IJA and RJA is a common cognitive mechanism for shared intentionality then one would expect individual differences in the development of these behaviors to be accounted for by variance in social-cognitive development (Carpenter et al., 1998; Tomasello et al., 2005). Some behavioral evidence offers support for this prediction (Carpenter et al., 1998; Osório et al., 2011). For example

between ages 9 and 15 months sharing attention, following attention, and initiating attention behaviors emerge quickly and in a reliable order (Carpenter et al., 1998), but see Slaughter and McConnell (2003). An alternative model, however, suggests that distinct processes underlie development of IJA and RJA (Mundy and Newell, 2007; Mundy et al., 2007): IJA development is mediated by developments in volitional attention and control while RJA development is mediated by automatic attention orienting. Support for this hypothesis is found in longitudinal studies in which individual differences within RJA and IJA behaviors are stable over development (9–18 months) but individual differences in RJA do not predict development of IJA behaviors and vice versa (Mundy and Newell, 2007).

Neuroimaging measures offer a complementary tool to examine the common and distinct cognitive processes underlying RJA and IJA. The common mechanism should be reflected in a common neural substrate, whereas distinct mechanisms should be reflected in distinct neural substrates. Currently, the neural correlates of joint attention behaviors remain unclear. Neuroimaging studies have characterized the neural bases of components of RJA: especially observing someone else's gaze or point, shifting of attention, and sharing attention on an object at which another person looked. These studies have primarily required participants to view images or movies of real or virtual people shifting gaze toward or away from an object. In general, these studies report that the posterior superior temporal sulcus (pSTS) (Morris et al., 2005; Materna et al., 2008) and/or the medial prefrontal cortex (MPFC) (Bristow et al., 2007; Schilbach et al., 2010) are recruited during components of RJA (review, Redcay and Saxe, in press).

While these behaviors are part of responding to joint attention (RJA), the "joint" aspect of joint attention is typically not examined. To achieve full joint attention, both members of the dyad must know they are jointly attending to the same thing and have reached the state of joint attention through mutual coordination (Carpenter and Liebal, 2011). Experimental manipulations of *IJA* are even more rarer, because the participant must perceive that his or her bids for joint attention are met with a contingent response. Given the constrained environment of MRI scanners, acquiring neuroimaging data during a real-time contingent social interaction poses technical challenges.

A previous study (Schilbach et al., 2010) has examined IJA and RJA, using a gaze-contingent interaction paradigm with an avatar that was supposed to represent a real person. Participants were told they were playing an interactive game in which the participant would follow the avatar's gaze shifts (RJA conditions) and pay attention to the avatar's tendency to follow the participant's gaze shifts (IJA conditions). In the initiating condition, participants initiated a gaze shift to a chosen location that was (joint attention) or was not (non-joint attention) followed by the avatar. In the responding condition, participants responded to a gaze shift from the avatar by following gaze to the chosen location (joint attention) or choosing a non-target location (non-joint attention). The goal was not explicitly to coordinate and share attention on an object, but rather to learn about the gaze or response patterns of another person. In this experiment, both IJA and RJA recruit the MPFC relative to the matched non-joint conditions, and additional distinct regions are recruited for each behavior (Schilbach et al., 2010). Specifically, initiating a bid for joint attention recruits ventral striatum while responding to a bid for joint attention recruits MPFC.

The current study extends the previous study by using a novel design to examine two aspects of joint attention that were not examined in the previous study. First, the previous study did not require the intentional coordination of attention between two people for the purpose of communication. For example, in the joint attention scenario in the zoo, the girl requests that her dad share attention with her on the bear. The father coordinates his attention between her and the object and labels the object: "Yes, bear!" This active coordination toward a communicative goal is why joint attention is such a powerful learning tool. Additionally, this intentional coordination is the aspect of joint attention in the second year of life that correlates with later theory of mind abilities (Charman, 2000). Second, the previous study used an anti-saccade condition as a control for the joint attention conditions to control for the perception of eye movements (e.g., if the avatar looks left, look to the opposite side). One limitation of this control condition, however, is that it contains an important component of joint attention: namely using another person's gaze to cue your attention. Because gaze cueing is rapid and automatic the participants are likely cued by the gaze shift and then have to reorient to another location (review, Frischen et al., 2007).

In order to examine shared and distinct brain networks involved in IJA and RJA, we developed a novel communicative paradigm in which the subject and experimenter participate in a face-to-face real-time interaction while the subject is in the scanner (Redcay et al., 2010). During scanning, the experimenter and subject played a game in which both had to use gaze cues to communicate information about the location of a target object, and then share attention on the object. In each trial, participants either followed the experimenter's gaze to a target (RJA) or cued the experimenter to look at the target (IJA). A control condition, solo attention (SA), was included in which the subject shifted gaze to a target while the experimenter closed her eyes, thus eliminating the anti-saccade task in the control condition. We examined (1) the extent to which IJA and RJA recruit common and distinct regions during joint attention and (2) the extent to which regions recruited during IJA and RJA are part of distinct functional networks, measured by correlations during resting baseline periods. We predicted that IJA would require greater coordination of attention between the participant and object, and thus recruit attention orienting and cognitive control regions to a greater extent than RJA. Additionally, we predicted that RJA would require greater attention to another's intentions behind their actions (i.e., gaze shift) and thus, recruit the posterior STS to a greater extent. Finally, based on previous research on the role of the dorsal medial prefrontal cortex (dMPFC) in the representation of self and other (review: Amodio and Frith, 2006; Saxe, 2006) and joint attention (Williams et al., 2005; Schilbach et al., 2010) we predicted that engaging in joint attention, whether one is the initiator or responder, would recruit a shared region within the dMPFC.

METHODS

PARTICIPANTS

Neuroimaging data were collected from 41 healthy, typical adults. All participants gave informed written consent and were paid for their participation in the study as approved by the committee on the use of humans as experimental subjects (COUHES) at MIT. Participants were screened for neurological or psychiatric conditions as well as any contraindications for MRI scanning. Four participants were excluded from further analyses due to excessive motion during the imaging session (criteria described below). Five were excluded due to a failure to record behavioral data during the session. Thus, the final sample consisted of 32 participants (19 male, age 24.5 ± 5 years). Data from eight of these participants have been published previously for the RJA condition only (Redcay et al., 2010).

JOINT ATTENTION TASK

Participants engaged in a game designed to elicit both IJA and RJA behaviors during a real-time interaction with an experimenter via live video feed. Participants were instructed that the goal of the game was to find the location of a hidden mouse. The mouse was "hiding" in a box within one of the four corners of the screen. On each trial, a clue (a mouse tail) would appear in one of the four corners to indicate where the mouse was hiding (**Figure 1**). During joint attention conditions (initiating and responding) participants were playing the game with the experimenter in order to find the mouse together. On IJA trials the participant saw the mouse tail clue on his or her screen and had to direct the experimenter's attention to the correct location using gaze cues. During RJA events, the experimenter received the clue on her screen and had to direct the participant to the location of the mouse. The experimenter directed the participant by shifting her gaze to the correct location. She maintained her gaze there until the participant matched her gaze. For both conditions, only when both experimenter and participant were fixating on the target location did the mouse appear. During the SA condition the participant's goal was to find the mouse alone while the experimenter simply opened and closed her eyes to indicate that she was not participating in the game.

JOINT ATTENTION DESIGN

The joint attention task was performed during four separate runs of functional MRI data acquisition[1]. Joint attention trials were presented in a blocked design with each block containing five trials of the same condition in a row. Each block was preceded by a 4 s period of instructions to inform participants of the upcoming condition. Each functional run contained a 30 s rest period at the beginning, middle, and end of the run and contained six experimental blocks (two of each condition) in a semi-counterbalanced order. Each trial was 6 s and consisted of a variable delay between 0 and 1 s before the cue (mousetail) onset to either the participants (IJA and SA) or experimenter's (RJA) screen. The experimenter and participant determined the timing of the rest of the trial, with a maximum length of 6 s.

---

[1] For one participant, behavioral data were available for only three of the four runs and thus only three were included in the analysis.

