

Social, reward, and attention brain networks are involved when online bids for joint attention are met with congruent versus incongruent responses

Ilanit Gordon, Jeffrey A. Eilbott, Ruth Feldman, Kevin A. Pelphrey & Brent C. Vander Wyk

To cite this article: Ilanit Gordon, Jeffrey A. Eilbott, Ruth Feldman, Kevin A. Pelphrey & Brent C. Vander Wyk (2013) Social, reward, and attention brain networks are involved when online bids for joint attention are met with congruent versus incongruent responses, *Social Neuroscience*, 8:6, 544-554, DOI: [10.1080/17470919.2013.832374](https://doi.org/10.1080/17470919.2013.832374)

To link to this article: <http://dx.doi.org/10.1080/17470919.2013.832374>

 Published online: 17 Sep 2013.

 Submit your article to this journal [↗](#)

 Article views: 211

 View related articles [↗](#)

 Citing articles: 8 View citing articles [↗](#)

Social, reward, and attention brain networks are involved when online bids for joint attention are met with congruent versus incongruent responses

Ilanit Gordon¹, Jeffrey A. Eilbott¹, Ruth Feldman², Kevin A. Pelphrey¹, and Brent C. Vander Wyk¹

¹Child Study Center, Yale University, New Haven, CT 06520, USA

²Psychology Department and Gonda Brain Research Center, Bar-Ilan University, Ramat Gan 52900, Israel

Joint attention (JA) is a cornerstone of adaptive human social functioning. Little functional magnetic resonance imaging (fMRI) research has examined, in interactive paradigms, neural activation underlying bids for JA, met with a congruent or an incongruent social response. We developed a highly naturalistic fMRI paradigm utilizing eye-tracking to create real-time, contingent social responses to participant-initiated JA. During congruent responses to JA bids, we observed increased activation in the right amygdala, the right fusiform gyrus, anterior and dorsal anterior cingulate cortices, striatum, ventral tegmental area, and posterior parietal cortices. Incongruent responses to JA bids elicited increased activity localized to the right temporoparietal junction (TPJ) and bilateral cerebellum. No differences in eye-gaze patterns were observed during congruent or incongruent trials. Our results highlight the importance of utilizing interactive fMRI paradigms in social neuroscience and the impact of congruency in recruiting integrated social, reward, and attention circuits for processing JA.

Keywords: Joint attention; fMRI; Eye tracking.

A fundamental building block of adaptive social functioning lies in our ability to use gaze to first capture others' attention to ourselves and then direct their attention to a shared location. This allows us to establish joint attention (JA), considered to be a cornerstone of socio-cognitive development, self-organization, learning, and the participation in shared interactions with other intentional agents (Charman, 2003; Moore & Dunham, 1995; Mundy & Newell, 2007; Senju & Johnson, 2009; Tomasello, 1995). JA states allow for a fundamental propensity to detect intention or mental states, involving inherent hedonic properties related to shared attentional states and congruent responses from others (Baron-Cohen, 2005; Gopnik, Slaughter, & Meltzoff, 1994; Mundy, Sullivan, & Mastergeorge, 2009; Tomasello,

Carpenter, Call, Behne, & Moll, 2005). In healthy adulthood, neural processes underlying JA have been examined (see Redcay et al., 2010; Saito et al., 2010; Schilbach et al., 2010; Williams, Waiter, Perra, Perrett, & Whiten, 2005). Yet, little functional magnetic resonance imaging (fMRI) research has been done to examine the impact of congruent and incongruent responses to bids for JA on brain function in truly interactive, ecologically valid paradigms that utilize dynamic stimuli reactive to participants' gaze.

Prior studies of JA implicated a well-characterized system of neural regions known for processing social information, especially the anterior cingulate cortex (ACC), the orbitofrontal cortex, posterior parietal association area, and superior temporal cortex (Mundy & Newell, 2007; Williams et al., 2005). JA states also

Correspondence should be addressed to: Ilanit Gordon, Child Study Center, Yale University, 230 South Frontage Road, New Haven, CT 06520, USA. E-mail: ilanit.gordon@yale.edu

engage the frontal–parietal attention network (Mundy et al., 2009). Moreover, the successful direction of someone else’s gaze toward an object activates reward regions such as the ventral striatum (Shilbach et al., 2010). Together, the available results suggest that reward, attentional, and social perception circuitries are involved in initiating JA. In the current study, we wanted to focus on the neurofunctional underpinnings of a certain aspect of the JA process: reaction to congruent versus incongruent responses following online bids for JA. This process specifically has received little attention in earlier online eye-tracking fMRI studies (in Schilbach et al. (2010), it was termed “Self-NOJA” and was found to strictly activate the striatum) and we hoped to be able to improve on earlier work in regard to the social setting of initiating JA. We hypothesized that if we could create a paradigm that focuses on online social bidding for JA, we could uncover the involvement of brain regions beyond the striatum—that have to do with attention and social functioning as well.

Few fMRI paradigms utilize interactive designs (the notable exceptions are innovative paradigms by Redcay et al. (2010), Saito et al. (2010), and Shilbach et al. (2010)), mostly due to technical challenges in incorporating eye-tracking technology to establish truly interactive paradigms during an fMRI scan. Recent work has emphasized the importance of using live interactive paradigms in neuroscience in order to identify brain regions that activate in naturalistic settings. (Gallagher, 2008; Legrand & Iacobini, 2010). Interactions that were perceived as “live” and not “recorded” yielded increased activations in brain regions involved in social cognition and reward, including the right temporoparietal junction (TPJ), ACC, right superior temporal sulcus, ventral striatum, and amygdala (Redcay et al., 2012; Redcay, Kleiner, & Saxe, 2012). Guided by suggestions for future studies, we aimed to use a platform similar to that used in state-of-the-art eye-tracking fMRI paradigms and to focus on the online aspects of bidding for JA, so that we would be able to assess the impact of congruent response to such “real-life” attempts to initiation of JA. We also wanted to strictly focus on social JA states, so instead of JA to an object (for instance, houses and a mouse in Redcay et al. (2010) paradigm or rectangles in Schilbach et al. (2010) paradigm), we asked participants to choose a human avatar to direct attention toward (see Figure 1). We believed this would increase ecological validity and recruit the integrated involvement of discrete brain regions that have previously been implicated in JA. Considering the importance of examining live interactive behavior, we were interested in creating an online reactive fMRI paradigm that would not rely on an experimenter’s behavior

