



Published in final edited form as:

Neuroimage. 2010 May 1; 50(4): 1639–1647. doi:10.1016/j.neuroimage.2010.01.052.

Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience

Elizabeth Redcay^{1,*}, David Dodell-Feder¹, Mark J. Pearrow¹, Penelope L. Mavros¹, Mario Kleiner², John D.E. Gabrieli¹, and Rebecca Saxe¹

¹ Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

² Max Planck Institute for Biological Cybernetics

Abstract

Cooperative social interaction is critical for human social development and learning. Despite the importance of social interaction, previous neuroimaging studies lack two fundamental components of everyday face-to-face interactions: contingent responding and joint attention. In the current studies, functional MRI data were collected while participants interacted with a human experimenter face-to-face via live video feed as they engaged in simple cooperative games. In Experiment 1, participants engaged in a live interaction with the experimenter (“Live”) or watched a video of the same interaction (“Recorded”). During the “Live” interaction, as compared to the Recorded conditions, greater activation was seen in brain regions involved in social cognition and reward, including the right temporo-parietal junction (rTPJ), anterior cingulate cortex (ACC), right superior temporal sulcus (rSTS), ventral striatum, and amygdala. Experiment 2 isolated joint attention, a critical component of social interaction. Participants either followed the gaze of the live experimenter to a shared target of attention (“Joint Attention”) or found the target of attention alone while the experimenter was visible but not sharing attention (“Solo Attention”). The right temporo-parietal junction and right posterior STS were differentially recruited during Joint, as compared to Solo, attention. These findings suggest the rpSTS and rTPJ are key regions for both social interaction and joint attention. This method of allowing online, contingent social interactions in the scanner could open up new avenues of research in social cognitive neuroscience, both in typical and atypical populations.

Introduction

Humans seek social interactions, from birth onwards. Within only two months after birth, typically developing infants prefer the subtle patterns of contingency in face-to-face interactions, including turn-taking and correlated affect (Gergely and Watson, 1999; Murray and Trevarthen, 1985). By 9-months infants are able to follow another person’s gaze to a location outside of their visual field: a key first step in establishing joint attention (for review, Moore 2007). Joint, also called “triadic”, attention provides a platform by which two or more people coordinate and communicate their intentions, desires, emotions, beliefs, and/or knowledge about a third entity (e.g. an object or a common goal) (Tomasello et al., 2005). Joint attention is distinct from shared attention or mutual gaze, in which two people share attention

*Corresponding Author: Elizabeth Redcay, Massachusetts Institute of Technology, 43 Vassar St, Rm 46-4033D, Cambridge, MA 02139, Phone: 617-324-3981, Fax: 617-324-5311, redcay@mit.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

by looking at each other, rather than coordinating their attention on a third entity. Despite the centrality of contingent responding and joint attention in human social interactions, the neural mechanisms of these key features of social interactions remain understudied.

Previous research has investigated the neural bases of various aspects of social interactions in adults via several approaches: (1) the participant observes a recorded social interaction between two other people in a story, cartoon, or movie (Iacoboni et al., 2004; Pierno et al., 2008; Saxe and Kanwisher, 2003; Walter et al., 2004) (2) the participant plays an online game with an alleged, but invisible, human partner (Fukui et al., 2006; Gallagher et al., 2002; Kircher et al., 2009; Rilling et al., 2002; Rilling et al., 2004) or (3) the participant views a virtual character who shifts gaze towards or away from the participant (Pelphrey et al., 2004b; Schilbach et al., 2006). These approaches provide important indications of possible neural mechanisms for social interaction, but are missing key components of everyday social interactions: contingent responding and joint attention.

These two features of social interaction are difficult to examine with functional MRI due to several methodological challenges. The first challenge is common to examining both contingent interaction and joint attention. That is, to create live, face-to-face contact with minimum temporal delay, while at least one of the people is lying inside the bore of a scanner. To address this challenge, we used a dual-video presentation to allow two people to interact face-to-face with minimal temporal delays. A second challenge, which is specific to identifying the neural correlates of a live interaction, is to design a control condition that would capture the visual complexity of a live interaction, thus isolating the social, contingent aspect of the interaction.

To address these challenges, inspired by a paradigm from infant research (Murray and Trevarthen, 1985), we compared a live social interaction to a recorded video of the same interaction in Experiment 1. By comparing a live interaction to a recording of the same interaction, we controlled for the perception of a person speaking and moving. The key difference between the two conditions is thus contingency and/or self-relevance: only during the live condition are the participant's and the experimenter's actions contingent on one another. Thus, the live condition should differentially recruit brain regions that are sensitive to interpreting another person's actions and speech in a self-relevant context, like an online face-to-face social interaction. We hypothesize that these regions will include those involved in reasoning about another person's actions or intentions and representing another's mental state, including dorsal medial prefrontal cortex (dmPFC), right posterior superior temporal sulcus (rpSTS), and right temporo-parietal junction (rTPJ) (review, Saxe 2006). We additionally predicted that contingent interaction would recruit regions involved in attention, or goal-directed tasks, including dorsal anterior cingulate cortex (dACC) (Dosenbach et al., 2007).

A full understanding of the contributions of multiple brain regions to social interaction will thus require breaking social interaction into its component parts. In Experiment 2, we utilized the same live set-up to isolate the neural bases of one component of a social interaction, namely joint attention. In the "joint attention game", two players coordinate their visual attention in order to jointly discover a target (joint attention); in the control condition, the two players deployed their attention independently (solo attention). Thus, both conditions involved 'face-to-face' interactions, but only one (joint attention) required coordinating attention with another person. Given previous findings of right posterior STS (Materna et al., 2008) and dorsal medial prefrontal cortex (Williams et al., 2005) involvement in joint attention, we predicted these regions would be selectively recruited during joint attention, as compared to solo attention, trials. Further, as joint attention is a key social component of a live interaction, we predicted that parts of the social brain areas (e.g. pSTS, RTPJ, and dmPFC), but not attention-related areas,

identified in Experiment 1 would be selectively recruited during joint as compared to solo attention.

Methods

Participants

Participants for both experiments were recruited from the Boston community. Participants had never met the experimenter prior to the scan date. All participants gave informed written consent as approved by the MIT human subjects committee (COUHES) and were compensated for their participation. Participants were excluded if they self-reported any psychiatric or neurological diagnosis on a medical questionnaire. In addition, all participants completed a Social Communication Questionnaire (Berument et al., 1999) and scored below 15 (i.e. in the normal range). For Experiment 1, data were collected from 23 participants. Seven were excluded for the following reasons: head movement exceeding the criteria detailed below ($n=4$), falling asleep (2), and equipment malfunction (1). The final sample consisted of 16 adults (7 males, age 18–29 years, mean 22.8 ± 3.1 years). For Experiment 2, data were collected from 14 participants. One was excluded for exceeding the motion criteria, leaving a sample of 13 adults (6 males, age 18–29 years, mean 22.6 ± 3.5 years).