The experimenter controlled the appearance of the mouse when both she and the participant were determined to be looking at the appropriate corner of the screen (with assistance from a second experimenter who was out of sight from the participant). Discrepancies between joint attention events and mouse appearance were quantified through comparison of recorded key presses and *post-hoc* video coding (see below).
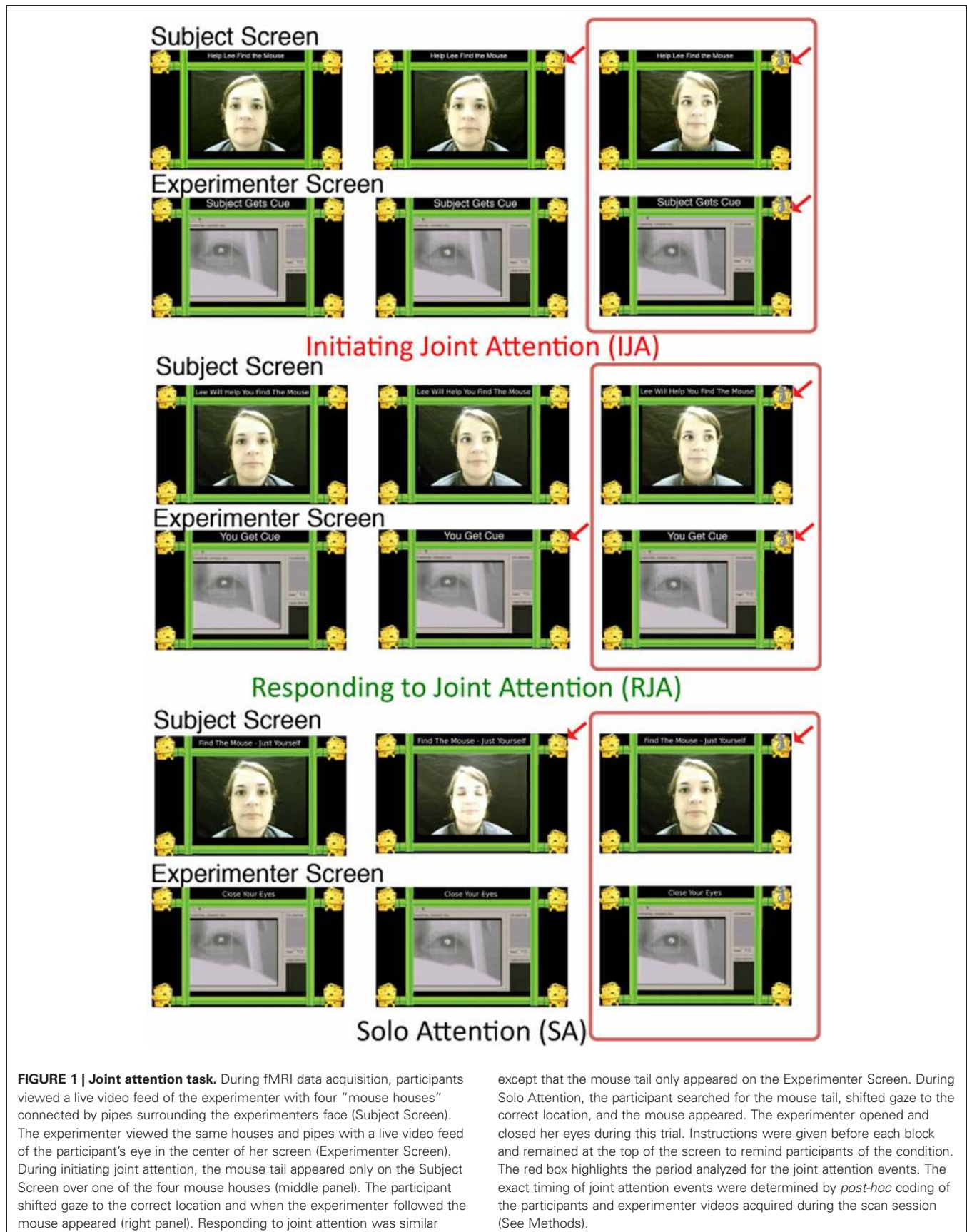
EXPERIMENTAL SET-UP

Extensive details on the experimental set-up can be found in a previously published paper (Redcay et al., 2010). During joint and SA trials, the participant viewed a live video-feed of the experimenter's face surrounded by an image that contained a "cheese house" in each corner of the screen connected by pipes. During rest periods, only a fixation cross was presented on the screen. A camera was positioned at the end of the bore of the scanner to acquire a picture of the participant's eye. This video of the eye was provided in real-time with minimal delays to a MacbookPro laptop that was positioned in front of the experimenter in the MRI control room. The experimenter also had an image of four "cheese houses" connected by pipes surrounding the live video-feed of the participant's eye. This dual video-feed set-up allowed for real-time monitoring of gaze cues by both participant and experimenter. Additionally, this set-up gave the illusion that the participant and experimenter were looking at different sides of the same image (see **Figure 1**). Video recording of the experimenter and participant during the task (referred to as behavioral data) allowed for *post-hoc* coding of event timing during the trial.

All stimuli were programmed and recorded in Matlab 7.8 using the Psychophysics Toolbox Extensions (PTB-3) (Brainard, 1997; Pelli, 1997) on an Apple MacbookPro running OSX 10.5.6.

BEHAVIORAL VIDEO CODING

Videos from the participant and experimenter during each functional run were coded offline using VCode software (http://social.cs.uiuc.edu/projects/vcode.html). Each timepoint in which a participant shifted gaze toward or away from one of the four corners of the screen was recorded. Additionally, each time the experimenter shifted her gaze toward the target (joint attention trials) or closed her eyes (SA) was recorded. The onset of a joint attention event was calculated as the time at which either experimenter (initiating) or participant (responding) shifted gaze to the location at which the other member of the dyad was already looking. The end of the joint attention event was marked by one member of the dyad shifting gaze away from the target location. During SA, the onset was defined as the time at which the participant shifted gaze to the target and the end of the event was defined as the time at which the participant shifted gaze away from the target or the trial ended. The onset and duration of each (joint or solo) attention event were used as regressors for the event-related analyses described below. Trials in which experimenter and participant did not share attention on the same location (for joint attention) or in which the participant did not shift gaze to the target (for SA) were noted as incorrect trials. Using JMP statistical software, three One-Way ANOVAs were conducted to examine the effect of condition (IJA, RJA, SA) on accuracy (% correct), event duration, and

FIGURE 1 | Joint attention task. During fMRI data acquisition, participants viewed a live video feed of the experimenter with four "mouse houses" connected by pipes surrounding the experimenters face (Subject Screen). The experimenter viewed the same houses and pipes with a live video feed of the participant's eye in the center of her screen (Experimenter Screen). During initiating joint attention, the mouse tail appeared only on the Subject Screen over one of the four mouse houses (middle panel). The participant shifted gaze to the correct location and when the experimenter followed the mouse appeared (right panel). Responding to joint attention was similar except that the mouse tail only appeared on the Experimenter Screen. During Solo Attention, the participant searched for the mouse tail, shifted gaze to the correct location, and the mouse appeared. The experimenter opened and closed her eyes during this trial. Instructions were given before each block and remained at the top of the screen to remind participants of the condition. The red box highlights the period analyzed for the joint attention events. The exact timing of joint attention events were determined by *post-hoc* coding of the participants and experimenter videos acquired during the scan session (See Methods).

total number of subject eye movements. For significant effects, follow-up contrasts were conducted using Tukey's HSD.

DATA ACQUISITION AND ANALYSES

Data were collected on a 3T Siemens scanner at the Athinoula A. Martinos Imaging Center at the McGovern Institute for Brain Research at the Massachusetts Institute of Technology. T1-weighted structural images were collected in the axial plane (128 slices, TE = 3.39 ms; TR = 2350 ms; 1.3 mM isotropic voxels). During the joint attention task, T2*-weighted gradient echo-planar images (EPI) were acquired (TR = 2 s; TE = 30 ms; $3.1 \times 3.1 \times 4$ mM; 30 slices). The EPI sequences used Siemens online pace motion correction, which corrected for motion less than 8 mM per volume acquisition. The first four images of each run were discarded.