(as in Redcay et al. (2010)), to increase control and replicability. We thus developed a paradigm that uses online eye-tracking information to create contingent reactions to individuals’ gaze patterns. We focused on neural activations underlying the processing of different responses to JA bids, an important aspect of social-gaze contingencies. This highly naturalistic fMRI paradigm incorporates eye-tracking technology to create a real-time response to initiating JA, in order to assess the neural processes associated with congruent versus incongruent gaze responses following bids for JA. We asked participants to direct an actors’ attention to one of two men on the screen, which resulted in either successful JA (congruent—the actor on the screen followed the gaze of the participant to the same avatar the participant was looking at) or failed JA (incongruent—when the actor on the screen looked at the other avatar). We employed a whole-brain analysis approach in order to localize brain regions exhibiting increased activation to congruent versus incongruent trials. On the basis of the available literature, we expected that bids for JA that received congruent responses would increase activity in brain regions previously implicated in social perception (e.g., amygdala, lateral prefrontal cortex, and ACC activations), reward circuitry (e.g., the striatum), and attention circuits (e.g., cingulate and posterior parietal cortices (PPCs)). We also hypothesized that activations in TPJ and posterior superior temporal sulcus (pSTS) regions, known for increased activations when social expectations are violated would appear as a result of incongruent responses to JA bids.

MATERIALS AND METHODS

Participants

Twenty-two healthy adults (12 men; ages: 22–35 years) participated in the study. We registered eye-tracking data successfully from 14 individuals. The remaining eight participants could not be eye-tracked as they failed the calibration process (see details in Calibration section) owing mostly to technical aspects of eye-tracking technology; certain eyelid structures and iris colors resulted in reduced contrast between iris and sclera necessary for eye-tracking. Demographically (ethnicity, age, handedness), we found no systematic feature that distinguished the 14 people whose gaze direction was successfully calibrated (thus able to take part in the fMRI paradigms) from those whose gaze direction could not be calibrated and did not participate.

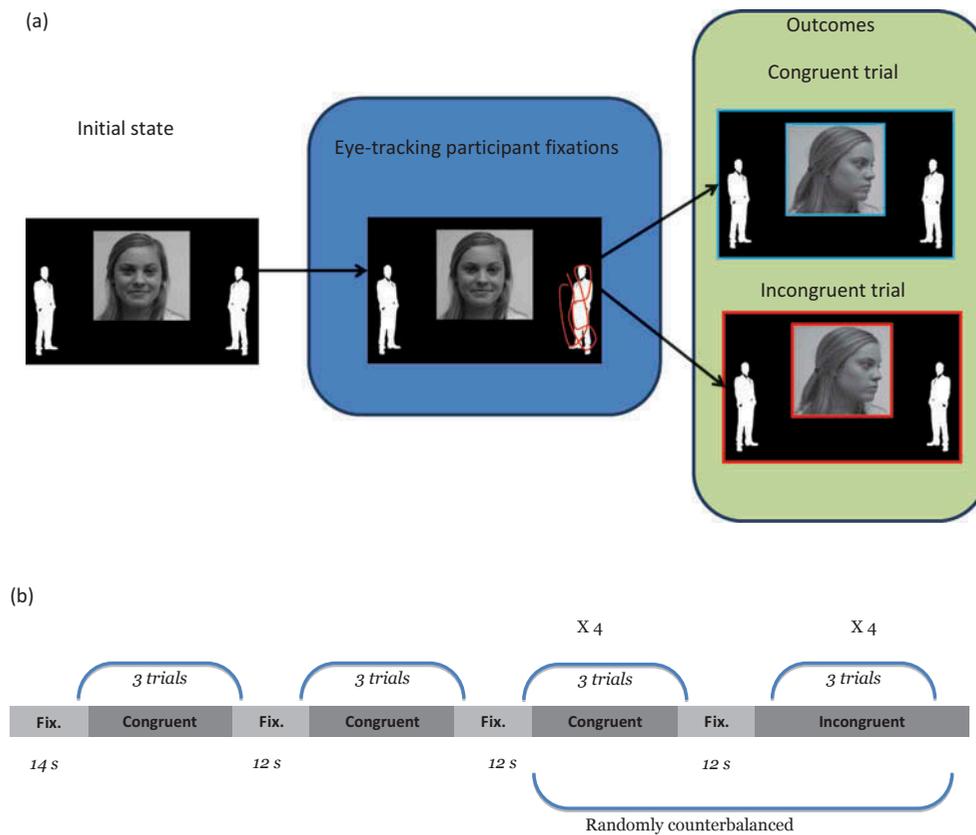


Figure 1. fMRI paradigm: (a) Flow and possible outcomes and (b) design.

Experimental design

Once positioned in the fMRI and following structural scans, participants were shown an introductory screen explaining the task they were to perform, which displayed the actual stimuli they would be viewing. Participants were simply instructed to “get” a prerecorded on-screen character (“Sally”) to follow their gaze to one of two identical male avatars located on the bottom left and right of the screen. When a participant’s bid for JA was successful, “Sally” reacted in one of two ways: either following the participant’s gaze to the corresponding avatar (congruent trial), or looking at the opposite avatar (incongruent trial)—see [Figure 1a](#). The incongruent or congruent sequence lasted for 3 seconds across all trials and all participants.

