Author manuscript. Accepted for publication in *Human Brain Mapping, 39*(10). DOI:10.1002/hbm.24221

Title: Social interaction recruits mentalizing and reward systems in middle childhood

Running title: Mentalizing during social interaction

Authors: Diana Alkire[1,2*], Daniel Levitas[3], Katherine Rice Warnell[4], & Elizabeth Redcay[1,2]

[1] Department of Psychology, University of Maryland, College Park, MD, 20742
[2] Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD, 20742
[3] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, 47405
[4] Department of Psychology, Texas State University, San Marcos, TX, 78666

* Corresponding author

Correspondence:
Diana Alkire
Department of Psychology
University of Maryland
College Park, MD 20742
Tel.: 240-535-6862
Fax: 301-314-9566
Email: diana@umd.edu

Conflict of Interest: None

Keywords: social interaction, mentalizing, theory of mind, fMRI, middle childhood, social reward

Funding: This work was supported by the University of Maryland and a grant from the National Institute of Mental Health to E. R. [R01-MH107441].

Abstract

Social cognition develops in the context of reciprocal social interaction. However, most neuroimaging studies of mentalizing have used non-interactive tasks that may fail to capture important aspects of real-world mentalizing. In adults, social-interactive context modulates activity in regions linked to social cognition and reward, but few interactive studies have been done with children. The current fMRI study examines children aged 8–12 using a novel paradigm in which children believed they were interacting online with a peer. We compared mental and non-mental state reasoning about a live partner (Peer) versus a story character (Character), testing the effects of mentalizing and social interaction in a 2x2 design. Mental versus non-mental reasoning engaged regions identified in prior mentalizing studies, including the temporoparietal junction, superior temporal sulcus, and dorsomedial prefrontal cortex. Moreover, peer interaction, even in conditions without explicit mentalizing demands, activated many of the same mentalizing regions. Peer interaction also activated areas outside the traditional mentalizing network, including the reward system. Our results demonstrate that social interaction engages multiple neural systems during middle childhood and contribute further evidence that social-interactive paradigms are needed to fully capture how the brain supports social processing in the real world.

Introduction

Social interaction shapes our daily experiences, personalities, and wellbeing throughout the lifespan, yet its biological mechanisms are underexplored. Mentalizing—the process of attributing mental states to others, also known as theory of mind—is necessary for successful social interactions, and thus there has been considerable effort in the last two decades to explicate its neural bases. Functional magnetic resonance imaging (fMRI) studies have identified several regions that consistently show greater activation during tasks that require mental state reasoning, including the temporoparietal junction (TPJ), superior temporal sulcus (STS), anterior temporal lobes (ATL), dorsomedial prefrontal cortex (dMPFC), inferior frontal gyrus (IFG), posterior cingulate cortex (PCC), and precuneus (meta-analyses: Mar, 2011; Molenberghs et al., 2016; Schurz et al., 2014). However, this "mentalizing network" has been characterized mainly by non-interactive tasks that use artificial stimuli such as photographs of faces, animated shapes, or stories about fictional characters. This lack of engagement with a live social partner is a crucial limitation in light of recent work suggesting that participating in social interaction profoundly alters social-cognitive processes (reviewed in Schilbach et al., 2013). Conversely, extant studies on the effect of social interaction on brain function lack the experimental controls needed to directly examine whether and how brain activity differs when mentalizing occurs within social interaction versus observation ("offline").

Prior neuroimaging research has shown that components of social interaction activate regions within the mentalizing network. For instance, dMPFC and STS have been found in studies examining communicative intent via eye gaze or gestures directed at the participant versus a third party (Ciaramidaro et al., 2014; Kampe et al., 2003; Redcay et al., 2016; Schilbach et al., 2006; cf. Calder et al., 2002; Conty et al., 2007). Furthermore, joint attention, in which two

people coordinate attention to a shared target, compared with solo attention activates posterior STS, TPJ, and dMPFC (Caruana et al., 2015; Redcay et al., 2010; Redcay et al., 2012; Schilbach et al., 2010). Recent work from our group has shown a similar effect even when participants do not engage in reciprocal interaction. Simply hearing short spoken vignettes with no explicit social-cognitive demands activated left TPJ and right dMPFC more when the participants believed the speech was live than when they knew it was pre-recorded (Rice & Redcay, 2016), suggesting that the mere presence of a potential social partner is sufficient to automatically engage the mentalizing network.

A compelling interpretation of these findings is that each task, though not requiring overt mental state reasoning, nevertheless evoked spontaneous mentalizing. However, the validity of this reverse inference is threatened by the apparent heterogeneity of function of the brain regions in question, particularly the TPJ (Corbetta et al., 2008; Lee & McCarthy, 2016; Schuwerk et al., 2016), STS (Redcay, 2008), and dMPFC (Isoda & Noritake, 2013), all of which have been linked to domain-general processes in addition to social cognition. Moreover, it remains unclear whether regions engaged in offline mentalizing are precisely the same as those recruited during social interaction. Given the evidence of functional segregation within regions broadly implicated in social cognition (e.g., Gilbert et al., 2007; Krall et al., 2015; Mars et al., 2012), we cannot rule out the possibility that adjacent areas are differentially involved in social interaction versus offline mentalizing, and such distinctions may be obscured when comparing activation across samples and task designs. The gap in our understanding of how the brain's mentalizing system is affected by interactive context can only be bridged by paradigms that manipulate both social interaction and mentalizing demands within the same task and participants.