Data were analyzed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm/) and in-house matlab scripts. Data from all functional runs were realigned to the first volume of the first run using a 6-degree rigid spatial transformation. Images were then spatially normalized to Montreal Neurological Institute (MNI) space using a 12-parameter affine transformation and spatially smoothed (fwhm = 5 mM). Data were high pass filtered at 264 Hz, a frequency corresponding to the length of each functional run (i.e., 264 s). Motion artifacts were examined using an artifact detection toolbox (ART) (http://www.nitrc.org/projects/artifact_detect/). Timepoints (volumes) in which global signal deviated more than three standard deviations from the mean signal or in which the difference in motion between two neighboring timepoints exceeded 1 mM (across rotational or translation directions) were marked as outlier timepoints. Participants who had outlier timepoints for greater than 20% of their functional data were excluded from analyses. As noted above, four participants were excluded due to motion artifact.

Two separate first-level analyses were conducted within each subject. One examined activation across the full block for each condition (Block analyses) and one modeled the periods of joint, or solo, attention separately as events (Event analyses) (see above "*Behavioral Video Coding*" for details). For both analyses, General Linear Model analyses were used to estimate parameter values for each condition (IJA, RJA, and SA) of interest as well as the instruction period. The model additionally included a separate regressor for every outlier timepoint. In the Block analyses the condition events included the full 30 s period. In the Event analyses the condition events included only the time period in which the participant was engaged in joint (or solo) attention. The Event analyses also contained a regressor that modeled all blocks in order to account for variance associated with generic aspects of the task (as compared to rest). For both Block and Event analyses, contrasts were modeled to compare each condition (IJA vs. RJA, IJA vs. SA, RJA vs. SA, JA(IJA + RJA) vs. SA, and reverse contrasts). A brain mask was created for each participant using FSL's brain extraction tool (BET) (Smith, 2002) to restrict analyses to voxels within the brain.

Voxel-wise whole brain two-tailed *t*-tests were conducted separately for each condition and contrast of interest. Data were corrected for multiple comparisons at the voxel and cluster level ($p < 0.05$) using nonparametric permutation analyses

(SnPM5b), except where noted. In order to examine the extent to which IJA and RJA engage overlapping regions, conjunction analyses were run for both Block and Event analyses, which identified regions which showed an above-threshold response to both IJA vs. SA and RJA vs. SA across the whole-brain. In order to identify regions that were recruited to a greater extent for IJA than RJA the contrast of IJA vs. RJA was masked by the comparison of IJA vs. SA ($p < 0.001$, cluster-corrected at $p < 0.05$) and similarly the contrast of RJA vs. IJA was masked by the comparison of RJA vs. SA ($p < 0.001$, cluster-corrected at $p < 0.05$). Each comparison was masked in order to eliminate differences between tasks that are accounted for by the SA control condition. A more liberal threshold (i.e., cluster-correction only) was used for the masks in order to avoid type II errors that may arise from examining a contrast within a contrast. Cluster correction for the condition masks was calculated using AFNIs AlphaSim program (Cox, 1996), which suggested that a minimum cluster size of 384 mm^3 with a voxel threshold of $p < 0.001$ was necessary in order to maintain a cluster-corrected alpha of 0.05. All statistical parametric maps are displayed on a standard template brain in MNI space using mricron software.

FUNCTIONAL CONNECTIVITY ANALYSES

Functional connectivity and hierarchical clustering analyses were conducted in order to examine the extent to which regions recruited during joint attention are part of shared and distinct functional networks. Functional connectivity was examined during the 20 s rest periods, which occurred at the beginning, end, and middle of each run in order to identify task-independent network organization. Seed regions for the functional connectivity analyses were identified from the contrast of JA (IJA + RJA) > SA in the event analyses ($p < 0.001$, cluster-correction at $p < 0.05$) (**Table 1**). Event analyses were used so that differences between conditions would be minimized since the period of analyses was focused to periods with more similar behaviors (i.e., sharing attention). Seed regions were created to include all voxels within a 6 mm radius sphere surrounding the peak voxel of each region identified for the JA > SA contrast (**Table 1**). In addition to the preprocessing described above, data were band-pass filtered ($0.001 < f < 0.08$) to examine low-frequency oscillations characteristic of resting-state networks. Pair-wise partial correlation analyses were run for each seed region of interest (with every other seed) that included the timecourse from that seed region as a regressor of interest. Regressors of no interest included the first-order derivatives of the six motion parameters (from realignment, above), and eigenvectors from a principal component analysis on the white matter and cerebrospinal fluid voxels (separately). Additionally, beginning and ends of blocks were weighted down (using a Hanning filter) in order to minimize any residual effects of the preceding task on the rest blocks. Connectivity analyses were conducted using the CONN-fMRI functional connectivity toolbox for SPM (ver 12) (http://web.mit.edu/swg/software.htm). Correlation values were submitted to a hierarchical cluster analysis in JMP statistical software (ver 9) using Ward's method to identify clusters of regions with similar pair-wise correlation patterns. The number of clusters identified was based on visual inspection of a scree plot. The

Table 1 | Comparisons between joint and solo attention blocks.

Contrast region	Hemi	*x*	*y*	*z*	*T*	*k*
INITIATING JOINT ATTENTION > SOLO ATTENTION						
Inferior frontal gyrus	R	44	24	−16	6.71	10304
Posterior superior temporal sulcus	R	66	−46	16	6.26	20896
Intraparietal sulcus	R	54	−48	50	5.78	20896*
Posterior superior temporal sulcus	L	−50	−56	12	6.09	21824
Intraparietal sulcus	L	−46	−50	52	5.40	21824*
Middle frontal gyrus	R	50	14	46	5.67	5648
Posterior medial frontal gyrus	R	2	20	64	5.58	5808
Inferior frontal gyrus	L	−54	8	0	5.11	2800
SOLO ATTENTION > INITIATING JOINT ATTENTION						
Middle occipital gyrus	L	−26	−96	0	10.56	263200
Middle occipital gyrus	R	26	−92	2	8.80	263200*
Cingulate gyrus	R	20	18	42	4.53	3392
Cingulate gyrus	L	−26	10	32	4.46	3136
RESPONDING TO JOINT ATTENTION > SOLO ATTENTION						
Temporoparietal junction	L	−52	−68	30	8.18	43136
Posterior superior temporal sulcus	L	−52	−66	18	6.78	43126*
Posterior cingulate	R	2	−50	28	7.65	9488
Middle occipital gyrus	L	−10	−98	2	7.46	9120
Middle temporal gyrus	R	58	−68	14	7.27	44928
Posterior superior temporal sulcus	R	60	−52	14	6.66	44928*
Inferior frontal gyrus	L	−50	34	−12	7.04	8016
Dorsal medial prefrontal cortex	L	−2	60	34	5.90	31088
Ventral medial prefrontal cortex	R	2	42	18	5.72	31088*
Fusiform gyrus	R	46	−54	−26	5.82	2288
Inferior frontal gyrus	R	48	32	−16	5.31	4720
SOLO ATTENTION > RESPONDING TO INITIATING JOINT ATTENTION						
Middle occipital gyrus	L	−32	−86	10	10.91	228640
Thalamus	L	−18	−28	14	5.02	2944
INITIATING > RESPONDING TO JOINT ATTENTION MASKED BY IJA > SA						
Middle frontal gyrus	R	34	46	34	7.56	2096
Superior frontal gyrus	R	24	4	66	7.43	688
Middle frontal gyrus	L	−32	48	28	5.86	784
Precuneus	R	8	−64	60	5.86	672
Thalamus	L	−8	−16	10	5.71	736
Inferior frontal gyrus	R	50	14	2	5.54	2928
Caudate	R	16	−10	18	5.23	432
Superior frontal gyrus	L	−2	2	62	5.08	1104
Intraparietal sulcus	L	−36	−54	58	4.90	752
Thalamus	R	8	−16	10	4.89	384
Dorsal anterior cingulate	R	10	16	38	4.66	624
Supramarginal gyrus	R	56	−46	36	3.95	1312
RESPONDING TO > INITIATING JOINT ATTENTION MASKED BY RJA > SA						
Occipital lobe	L	−12	−98	4	10.49	6704
Middle temporal gyrus	R	46	−66	4	6.47	5520
Posterior superior temporal sulcus	L	−60	−42	4	6.10	3824
Inferior frontal gyrus	L	−38	38	−18	5.95	560
Posterior superior temporal sulcus	R	48	−38	4	5.88	2624
Ventral medial prefrontal cortex	L	0	34	−14	5.85	8752
Middle temporal gyrus	R	50	4	−26	5.71	416
Posterior cingulate	R	6	−46	28	5.64	7376
Cuneus	R	12	−102	18	5.59	464
Fusiform gyrus	R	44	−46	−24	5.57	1120
Temporoparietal junction	L	−48	−68	30	5.57	6960
Temporal pole	R	38	16	−42	4.80	576

Regions were identified using voxel- and cluster-correction of p < 0.05. Coordinates are given in MNI space. T-values from the peak voxel of the cluster and size (k) of the cluster are given. * Indicates a subregion of the cluster listed with the same k.