Participants were asked to get “Sally’s” attention toward one of two avatars of their choice. Getting “Sally’s” attention meant getting her to move her head toward an avatar by looking at the avatar. Avatars were identical, so that the choice was not skewed due to their attributes. Participants knew that “Sally” was prerecorded and not live. The live aspect of the paradigm was twofold: (1) “Sally” is a recording of a

live person and not an animation. (2) “Sally” begins to direct attention toward an avatar only after she has “picked up” on an online bid for JA, thus her recording starts to move only after live JA has been initiated by the participant. The prerecorded aspects of the paradigm (which participants were well aware of) were meant to increase controllability and consistency of the paradigm, the live online aspects of our paradigm were meant to increase the naturalistic nature of initiating JA by participants.

We utilized a pseudo-randomized block design ([Figure 1b](#)) with five blocks of congruent reactions and five blocks of incongruent reactions. The first two blocks were always congruent to establish a sense of contingency between participants’ eye gaze and “Sally’s” response, while the order of subsequent blocks was randomized. A 12-second fixation was displayed between blocks. Each block consisted of three trials. A trial began with an “Initiating JA” period (a video of “Sally” looking straight toward the participant), during which participants were expected to initiate JA by gazing at either the left or right avatar. If, as determined by a gaze-area-of-interest (GAOI)-based algorithm, participants elected an avatar, “Sally” would respond

in accordance with the current block type (congruent or incongruent). If participants were unable to establish JA, as determined by the GAOI-based algorithm, the “Initiating JA” video clip would repeat up to three times to give participants more chances before ending the trial without “Sally” reacting. Due to individual differences in the number of “Initiating JA” loops played before JA was obtained, the length of individual trials varied both within and across participants. The minimum time for a trial was 7 seconds, while the maximum was 11 seconds.

Gaze data acquisition

Participants’ gaze location was monitored using an infrared camera and an infrared light source. Raw gaze data was collected at 60 Hz using ISCAN’s DQW software (Woburn, MA, USA). Participants’ point of regard (POR) with respect to the stimulus display screen was calculated by determining the vector between the center of the pupil and the infrared light reflected from the cornea. Calibration preceded every experiment, enabling us to interpolate gaze positions as they occur in real time and outputting coordinates through a serial connection to a stimulus display laptop running E-Prime 2.0 (Psychology Software Tools, Inc, Sharpsburg, PA, USA), where further real-time analysis was performed.

Calibration

In order to accurately determine gaze positions, as they occurred in real time, a calibration preceded every experiment. Participants were instructed to view a black stimulus display screen and fixate on a series of five pulsating dots. Calibration dots were localized to the upper-left, upper-right, bottom-left, and bottom-right corners of the display. A center calibration dot was also displayed. These calibration dots were accompanied by a beeping sound synchronous with their visual pulsation and were displayed one at a time to the participant. The experimenter was able to view the participants’ eyes with two crosshairs overlaid and calculated by the DQW software: one over the corneal reflection (CR) and the other over the center of the pupil. Taking into account blinks, saccades, and the stability of these crosshairs, the experimenter pressed a button to bind the vector between the crosshairs to the spatial location of the dot at an optimal point an time. The combination of the five vector-position pairs allowed the DQW software to maximize accuracy interpolating all possible PORs within the stimulus display screen as they occurred in real time during the experiment.

The accuracy of each calibration was subsequently tested by displaying the same five dots, and visually inspecting the participants’ PORs (overlaid with a crosshair) on two dimensions: proximity to the dot and stability of the POR over a period of a few seconds. If the calibration was unsuccessful based on these metrics, either a recalibration was performed, or the participant did not continue with the experiment.

Real-time gaze analysis

We implemented a GAOI-based algorithm to determine whether the subject had selected an avatar to establish JA. During the 2-second “Initiating JA” period, POR was recorded and tallied based upon predetermined GAOIs (see [Figure 2](#)). Data points for which no POR could be calculated were tallied as “lost.”

After the 2-second “Initiating JA” clip had completed, the algorithm calculated the percentage of data that had been characterized as lost; and if this exceeded 50%, the bid for JA was deemed unsuccessful. However, if the initial quality test passed (i.e., 50% or less of the data was “lost”), the algorithm went on to calculate the percentage of the remaining valid data points falling in the left or right avatar GAOIs. If a threshold of at least 25% was met, the algorithm then determined which avatar was selected by calculating whether a certain side comprised at least 85% of the combined left and right POR data. If all these criteria were met, the “Initiating JA” period was considered a successful bid for JA and the proper reaction played. If not, the “Initiating JA” video played again, and the same algorithm was applied, until the maximum of three JA bids was reached. All GAOI tallies and percentage data were recorded.

Imaging protocol

Images were collected on a Siemens 3T Tim Trio scanner (Siemens Medical Solutions USA Inc., Malvern, PA, USA) located in the Yale University Magnetic Resonance Research Center. High-resolution T1-weighted anatomical images were acquired using an magnetization prepared rapid acquisition with gradient echo (MPRAGE) sequence (TR = 1230 ms; TE = 1.73 ms; field of view (FOV) = 256 mm; image matrix = 256²; 1 × 1 × 1 mm). Whole-brain functional images were acquired using a single-shot, gradient-recalled echo planar pulse sequence (TR = 2000 ms; TE = 25 ms; flip angle = 60°; FOV = 220 mm; image matrix = 64²; voxel size = 3.4 × 3.4 × 4.0 mm; 34 slices) sensitive to blood oxygen level dependence (BOLD) contrast.