Task Designs

The goal of Experiment 1 was to identify brain regions in which the response to another person's actions and speech was increased when those actions were contingent on the subject's own actions. To isolate social contingency in the experimental design, subjects participated in three conditions: (1) "Live", in which they interacted with the experimenter via a live video feed, (2) "Recorded-Same" (Rec-S), in which a video of the experimenter from a previous live interaction with that *same* subject was replayed and (3) "Recorded-Different" (Rec-D), in which a video of the experimenter from a previous live interaction with a *different* subject was replayed. The Recorded-Same condition was included because the audio and visual inputs were identical to those in the Live condition. The Recorded-Different condition was included to ensure that any differences between Live and Recorded-Same conditions were not due to repetition suppression of the neural response. Participants were told that they would interact with a live experimenter via video feed during "live" conditions and would view a recorded version of the same interaction during "recorded" conditions (Figure 1a). A colored frame around the video (green for live, red for recorded) and a label informed the participant of the condition in each episode. The interactions were scripted such that events, timing, and dialogue were naturalistic but also consistent within and across scan sessions (Supplemental Information). Participants were instructed to follow the experimenter's instructions to look at a particular object when prompted in both live and recorded conditions, but that the experimenter would not be able to see him or her in the recorded conditions. In order to maintain participant interest, and reduce cross-trial habituation of the neural response, two kinds of interaction were alternated: matching events (participants were asked to sort toys into bins), and showing events (participants chose a bin, and the experimenter showed them the toy in that bin). Each run consisted of 6 video events (40 s each, 2 per condition, in counterbalanced order) and 3 rest events (20 s each, at the beginning, middle, and end of the run). Conditions were semi-counterbalanced given the constraint that a live interaction must precede its subsequent replay (Recorded-Same). Data from four runs were collected per participant.

For Experiment 2, Joint Attention (JA) subjects participated in a game called "Catch the Mouse" (Figure 1b). Participants were told they would be playing a game with a live experimenter. The goal was to catch the mouse by looking at the location the mouse was "hiding". At the beginning of each trial, the mouse was "hiding" in one of four houses (there is one house in each corner of the screen). During the Joint Attention (JA) condition, the

experimenter was given a cue to the location of the mouse (a tail appeared). The participant followed the experimenter's gaze to the correct house. When both subjects looked to the correct house, the mouse was "caught" and thus appeared. During the Solo Attention (SA) condition, the subject received the cue and had to look at the house corresponding to the cue in order for the mouse to appear. The experimenter remained on the screen and closed and opened her eyes at the start and end of each trial in order to keep the amount of perceived eye movement constant across both conditions. The subject was informed that the experimenter was not "playing the game" during these trials. For each condition, a second experimenter (not visible to the subject) monitored the subject's gaze. When the subject looked at the correct target, the experimenter pushed the space bar to trigger the mouse to appear. Trials were presented in blocks (5 trials per block). At the start of each block, 4 seconds of instructions were displayed to let them know their role in the game for the next 5 trials. The delay before cue onset was jittered between 0 and 1 second across trials. Total trial time was 6 seconds. Each run consisted of 6 blocks (30 s each, 2 per condition, in counterbalanced order) and 3 rest events (20 s each, at the beginning, middle, and end of the run). A third condition was included in the game, but is not analyzed here. Data from four runs were collected per participant.

Audio, Video, and Computing Systems Set-up

The experimental paradigms for live, face-to-face interaction between an experimenter and the subject required a flexible audio/video routing and recording system that integrated with the stimulus presentation software and scanner triggering interface. Our system was built from off-the-shelf hardware and a mixture of off-the-shelf and custom software. A diagram of these details can be found in Supplementary Information Figure S1.

For subject monitoring, an iScan camera positioned at the head-end of the scanner was used. This camera had built-in infrared illuminators and was aimed at the headcoil-mounted 45° mirror. The camera provided a clear view of one of the subject's eyes to allow for accurate monitoring of the subject's gaze shifts. Live, online video sessions, as well as recorded sessions, were routed to the subject screen via a rear-projection LCD system located in an adjacent room.

Video set-up for Experiment 1 and Experiment 2 were nearly identical. The video feed from the eye-tracking camera was intercepted at the eye tracking control PC, where the VGA output of the PC was split by way of an Avermedia scan converter. The VGA signal was passed through to the monitor and also converted to an NTSC video feed over the S-Video out port on the scan converter. This was fed into a Canopus ADVC-110 Video-to-DV converter and then to the stimulus presentation laptop. The subject's responses were monitored via a DV monitoring application.

Experiment 1 required audio to be presented to the subject that was synchronized with live and recorded video presentations. Stimulus audio was piped into the bore by way of MRI-compatible Sensimetrics earphones. Audio from the live experimenter, from the freshly-recorded video of live within-session presentations, and from stock recordings of the experimenter were mixed so that there was no discernible difference between audio levels or quality of the various modes. Audio output from the stimulus presentation laptop was mixed with output from a Shure SM93 lavalier microphone via a Mackie VLZ-1604 audio mixer. This mixed audio feed was then fed into the stimulus presentation laptop via an M-Audio MobilePre USB audio-to-USB interface. The same audio signal was also sent via a second bus on the Mackie mixer to the bore's audio system, which in turn fed the Sensimetrics earphones. Audio levels from the live (microphone) feed and the recorded video feed were compared before each participant.

All stimuli were programmed in Matlab 7.8 using Psychophysics extensions toolbox (PTB-3) (Brainard, 1997; Pelli, 1997) on an Apple MacbookPro running OS \times 10.5.6. For Experiment

2, dual video recording was utilized to obtain a recording of both the experimenter video and the subject video during the “catch the mouse” game.

Data Acquisition & Analysis

Data were collected on a 3T Siemens Tim Trio scanner at the Athinoula A. Martinos Imaging Center at McGovern Institute for Brain Research at Massachusetts Institute of Technology. For both experiments, whole-brain T2*-weighted gradient echo-planar images were collected at a resolution of 3.1×3.1×4.0mm voxels (TR=2sec, TE=30ms, slices=32). These sequences used Siemens PACE online motion correction for movement < 8 mm. For a subset of the data from Experiment 2, a multi-PACE sequence adjusted for movement 3 times per volume acquisition. T1-weighted structural images were collected with 128 slices in the axial plane (TE=3.39 ms, TR=2530 ms, 1.3 mm isotropic voxels).

Data were analyzed using SPM5 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm5/>) and in-house software. Data from all runs were realigned to the mean volume of the first run using a least squares approach with a 6-degree rigid spatial transformation. Images were stereotactically normalized to Montreal Neurological Institute (MNI) space and spatially smoothed (fwhm = 6 mm). Data were high pass filtered at 128 Hz, and inspected for motion artifact using an artifact detection toolbox (http://www.nitrc.org/projects/artifact_detect/). A timepoint which deviated from the previous one by greater than 3 SD, 1 mm, or .01 degrees was marked as an outlier timepoint. A functional run containing greater than 20% outlier time points was excluded from the analysis. Subjects who had more than 1 run with 20% outliers were excluded from the analyses. These conditions led to exclusion of 4 participants in Experiment 1, and 1 participant in Experiment 2.