Scree plot displays the dissimilarity value between clusters (y) by number of clusters (x). The point at which the dissimilarity values begin to level defines the optimal number of clusters identified (Catell, 1966).

RESULTS

BEHAVIORAL DATA

Accuracy, joint (or solo) attention event duration, and number of eye movements per block all showed significant effects of condition (p's < 0.05) (**Figure 2**). Mean accuracy for all conditions was above 98%; however, an effect of accuracy was found in that accuracy was slightly lower in IJA than RJA trials. Duration of attention events (i.e., time spent looking at the mouse) varied by condition: the events were longer in SA than joint attention trials; and longer when participants responded to rather than initiated joint attention. Finally, more eye movements were seen in SA than in joint attention conditions and in RJA than IJA conditions.

Experimenter error (i.e., discrepancy between mouse appearance and successful joint (or solo) attention to the correct location was minimal and not significantly different across conditions [$F_{(2, 93)} = 0.49$, $p > 0.62$; IJA: 2.9%; RJA: 2.6%; SA: 2%].

BLOCK ANALYSES

In this first analysis, we were interested in examining the response to the joint attention conditions as compared to the SA control across the full 30 s block. This analysis gives regions involved in the full process of joint attention, as elicited in our communicative game.

Responding to Joint Attention (RJA)

RJA recruited a greater BOLD response than SA within midline regions, including ventral and dorsal medial prefrontal cortex, and posterior cingulate cortex, as well as bilateral inferior frontal gyrus extending into the insula and bilateral superior temporal sulcus extending into middle temporal gyrus and the temporoparietal junction (**Table 2** for full list).

Initiating Joint Attention (IJA)

IJA also showed greater activation than SA within bilateral superior temporal sulcus and left inferior parietal lobe and bilateral inferior frontal gyrus and right middle frontal gyrus. Additionally, activation was seen in the posterior medial frontal cortex/supplementary motor area, middle frontal gyrus, and right inferior parietal lobe (**Figure 3A**).

Common regions

A conjunction analysis revealed five regions of significantly overlapping activation between IJA vs. SA and RJA vs. SA. These regions were bilateral pSTS, left intraparietal sulcus, right inferior frontal gyrus, and posterior medial frontal cortex (**Figure 3A**).

Distinct regions

Distinct regions were recruited for IJA and RJA (**Figure 3B**). IJA recruited regions often associated with cognitive control and attention shifting including bilateral middle frontal gyri, bilateral intraparietal sulci, and dorsal anterior cingulate to a greater extent than RJA. RJA, however, showed a greater response in regions associated with social perception and social cognition including posterior STS, as well as ventral MPFC and posterior cingulate.

EVENT ANALYSES

One possibility for these distinct regions may be due to the different behaviors necessary to perform the initiating vs. responding conditions. For example, in the initiating trials the beginning of the trial is spent searching for the clue and then shifting attention, whereas in RJA the beginning of the trial is spent looking at the experimenter's face for a gaze cue. In order to reduce the differences due to early portions of the trial, we conducted a second analysis in which the period of joint or SA on the mouse was used as an event regressor. During these events, across all conditions, the participant is simply looking at the mouse. What differs across conditions is whether the experimenter is also looking at the mouse (joint vs. solo conditions) and whether the participant initiated or responded to the bid to share attention. Because it is not possible to systematically jitter the time between identification of the cue and the shared attention period, these analyses should not be thought of as strictly isolating the joint attention event. Rather, this method prioritizes the periods of shared attention.
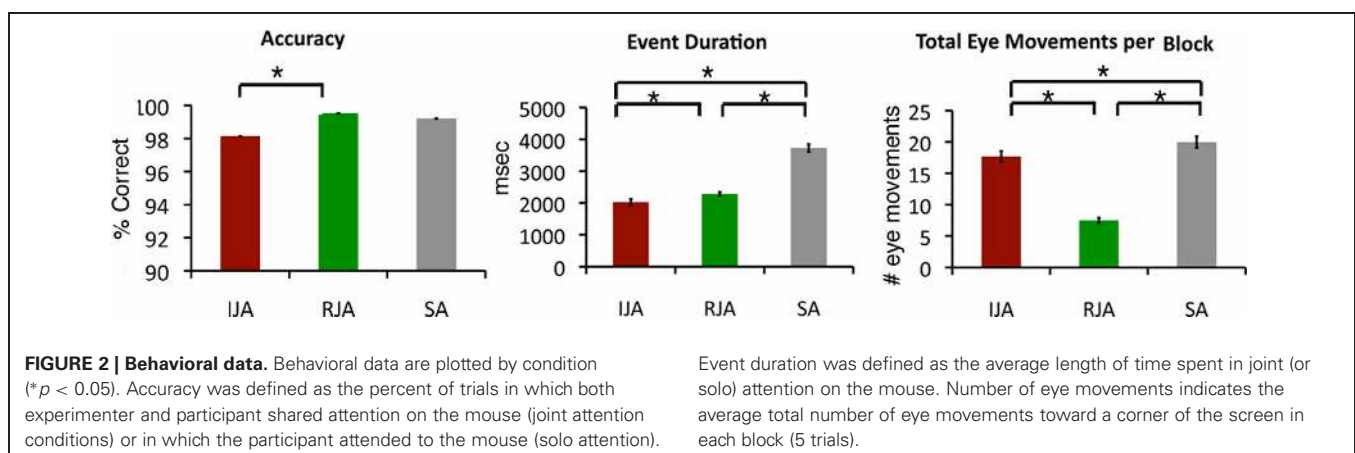

FIGURE 2 | Behavioral data. Behavioral data are plotted by condition (*$p < 0.05$). Accuracy was defined as the percent of trials in which both experimenter and participant shared attention on the mouse (joint attention conditions) or in which the participant attended to the mouse (solo attention). Event duration was defined as the average length of time spent in joint (or solo) attention on the mouse. Number of eye movements indicates the average total number of eye movements toward a corner of the screen in each block (5 trials).

Table 2 | Comparisons between joint and solo attention events.

Contrast region	Hemi	x	y	z	T	k
INITIATING JOINT ATTENTION > SOLO ATTENTION						
Cuneus	L	−16	−68	6	6.50	4896
Inferior parietal lobule	R	52	−50	48	6.23	5632
Caudate	L	−12	−8	18	5.94	464
Inferior frontal gyrus	R	40	26	−20	5.52	592
Intraparietal sulcus	L	−38	−62	54	5.48	4656
Dorsal medial prefrontal cortex	R	2	58	28	5.32	4304
Cuneus	R	14	−72	18	5.14	12560
Posterior superior temporal sulcus	R	58	−50	16	4.09	2528
SOLO ATTENTION > INITIATING JOINT ATTENTION						
Fusiform gyrus	L	−36	−64	−6	5.69	20736
Middle occipital gyrus	L	−32	−86	16	3.93	20736*
Middle occipital gyrus	R	34	−76	6	5.50	4480
RESPONDING TO JOINT ATTENTION > SOLO ATTENTION						
Middle occipital gyrus	L	−16	−102	12	7.34	3968
Temporal pole	L	−36	24	−34	6.40	512
Ventral medial prefrontal cortex	R	2	38	−18	6.36	1250
Dorsal medial prefrontal cortex		0	64	22	4.82	1250*
Middle temporal gyrus	R	54	−60	10	6.15	15088
Posterior superior temporal sulcus	R	54	−38	2	4.93	15088*
Posterior cingulate	R	4	−48	28	5.80	4672
Posterior superior temporal sulcus	L	−50	−68	14	5.79	7088
Lingual gyrus	R	2	−82	−2	5.77	1632
Temporoparietal junction	L	−48	−70	44	5.14	4112
SOLO ATTENTION > RESPONDING TO JOINT ATTENTION						
Middle occipital gyrus	L	−30	−86	12	9.74	53600
Precuneus	R	20	−64	52	7.14	40912
Insula	L	−32	16	−2	6.72	6096
Cerebellum	R	32	−62	−34	5.51	3264
Anterior cingulate gyrus	R	12	14	44	5.45	2416
Supplementary motor area	R	14	2	66	5.38	7920
INITIATING > RESPONDING TO JOINT ATTENTION MASKED BY IJA > SA						
Superior frontal gyrus	R	34	48	32	5.98	
Middle occipital gyrus	L	−32	−86	12	5.47	528
Cuneus	L	−10	−72	4	4.83	2576
Middle frontal gyrus	R	42	42	26	4.80	688
Intraparietal sulcus	L	−36	−56	52	4.71	768
RESPONDING TO > INITIATING JOINT ATTENTION MASKED BY RJA > SA						
Middle occipital gyrus	L	−16	−102	14	8.22	3232
Ventral medial prefrontal cortex	R	4	34	−14	5.34	4480