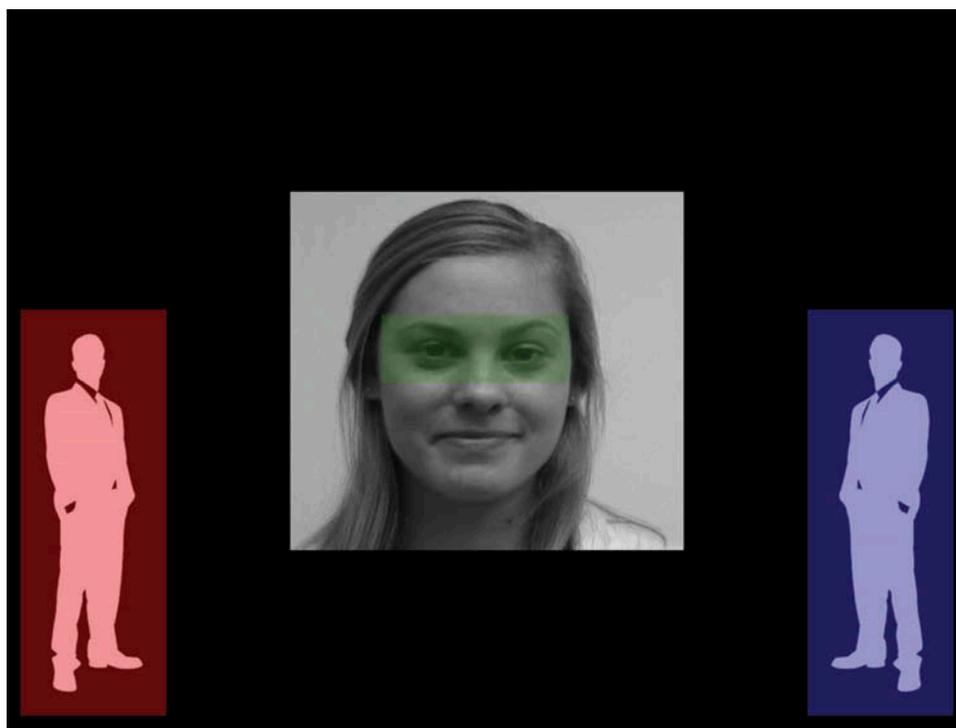


Figure 2. Regions of interest used by the algorithm to determine the success of interaction: green—“eyes,” red—“left,” blue—“right,” grayscale—“outer.” If individuals were not looking at the screen at all, it was considered “lost.” Gaze data was recorded as proportions of time spent (in percent) looking at GAOIs.

fMRI analysis

Data were processed and analyzed using BrainVoyager QX 2.0.08 software (Brain Innovation, Maastricht, The Netherlands). Preprocessing of functional data included sinc interpolation slice-time-correction, trilinear-sinc interpolation 3D rigid-body motion correction, spatial smoothing with a full width at half maximum (FWHM) 7-mm Gaussian kernel, linear trend removal, and temporal high-pass filtering (general linear model (GLM) with Fourier basis set, using two cycles/time course). Functional data sets were coregistered to within-session anatomical images, which were normalized to Talairach space. For each participant, we assessed estimated motion plots and cine-loops. For all scan sessions, participant’s head position had not deviated from first volume acquisition by more than 2 mm of translation in any direction or two degrees of rotation about any axis. No participant had greater than 1 mm/degree of translation or rotation between two consecutive volumes or greater than a total of 2 mm/degrees of translation or rotation integrated over four consecutive volumes.

GLM-based analyses were conducted for each participant to assess task-related BOLD responses. To create predictors for “Congruent”

and “Incongruent” conditions (each “Congruent” or “Incongruent” trial lasting exactly 3 seconds and having comparable physical attributes—“Sally” moving her head to look at one of the avatars), a boxcar function with a value of 1 during the condition and 0 otherwise was convolved with a double-gamma hemodynamic response function (HRF). “Initiating JA” periods as well as predictors depicting motion in all six parameters were included as predictors of no interest. Fixation was not modeled.

Methodologically, “Initiating JA” periods were highly variable, individualized and not balanced in the design. For instance, some individuals succeeded to initiate JA every trial pretty quickly, versus others who only succeeded to achieve JA states 75% of the time and it took them longer to get “Sally’s” attention. We thus modeled periods of “Initiating JA” in our GLM as events of no interest to account for the variance associated with those events.

“Initiating JA” was modeled yet not used as a comparison condition in our study for two main reasons: First, our a priori hypotheses had to do with the neural processes underlying congruent versus incongruent responses following bids for JA. Second, during “Initiating JA” periods, the stimulus on the screen (“Sally”) is looking directly toward the participant who

is required to direct “Sally’s” gaze toward an avatar. During the “Congruent” and “Incongruent” periods of time, JA has been initiated and “Sally” moves her head similarly, either toward the participants’ chosen avatar or toward the second avatar. As soon as Sally’s gaze response was initiated, participants’ gaze patterns were no longer required to impact or shift Sally’s gaze. In short, “Initiating JA” periods had unique physical attributes of the stimulus on screen (motion versus no motion) as well as unique task demand (initiating JA versus passive viewing).

All group-level analyses were limited to only voxels within the extent of the Montreal Neurological Institute (MNI) brain normalized to Talairach space. This whole-brain mask consisted of 1,449,746 ($1 \times 1 \times 1$ mm) voxels. Whole-brain investigations were conducted using random-effects (RFX) GLM-based analyses. Analyses of the congruent versus incongruent conditions were assessed at a threshold of $p < .05$, and were corrected for multiple comparisons using cluster thresholds determined by the Brain Voyager QX Cluster-level Statistical Threshold Estimator (Brain Innovation, Maastricht, The Netherlands) plug-in (Forman et al., 1995; Goebel, Esposito, & Formisano, 2006). After 1000 iterations of a Monte Carlo simulation, the cluster size corresponding to a corrected threshold of $\alpha < .05$ was determined. For the main analyses, contrast (congruent > incongruent) with a gray matter mask that included only voxels within the brain, a cluster threshold of 31 voxels was calculated.

RESULTS

Eye tracking

First, we present eye-tracking results describing the raw data acquired, specifically the proportions (in percent) of gazing at each GAOI (as presented in Figure 3) for our participants averaged across incongruent and congruent trials, and results of

pairwise comparisons (t -tests) to assess differences across conditions. To correct for multiple comparisons, we employed a Bonferroni correction and used an adjusted alpha level ($\alpha = .00625$). As can be seen from Table 1, there were no significant differences between gaze patterns in the congruent and incongruent trials. It is of importance to note here that the number of “Initiating JA” loops preceding congruent and incongruent trials were also not significantly different. These behavioral results affirm that brain activation differentiated by conditions were not due to significantly differential eye-gaze patterns.