General Linear Models were used to estimate the parameters for each condition. Both models included conditions of interest (i.e. Live, R-Same, R-Diff for Experiment 1 and JA & SA for Experiment 2) and their temporal derivatives to account for shifts in the hemodynamic response. Six directions of motion parameters from the realignment step as well as outlier time points were included as nuisance regressors. For Experiment 1, a nuisance regressor containing the timeseries extracted from a region of interest in each individual’s left frontal eye field (FEF) was included to account for activity related to eye movements. Individual FEF regions were identified in each individual from the contrast of all conditions as compared to baseline. The FEF ROI comprised all significant activity within a 10 mm sphere centered at the nearest local maximum to an FEF coordinate [−30, −8,50] identified in a meta-analysis as maximally involved in saccades and attention-related tasks (Grosbras et al., 2005).

Whole brain two-tailed t-tests were run for each main effect of interest and comparison of interest for each experiment separately. For Experiment 1, the contrasts of interest were Live vs. Rec-D, Live vs Rec-S, Live vs. Rec [Live – (Rec-S + Rec-D)] and Rec-S vs Rec-D. For Experiment 2, the contrast of interest was Joint Attention vs. Solo Attention. Whole brain contrasts were corrected for multiple comparisons at both the voxel and cluster level (p<.05) using non-parametric permutation analyses (Statistical non-parametric mapping toolbox, SnPM5b). Additional analyses were conducted to examine the effect of age and gender on brain activation. These are reported in Supplementary Information.

In order to examine the degree to which Joint Attention is a component of the Social Interaction network, regions of interest (ROIs) were identified from the group contrast Live-Rec. Note that native space analyses could not be used because the subjects in the two experiments were largely non-overlapping. The region of interest comprised all voxels showing significant activation within a 3mm radius sphere centered on the coordinate with the peak intensity for that region. Contrast values for the JA and SA conditions in Experiment 2 were scaled by the global mean signal to give a measure of percent signal change. Percent signal change values

were then extracted within each ROI for both Joint Attention and Solo Attention conditions. Because the participants in the Joint Attention experiment were largely distinct from those in the Social Interaction Experiment, the same ROIs from the group analysis in Experiment 1 were used to extract individual data from Experiment 2. Regions of interest from the Live-Rec contrast were grouped into social, reward, and attention brain areas. A repeated-measures ANOVA with condition (JA, SA) as the repeated measure and network (social, reward, attention) as the non-repeated measure, with individual ROIs nested under each network, was run to identify which type of brain areas showed differential modulation during joint attention.

Results

Experiment 1: Social Interaction Experiment

Whole brain comparisons between Live and Recorded conditions reveals significantly greater Blood Oxygenated Level Dependent (BOLD) signal for the Live condition in a large number of brain regions involved in social cognition, attention, and reward processing (See Figure 2 and Table 1). Regions typically identified in studies of social perception and social cognition (Allison et al., 2000; Saxe, 2006) include the right posterior superior temporal sulcus (rpSTS) [(54, -48, 12), $t = 7.22$], right temporoparietal junction (rTPJ) [(60, -46, 24), $t = 5.23$], and right anterior superior temporal sulcus (raSTS) [(48, 12, -30), $t = 7.08$]. Brain regions which can be broadly construed to be involved in attention (e.g. goal-directed and visual attention) included the dorsal anterior cingulate cortex (dACC) [(-4, 22, 26), $t = 9.16$], left cerebellum [(-42, -56, -28), $t = 6.92$], and right cuneus/lingual gyrus [(10, -74, -12), $t = 9.72$] (Dosenbach et al., 2007). Regions typically identified during tasks involving reward (Walter et al., 2005) included regions within ventral striatum, including right [(18, 14, -8), $t = 4.85$] and left putamen [(-12, 12, -4), $t = 4.76$] and right amygdala [(20, -2, -18), $t = 4.44$]. Recorded conditions revealed greater activation than live conditions in motor areas, including precentral gyrus [L:(-38, -20, 58), $t = 6.11$; R:(46, -24, 58), $t = 5.22$] and supplementary motor area (SMA) [(8, -18, 64), $t = 5.55$].

A direct contrast of Recorded-Different (novel video condition) and Recorded-Same (exact replay condition) revealed only a small cluster within right angular gyrus [(46, -60, 44), $t = 5.55$], which showed more activity for Rec-S than Rec-D. There was no difference in the response to the two recorded conditions near any of the brain regions recruited by Live>Recorded. Paired samples *t*-tests showed that of all the ROIs identified by the Live>Recorded contrast, only IMT and ACC showed differential response to Rec-D versus Rec-S. These results suggest that activation during the Live condition in social, attention, and reward regions was not due to selective reduction in response during the Recorded-Same condition (e.g. due to repetition suppression).

Experiment 2: Joint Attention Experiment

Comparison of Joint to Solo Attention revealed greater activation within regions involved in perception of biological motion and inferring intentions of another person (Figure 3 and Table 2). Specifically, the largest cluster of activation included the right pSTS [(48, -40, 6), $t = 8.36$] which extended into the rTPJ [(54, -58, 14), $t = 8.13$] and right middle temporal gyrus [(50, -70, 0)]. This posterior STS/TPJ activity was also seen on the left, but to a smaller extent [(-48, -56, 14), $t = 6.97$]. At a more lenient threshold ($p < .001$, uncorrected), greater activation to Joint than Solo Attention was seen within right inferior frontal gyrus/frontal operculum [(46, 30, -6), $t = 6.46$], right aSTS [(58, 2, -24), $t = 5.88$], and dorsal medial prefrontal cortex (dMPFC) [(6, 50, 44), $t = 5.47$]. No regions showed significantly greater activation in the Solo Attention than Joint Attention conditions.

Regions of the Social Interaction Network involved in Joint Attention

Group whole-brain maps revealed substantial overlap for regions engaged during both the Social Interaction (Live>Rec) and Joint Attention (JA>SA), particularly within bilateral TPJ and STS (Figure 4a). To determine the extent to which Joint Attention and Social Interaction are engaging the same regions, follow-up analyses were conducted. We used the Live-Rec contrast from Experiment 1 to identify regions of interest (ROI), in which we then measured activity during Joint Attention and Solo Attention conditions. We specifically hypothesized that brain regions identified in the “Live” condition because of their role in social cognition, would be differentially recruited during Joint Attention. As predicted, a significant interaction was seen for condition by brain network (social, attention, reward; $F(2,15)=2.06$, $p<.01$), indicating differences between the Joint and Solo attention conditions were specific to the social brain areas; these differences were not seen in attention or reward regions. Follow-up paired sample t-tests, tested the difference between joint and solo attention in each ROI of the “social” network. The only regions showing significantly greater percent signal change during the JA than the SA conditions were the rpSTS [$t(12)=4.14$, $p=.001$], rTPJ [$t(12)=3.71$, $p=.003$], and raSTS [$t(12)=2.45$, $p=.03$] (Figure 4b).

Discussion

In this first study of face-to-face interactions in the scanner we found greater activation in several social-cognitive, attention, and reward processing brain areas when participants interacted with a live experimenter, as compared to a video replay of the experimenter (Experiment 1). We then utilized this method of face-to-face interactions to examine the neural bases of joint attention in a live interactive game involving joint and solo attention (Experiment 2). Only social-cognitive brain regions were differentially engaged during joint attention, specifically the rSTS and rTPJ. In contrast, the attention and reward areas identified in the live interaction were not differentially modulated by joint attention. These findings suggest that the rSTS and rTPJ are critical to one key component of real social interactions, namely joint attention.