One commonly used paradigm does incorporate both elements within the same task: a strategic game in which participants play against a supposedly human partner and must ascribe mental states to their opponents to predict their next move (e.g., Gallagher et al., 2002). Human conditions are contrasted with conditions in which responses are computer generated; thus, the computer conditions are neither socially interactive nor do they contain explicit mentalizing demands. Although such tasks suggest that social-interactive and offline mentalizing involve similar regions (e.g., Coricelli & Nagel, 2009; Gallagher et al., 2002; Kircher et al., 2009; McCabe et al., 2001), they cannot directly speak to any differences between types of mentalizing because they conflate social interaction and mentalizing within the same condition. Identifying the role of mentalizing regions in social interaction more broadly necessitates closely matched conditions contrasting mental and non-mental reasoning within both the social interaction and an offline control task.

Furthermore, mentalizing during social interaction may involve brain systems beyond the mentalizing network. In line with evidence that social interactions are inherently rewarding (Chevallier et al., 2012), Redcay et al. (2010) found greater activation of the reward system (including ventral striatum and amygdala) when participants interacted with an experimenter through a live video feed versus watching a recording of the same interaction. Other studies have also shown that reward-related regions respond to social-interactive context, such as gaze-based interactions (Pfeiffer et al., 2014), initiating joint attention (Schilbach et al., 2010), and considering whether to share information with others (Baek et al., 2017). Paradigms that elicit mentalizing while simultaneously capturing the motivational processes that likely differ between interactive and non-interactive contexts will provide a more holistic understanding of how we perceive other minds in real time.

Previous neuroimaging work in this area has also focused overwhelmingly on adults. Middle childhood (roughly ages 7 to 13) is particularly understudied, despite evidence of significant social and neurocognitive development in this age range. Peer interactions become more complex (Bigelow, 1977; Farmer et al., 2015; Feiring & Lewis, 1991), and this increasing sophistication in social behavior may be accompanied by advances in social cognition (reviewed in Miller, 2009; Devine & Hughes, 2013). There is also evidence that across middle childhood, the TPJ becomes increasingly selective for representing mental states as opposed to more general social information, as revealed by an offline story-based task (Gweon et al., 2012). Still, as in the adult literature, neuroimaging studies on the effect of social interaction on social cognition in middle childhood are scarce. In one such study, similar to the aforementioned study in adults (Rice & Redcay, 2016), perceived live versus recorded speech engaged the TPJ and precuneus in children aged 7–13 (Rice et al., 2016). In a separate experiment in a similar age group, receiving feedback from a peer after sharing information about oneself activated social-cognitive and reward regions, and the magnitude of the social-interactive effect in social-cognitive regions increased with age (Warnell, Sadikova, & Redcay, 2018). However, as discussed above, because these tasks lacked explicit mentalizing demands, we cannot definitively infer that mentalizing (and not some other computation relevant to social processing) occurred during social-interactive conditions, nor can we directly compare activation patterns associated with social-interactive versus offline mentalizing.

The present study is the first (to our knowledge) to employ a two-by-two factorial design in which the effects of social context and mentalizing can be simultaneously examined. Inside the MRI scanner, children aged 8–12 engaged in a social prediction task in which they believed they were interacting with a peer in another laboratory (Peer condition) and answering questions

about a fictional character (Character condition). Across Peer and Character conditions, half the trials required the children to use mental state information when making predictions (Mental condition), while the other half did not (Non-Mental condition).

We hypothesized that regions of the traditional mentalizing network would be activated by the Mental versus Non-Mental contrast regardless of social-interactive context. We further hypothesized that mentalizing regions would be activated more in Peer than in Character conditions, suggestive of spontaneous mentalizing during social interaction regardless of explicit task demands, as in our previous studies (Rice et al., 2016; Rice & Redcay, 2016). Further, through conjunction analysis, we determined the extent to which engagement in social interaction recruits the same neural resources as mentalizing did in the offline task.

The 2x2 factorial design also allowed us to assess whether there is an interaction effect between social interaction (Peer vs. Character) and explicit mentalizing demands (Mental vs. Non-Mental), though we considered several possible hypotheses. One possibility is that mentalizing regions show a greater difference in activation between Mental and Non-Mental conditions in the Peer as opposed to Character conditions, with the Peer Mental condition showing the greatest activation, which would suggest an additive effect of social interaction and explicit mentalizing demands. On the other hand, there may be less difference in activation of mentalizing regions between the two Peer conditions relative to the Character conditions. In other words, while we expect certain regions to show significantly more activation in Character Mental than in Character Non-Mental conditions, the Peer conditions might elicit a similar amount of activation in these regions regardless of whether the task contains explicit mentalizing demands, again suggesting that engaging with a social partner is sufficient to induce spontaneous mentalizing.

Beyond mentalizing regions, we predicted that the Peer versus Character contrast would activate reward regions such as the striatum and orbitofrontal cortex (OFC), in line with previous social-interactive experiments (Pfeiffer et al., 2014; Warnell et al., 2018). Lastly, we examined whether our results would replicate previous findings that social-cognitive regions become increasingly specialized for mentalizing (Gweon et al., 2012) and social interaction (Warnell et al., 2018) across middle childhood. Altogether, the present study aims to capture the neural effects of social interaction during a dynamic yet understudied period of social development.

Materials and Methods

Participants

Children were recruited using a database of families in the Washington, D.C., metropolitan area. Exclusionary criteria were any MRI contraindications, diagnosis of neurological or psychiatric disorders, or first-degree relatives with autism or schizophrenia. All participants were full-term, native English speakers. Thirty-five typically developing children aged 8–12 years participated in the study. Seven children were excluded from data analysis—two for excessive motion in the scanner, one due to a technical error during scanning, three for not believing the live illusion, and one who scored in the "moderate" range on the Social Responsiveness Scale, indicating clinically significant deficits in social interaction (Constantino & Todd, 2003)—leaving a final sample of 28 children (14 females; mean age = 10.41 years, SD = 1.46 years, range = 8.18–12.98 years). We obtained informed assent from all participants and informed consent from their parents or guardians. All procedures were approved by the University of Maryland Institutional Review Board.