Functional magnetic resonance imaging

Following, we present results of a multi-participant (Random Effects General Linear Model) RFX GLM-based analysis for the contrast congruent > incongruent corrected for multiple comparisons with a cluster size of 31 contiguous voxels (837 cubic mm) corresponding to $\alpha < .05$. The results of this analysis are shown in Table 2 and Figure 3. As illustrated in Figure 3, this contrast resulted in bilateral activations in the ACC/dorsal ACC (dACC), right fusiform gyrus (FG), PPCs, right amygdala, and bilateral striatum and parahippocampal regions. In contrast, bilateral cerebellar regions as well as right posterior TPJ regions activated more during incongruent conditions versus congruent conditions.

In Figure 4, we present comparisons of beta coefficients for the main regions depicted in Figure 3 in the incongruent and congruent trials.

DISCUSSION

During congruent (versus incongruent) trials, there was increased fMRI activation in key nodes of neural circuits supporting social perception, response to



Figure 3. Results of contrasting congruent versus incongruent trials in a JA task; $q < .05$, $k > 31$. A minimum cluster threshold of 31 contiguous functional voxels was used. Activations appear in an orange color if they were stronger during congruent trials and in blue if they were stronger in incongruent trials.

TABLE 1

Means, standard deviations, and results of pairwise *t*-tests comparing eye gaze behavior in congruent and incongruent conditions

| | Congruent trials <i>M</i> (<i>SD</i>) | Incongruent trials <i>M</i> (<i>SD</i>) | <i>t</i> (13), sig.* |
|--|--|--|-------------------------|
| Overall participant success to initiate JA (% from overall number of attempts to bid for JA) | 0.7 (.282) | 0.738 (.236) | -1.075, <i>p</i> = .302 |
| Number of "Initiating JA" loops prior to successful bid for JA | 1.391 (.083) | 1.515 (.098) | -1.447, <i>p</i> = .172 |
| Proportion of time participant spent gazing toward GAOIs (%) | | | |
| Sally's eyes | .125 (.181) | .118 (.156) | 0.561, <i>p</i> = .584 |
| An avatar that sally is also gazing toward at the same time | 0.396 (.233) | 0.419 (.205) | -0.929, <i>p</i> = .37 |
| Right avatar | 0.513 (.084) | 0.599 (.089) | -2.545, <i>p</i> = .024 |
| Left avatar | 0.429 (.141) | 0.344 (.124) | 2.336, <i>p</i> = .036 |
| Outer (on screen but not on any of the GAOIs) | 0.478 (.214) | 0.461 (.197) | 0.71, <i>p</i> = .584 |
| Lost (not registered on screen) | 0.333 (.195) | 0.34 (.202) | -0.192, <i>p</i> = .851 |

Note: *Following a Bonferroni correction ($\alpha = .00625$), none of the *p*-values presented here remain statistically significant.

TABLE 2

Results from congruent versus incongruent contrast

| Region | Peak X | Peak Y | Peak Z | <i>T</i> | <i>p</i> | Extent in mm ³ |
|---|--------|--------|--------|----------|----------|---------------------------|
| Congruency > incongruency | | | | | | |
| L PPC | -21 | -64 | 28 | 5.312 | .000141 | 3827 |
| R PPC | 24 | -61 | 46 | 5.439 | .000113 | 2896 |
| R postcentral gyrus | 42 | -31 | 31 | 4.649 | .000455 | 1193 |
| L postcentral gyrus | -54 | -22 | 37 | 4.187 | .001065 | 1218 |
| R fusiform gyrus | 48 | -52 | -8 | 4.632 | .000469 | 1534 |
| ACC/dACC | 0 | 35 | 16 | 4.960 | .000261 | 2562 |
| Mid-cingulate gyrus | 0 | -4 | 37 | 3.880 | .001895 | 962 |
| R subthalamic region (including VTA, amygdala) | 21 | -31 | 4 | 5.912 | .000051 | 4220 |
| L subthalamic region | -15 | -28 | 13 | 4.399 | .000718 | 1709 |
| L hippocampus | -27 | -31 | -5 | 6.288 | .000028 | 947 |
| L inferior parietal lobule | -54 | -22 | 37 | 4.187 | .001065 | 2290 |
| Incongruency > congruency | | | | | | |
| R caudate | 26 | -37 | 10 | -4.567 | .000528 | 1411 |
| R cerebellum | 39 | -70 | -32 | -3.925 | .001740 | 1662 |
| L cerebellum | -27 | -79 | -35 | -5.331 | .000718 | 3474 |
| R TPJ | 57 | -49 | 31 | -4.221 | .00100 | 1509 |

Note: R = right; L = left.

reward, and the allocation of attention. Specifically, social perception regions included the right FG. Reward regions included subthalamic areas such as the striatum and ventral tegmental area (VTA). Attention regions included bilateral mid-cingulate and PPCs. The right amygdala may serve to integrate reward and social processing. Activation in the ACC may represent multiple functions, as this region probably integrates or modulates neural circuits involved in attention, reward, and social processing. For incongruent JA states (versus congruent), there was increased activation in right TPJ and bilateral cerebellar cortices.