*Regions were identified using voxel- and cluster-correction of p < 0.05. Coordinates are given in MNI space. T-values from the peak voxel of the cluster and size (k) of the cluster are given. * Indicates a subregion of the cluster listed with the same k.*

In the event-related analysis, *RJA* recruited a greater response than SA in bilateral posterior STS extending into the temporoparietal junction on the left side, posterior cingulate cortex, and ventral and dorsal MPFC. *IJA* as compared to SA revealed a greater response in right posterior STS, bilateral intraparietal sulcus, and dMPFC (**Figure 4A**, **Table 3** for full list).
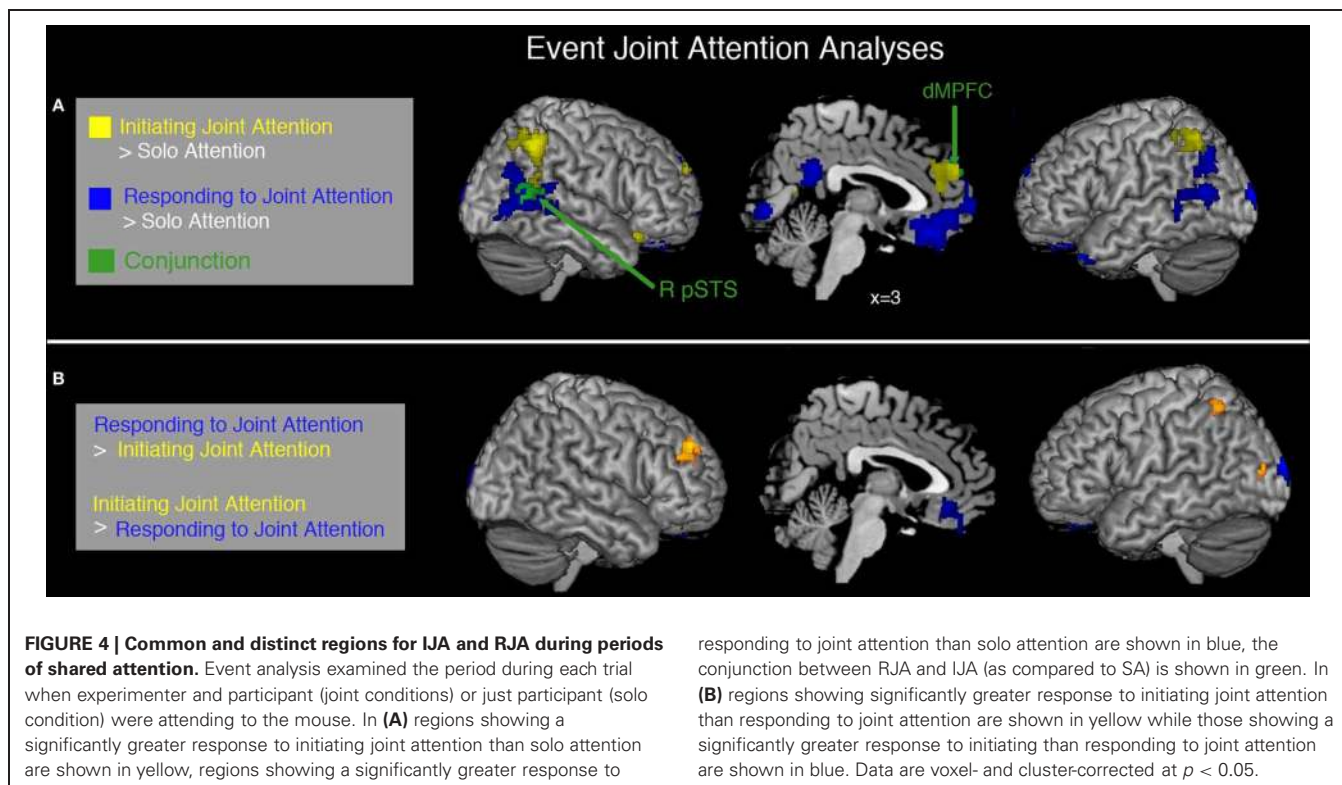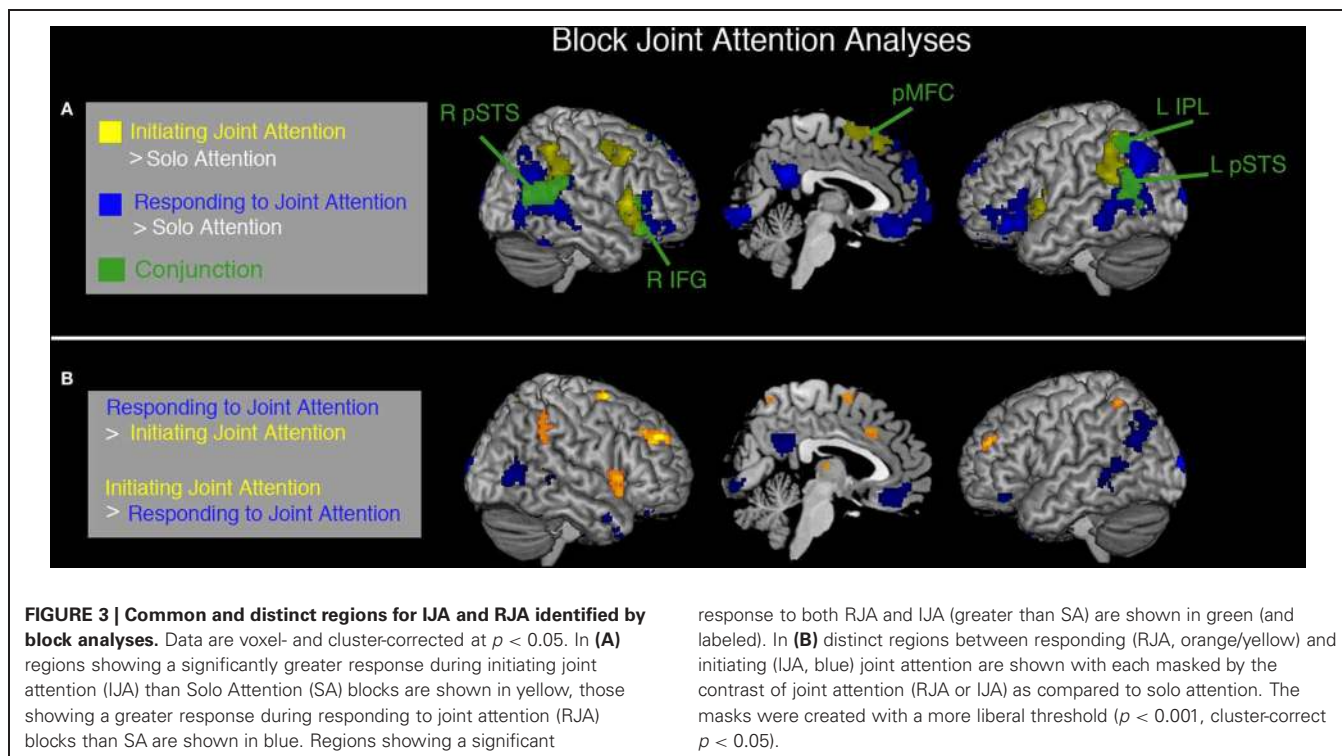
Common regions
Conjunction analyses revealed a greater response to IJA vs. SA and RJA vs. SA within dMPFC and right posterior STS only.