Task procedures

Creating the live illusion. Before the scan, children were told they would be interacting ("chatting") with a peer in a different laboratory who would also be undergoing an MRI scan. During a demonstration of the chat (see Supplementary Materials), children learned they would chat with their partners only half the time; for the other half, they would answer questions provided by a computer about a fictional character of the same gender and age as the participant. Participants were then shown photos of two children (and had their own photo taken to enhance the live illusion), both matched to the participant's age and gender, and were told to choose one to be their chat partner (Supplementary Figure S1). Photos were selected from the NIMH Child Emotional Faces Pictures Set (smiling, direct gaze only; Egger et al., 2011), as well as from Getty Images (www.gettyimages.com) and Google Images search to attain racial and ethnic diversity reflective of our participant population.

fMRI task design. In the scanner, children played the role of the "guesser" in a social prediction game. In each trial they received a one-sentence hint about either their chat partner or a fictional character in a story (see Supplementary Materials for examples), then answered either "Which will I/she/he pick?" (Mental) or "Which of these match?" (Non-Mental) by choosing via button-press between two choices. Each trial was divided into two phases: "Guess" (8 s), including the hint and choice periods, and "Feedback" (2 s), in which participants learned whether their choices matched those of the chat partner or the computer (Figure 1). The task contained 96 trials. In 48 trials, the hints described mental states such as knowledge, beliefs, desires, preferences, and emotions (Mental). The other 48 hints described facts or situations about the peer or character but made no reference to mental states (Non-Mental). Furthermore, 48 trials (24 Mental, 24 Non-Mental) were presented in the first-person (Peer) and the other 48 in the third-person perspective (Character), yielding four conditions: Peer Mental, Peer Non-Mental,

Character Mental, and Character Non-Mental. Individual trials were counterbalanced across participants between Peer and Character conditions. Throughout each trial, either the chat partner's name (Peer) or the word "Computer" (Character) was displayed at the top of the screen.

Stimuli presentation. The task was presented using PsychoPy (Peirce, 2009) in four runs of 24 trials (24 trials per condition total). Guess and Feedback periods were separated by a fixation cross presented for a jittered 2-6 s, centered around 3.5 s and distributed exponentially. Trials were separated by a fixation cross with the same jittered parameters. Trial distribution and inter-stimulus/trial intervals were optimized using Design Explorer (Moraczewski et al., unpublished software), which minimizes collinearity between events in the design matrix. The resulting matrix was submitted to AFNI's 1d_tool program (Cox, 1996) to confirm that correlations between regressors of interest were minimal. A fixation cross was presented for 10 s at the beginning and 15 s at the end of each run. To maintain the live illusion, the chat partner's photo appeared at the end of every run.

Post-test questionnaire. After the scan, participants answered a series of questions in which they rated on a scale of 1 to 5 their preference for and attention to the live partner versus the computer. The post-test also probed participants' belief in the live illusion (see Supplementary Materials). Three participants who expressed disbelief in the live illusion during the post-test or debriefing were excluded from analysis.

Image acquisition & preprocessing

fMRI data were acquired at the Maryland Neuroimaging Center on a 3.0 Tesla scanner with a 32-channel head coil (MAGNETOM Trio Tim System, Siemens Medical Solutions). Four runs of the task were acquired using multiband-accelerated echo-planar imaging (66 interleaved axial slices, multiband factor = 6, voxel size = 2.19 x 2.19 x 2.20 mm, repetition time = 1250 ms,

echo time = 39.4 ms, flip angle = 90°, pixel matrix = 96 x 96) followed by a structural scan (three-dimensional T1 magnetization-prepared rapid gradient-echo sequence, 192 contiguous sagittal slices, voxel size = 0.45 x 0.45 x 0.90 mm, repetition time = 1900 ms, echo time = 2.32 ms, flip angle = 9°, pixel matrix = 512 x 512). Data were preprocessed using AFNI (Cox, 1996). Functional scans were slice-time corrected. The structural scan was aligned to the first volume of a functional run and normalized to the Haskins pediatric template (nonlinear; Molfese et al., 2015) using a 12-parameter affine transformation, which was then applied to all functional volumes. Finally, functional data were spatially smoothed with a 5 mm full-width half-maximum Gaussian kernel and intensity normalized to a mean of 100 per voxel.

Time points for which framewise displacement (FD) of two consecutive volumes exceeded 1 mm were censored in subsequent analyses, and runs were excluded if 10% or more of the volumes would be censored or if mean FD was 0.50 mm or greater. Two participants with fewer than three usable runs were excluded from analyses, leaving a final sample of 20 children with four runs and eight with three runs.

Data analysis

fMRI data were analyzed in AFNI using general linear models. At the first level, events of interest (Guess periods for Peer Mental, Peer Non-Mental, Character Mental, Character Non-Mental conditions) were convolved with the canonical hemodynamic response function using a duration modulated response function (AFNI's dmBlock). Guess and Feedback were modeled as separate events, with only the Guess periods analyzed as events of interest, as they were designed to capture the mentalizing processes relevant to the current study. To exclude task-irrelevant cognition that might have occurred between the participant's response and the end of the response window, duration modulation was performed based on the reaction time (RT) at

each Guess event, such that each modeled Guess period only lasted until the child responded. Regressors of no interest included the four Feedback conditions, the six motion parameters (x, y, z, roll, pitch, and yaw) and their derivatives, time points censored due to FD greater than 1 mm, and polynomial terms (constant, linear, quadratic, and cubic) to model baseline and scanner drift.