Social interaction recruits reward systems

Regions within the ventral striatum, amygdala, and anterior cingulate cortex (ACC) were engaged during a live interaction to a greater extent than during viewing of exactly the same complex, dynamic, and social stimuli via recorded video. These regions have been consistently activated in studies of social reward including viewing an attractive face, playing a game with an alleged human partner, and experiencing pleasant touch (Walter et al., 2005). The current findings offer a neural mechanism for the early-emerging, powerful, and pervasive drive for humans to seek out social interactions; contingent interactions with a live person activate reward systems. This activation of a combination of reward, attention, and social-cognitive systems during a live interaction may explain why live social interactions facilitate early human learning. For example, infants learn a novel language from a live interaction with a speaker, but not from the same information presented via a video recording (Kuhl et al., 2003).

dMPFC is not recruited during live social interaction

Previous studies of social interaction have utilized varied approaches, including having the participant perceive an image, video, or virtual character or interact with an invisible, but imagined, person. The most consistently identified region during either observation of social interactions or engagement in a social interaction with an imagined other is the dMPFC (Fukui et al., 2006; Gallagher et al., 2002; Iacoboni et al., 2004; Kircher et al., 2009; Pierno et al., 2008; Rilling et al., 2008; Schilbach et al., 2006; Walter et al., 2004). In the current study, the dMPFC was not differentially recruited during a live social interaction (relative to a video) and was recruited only weakly during joint attention (relative to solo attention). These results

suggest that social interaction alone may not be sufficient to engage dMPFC. Instead, dMPFC may be recruited when social interaction includes a strategic or competitive component (Rilling et al., 2008) or for making judgments about oneself and others (e.g., (Saxe et al., 2006; Schmitz et al., 2004). The current episodes involved only simple, cooperative interactions with a single, stable partner. Future studies using face-to-face interactions while manipulating the complexity and demand of the cooperative game may elicit dMPFC to a greater extent.

The pSTS is modulated by actions of social relevance

The key node for contingent, social interactions was the rpSTS. The rpSTS is known to be a critical site for the perception of biological motion, such as shifts in eye gaze, human walking, or reaching (reviewed in, (Allison et al., 2000). More recent evidence has shown that the pSTS is not simply playing a perceptual role in biological motion detection, but rather is involved in a higher-order conceptual representation of the social significance of that motion (reviewed in, (Pelphrey and Morris, 2006). Until now, the evidence to support this claim relied on findings demonstrating that the STS responds more when the same observed biological motion violates the observer's expectations about rational actions (Pelphrey et al., 2004a). For example, the pSTS showed greater activity when a virtual character smiled at one object and then reached for the other, compared to smiling at and then reaching for the same object (Wyk et al., 2009). The current study provides convergent evidence that the pSTS is differentially recruited for socially significant actions by manipulating significance without violating expectations. In Experiment 1, the visible biological motion of the experimenter was identical in the live and recorded conditions; in the live condition the experimenter's actions were always congruent with the observer's expectations. Nevertheless, the pSTS showed greater response to the live than the recorded events. We suggest that the key dimension in pSTS activation is social relevance; only during a contingent social interaction is the action socially relevant to the observer. One limitation of the current study is that subject eye-movement was not recorded and thus differences in behavior between live and recorded conditions cannot be characterized. However, to control for possible differences in eye-movement, we used BOLD signal within the frontal eye fields (FEF) as a regressor in the analysis (see Methods). Future studies would benefit from recording and characterizing subject eye-movements as well as examining only auditorily or only visually contingent input to see whether involvement of the STS in detecting social relevance is modality-independent.

The RTPJ is recruited for contingent interaction

Activation in the right temporal lobe for live, as compared to recorded, conditions extended from right posterior STS into the TPJ. The rTPJ is adjacent to but functionally distinct from the pSTS: The rTPJ, but not the pSTS, is selectively recruited by verbal stories describing another person's thoughts or beliefs (Saxe and Kanwisher, 2003). The simple engagement in a contingent interaction with a live person may be sufficient to engage this region. That is, we may be automatically predisposed to think about another's thoughts or beliefs when involved in a cooperative live interaction. However, given that the rTPJ is difficult to define anatomically, this claim would need to be verified through the use of a functional localizer to isolate the rTPJ region that is recruited for belief representation within each individual.

Right pSTS and TPJ are key sites of 'online' joint attention

The rpSTS and rTPJ were differentially engaged during an interactive game designed to isolate regions involved in joint attention as compared to solo attention. These two regions were the same as those engaged during a live, contingent interaction. These findings provide evidence that the rpSTS is key to the representation of an actor's intentions behind an action, particularly when those intentions are relevant to the observer. During joint attention conditions, the subject can only succeed through utilizing the experimenter's gaze cue to find the target.

Two previous studies have examined the neural bases of joint attention, or the experience of sharing attention with another person on an object. In one study (Williams et al., 2005), participants attended to a ball that moved horizontally to one of four positions on a screen. Above the ball, an animated character's face was visible. In both conditions the participant was instructed to follow the ball to its new location. The character either also made a gaze shift to the ball (joint attention) or made a gaze shift away from the ball (nonjoint attention). Thus, "joint attention" was effectively coincidental; the participants did not deliberately follow or share the character's gaze. In this study, the joint attention condition led to activation of the dMPFC.

In the second study (Materna et al., 2008), participants not only "shared" attention on an object with a person (actually a static photo of a face), but specifically used the face's gaze shifts to direct their own attention. In the "joint attention" condition, the eyes in the image moved toward an object; participants were instructed to follow the gaze, in order to identify the color of the object. In the control condition, the face's pupils changed colors and participants were instructed to shift attention to the object that matched the color of the pupils. As in the current study, the key region recruited during gaze following for joint attention was the right posterior STS. Thus, we suggest that full joint attention requires more than just simultaneously gazing at the same object. Instead, people must deliberately coordinate attention on the object, usually with the expectation that the object will be rewarding (for cooperative exchanges) and/or relevant (for communicative exchanging). When participants engage in full joint attention, the pSTS rather than the dMPFC is the critical site of activation.

One region did appear to be recruited in our study, but not in the previous one (Materna et al., 2008): the rTPJ. The key difference between the two studies is that in our experiment, participants engaged in joint attention with a live human, rather than an image of a face. Thus, cooperation with a real person may be sufficient to recruit brain regions critical to thinking about another person's thoughts (i.e. the rTPJ). However, further studies using online and offline joint attention tasks with the same participant are needed to determine the specific difference in both location and intensity of activity.

Live face-to-face social interactions: A novel method for social cognitive neuroscience

This novel method of live social interactions in a scanner offers both strengths and limitations over current methods in social neuroscience. Social interaction in the presence of a live person (compared to a visually identical recording) resulted in activation of multiple neural systems which may be critical to real-world social interactions but are missed in more constrained, offline experiments. Of course, a limitation is that some experimental control is lost when studying even relatively simple live, naturalistic social interactions. To address this limitation, we used the face-to-face method in a second experiment to investigate one component of social interaction: joint attention. In the joint attention experiment, the presence of a live person was kept constant across both the condition of interest (Joint Attention) and the control condition (Solo Attention), allowing for greater experimental control. However, a limitation of both experiments was that the social exchange was simple, heavily scripted, and very predictable. Future paradigms using more complex and unpredictable live interactions may come closer to simulating natural, real-world social interactions.