Distinct regions
IJA recruited the right middle frontal gyrus, left inferior parietal lobe, and left occipital regions to a greater extent than RJC. RJA, showed greater activation in ventral MPFC and middle occipital gyrus as compared to IJA (**Figure 4B**, **Table 3** for full list).

FUNCTIONAL CONNECTIVITY ANALYSES
Hierarchical cluster analyses were performed on the pair-wise correlations between each joint attention region (**Figure 5**). Visual inspection of the scree plot suggests that the optimal number of

FIGURE 3 | Common and distinct regions for IJA and RJA identified by block analyses. Data are voxel- and cluster-corrected at $p < 0.05$. In **(A)** regions showing a significantly greater response during initiating joint attention (IJA) than Solo Attention (SA) blocks are shown in yellow, those showing a greater response during responding to joint attention (RJA) blocks than SA are shown in blue. Regions showing a significant response to both RJA and IJA (greater than SA) are shown in green (and labeled). In **(B)** distinct regions between responding (RJA, orange/yellow) and initiating (IJA, blue) joint attention are shown with each masked by the contrast of joint attention (RJA or IJA) as compared to solo attention. The masks were created with a more liberal threshold ($p < 0.001$, cluster-correct $p < 0.05$).

FIGURE 4 | Common and distinct regions for IJA and RJA during periods of shared attention. Event analysis examined the period during each trial when experimenter and participant (joint conditions) or just participant (solo condition) were attending to the mouse. In **(A)** regions showing a significantly greater response to initiating joint attention than solo attention are shown in yellow, regions showing a significantly greater response to responding to joint attention than solo attention are shown in blue, the conjunction between RJA and IJA (as compared to SA) is shown in green. In **(B)** regions showing significantly greater response to initiating joint attention than responding to joint attention are shown in yellow while those showing a significantly greater response to initiating than responding to joint attention are shown in blue. Data are voxel- and cluster-corrected at $p < 0.05$.

clusters is 3. The first cluster was comprised of social-cognitive regions including MPFC (dorsal, ventral, and orbital), posterior cingulate, and bilateral pSTS. These regions corresponded to those recruited during RJC and the conjunction between RJA and IJA. The second cluster contained regions typically associated with voluntary attention orienting (e.g., right and left intraparietal sulcus and middle frontal gyrus) and cognitive control (e.g., supplementary motor area, right inferior frontal gyrus). Most

Table 3 | Seed regions for functional connectivity analyses.

Contrast region	Hemi	x	y	z	T	k
JOINT ATTENTION > SOLO ATTENTION						
Inferior frontal gyrus	R	40	24	−20	6.78	944
Dorsal medial frontal cortex	R	4	60	28	6.26	4992
Posterior superior temporal sulcus	L	−50	−66	14	5.96	13072
Intraparietal sulcus	L	−42	−62	52	4.78	13072*
Cerebellum	L	−30	−80	−32	5.60	3664
Posterior superior temporal sulcus	R	50	−56	10	5.25	3664*
Temporoparietal junction	R	56	−62	24	5.59	20560
Intraparietal sulcus	R	50	−56	50	5.30	2448
Posterior cingulate	R	2	−44	26	5.25	2560
#Temporal pole	L	−48	20	−14	4.83	720
#Superior frontal gyrus	R	20	66	18	4.68	752
#Ventral medial prefrontal cortex	R	2	52	−6	4.67	1440
#Supplementary motor area	R	4	26	62	4.64	1216
#Cuneus	L	−16	−68	6	4.58	1264
#Lingual gyrus	R	20	−60	2	4.47	1200
#Orbitofrontal cortex	R	6	30	−12	4.35	1568
#Middle frontal gyrus	L	−44	6	52	4.12	448
#Calcarine sulcus	R	2	−84	2	4.09	1360
#Middle frontal gyrus	R	48	16	48	4.08	672
#Inferior frontal gyrus	R	60	24	8	4.00	448

*Regions listed above $p < 0.001$, uncorrected were identified using voxel- and cluster-correction of $p < 0.05$. Regions marked with # were identified using $p < 0.001$, uncorrected with cluster-correction at $p < 0.05$, 384 mm^3. Coordinates are given in MNI space. T-values from the peak voxel of the cluster and size (k) of the cluster are given. * Indicates a subregion of the cluster listed with the same k.*

of these regions were recruited specifically in the IJA condition. The third cluster consisted of regions within visual cortex, which were recruited differentially during responding to and IJA conditions when viewed at a liberal threshold ($p < 0.001$, uncorrected)[2].

DISCUSSION

This study examined the neural correlates of both initiating and responding to a bid for joint attention in the context of a face-to-face communicative game. By allowing the participant to play the role of both initiator and responder in a face-to-face social interaction, this paradigm allowed for identification of brain regions during a "meeting of the minds" from both a

[2]In a *post-hoc* analysis, we examined whether networks identified via cluster analyses on functional connectivity data would differ during task periods. Hierarchical cluster analyses with this matrix revealed a broadly similar pattern as that obtained during rest. However, unlike during rest, the right and left posterior superior temporal sulcus (RpSTS and LpSTS) and right temporoparietal junction (RTPJ) were part of the "attention orienting" cluster. Thus, while these posterior temporal regions show more similar functional patterns to midline social-cognitive regions during rest, their fluctuations during joint and solo attention are more similar to regions associated with "attention-orienting and cognitive control." This may reflect integration across these two networks during task performance. However, caution should be noted in interpreting strong differences between rest and task analyses as the optimal cluster number is subjective and based on visual inspection of the scree plot.

first- and second-person perspective (see also Saito et al., 2010; Schilbach et al., 2010). Additionally, this method allowed the participant to coordinate his or her attention with a real person and achieve a state of "knowing together" that both (s)he and the experimenter are attending to the same object—this "knowing together" (also called shared intentionality) allows for true joint attention (Carpenter and Liebal, 2011).

With this method, we identified a number of regions that are involved in joint attention with another person during a live interaction. These included regions that are part of a social-cognitive network, including medial prefrontal regions, posterior cingulate, and bilateral posterior superior temporal sulcus (STS) (Saxe, 2006) as well as those often associated with voluntary attentional control including bilateral intraparietal sulcus, middle frontal gyrus, and inferior frontal gyrus (Corbetta and Shulman, 2002). Consistent with our hypotheses, both common and distinct networks were engaged during joint attention when one was the initiator or the responder (as compared to SA). Whether the participant was playing the role of initiator or responder during joint attention, the dMPFC and right posterior STS were engaged to a greater extent during periods of shared attention than SA on the mouse, suggesting these regions form part of a core neural system in joint attention processes. These core regions are part of the social-cognitive network, as identified using resting-state connectivity analyses. Thus, these data suggest a key role of the social-cognitive network in both IJA and RJA.

FIGURE 5 | Regions identified in the JA > SA contrast ($p < 0.001$, cluster-corrected $p < 0.05$) are displayed on a reference brain in (A). Spheres surrounding the peak coordinates from each region were used as seed regions in the connectivity analyses. These spheres are shown on a reference brain color-coded by the cluster in which they were identified. Clusters are labeled social-cognitive (pink), attention and control (green), and visual (blue) based on the functions associated with the set of regions within each cluster. In **(B)** a correlation matrix displays the region–region correlation values from the resting baseline periods with blue colors representing negative correlation and red/yellow positive. A dendrogram shows the results of the hierarchical cluster analysis and the scree plot depicts the dissimilarity value plotted by number of clusters identified.