At the second level, whole-brain comparisons between the four conditions were generated using mixed-effects multilevel analysis (3dMEMA; Chen et al., 2012) to model within- and between-subject variability. In addition to the main effect of mentalizing ([Peer Mental + Character Mental] vs. [Peer Non-Mental + Character Non-Mental]), the main effect of social interaction ([Peer Mental + Peer Non-Mental] vs. [Character Mental + Character Non-Mental]), and their interaction, we conducted pairwise comparisons to isolate the effect of mentalizing in the offline (Character Mental vs. Character Non-Mental) and social-interactive (Peer Mental vs. Peer Non-Mental) contexts separately, as well as the effect of social interaction within Mental and Non-Mental conditions respectively (Peer Mental vs. Character Mental; Peer Non-Mental vs. Character Non-Mental). Each model included age and mean FD as covariates. The same Haskins pediatric template used to normalize the data was resampled to match the functional data and then used as a structural mask (i.e., only voxels within this mask were analyzed). Contrast maps were first thresholded at $p < 0.005$ (2-tailed), then cluster corrected at alpha = 0.05 ($k = 86$, bi-sided, second nearest-neighbor). The cluster-size threshold was determined by averaging individual participants' non-Gaussian spatial autocorrelation function parameters and inputting these values (a = 0.51, b = 2.91, c = 7.26) to 3dClustSim according to recent recommendations (Cox et al., 2017).

To determine regions active during both offline mentalizing and social interaction without explicit mentalizing demands, we performed a conjunction analysis by multiplying the

binarized, corrected group maps for the Character Mental > Character Non-Mental and Peer Non-Mental > Character Non-Mental contrasts to identify voxels significant for both contrasts (Nichols et al., 2005).

Analysis of behavioral performance and regions of interest (ROI) were conducted in R (R Core Team, 2016). RT in seconds and accuracy (percent correct responses) were each entered into a two-way repeated measures ANOVA to determine the main effects of mentalizing (Mental vs. Non-Mental) and social interaction (Peer vs. Character), and their interaction. Significant results were followed up with paired *t*-tests. Post-test questionnaire data were analyzed using Wilcoxon signed rank tests to compare ordinal ratings between Peer and Character conditions.

As a post-hoc exploration of activation within the mentalizing network during social interaction, we used the Character Mental > Character Non-Mental contrast to define "offline mentalizing" ROIs, then extracted individual beta values for each condition versus baseline. To examine the relationship between age and activation of mentalizing regions, we extracted individual beta values for each condition versus baseline from ROIs defined by the Mental > Non-Mental contrast (without age as a covariate), then created difference scores for Mental versus Non-Mental and Peer versus Character conditions, respectively. We conducted partial correlations between these scores and age, controlling for mean FD, which was significantly correlated with age ($r = -.38$, $p < 0.05$). For the ROI analyses, *p*-values were corrected for multiple comparisons using Holm's sequentially rejective Bonferroni test, which is more powerful than the classical Bonferroni test (Holm, 1979).

Results

Behavioral

Overall in-scanner performance was high (mean accuracy = 91% correct, SD = 7%; mean RT = 2.04 s, SD = 0.27 s). A repeated measures ANOVA indicated a significant main effect of social interaction (Peer vs. Character) on RT ($F(1, 27) = 11.61$, $p < 0.005$; Figure 2A); a paired *t*-test revealed that children responded more quickly in Peer than in Character conditions (mean difference = 0.07 s, $t(55) = 3.74$, $p < 0.001$). The main effect of mentalizing (Mental vs. Non-Mental) and the interaction term were not statistically significant for RT ($p > 0.05$). No statistically significant effects were found for accuracy.

Post-test questionnaires indicated a general preference for Peer over Character conditions (Figure 2B). Specifically, participants gave significantly higher ratings (on an ordinal scale of 1–5) for how much they liked interacting with their partners versus answering questions from the computer (median Peer = 5, median Character = 3, $p < 0.001$) and how much they liked guessing what their partners would pick versus guessing what would come next in the story (median Peer = 4, median Character = 3, $p < 0.005$). There was a trend of children reporting that they paid more attention during Peer than Character conditions (median Peer = 4, median Character = 4, $p = 0.05$).

Neuroimaging

Effect of mentalizing. Whole-brain analyses revealed a main effect of mentalizing ([Peer Mental + Character Mental] vs. [Peer Non-Mental + Character Non-Mental]) in several regions identified in previous mentalizing studies, including right dMPFC**,** left TPJ, and bilateral STS and ATL (Figure 3, Table 1). A similar pattern of activation emerged for the pairwise comparison of Character Mental versus Character Non-Mental, albeit in smaller clusters and without TPJ or dMPFC (Figure 4, Table 1). In contrast, no regions were significantly more active for Peer Mental than Peer Non-Mental.

Effect of social interaction. A test of the main effect of social interaction ([Peer Mental + Peer Non-Mental] vs. [Character Mental + Character Non-Mental]) revealed extensive activation, including anterior and posterior midline regions (dMPFC, medial OFC, ACC, PCC, precuneus); bilateral IFG and lateral OFC; bilateral insula; bilateral STS and ATL; bilateral inferior parietal cortex extending into TPJ; medial occipital regions (cuneus, pericalcarine, and lingual cortex) extending into the fusiform gyri, parahippocampal gyri, and hippocampus; bilateral middle and left inferior temporal cortex; and subcortical structures (striatum, amygdala, thalamus, and cerebellum; Figure 3, Table 1). Most of the same regions were activated to a lesser extent by the pairwise comparison of Peer Non-Mental vs. Character Non-Mental (Figure 4, Table 1). The contrast of Peer Mental vs. Character Mental yielded still more limited activation along the same general patterns, with notably less activation in bilateral STS and ATL and no activation in right TPJ or bilateral IFG (Figure 4, Table 1).