This method of engaging in live interactions in the scanner may be useful in understanding social cognition in disorders of social communication, particularly Autism Spectrum Disorder (ASD). Individuals with ASD show marked deficits in reciprocal social interaction. Even in high-functioning individuals, these impairments are manifest in everyday social interactions, including inability to infer a speaker's intention, grasp metaphor, utilize prosody, or initiate appropriate conversation or eye contact (Klin et al., 2007; Paul et al., 2009). Furthermore, a specific abnormality in the detection of contingency, like those in social interactions, has been

hypothesized to be a core impairment in autism (Gergely, 2001). Despite these notable differences in person, laboratory-controlled tasks examining social behaviors and their neural bases often reveal mixed results. That is, offline “social cognition tests” fail to capture the central and persistent social deficits characteristic of ASD. Anecdotal reports from individuals with ASD suggest the difficulties in real-time social interaction arise from the need to integrate input from multiple modalities that are rapidly changing and unpredictable (Redcay, 2008). Thus, a neuroimaging task that includes the complexity of dynamic, multimodal social interactions may provide a more sensitive measure of the neural basis of social and communicative impairments in ASD.

Conclusion

The current study utilizes a novel method to examine the neural bases of social interaction and joint attention: Face-to-face interactions in the scanner. Prior tasks in social cognitive neuroscience generally used stimuli which lack fundamental properties of real-world interactions: Perhaps most importantly, a live, visible person with whom to interact. Some authors suggest that this tradition of utilizing simplified and decontextualized stimuli has produced overly constrained theories that may not bear directly on real-world social cognition (Zaki and Ochsner, 2009). If so, the novel method used in the current study could provide new insights into the biological mechanisms underlying our everyday social interactions. Furthermore, this tool may be particularly critical in the study of autism where the use of offline, simplified images may not capture true deficits in reciprocal social interaction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Rebecca Cox for help with data collection, Jasmine Wang for assistance with figures, and Dr. Marina Bedny, Mike Frank, Dr. Frida Polli, Zeynep Saygin, and Todd Thompson for discussion of the analyses and manuscript. We also thank Steve Shannon, Dr. Oliver Hinds and Todd Thompson for technical advice and Dr. Christina Triantafyllou for assistance with scanner protocols. We are grateful to the Simons Foundation for funding awarded to RS for this project and for an NIH Postdoctoral National Research Service Award for support of ER.

References

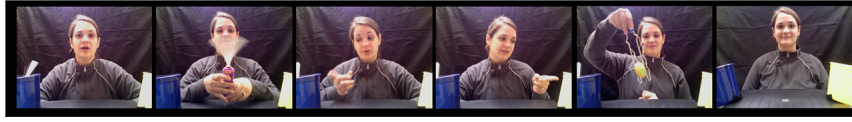
- Allison T, Puce A, McCarthy G. Social perception from visual cues: role of the STS region. *Trends Cogn Sci* 2000;4:267–278. [PubMed: 10859571]
- Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: diagnostic validity. *Br J Psychiatry* 1999;175:444–451. [PubMed: 10789276]
- Brainard DH. The Psychophysics Toolbox. *Spatial Vision* 1997;10:433–436. [PubMed: 9176952]
- Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, Fox MD, Snyder AZ, Vincent JL, Raichle ME, Schlaggar BL, Petersen SE. Distinct brain networks for adaptive and stable task control in humans. *Proc Natl Acad Sci U S A* 2007;104:11073–11078. [PubMed: 17576922]
- Fukui H, Murai T, Shinozaki J, Aso T, Fukuyama H, Hayashi T, Hanakawa T. The neural basis of social tactics: An fMRI study. *Neuroimage* 2006;32:913–920. [PubMed: 16682235]
- Gallagher HL, Jack AI, Roepstorff A, Frith CD. Imaging the intentional stance in a competitive game. *Neuroimage* 2002;16:814–821. [PubMed: 12169265]
- Gergely G. The obscure object of desire: ‘nearly, but clearly not, like me’: contingency preference in normal children versus children with autism. *Bull Menninger Clin* 2001;65:411–426. [PubMed: 11531136]
- Gergely, G.; Watson, JS. *Early Socialization*. Lawrence Erlbaum Associates Inc; Mahwah, NJ: 1999. *Early social-emotional development: Contingency perception and the social-biofeedback model*.

- Grosbras MH, Laird AR, Paus T. Cortical regions involved in eye movements, shifts of attention, and gaze perception. *Hum Brain Mapp* 2005;25:140–154. [PubMed: 15846814]
- Iacoboni M, Lieberman MD, Knowlton BJ, Molnar-Szakacs I, Moritz M, Throop CJ, Fiske AP. Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *Neuroimage* 2004;21:1167–1173. [PubMed: 15006683]
- Kircher T, Blumel I, Marjoram D, Lataster T, Krabbendam L, Weber J, van Os J, Krach S. Online mentalising investigated with functional MRI. *Neurosci Lett* 2009;454:176–181. [PubMed: 19429079]
- Klin A, Saulnier CA, Sparrow SS, Cicchetti DV, Volkmar FR, Lord C. Social and communication abilities and disabilities in higher functioning individuals with autism spectrum disorders: the Vineland and the ADOS. *J Autism Dev Disord* 2007;37:748–759. [PubMed: 17146708]
- Kuhl PK, Tsao FM, Liu HM. Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. *Proc Natl Acad Sci USA* 2003;100:9096–9101. [PubMed: 12861072]
- Materna S, Dicke PW, Thier P. Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: a functional magnetic resonance imaging study. *J Cogn Neurosci* 2008;20:108–119. [PubMed: 18095789]
- Murray, L.; Trevarthen, C. Emotional regulation of interactions between two-month-olds and their mothers. In: Field, TM.; Fox, NA., editors. *Social perception in Infancy*. ABlex; Norwood, NJ: 1985. p. 177-198.
- Paul R, Orlovski SM, Marcinko HC, Volkmar F. Conversational behaviors in youth with high-functioning ASD and Asperger syndrome. *J Autism Dev Disord* 2009;39:115–125. [PubMed: 18607708]
- Pelli DG. The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision* 1997;10:437–442. [PubMed: 9176953]
- Pelphrey KA, Morris JP. Brain Mechanisms for Interpreting the Actions of Others From Biological-Motion Cues. *Curr Dir Psychol Sci* 2006;15:136–140. [PubMed: 18079992]
- Pelphrey KA, Morris JP, McCarthy G. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J Cogn Neurosci* 2004a;16:1706–1716. [PubMed: 15701223]
- Pelphrey KA, Viola RJ, McCarthy G. When strangers pass: processing of mutual and averted social gaze in the superior temporal sulcus. *Psychol Sci* 2004b;15:598–603. [PubMed: 15327630]
- Pierro AC, Becchio C, Turella L, Tubaldi F, Castiello U. Observing social interactions: the effect of gaze. *Soc Neurosci* 2008;3:51–59. [PubMed: 18633846]
- Redcay E. The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neurosci Biobehav Rev* 2008;32:123–142. [PubMed: 17706781]
- Rilling J, Gutman D, Zeh T, Pagnoni G, Berns G, Kilts C. A neural basis for social cooperation. *Neuron* 2002;35:395–405. [PubMed: 12160756]
- Rilling JK, King-Casas B, Sanfey AG. The neurobiology of social decision-making. *Curr Opin Neurobiol* 2008;18:159–165. [PubMed: 18639633]
- Rilling JK, Sanfey AG, Aronson JA, Nystrom LE, Cohen JD. The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* 2004;22:1694–1703. [PubMed: 15275925]
- Saxe R. Uniquely human social cognition. *Curr Opin Neurobiol* 2006;16:235–239. [PubMed: 16546372]
- Saxe R, Kanwisher N. People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind”. *Neuroimage* 2003;19:1835–1842. [PubMed: 12948738]
- Saxe R, Moran JM, Scholz J, Gabrieli J. Overlapping and non-overlapping brain regions for theory of mind and self reflection in individual subjects. *Soc Cogn Affect Neurosci* 2006;1:229–234. [PubMed: 18985110]
- Schilbach L, Wohlschlaeger AM, Kraemer NC, Newen A, Shah NJ, Fink GR, Vogeley K. Being with virtual others: Neural correlates of social interaction. *Neuropsychologia* 2006;44:718–730. [PubMed: 16171833]
- Schmitz TW, Kawahara-Baccus TN, Johnson SC. Metacognitive evaluation, self-relevance, and the right prefrontal cortex. *Neuroimage* 2004;22:941–947. [PubMed: 15193625]