REGIONS OF MEDIAL PREFRONTAL CORTEX PLAY DIFFERENTIAL ROLES DURING JOINT ATTENTION

The dMPFC was recruited during RJA to a greater extent than SA in both block and event-related analyses. This region was also recruited more during IJA events as compared to SA events in the event-related analyses. Previous research has identified the dMPFC as associated with perception of a social partner (Kampe et al., 2003; Schilbach et al., 2006; Pierno et al., 2008), making

judgments about others and oneself (Mitchell et al., 2006; Moran et al., 2011), reasoning about others' mental states (Saxe and Kanwisher, 2003) and coincidental shared attention on an object with a virtual character (Williams et al., 2005). This shared self- and other-representation led some to suggest that this region may be involved in "triadic" interactions (Saxe, 2006) and a "meeting of the minds" (Amodio and Frith, 2006). These data, and converging evidence from other studies (Schilbach et al., 2010),

provide more direct support for this hypothesis that the dMPFC is involved in shared attention between you, me, and this (Saxe, 2006).

The ventral MPFC, on the other hand, was selectively responsive to responding to a bid for joint attention, but not initiating (in both block and event analyses). The selectivity of the ventral MPFC (vMPFC) in RJA is consistent with a previous study (Schilbach et al., 2010), however, the cluster in the current study extended more inferiorly into medial orbitofrontal cortex (OFC). The medial OFC has been associated with reward expectancies based on an associated cue (e.g., Elliot et al., 2000; Kahnt et al., 2010). In the current paradigm, the gaze shift from the experimenter helped the participant achieve the goal of catching the mouse with less effort on the part of the participant. Accuracy is higher in this condition and the duration of joint attention events are longer. Thus, experimenter's gaze cue may have signaled the anticipation of a reward (i.e., successful trial completion). This paradigm is distinct from previous experimental paradigms of joint attention (Saito et al., 2010; Schilbach et al., 2010) in that the participant and experimenter had a joint goal and needed to use gaze cues to help each other achieve a joint goal—thus, in this context, assistance from a partner via gaze cues may be more rewarding. Without corroborating behavioral reports though this conclusion remains speculative.

One alternative explanation for ventral MPFC activation during RJA is that this condition required less goal-directed attention (as reflected in greater accuracy and fewer eye movements). These differences could have allowed for greater "default mode" activity within the medial prefrontal cortex (e.g., Grecius and Menon, 2004). Given the consistency between our findings and previous studies of joint attention (e.g., Williams et al., 2005; Schilbach et al., 2010), which did not have differences in accuracy or total number of eye movements, we believe this interpretation is unlikely. However, future designs should match accuracy and total number of eye movements across conditions to be able to tease out the specific contributions of the ventral medial prefrontal cortex to joint attention.

RIGHT POSTERIOR STS IS INVOLVED IN BOTH RESPONDING TO AND INITIATING JOINT ATTENTION

In the current study, the region that was most robustly engaged during both RJA and IJA across both block and event-related analyses was the right posterior STS, suggesting that like the dMPFC, it plays a core role in both initiating and responding to joint attention. The STS is sensitive to the direction of another person's gaze and attention as well as the intention behind a gaze shift (Pelphrey et al., 2003; Nummenmaa and Calder, 2009). Greater activation is seen in the pSTS when a gaze shift occurs in a self-relevant context, for example in the context of a social interaction (Morris et al., 2005; Redcay et al., 2010). Additionally, two previous studies[3] have revealed a key role of this region in RJA (Materna et al., 2008; Redcay et al., 2010). Thus, we predicted, and found, that the pSTS would be recruited during RJA. Interestingly, IJA also recruited the right pSTS. These findings suggest a broader

role of the pSTS beyond simply interpreting another person's gaze cues; however, a leaner interpretation is that gaze shifts alone, which were present in both IJA and RJA, drove the response in the pSTS. One possibility is that the pSTS is differentially engaged during the coordination of attention (using gaze or other biological motion cue) while the dMPFC is more engaged during the sharing of attention. Given that coordination always immediately precedes sharing it is challenging to disentangle coordinating vs. sharing attention using fMRI methods, which have poor temporal resolution.

While the pSTS region has been reported in some studies examining joint attention (Materna et al., 2008; Redcay et al., 2010), others have not found evidence for a role of the pSTS (Williams et al., 2005; Saito et al., 2010; Schilbach et al., 2010). These discrepancies are likely due to the choice of control condition for the joint attention conditions. We used a control condition in which the experimenter disengaged, so the participant's attention was no longer related to the experiment's attention. In other studies, in the nonjoint attention condition participants are instructed to look in the *opposite* direction of the experimenter's gaze shift. In other words, they are still cued by another person's gaze but in the opposite direction. If the pSTS is recruited for coordinating gaze with another person, the anti-contingent control condition may still elicit activity in the pSTS, compared to a non-contingent condition[4]. In a previous fMRI study (Materna et al., 2008), the bilateral posterior STS were selectively recruited for joint attention events. In that study gaze shifts were present in both joint and non-joint attention conditions, but only in the joint conditions were the gaze shifts communicative—adding support for a role of the STS in coordinating attention through gaze cues. An exciting future direction is to determine the extent to which the STS is involved in coordination of attention through visual cues explicitly or whether this region is involved in coordination of attention via amodal communicative cues (e.g., auditory cues through spoken language) (e.g., Redcay, 2008; Noordjiz et al., 2009)

FRONTAL-PARIETAL ATTENTION REGIONS ARE RECRUITED DURING INITIATING JOINT ATTENTION

Initiating, but not responding to, joint attention differentially recruited portions of the fronto-parietal attention network including the intraparietal sulcus and middle frontal gyrus which have been shown to be involved in voluntary shifts of spatial attention attention (Corbetta and Shulman, 2002; Kincade et al., 2005). IJA requires greater voluntary attention than RJA. Note that IJA also involved more eye movements than RJA. Nevertheless, the observed activation is unlikely to be due to more frequent gaze shifts, because participants made more eye movements in SA control trials than during IJA, but these regions showed greater activity during IJA than SA control trials. Involvement of frontal and parietal cortices is therefore consistent with previous suggestions that a mechanism for goal-directed attention orienting is a necessary component of IJA

---

[3]In one study (Redcay et al., 2010), eight of the participants were the same as the current study.

[4]In fact, in pilot versions of the current task in which we included this same control condition, participants found it very difficult, if not impossible, to do so in the context of a live face-to-face interaction.

(Mundy and Jarrold, 2010). Further, these data reveal that goal-directed attention orienting in a social joint attention context recruits frontal-parietal regions to a greater extent than just goal-directed attention orienting without a social context (i.e., SA).

While both social-cognitive and goal-directed attention systems were recruited during IJA, these regions do not seem to part of the same functional network. Functional connectivity and hierarchical clustering analyses on data during a no-task resting baseline revealed clustering of joint attention regions into three networks: a social-cognitive, attention orienting, and visual network. The attention orienting network was recruited to a greater extent during IJA than responding, whereas the regions involved in RJA were part of the social-cognitive network that was overlapping with IJA.

While the current data cannot directly inform the development of these behaviors, they offer support for a core role of the social-cognitive system (e.g., pSTS and MPFC) in both responding and IJA behaviors, at least in adults. We find it intriguing that a study of 5-month-old infants revealed selective recruitment of the dMPFC during RJA (Grossmann and Johnson, 2010). This study used functional Near Infrared Spectroscopy (fNIRS) which has lower spatial resolution than fMRI, but, nonetheless, suggests an early role of dMPFC in the development of joint attention. That study only examined activation over the dMPFC; so, early involvement of other regions (e.g., the pSTS) in joint attention at 5-months cannot be determined. Interestingly, EEG studies in the second year of life reveal a positive correlation between alpha coherence (an index of functional maturation) over left frontal and left and right central electrode sites and IJA behaviors (Mundy et al., 2000). These scalp locations could correspond to regions of the social-cognitive and attention orienting systems. Thus, one possibility that remains speculative is that portions of the social-cognitive system underlie the early development of IJA and RJA but the emergence of IJA may be due to the later development of a frontal network involved in attention orienting and cognitive control.

LIMITATIONS AND FUTURE DIRECTIONS

Our protocol was designed to capture the communicative dimension of natural joint attention interactions. Bids for joint attention via gaze cues were communicative and in the service of achieving a shared goal (i.e., catch the mouse). On the other hand, our paradigm lacked the motivational aspect of natural joint attention. Specifically, in our paradigm the endogenous desire to share attention is not necessarily invoked. Participants are instructed that the goal is to share attention on the mouse with the experimenter (or alone in the case of SA). Future studies tackling the

spontaneous and communicative aspects of joint attention will prove fruitful in elucidating the neural correlates of this pivotal behavior.