Interaction effect. Whole-brain analysis of the interaction term ([Peer Mental vs. Peer Non-Mental] vs. [Character Mental vs. Character Non-Mental]) revealed no significant activation.

Shared regions for mentalizing and social interaction. To examine shared regions for mentalizing and social interaction, we conducted a conjunction analysis to identify voxels that were significantly activated for both Character Mental > Character Non-Mental (i.e., the offline mentalizing task) and Peer Non-Mental > Character Non-Mental (i.e., social interaction with no explicit mentalizing demands). This analysis revealed overlapping activation in bilateral ATL, right posterior STS, left lateral OFC and insula, and right IFG (Figure 5, Table 2).

We next examined activation within the offline mentalizing ROIs (Character Mental > Character Non-Mental; Figure 6). Paired *t*-tests indicated non-significant differences between

Peer Mental and Peer Non-Mental conditions in all ROIs except right ATL (mean difference = 0.10, $t(27) = 3.60$, $p < 0.01$ corrected). Comparison between Character Mental and Peer Non-Mental conditions revealed no significant differences in activation in any regions. Altogether, this ROI analysis suggests that the two Peer conditions elicited similar activation of mentalizing regions regardless of task demands.

Age effects. Whole-brain analysis showed no significant effects of age on the Mental versus Non-Mental contrast. A follow-up ROI analysis found no significant correlations between age and mentalizing activity within mentalizing ROIs (Mental > Non-Mental).

Conversely, the whole-brain Peer versus Character contrast revealed a negative effect of age in many frontal, temporal, insular, and subcortical areas (Supplementary Table S1, Figure 7A). Analysis of the same mentalizing ROIs as above found that age was significantly ($p < 0.05$ corrected) negatively correlated with activation to Peer versus Character conditions in right ATL (r = -0.47), left ATL/lateral OFC/insula (r = -0.51), dMPFC (r = -0.43), right STS (r = -0.59), left STS (r = -0.42), and left TPJ (r = -0.55). However, correlations between age and average activation to Peer and Character conditions, respectively, did not reach significance.

Discussion

This study examined the effect of perceived social interaction on brain activation in the context of a mentalizing task performed by children aged 8–12. By manipulating both social interaction and mentalizing within the same participants, we were able to directly assess shared and distinct neural mechanisms associated with each factor. Social interaction engaged many of the same regions as the offline mentalizing task, even in the absence of explicit mentalizing demands from the task. Moreover, social interaction elicited more extensive activation in some regions associated with mentalizing, as well as regions outside the mentalizing network,

including the reward system. These results illuminate an understudied period of development and underscore the need for social-interactive paradigms to accurately characterize real-world social processing.

Our hypothesis regarding the main effect of mentalizing was broadly supported. That is, across Peer and Character conditions, the Mental versus Non-Mental contrast revealed a pattern of activation consistent with the prior literature (Schurz et al., 2014). These results add to the sparse literature on the neural correlates of social cognition in middle childhood by showing that the mentalizing system characterized in adults is generally established by ages 8–12.

Examination of the main effect of social interaction revealed greater activation for Peer versus Character conditions in all major components of the mentalizing network, including anterior and posterior midline and lateral temporal regions. In line with our hypothesis that spontaneous mentalizing occurs during social interaction in the absence of explicit mentalizing demands, a similar activation pattern emerged for the Peer Non-Mental versus Character Non-Mental contrast. As a stronger test of this interpretation, we performed a conjunction analysis to identify specific regions activated by both the offline mentalizing task and social interaction without mentalizing demands, which revealed several overlapping areas. Additionally, ROI analyses suggested that offline mentalizing regions were similarly activated by Character Mental, Peer Mental, and Peer Non-Mental conditions (Figure 6). Furthermore, the whole-brain contrast of Peer Mental > Peer Non-Mental revealed no significant activation, consistent with there being comparable recruitment of mentalizing regions in both Peer conditions. Finally, though dMPFC and bilateral TPJ—the regions most consistently activated across previous mentalizing studies (Schurz et al., 2014)—were not significantly activated by our offline mentalizing task (Character Mental > Character Non-Mental), they were engaged by social interaction (e.g., Peer Non-

Mental > Character Non-Mental; Figure 4). Together, these findings provide the strongest evidence to date that social interaction induces mentalizing even when the task does not explicitly require it. Additional support for this could come from future studies that link activation of mentalizing regions during social interaction to measures outside the scanner of mentalizing ability or propensity.

Our ROI analysis indicated that for most regions that showed a significant difference between Character Mental and Character Non-Mental conditions, activation was similar for Peer Mental and Peer Non-Mental conditions. Based on this, we might have expected mentalizing regions to show a 2 (Peer vs. Character) x 2 (Mental vs. Non-Mental) interaction effect at the whole-brain level, but this was not the case, probably due to a lack of statistical power. We also did not find the opposite interaction pattern, i.e., a greater difference in activation between Mental and Non-Mental in Peer versus Character conditions. Such a finding may have indicated an additive effect of social context and explicit mentalizing demands such that activation of certain regions would be greatest in the Peer Mental condition. This effect, if it exists, may be revealed by a future study with a larger sample size, or through analysis of a different set of ROIs than those examined in the current study.