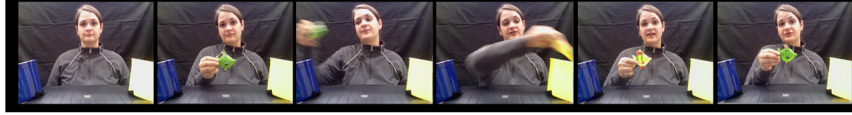
- Tomasello M, Carpenter M, Call J, Behne T, Moll H. Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci* 2005;28:675–691. discussion 691–735. [PubMed: 16262930]
- Walter H, Abler B, Ciaramidaro A, Erk S. Motivating forces of human actions. Neuroimaging reward and social interaction. *Brain Res Bull* 2005;67:368–381. [PubMed: 16216683]
- Walter H, Adenzato M, Ciaramidaro A, Enrici I, Pia L, Bara BG. Understanding intentions in social interaction: the role of the anterior paracingulate cortex. *J Cogn Neurosci* 2004;16:1854–1863. [PubMed: 15701234]
- Williams JH, Waiter GD, Perra O, Perrett DI, Whiten A. An fMRI study of joint attention experience. *Neuroimage* 2005;25:133–140. [PubMed: 15734350]
- Wyk BC, Hudac CM, Carter EJ, Sobel DM, Pelphrey KA. Action understanding in the superior temporal sulcus region. *Psychol Sci* 2009;20:771–777. [PubMed: 19422619]
- Zaki J, Ochsner K. The need for a cognitive neuroscience of naturalistic social cognition. *Ann N Y Acad Sci* 2009;1167:16–30. [PubMed: 19580548]

A) Social Interaction

“Showing”



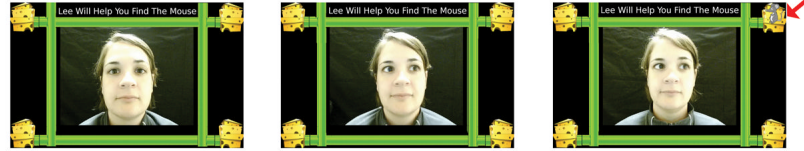
“Matching”



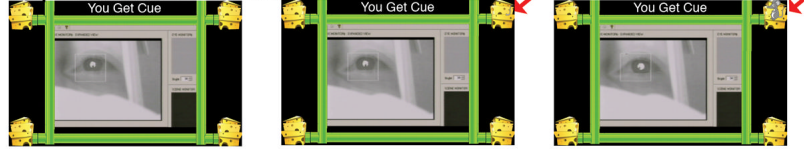
B) Joint Attention

Joint Attention (JA)

Subject Screen

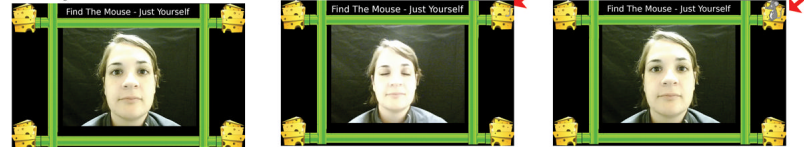


Experimenter Screen



Solo Attention (SA)

Subject Screen



Experimenter Screen



Figure 1.

Experimental Design. Still frames are shown from movies recorded for Experiment 1 (A) and Experiment 2 (B). In Experiment 1, 40 second live and recorded video feeds were played to participants. Two types of interactions occurred during these videos. The first, “showing”, displayed in the top panel of (A), entailed showing dynamic toys from either a yellow or blue bucket depending on which bucket the participant chose to look at. The second, “matching”, displayed in the bottom panel of (A), entailed presenting objects that differed along one dimension and asking participants to look at the bucket in which the object should be placed (See Supplemental text for script). In (B), an example of one experimental trial is shown for both the Joint Attention and Solo Attention condition types. Both the experimenter screen (the

screen the experimenter viewed during the trial) and the subject screen (the screen the subject viewed while in the scanner) are displayed. The center of each frame displays a video feed either from the video camera pointed at the experimenter (in the subject screen) or the eye-tracking camera in the scanner (in the experimenter screen). Each trial begins with the experimenter looking straight ahead. Instructions at the top of the screen are there to remind the participant and experimenter of the rules of the condition block (5 trials per block). In JA, the experimenter receives the cue of a mouse tail and looks at the tail. The subject follows the experimenter's gaze. When both are looking at the correct house, the mouse appears. In SA, the subject receives the cue and looks at the tail. When the subject looks at the tail, the mouse appears. The experimenter closes and opens her eyes once during each trial to signal that she is not playing the game.