Our results represent an integration of several current schools of thought regarding the brain

mechanisms supporting JA processes. Previous work has separately implicated neural circuits supporting social perception (see Mundy & Newell, 2007 for review), reward circuits (Shilbach et al., 2010), and attention circuits (Mundy et al., 2009). Here, we attempt to integrate and demonstrate the involvement of all of these different systems using an innovative, naturalistic, dynamic, and contingent fMRI paradigm. Our study adds to current efforts to overcome a lack of relatively naturalistic fMRI paradigms (Redcay et al., 2012). To our knowledge, there are very few demonstrations (Shilbach et al., 2010) of the use of dynamic, contingent eye-tracking technology to create real-time interactive JA paradigms. There are none that examine

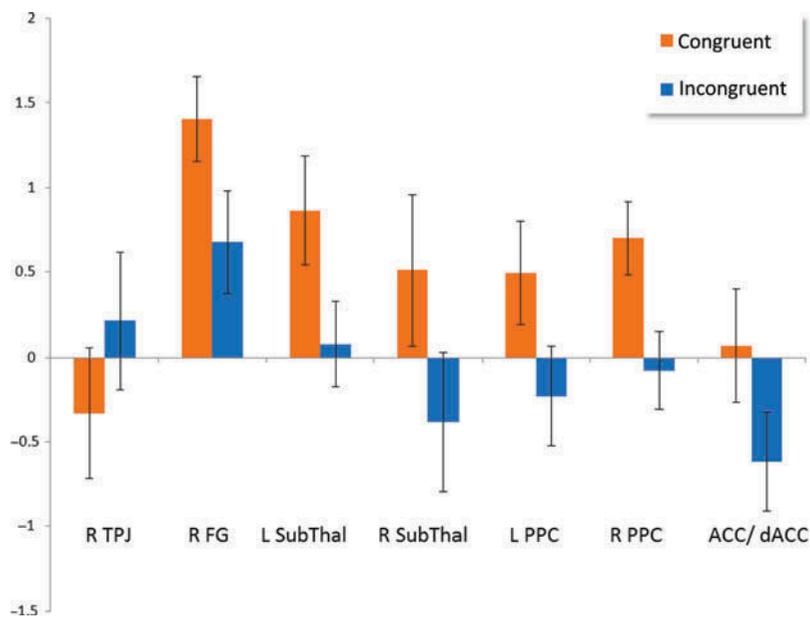


Figure 4. Bar graphs showing a comparison of beta values within brain regions, depicted in Figure 3, showing a significant difference between congruent and incongruent trials. Colors represent a possible model depicting each brain region's contribution to social perception, attention, or reward circuitry. All pairwise comparisons are significantly different at a $p < .005$. Error bars indicate standard errors of the means. R = right; L = left; FG = fusiform gyrus; PPC = posterior parietal cortex; ACC = anterior cingulate cortex; dACC = dorsal anterior cingulate cortex; TPJ = posterior temporoparietal junction; SubThal = subthalamic regions plotted includes the R amygdala, striatum and ventral tegmental area.

what occurs when either a congruent or an incongruent response follows bids for JA *specifically toward social (human avatars) stimuli*. We demonstrate here that one such attempt allows us to develop a more comprehensive understanding of the brain correlates of JA and indicate the integration of reward, social, and attentional mechanisms in initiating congruent JA states. In our paradigm, we chose to use identical avatars so that participants' choice would not be skewed by physical attributes. Future studies may introduce stimuli with increased real-life attributes and perhaps defined inherent or physical attributes (such as gender or emotional state), so that participants' choices even better resemble real-life JA processes.

Within a network of social processing, we include activations in the right FG which has been highly implicated in face perception (Hasson, Hendler, Ben Bashat, & Malach, 2001; McCarthy, Puce, Gore, & Allison, 1997; Rossion et al., 2003; Shah et al., 2001) and perhaps represents increased attention to "Sally's" face during congruent versus incongruent trials. The role of the amygdala in social processing is extensive, complex, and dynamic (for a review, see Adolphs, 2010). Most relevant for our study, the amygdala may play an integrative role in informing live social interactions with their concurrent hedonic value (Redcay et al., 2010). We suggest that amygdala activations may denote an increased rewarding value to congruent

responses from a social partner following online bids for shared attentional states. When Schilbach et al. (2010) compared congruent versus incongruent bids for JA, they captured brain activation only in the bilateral ventral striatum. These results, confined to a specific region of the reward network, are expanded in our results, which also include other brain regions involved in reward as well as attentional and social perception nodes. Perhaps our attempt at increased ecological validity, using a real face and not an animated one, as well as directing gaze toward human avatars, resulted in a more comprehensive pattern of brain activation. It should be noted here that the neural network underlying reward-related behavior is a complex one comprising many nodes, including ventral striatum, amygdala, midbrain regions (like the VTA), and ventromedial prefrontal cortex regions (for reviews, see Berridge & Robinson, 2003; Haber & Knuston, 2010). This may also explain why the above-mentioned study of JA reported activation in ventral striatum whereas we uncovered activations in more dorsal regions. Recently it has been suggested that ventral striatum activations are more likely in anticipation for reward (e.g., "wanting"), whereas ventromedial prefrontal activations are more involved in consumption ("liking") of rewarding stimuli (Kohls, Chevallier, Troiani, & Schultz, 2012). Indeed, it is also a possibility that the ACC/dACC region in the

current study is involved in integrating reward-related processes into attention reorientation during social perception. Considering we confined our GLM contrast to moments of shared attention (and not the entire sequence of bidding for JA), it is possible that activation of the VTA together with the ACC/dACC represents a signaling of the hedonic “liking” value of JA states. Future investigations into the neural mechanism of real-time JA are required in order to tease apart these important subtleties within the reward network.

The ACC has been previously implicated in attention orientation (Sarter, Gehring, & Kozak, 2006), social function (Bartels & Zeki, 2000; Gallagher & Frith, 2003; Mundy, 2003), and reward processing (Hampton & O’Doherty, 2007; Wallis & Kennerley, 2011). The ACC has widespread and multiple functions (Bissonette, Powell, & Roesch, 2013) and may be able to detect errors and utilize rewarding information to guide behavior and assign attention (Sarter et al., 2006). For instance, research has demonstrated that the ACC’s critical role in learned behavior is in directing voluntary choices based on the history of actions and outcomes (Kennerley, Walton, Behrens, Buckley, & Rushworth, 2006). In the current study, activations in PPC as well as ACC possibly represent an integration of posterior and anterior attentional brain regions (Corbetta & Shulman, 2002; Petersen & Posner, 2012), previously implicated in discrete aspects of JA processes (Mundy & Newell, 2007). Although these authors postulate that posterior regions are more reflexive attentional systems involved in responding to JA bids whereas anterior regions take part in initiating JA intentionally, they also state that these two regions interact and integrate in JA processes, especially as the JA system undergoes development and repeated experience (Mundy et al., 2009).