We were also interested in how social interaction modulates the reward network. Taken together, our neuroimaging and behavioral results suggest that participants found Peer conditions more rewarding or motivating than Character conditions. The main effect of social interaction revealed activation in several components of the reward system, including medial OFC, dorsal and ventral striatum, thalamus, and amygdala (Berridge & Kringelbach, 2015; Schultz, 2015). Supporting our interpretation of this activation as reflecting subjective feelings of motivation and reward, participants' responses to the post-test questionnaire indicated greater enjoyment for

Peer than Character conditions. Because the Feedback period was modeled as a covariate of no interest, it is unlikely that this activation reflects positive feelings directly resulting from participants learning that their responses matched those of their peers. Rather, our results could be driven by anticipation of such a reward in the Feedback period, hedonic response to the Guess period itself, or both. Additionally, faster responses in Peer than in Character trials may reflect the participants' heightened motivation to interact with their partners. Overall, our results add to extant evidence that social interaction is intrinsically rewarding (Chevallier et al., 2012; Pfeiffer et al., 2014; Schilbach et al., 2010; Warnell et al., 2018).

We also found that social interaction recruits areas outside both the mentalizing and reward networks. Peer more than Character conditions activated large portions of medial occipital cortex, which has been associated with mental imagery (Kosslyn et al., 1999; Kosslyn et al., 1995), and medial temporal regions linked to memory processes (Eichenbaum et al., 2007). Because participants were shown photos of their partners before the scan and after each run, this activation could result from their recollection of these images during Peer trials. Future studies should explore whether visualization of one's social partner is inherent to social interaction (especially when one's partner is physically remote), as well as how social processing interacts with memory encoding and retrieval.

We found no effect of age on activation of mentalizing regions to Mental versus Non-Mental conditions. Instead, activation of these regions to Peer versus Character conditions decreased with age. These results are at odds with previous findings that over middle childhood, mentalizing regions become increasingly selective for belief representation (although our age range is narrower and slightly older than that of Gweon et al., 2012, which included ages 5–11) and social interaction (Warnell et al., 2018). However, these findings should be interpreted with

caution given the possibility that the current and previous studies of this nature are underpowered to detect what may be subtle between-subjects effects (e.g., Cremers et al., 2017). Nevertheless, our analysis of mentalizing ROIs showed a consistent pattern, with differences in activation to Peer versus Character conditions decreasing with age in all ROIs. Though correlations between age and activation to Peer and Character conditions, respectively, did not reach significance, the decreasing difference may have been driven by increasing mentalizing in response to Character but not Peer conditions, which would accord with previous findings of increasing activation of dMPFC to non-interactive social stimuli across middle childhood (Rice et al., 2016) and in adolescence relative to adulthood (reviewed in Blakemore, 2008). It is also possible that our task is more similar to the real-life peer interactions of younger than older children. Prior research suggests that while younger children's friendships are based around common activities and other superficial aspects, children approaching adolescence increasingly value "empathy, understanding, and self-disclosure" (Bigelow, 1977)—in other words, a level of intimacy unattainable within the constraints of our paradigm and with an unfamiliar peer. Still, these results warrant further investigation using larger—and ideally, longitudinal—samples to more firmly establish how the social-interactive brain develops from childhood through adolescence.

Another limitation of our modest sample size is that we were unable to assess gender differences in brain activation related to mentalizing or social interaction. In adults, there is evidence of gender differences in the neural correlates of social cognition, though the direction of effects and the specific brain regions involved vary across studies (Adenzato et al., 2017; Frank, Baron-Cohen, & Ganzel, 2015; Krach et al., 2009; Veroude, Jolles, Croiset, & Krabbendam, 2014). In middle childhood, some behavioral studies indicate a female advantage for mentalizing (e.g., Devine & Hughes, 2013), which may relate to differential styles of

interacting with peers, with girls more likely to form intimate relationships that demand perspective-taking (Maccoby, 1990, as cited in Devine & Hughs, 2013). Whether these behavioral differences are mirrored by differences in brain activation during social interaction in middle childhood is yet unknown. Also unclear is whether the apparent gender differences pertain to mentalizing ability—which may be captured by offline tasks with explicit mentalizing demands—or the propensity to spontaneously mentalize in the context of a real-time social interaction. With a larger sample, our interactive mentalizing task may be particularly well-suited to answering these questions.

In sum, this study provides direct evidence that mentalizing and engagement with a social partner recruit many of the same neural substrates. Furthermore, social interaction elicits activation well beyond these offline mentalizing regions, including the reward system. Beyond advancing our nascent understanding of the social brain in middle childhood, the findings of this and other social-interactive studies may enable important insights into disorders such as autism spectrum disorder and social anxiety, which are defined by difficulties in real-world social interactions. Our ability to characterize these difficulties at the neural level hinges on developing an ecologically valid model of how the typical brain functions in the presence of other minds.

Acknowledgements

The authors wish to thank Ruth Ludlum, Sydney Maniscalco, Tova Rosenthal, Sabine Huber, Alex Mangerian, Zoey Maggid, Hunter Rogoff, Morgan Biggs, Alexandra Hickey, Laura Anderson Kirby, Dustin Moraczewski, and the Maryland Neuroimaging Center staff for their assistance with data collection and analysis. We also thank Drs. Peter Carruthers and Donald J. Bolger for their advice on task design and discussion of an early version of the manuscript.

References

Adenzato, M., Brambilla, M., Manenti, R., De Lucia, L., Trojano, L., Garofalo, S., … Cotelli, M. (2017). Gender differences in cognitive Theory of Mind revealed by transcranial direct current stimulation on medial prefrontal cortex. *Scientific Reports, 7*, 41219. https://doi.org/10.1038/srep41219

Baek, E. C., Scholz, C., O'Donnell, M. B., & Falk, E. B. (2017). The value of sharing information: a neural account of information transmission. *Psychological Science, 28*(7), 851–861..