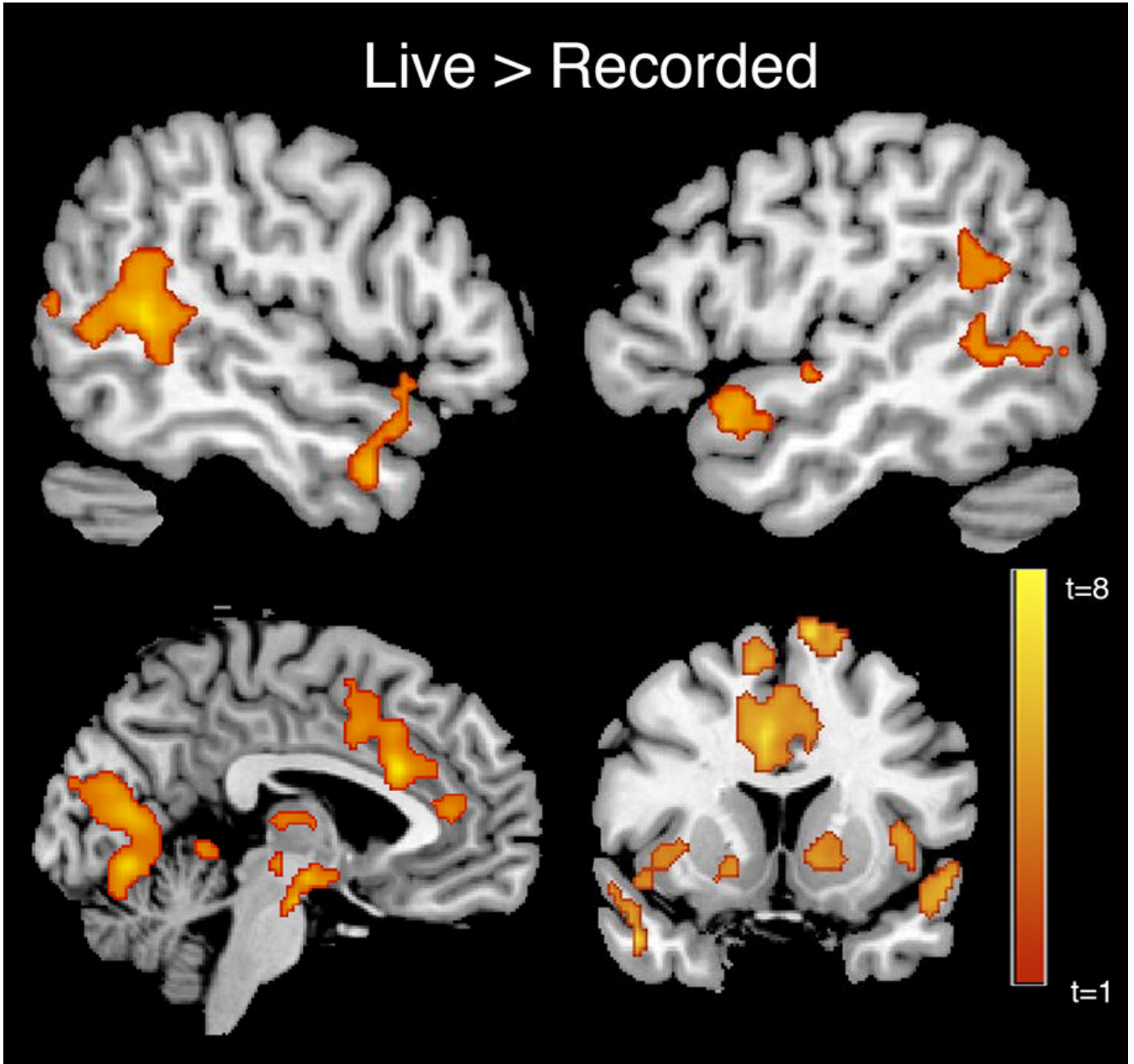


Figure 2. Experiment 1: Live vs. Recorded conditions. T-values for regions showing significantly greater BOLD signal during live presentation of a person as compared to a video recording are displayed on a template brain. Left sagittal ($x=-52$), Right sagittal ($x=52$), medial ($x=-4$), and coronal ($y=12$) sections are displayed.

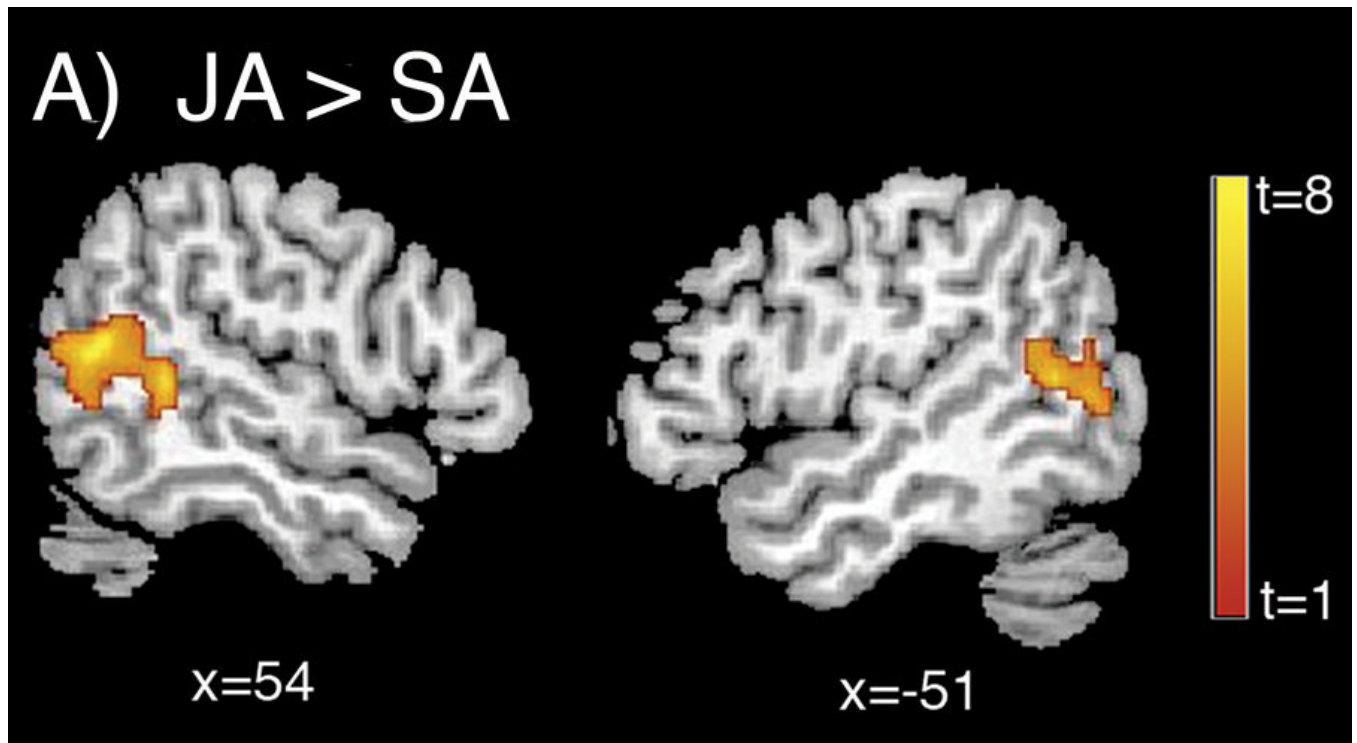


Figure 3. Experiment 2: Joint Attention. A) T-values for regions showing significantly greater BOLD signal during conditions of joint attention as compared to solo attention are displayed on a template brain. Left sagittal ($x=-51$) and Right sagittal ($x=54$) sections are displayed.