A final limitation is that in this interactive task events of interest occur on the timeline of real-world interactions, making them very difficult to isolate in time. For example, the appropriate randomized jitter between a gaze shift and shared attention could not be introduced while keeping the behavior naturalistic. Future paradigms using converging methods with better temporal resolution, such as event-related potentials or magnetoencephalography, could provide insights into shared and distinct mechanisms underlying the perception of gaze shifts, eye contact, and shared attention in a naturalistic joint attention context.

Despite inherent difficulties in the study of real-time social interactions, we are optimistic that this new era of interactive social neuroscience will bring converging evidence from a diverse set of paradigms. The current study, similar to Schilbach et al., 2010, reported a key role for the dMPFC in real-time shared attention for both the initiator and responder. Furthermore, IJA, specifically, recruits regions associated with attention orienting and cognitive control systems. Finally, functional connectivity analyses demonstrated that these joint attention interactions draw on multiple overlapping and distinct networks, including social-cognitive, attention orienting, and visual networks. This convergence of information from these and subsequent studies will provide for significant advances in our understanding of how we achieve a fundamental and critical aspect of human behavior and survival: namely, coordinated social interactions.

REFERENCES

Amodio, D., and Frith, C. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat. Vis.* 10, 433–436.

Bristow, D., Rees, G., and Frith, C. (2007). Social interaction modifies neural response to gaze shifts. *Soc. Cogn. Affect. Neurosci.* 2, 52–61.

Brooks, R., (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal,

growth curve modeling study. *J. Child. Lang.* 35, 207–220.

Brooks, R., and Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Dev. Sci.* 6, 535–543.

Carpenter, M., and Liebal, K. (2011). "Joint attention, communication,

and knowing together in infancy," in *Joint Attention: New Developments in Psychology, Philosphy of Mind, and Social Neuroscience*, ed A. Seemann (Cambridge, MA: MIT Press), 159–182.

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social

cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63 (4, Serial No. 255).

Catell, R. B. (1966). The scree test for the number of factors. *Multivariate Behav. Res.* 1, 245–276.

Charman, T. (2000). "Theory of mind and the early diagnosis of autism," in *Understanding Other Minds: Perspectives from Autism and Developmental Cognitive Neuroscience*, 2nd Edn. eds S. Baron-Cohen and H. Tager-Flusberg (Oxford: Oxford University Press), 422–441.

Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 315–324.

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173.

Elliot, R., Dolan, R. J., and Frith, C. D. (2000). Dissociable functions in the medial lateral orbitofrontal cortex: evidence from human neuroimaging studies. *Cereb. Cortex* 10, 308–317.

Frischen, A., Bayliss, A. P., and Tipper, S. P. (2007). Gaze cueing of attention: social cognition, visual attention, and individual differences. *Psychol. Bull.* 133, 694–724.

Grecius, M. D., and Menon, V. (2004). Default-mode activity during a passive sensory task: uncoupled from deactivation but impacting activation. *J. Cogn. Neurosci.* 16, 1484–1492.

Grossmann, T., and Johnson, M. H. (2010). Selective prefrontal cortex responses to joint attention in early infancy. *Biol. Lett.* 6, 540–543.

Kahnt, T., Heinzle, J., Park, S. Q., and Haynes, J.-D. (2010). The neural code of reward anticipation in human orbitofrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6010–6015.

Kampe, K., Frith, C., 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.

Kincade, J. M., Abrams, R. A., Astafiev, S. V., Shulman, G. L., and Corbetta, M. (2005). An event-related functional magnetic resonance imaging study of voluntary and stimulus-driven orienting of attention. *J. Neurosci.* 25, 4593–4604.

Materna, S., Dicke, P. W., and Thier, P. (2008). Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. *J. Cogn. Neurosci.* 20, 108–119.

Mitchell, J. P., Macrae, C. N., and Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron* 50, 655–663.

Morales, M. (2000). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *J. Appl. Dev. Psychol.* 21, 283–298.

Moran, J. M., Lee, S. M., and Gabrieli, J. D. E. (2011). Dissociable neural systems supporting knowledge about human character and appearance in ourselves and others. *J. Cogn. Neurosci.* 23, 2222–2230.

Morris, J. P., Pelphrey, K. A., and McCarthy, G. (2005). Regional brain activation evoked when approaching a virtual human on a virtual walk. *J. Cogn. Neurosci.* 17, 1744–1752.

Mundy, P., and Jarrold, W. (2010). Infant joint attention, neural networks and social cognition. *Neural Netw.* 23, 985–997.

Mundy, P., and Newell, L. (2007). Attention, joint attention, and social cognition. *Curr. Dir. Psychol. Sci.* 16, 269–274.

Mundy, P., Block, J., Delgado, C., Pomares, Y., Vaughan Van Hecke, A., and Parlade, M. (2007). Individual differences and the development of joint attention in infancy. *Child Dev.* 78, 938–954.

Mundy, P., Card, J., and Fox, N. (2000). EEG correlates of the development of infant joint attention skills. *Dev. Psychobiol.* 36, 325–338.

Nelson, P. B., Adamson, L. B., and Bakeman, R. (2008). Toddlers' joint engagement experience facilitates preschoolers' acquisition of theory of mind. *Dev. Sci.* 11, 847–852.

Noordjiz, 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

Nummenmaa, L., and Calder, A. J. (2009). Neural mechanisms of social attention. *Trends Cogn. Sci.* 13, 135–143.

Osório, A., Martins, C., Meins, E., Martins, E. C., and Soares, I. (2011). Individual and relational contributions to parallel and joint attention in infancy. *Infant Behav. Dev.* 34, 515–524.

Pelli, D. (1997). The video toolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* 10, 437–442.

Pelphrey, K. A., Singerman, J. D., Allison, T., and McCarthy, G. (2003). Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia* 41, 156–170.

Pierno, A. C., Becchio, C., Turella, L., Tubaldi, F., and Castiello, U. (2008). Observing social interactions: the effect of gaze. *Soc. Neurosci.* 3, 51–59.

Redcay, E. (2008). The superior temporal sulcus performs a common function for speech and social perception: implications for the emergence of autism. *Neurosci. Biobehav. Rev.* 38, 123–142.

Redcay, E., and Saxe, R. (in press). "Do you see what I see? The neural bases of joint attention," in *Agency and Joint Attention*, eds H. S. Terrace and J. Metacalfe (New York, NY: Oxford University Press).

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

Saito, D. N., Tanabe, H. C., Izuma, K., Hayashi, M. J., Morito, Y., Komeda, H., Uchiyama, H., Kosaka, H., Okazawa, H., Fujibayashi, Y., and Sadato, N. (2010). "Stay tuned": inter-individual neural synchronization during mutual gaze and joint attention. *Front. Integr. Neurosci.* 4:127. doi: 10.3389/fnint.2010.00127

Saxe, R. (2006). Uniquely human social cognition. *Curr. Opin. Neurobiol.* 16, 235–239.

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind." *Neuroimage* 19, 1835–1842.

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., Shah, N. J., Fink, G. R., and Vogeley, K. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *J. Cogn. Neurosci.* 22, 2702–2715.

Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. Jon., Fink Gereon, R., and Vogeley, K. (2006). Being with virtual others: neural correlates of social interaction. *Neuropsychologia* 44, 718–730.

Slaughter, V., and McConnell, D. (2003). Emergence of joint attention: relationships between gaze following, social referencing, imitation, and naming in infancy. *J. Genet. Psychol.* 164, 54–71.

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain Mapp.* 17, 143–155.

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.

Vaughan Van Hecke, A., Mundy, P., Acra, C., Block, J. J., Delgado, C. E. F., Parlade, M., Meyer, J., and Pomares, Y. B. (2007). Infant joint attention, temperament, and social competence in preschool children. *Child Dev.* 78, 53–69.

Williams, J. H. G., Waiter, G. D., Perra, O., Perrett, D. I., and Whiten, A. (2005). An fMRI study of joint attention experience. *Neuroimage* 25, 133–140.