Berridge, K. C., & Kringelbach, M. L. (2015). Pleasure systems in the brain. *Neuron*, *86*(3), 646–664. https://doi.org/10.1016/j.neuron.2015.02.018

Bigelow, B. J. (1977). Children's friendship expectations: a cognitive-developmental study. *Child Development*, *48*(1), 246–253. https://doi.org/10.2307/1128905

Blakemore, S.-J. (2008). The social brain in adolescence. *Nature Reviews Neuroscience*, *9*(4), 267–277. https://doi.org/10.1038/nrn2353

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye gaze. *Neuropsychologia*, *40*(8), 1129–1138. https://doi.org/10.1016/S0028-3932(02)00008-8

Caruana, N., Brock, J., & Woolgar, A. (2015). A frontotemporoparietal network common to initiating and responding to joint attention bids. *NeuroImage*, *108*, 34–46. https://doi.org/10.1016/j.neuroimage.2014.12.041

Chen, G., Saad, Z. S., Nath, A. R., Beauchamp, M. S., & Cox, R. W. (2012). FMRI group analysis combining effect estimates and their variances. *Neuroimage*, *60*(1), 747–765.

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences*, *16*(4), 231–239. https://doi.org/10.1016/j.tics.2012.02.007

Ciaramidaro, A., Becchio, C., Colle, L., Bara, B. G., & Walter, H. (2014). Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. *Social Cognitive and Affective Neuroscience*, *9*(7), 909–916. https://doi.org/10.1093/scan/nst062

Constantino, J. N., & Todd, R. D. (2003). Autistic traits in the general population: a twin study. *Archives of General Psychiatry*, *60*(5), 524–530. https://doi.org/10.1001/archpsyc.60.5.524

Conty, L., N'Diaye, K., Tijus, C., & George, N. (2007). When eye creates the contact! ERP evidence for early dissociation between direct and averted gaze motion processing. *Neuropsychologia*, *45*(13), 3024–3037. https://doi.org/10.1016/j.neuropsychologia.2007.05.017

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron*, *58*(3), 306–324. https://doi.org/10.1016/j.neuron.2008.04.017

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(23), 9163–9168. https://doi.org/10.1073/pnas.0807721106

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017). FMRI clustering in afni: false positive rates redux. *Brain Connectivity, 7*(3), 152–171.

Cremers, H. R., Wager, T. D., & Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. *PLOS ONE*, *12*(11), e0184923. https://doi.org/10.1371/journal.pone.0184923

Devine, R. T., & Hughes, C. (2013). Silent Films and Strange Stories: Theory of mind, gender, and social experiences in middle childhood. *Child Development, 84*(3), 989–1003. https://doi.org/10.1111/cdev.12017

Egger, H. L., Pine, D. S., Nelson, E., Leibenluft, E., Ernst, M., Towbin, K. E., & Angold, A. (2011). The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): a new set of children's facial emotion stimuli. *International Journal of Methods in Psychiatric Research*, *20*(3), 145–156. https://doi.org/10.1002/mpr.343

Eichenbaum, H., Yonelinas, A. P., & Ranganath, C. (2007). The medial temporal lobe and recognition memory. *Annual Review of Neuroscience*, *30*(1), 123–152. https://doi.org/10.1146/annurev.neuro.30.051606.094328

Farmer, T. W., Irvin, M. J., Motoca, L. M., Leung, M.-C., Hutchins, B. C., Brooks, D. S., & Hall, C. M. (2015). Externalizing and internalizing behavior problems, peer affiliations, and bullying involvement across the transition to middle school. *Journal of Emotional and Behavioral Disorders*, *23*(1), 3–16. https://doi.org/10.1177/1063426613491286

Feiring, C., & Lewis, M. (1991). The transition from middle childhood to early adolescence: Sex differences in the social network and perceived self-competence. *Sex Roles*, *24*(7–8), 489–509.

Frank, C. K., Baron-Cohen, S., & Ganzel, B. L. (2015). Sex differences in the neural basis of false-belief and pragmatic language comprehension. *NeuroImage, 105*, 300–311. https://doi.org/10.1016/j.neuroimage.2014.09.041

Gallagher, H. L., Jack, A. I., Roepstorff, A., & Frith, C. D. (2002). Imaging the intentional stance in a competitive game. *NeuroImage*, *16*(3), 814–821. https://doi.org/10.1006/nimg.2002.1117

Gilbert, S. J., Williamson, I. D. M., Dumontheil, I., Simons, J. S., Frith, C. D., & Burgess, P. W. (2007). Distinct regions of medial rostral prefrontal cortex supporting social and nonsocial functions. *Social Cognitive and Affective Neuroscience*, *2*(3), 217–226. https://doi.org/10.1093/scan/nsm014

Gweon, H., Dodell-Feder, D., Bedny, M., & Saxe, R. (2012). Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Development*, *83*(6), 1853–1868. https://doi.org/10.1111/j.1467-8624.2012.01829.x

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

Isoda, M., & Noritake, A. (2013). What makes the dorsomedial frontal cortex active during reading the mental states of others? *Frontiers in Neuroscience*, *7*, 1–14. https://doi.org/10.3389/fnins.2013.00232

Kampe, K. K., Frith, C. D., & Frith, U. (2003). "Hey John": signals conveying communicative intention toward the self activate brain regions associated with "mentalizing," regardless of modality. *Journal of Neuroscience*, *23*(12), 5258–5263.