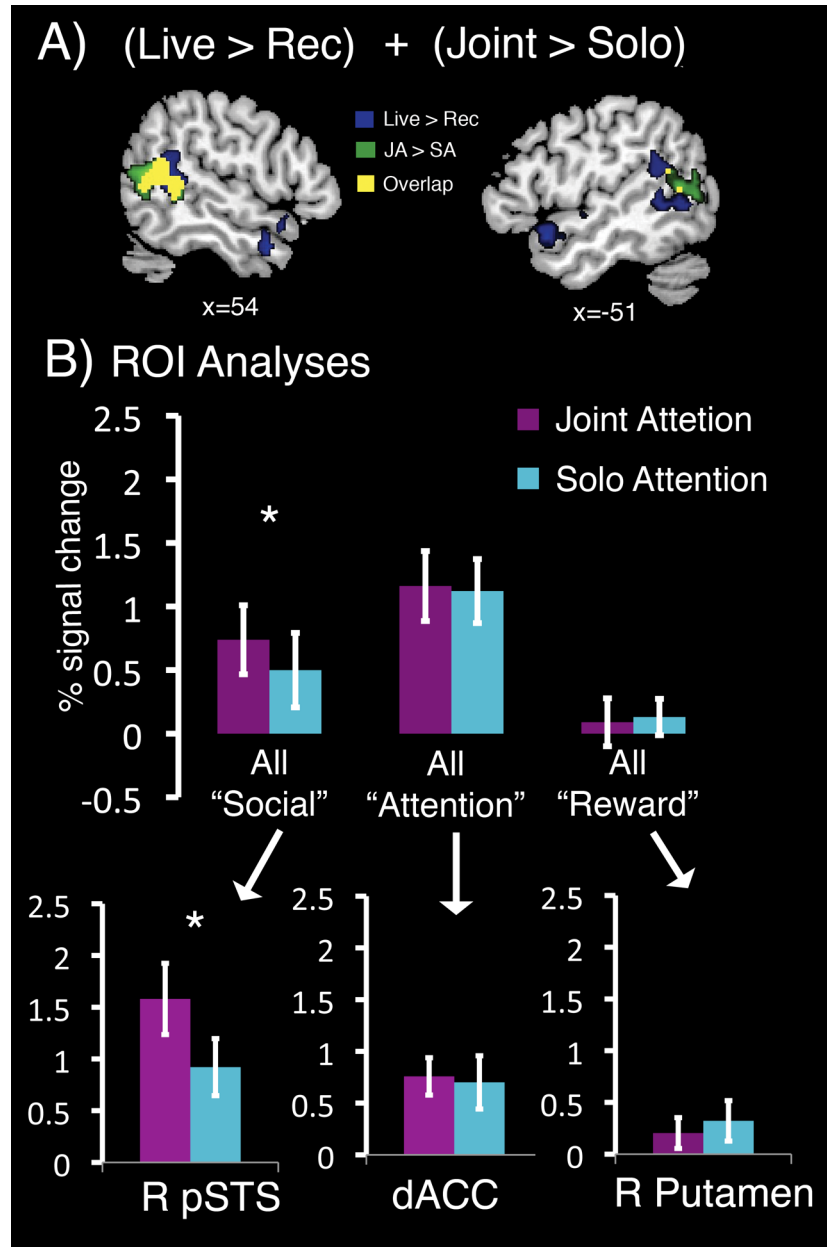


Figure 4. A) Conjunction Analyses. Maps from the contrasts Live vs. Recorded (blue) and Joint Attention vs. Solo Attention (green) are displayed on a template brain. The overlap (yellow) is a map of the voxels which were significant in both Live > Rec and JA > SA contrasts. B) ROI analyses of percent signal to Joint and Solo Attention (Experiment 2) within regions identified from the contrast Live>Recorded (Experiment 1). The top panel displays mean percent signal change for the Joint Attention (JA) (purple) and Solo Attention (SA) (turquoise) conditions, averaged across all regions within the “social”, “attention”, and “reward” networks. “Social” regions included rpSTS, rTPJ, raSTS, ITPJ, right Insula, and ACC. “Reward” regions included right and left Putamen, left Thalamus, left Caudate, and right Amygdala. “Attention” regions included dACC, left Cerebellum, Lingual Gyrus, and right Middle Occipital Gyrus. Standard error (SEM) bars are shown. Only social brain regions showed increased response during Joint compared to Solo, Attention. A representative graph of one region from each of the 3 networks

(i.e. social, attention, and reward) is displayed in the bottom panel. Each graph displays mean percent signal change values for the JA and SA conditions with standard error bars.

Table 1
 Experiment 1: Regions showing differences between Live and Recorded Conditions

Region	Side	x	y	z	t-value
Live > Recorded					
Cortical					
dAnterior Cingulate/Medial Prefrontal Cortex	R	6	14	36	9.16
Anterior Cingulate	R	2	36	14	4.90
Supplementary Motor Area	R	8	6	62	7.75
Insula	L	-40	14	4	6.22
Insula	R	38	16	0	4.46
pSuperior Temporal Sulcus	R	54	-48	12	7.22
Temporoparietal Junction	R	60	-46	24	5.23
Middle Temporal Gyrus	R	38	-64	8	6.94
aSuperior Temporal Sulcus	R	48	12	-30	7.08
Middle Occipital Gyrus	R	38	-70	20	5.24
Temporoparietal Junction	L	-48	-44	22	5.89
Temporal Pole	L	-46	16	-16	6.83
Middle Temporal Gyrus	L	-48	-60	4	5.20
Lingual Gyrus	L	10	-74	-12	9.72
Subcortical					
Cerebellum	L	-42	-56	-28	6.92
Amygdala	R	20	-2	-18	4.44
Hippocampus	R	26	-16	-14	7.14
Caudate	L	-18	18	4	6.84
Putamen	L	-12	12	-4	4.76
Putamen	R	22	8	-10	4.85
Thalamus	L	-14	-8	6	5.87
Recorded > Live					
Cortical					
Precentral Gyrus	L	-38	-20	58	6.11

Region	Side	x	y	z	t-value
Precentral Gyrus	R	46	-24	58	5.22
Supplementary Motor Area	R	8	-18	64	5.55

Note: p refers to posterior, d to dorsal, and a to anterior.

Coordinates are in Montreal Neurologic Institute Space

Table 2

Experiment 2: Regions showing differences between Joint Attention and Solo Attention

Region	Side	x	y	z	t-value
JA > SA					
p<.05 corrected					
pSuperior Temporal Sulcus	R	48	-40	6	8.36
Temporoparietal Junction	R	54	-58	14	8.13
Middle Temporal Gyrus	R	50	-70	0	7.55
Temporoparietal Junction	L	-48	-56	14	6.97
Middle Temporal Gyrus	L	-48	-68	6	5.67
Superior Occipital Gyrus	L	-10	-106	14	9.07
Calcarine Sulcus	R	2	-88	-2	8.51
p<.001, uncorrected					
Inferior Frontal Gyrus	R	46	30	-6	6.46
dMedial Prefrontal Cortex	R	6	50	44	5.47
aSuperior Temporal Sulcus	R	58	2	-24	5.88
SA > JA					
none					

Note: p refers to posterior, d to dorsal, and a to anterior.

Coordinates are in Montreal Neurologic Institute Space. JA refers to Joint Attention conditions. SA refers to Solo Attention conditions