In contrast, when initiating JA is met with an incongruent response, we can identify the right posterior TPJ and the cerebellum as having increased activation. The right TPJ has been studied for its role in social information processes like empathy, morality, theory-of-mind, and inferring other’s mental states (Saxe, Whitfield-Gabrieli, Scholz, & Pelphrey, 2009; Young, Dodell-Feder, & Saxe, 2010; Young & Saxe, 2008). This brain region has also been implicated in attention reorientation (Corbetta & Shulman, 2002) and assessing intentional agency (Saxe, Xiao, Kovacs, Perrett, & Kanwisher, 2004). Some research suggests there are distinct regions in the right TPJ that activate selectively to theory-of-mind and mentalizing and lower-level attentional processes (Lombardo, Chakrabarti, Bullmore, & Baron-Cohen, 2011; Saxe

& Kanwisher, 2003; Scholz, Triantafyllou, Whitfield-Gabrieli, Brown, & Saxe, 2009), and yet others claim there is no such distinction (Mitchell, 2008). In a meta-analysis of fMRI studies, authors found the “a high degree of activation overlap in and around the right TPJ region for low-level and high-level cognitive functions” (Decety & Lamm, 2007). We believe that states of incongruency in our paradigm recruited both processes of reorienting and mentalizing. The activation in right TPJ in the current study may therefore be indicative of the relatively high level of ecological validity and real-life relevance of our paradigm, resulting in the integration of low-level computational processes such as attention reorientation into higher-level mentalizing abilities in the context of violation of social expectations. Our findings point to an internal expectation for shared attention states and their rewarding value. When this expectation is not met, activity in the right TPJ comes into play. Earlier work from our lab has implicated a region bordering the right TPJ in social incongruencies—the right pSTS was activated when responding to gaze incongruencies and violation of social expectations (Pelphrey, Morris, & McCarthy, 2005; Vander Wyk, Hudac, Carter, Sobel, & Pelphrey, 2009; Vander Wyk, Voos, & Pelphrey, 2012). Although the right TPJ and right pSTS are neighboring brain regions, future studies may be required to explore why the current study yielded results exclusively in the right TPJ for incongruent conditions and not also in the right pSTS.

The use of online eye-tracking as part of the fMRI paradigm in the current study was technically challenging, which explains why relatively few studies do so. We were able to show that in our predefined GAOIs, there was no difference in the proportion of gaze patterns that may account for our brain function results differentiating congruent and incongruent states. Nevertheless, a limitation of the current study is that we were not able to collect a richer eye-tracking behavioral outcome data set. Unfortunately, we were unable to examine aspects of participants’ gaze patterns that may have been highly informative, such as point-by-point shifts in attention from avatars’ or participants’ gaze toward “Sally’s” entire face. Our algorithm was designed to collect data only from several predefined regions and to collapse across individual fixations. We programmed the paradigm in this manner, mainly to increase video’s presentation speed and smoothness. Writing individual gaze data would have caused perceptible lags in the stimulus view. With current technological improvements, we encourage future studies to use eye-tracking in fMRI paradigms and to increase the richness of data collected to allow for a better understanding of imaging results.

Most striking is how incongruent responses to bids for JA dramatically reduce brain activity in several reward, social, and attentional nodes. This emphasizes that neural strategies and coordination of several brain circuits are impacted by social context, specifically here—the congruency of the response of a social agent to our actions. Considering the well-established deficits in the initiation of JA in populations with social developmental dysfunctions, most notably individuals with autism spectrum disorders, the current study's results deepen our understanding of the neural circuitry supporting this critical aspect of social cognition and may, in future work, inform our understanding and treatment of social dysfunction.

Original manuscript received 6 November 2012

Revised manuscript accepted 30 July 2013

First published online 13 September 2013

REFERENCES

- Adolphs, R. (2010). What does the amygdala contribute to social cognition? *Annals of the New York Academy of Sciences*, 1191, 42–61.
- Baron-Cohen, S. (2005). The empathizing system: A revision of the 1994 model of the mindreading system. In B. Ellis & D. Bjorklund (Eds.), *Origins of the social mind*. New York, NY: Guilford Publications.
- Bartels, A., & Zeki, S. (2000). The neural basis of romantic love. *NeuroReport*, 11(17), 3829–3834.
- Berridge, K. C., & Robinson, T. E. (2003). Parsing reward. *Trends in Neuroscience*, 26(9), 507–513.
- Bissonette, G. B., Powell, E. M., & Roesch, M. R. (2013). Neural structures underlying set-shifting: Roles of medial prefrontal cortex and anterior cingulate cortex. *Behavioural Brain Research*, 250, 91–101.
- Charman, T. (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions of the Royal Society London B: Biological Sciences*, 358, 315–324.
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, 3(3), 201–215.
- Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: How low-level computational processes contribute to meta-cognition. *The Neuroscientist*, 13(6), 580–593.
- Forman, S. D., Cohen, J. D., Fitzgerald, M., Eddy, W. F., Mintun, M. A., & Noll, D. C. (1995). Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): Use of a cluster-size threshold. *Magnetic Resonance Medicine*, 33(5), 636–647.
- Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of “theory of mind.” *Trends in Cognitive Sciences*, 7(2), 77–83.
- Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, 17, 535–543.
- Goebel, R., Esposito, F., & Formisano, E. (2006). Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping*, 27(5), 392–401.
- Gopnik, A., Slaughter, V., & Meltzoff, A. (1994). Changing your views: How understanding of visual perception can lead to a new theory of mind. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind*. Hove: Lawrence Erlbaum Associates.
- Haber, S. N., & Knuston, B. (2010). The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology*, 35, 4–26.
- Hampton, A. N., & O'Doherty, J. P. (2007). Decoding the neural substrates of reward-related decision making with functional MRI. *Proceedings of the National Academy of Sciences of the USA*, 104(4), 1377–1382.
- Hasson, U., Hendler, T., Ben Bashat, D., & Malach, R. (2001). Vase or face? A neural correlate of shape-selective grouping processes in the human brain. *Journal of Cognitive Neuroscience*, 13, 744–753.
- Kennerley, W., Walton, M. E., Behrens, T. E., Buckley, M. J., & Rushworth, M. F. (2006). Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience*, 9, 940–947.
- Kohls, G., Chevallier, C., Troiani, V., & Schultz, R. T. (2012). Social “wanting” dysfunction in autism: Neurobiological underpinnings and treatment implications. *Journal of Neurodevelopmental Disorders*, 4(10), 1–20.
- Legrand, D., & Iacobini, M. (2010). Inter-subjective intentional actions. In F. Grammont, D. Legrand, & P. Livet (Eds.), *Naturalizing intention in action. An interdisciplinary approach*. Cambridge, MA: The MIT Press and ENS Editions.
- Lombardo, M. V., Chakrabarti, M., Bullmore, E. T., & Baron-Cohen, S. (2011). Specialization of right parietal junction for mentalizing and its relation to social impairments in autism. *NeuroImage*, 56(3), 1832–1838.
- McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience*, 9(5), 605–610.
- Mitchell, J. P. (2008). Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cerebral Cortex*, 18(2), 262–271.
- Moore, C., & Dunham, P. J. (1995). *Joint attention: Its origin and role in development*. Hillsdale, NJ: Lawrence Erlbaum.
- Mundy, P. (2003). Annotation: The neural basis of social impairments in autism: the role of the dorsal medial-frontal cortex and anterior cingulate system. *Journal of Child Psychology and Psychiatry*, 44(6), 793–809.
- Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, 16(5), 269–274.
- Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research*, 2(1), 2–21.
- Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2005). Neural basis of eye gaze processing deficits in autism. *Brain*, 128(5), 1038–1048.
- Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Reviews of Neuroscience*, 35, 37–89.
- Redcay, E., Dodel-Feder, D., Mavros, P. L., Kleiner, M., Pearrow, M. J., Triantafyllou, C., . . . Saxe, R. (2012). Atypical brain activation patterns during a face-to-face

- joint attention game in adults with autism spectrum disorder. *Human Brain Mapping*, doi:10.1002/hbm.22086.
- Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *NeuroImage*, *50*(4), 1639–1647.
- Redcay, E., Kleiner, M., & Saxe, R. (2012). Look at this: The neural correlates of initiating and responding to bids for joint attention. *Frontiers in Human Neuroscience*, *6*(169), 1–14.
- Rossion, B., Caldara, R., Seghier, M., Schuller, A. M., Lazeyras, F., & Mayer, E. (2003). A network of occipito-temporal face-sensitive areas besides the right middle fusiform gyrus is necessary for normal face processing. *Brain*, *126*(11), 2381–2395.
- Saito, D. S., Tanabe, H. C., Izuma, K., Hayashi, M. J., Morito, Y., Komeda, H., . . . Sadato, N. (2010). “Stay tuned”: Inter-individual neural synchronization during mutual gaze and joint attention. *Frontiers in Integrative Neuroscience*, *4*, 1–12.
- Sarter, M., Gehring, W. J., & Kozak, R. (2006). More attention must be paid: The neurobiology of attentional effort. *Brain Research Reviews*, *51*, 145–160.
- Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in theory of mind. *NeuroImage*, *19*, 1835–1842.
- Saxe, R., Xiao, D. K., Kovacs, G., Perrett, D. I., & Kanwisher, N. (2004). A region of right posterior superior temporal sulcus responds to observed intentional actions. *Neuropsychologia*, *42*, 1435–1446.
- Saxe, R. R., Whitfield-Gabrieli, S., Scholz, J., & Pelphrey, K. A. (2009). Brain regions for perceiving and reasoning about other people in school-aged children. *Child Development*, *80*(4), 1197–1209.
- Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., . . . Vogeley, K. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience*, *22*(12), 2702–2715.
- Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., & Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia*, *44*, 718–730.
- Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., & Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS ONE*, *4*(3), e4869. doi:10.1371/journal.pone.0004869.
- Senju, A., & Johnson, M. H. (2009). The eye contact effect: Mechanisms and development. *Trends in Cognitive Sciences*, *13*, 127–134.
- Shah, N. J., Marshall, J. C., Zafiris, O., Schwab, A., Zilles, K., & Markowitsch, H. J. (2001). The neural correlates of person familiarity, A functional magnetic resonance study with clinical implications. *Brain*, *124*(4), 804–815.
- Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. Dunham (Eds.), *Joint attention: Its origins and role in development*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–735.
- Vander Wyk, B. C., Hudac, C. M., Carter, E. J., Sobel, D. M., & Pelphrey, K. A. (2009). Action understanding in the superior temporal sulcus region. *Psychological Science*, *20*(6), 771–777.
- Vander Wyk, B. C., Voos, A., & Pelphrey, K. A. (2012). Action representation in the superior temporal sulcus in children and adults: An fMRI study. *Developmental Cognitive Neuroscience*, *2*(4), 409–416.
- Wallis, J. D., & Kennerley, S. W. (2011). Contrasting reward signals in the orbitofrontal cortex and anterior cingulate cortex. *Annals of the New York Academy of Sciences*, *1239*, 33–42.
- Williams, J. H. G., Waiter, G. D., Perra, O., Perrett, D. I., & Whiten, A. (2005). An fMRI study of joint attention experience. *NeuroImage*, *25*, 133–140.
- Young, L., Dodell-Feder, D., & Saxe, R. (2010). What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. *Neuropsychologia*, *48*(9), 2658–2664.
- Young, L., & Saxe, R. (2008). The neural basis of belief encoding and integration in moral judgment. *NeuroImage*, *40*(4), 1912–1920.