Kircher, T., Blümel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., … Krach, S. (2009). Online mentalising investigated with functional MRI. *Neuroscience Letters*, *454*(3), 176–181. https://doi.org/10.1016/j.neulet.2009.03.026

Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., L, W., … Alpert. (1999). The role of area 17 in visual imagery: convergent evidence from PET and rTMS. *Science*, *284*(5411), 167–170. https://doi.org/10.1126/science.284.5411.167

Kosslyn, S. M., Thompson, W. L., Klm, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, *378*(6556), 496–498. https://doi.org/10.1038/378496a0

Krach, S., Blümel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., … Kircher, T. (2009). Are women better mindreaders? Sex differences in neural correlates of mentalizing detected with functional MRI. *BMC Neuroscience, 10*, 9. https://doi.org/10.1186/1471-2202-10-9

Krall, S. C., Rottschy, C., Oberwelland, E., Bzdok, D., Fox, P. T., Eickhoff, S. B., … Konrad, K. (2015). The role of the right temporoparietal junction in attention and social interaction as revealed by ALE meta-analysis. *Brain Structure and Function*, *220*(2), 587–604. https://doi.org/10.1007/s00429-014-0803-z

Lee, S. M., & McCarthy, G. (2016). Functional heterogeneity and convergence in the right temporoparietal junction. *Cerebral Cortex*, *26*(3), 1108–1116. https://doi.org/10.1093/cercor/bhu292

Maccoby, E. E. (1990). Gender and relationships: A developmental account. *American Psychologist, 45*(4), 513.

Mar, R. A. (2011). The neural bases of social cognition and story comprehension. *Annual Review of Psychology*, *62*(1), 103–134. https://doi.org/10.1146/annurev-psych-120709-145406

Mars, R. B., Sallet, J., Schüffelgen, U., Jbabdi, S., Toni, I., & Rushworth, M. F. S. (2012). Connectivity-based subdivisions of the human right "temporoparietal junction area": evidence for different areas participating in different cortical networks. *Cerebral Cortex*, *22*(8), 1894–1903. https://doi.org/10.1093/cercor/bhr268

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, *98*(20), 11832–11835. https://doi.org/10.1073/pnas.211415698

Miller, S. A. (2009). Children's understanding of second-order mental states. *Psychological Bulletin*, *135*(5), 749–773. https://doi.org/10.1037/a0016854

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Molfese, P. J., Glen, D., Mesite, L., Pugh, K. R., & Cox, R. W. (2015). The Haskins pediatric brain atlas. Poster session presented at the 21st Annual Meeting of the Organization for Human Brain Mapping, Honolulu, HI.

Moraczewski, D., Kinnison, J., & Pessoa, L. (2016). Design Explorer. [Computer software].

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J.-B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, *25*(3), 653–660. https://doi.org/10.1016/j.neuroimage.2004.12.005

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics*, *2*. https://doi.org/10.3389/neuro.11.010.2008

Pfeiffer, U. J., Schilbach, L., Timmermans, B., Kuzmanovic, B., Georgescu, A. L., Bente, G., & Vogeley, K. (2014). Why we interact: On the functional role of the striatum in the subjective experience of social interaction. *NeuroImage*, *101*, 124–137. https://doi.org/10.1016/j.neuroimage.2014.06.061

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: Implications for the emergence of autism. *Neuroscience & Biobehavioral Reviews*, *32*(1), 123–142. https://doi.org/10.1016/j.neubiorev.2007.06.004

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., & Saxe, R. (2010). Live face-to-face interaction during fMRI: a new tool for social cognitive neuroscience. *NeuroImage*, *50*(4), 1639–1647. https://doi.org/10.1016/j.neuroimage.2010.01.052

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*. https://doi.org/10.3389/fnhum.2012.00169

Redcay, E., Velnoskey, K. R., & Rowe, M. L. (2016). Perceived communicative intent in gesture and language modulates the superior temporal sulcus: Shared Neural Systems for Gesture and Language. *Human Brain Mapping*, *37*(10), 3444–3461. https://doi.org/10.1002/hbm.23251

Rice, K., Moraczewski, D., & Redcay, E. (2016). Perceived live interaction modulates the developing social brain. *Social Cognitive and Affective Neuroscience*, *11*(9), 1354–1362. https://doi.org/10.1093/scan/nsw060

Rice, K., & Redcay, E. (2016). Interaction matters: A perceived social partner alters the neural processing of human speech. *NeuroImage*, *129*, 480–488. https://doi.org/10.1016/j.neuroimage.2015.11.041

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(04), 393–414.

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. https://doi.org/10.1162/jocn.2009.21401

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*(5), 718–730. https://doi.org/10.1016/j.neuropsychologia.2005.07.017

Schultz, W. (2015). Neuronal reward and decision signals: from theories to data. *Physiological Reviews*, *95*(3), 853–951. https://doi.org/10.1152/physrev.00023.2014

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Schuwerk, T., Schurz, M., Müller, F., Rupprecht, R., & Sommer, M. (2016). The rTPJ's overarching cognitive function in networks for attention and theory of mind. *Social Cognitive and Affective Neuroscience, 12*(1), 157–168. https://doi.org/10.1093/scan/nsw163

Veroude, K., Jolles, J., Croiset, G., & Krabbendam, L. (2014). Sex differences in the neural bases of social appraisals. *Social Cognitive and Affective Neuroscience, 9*(4), 513–519. https://doi.org/10.1093/scan/nst015

Warnell, K. R., Sadikova, E., & Redcay, E. (2018). Let's chat: developmental neural bases of social motivation during real-time peer interaction. *Developmental Science*, e12581. https://doi.org/10.1111/desc.12581

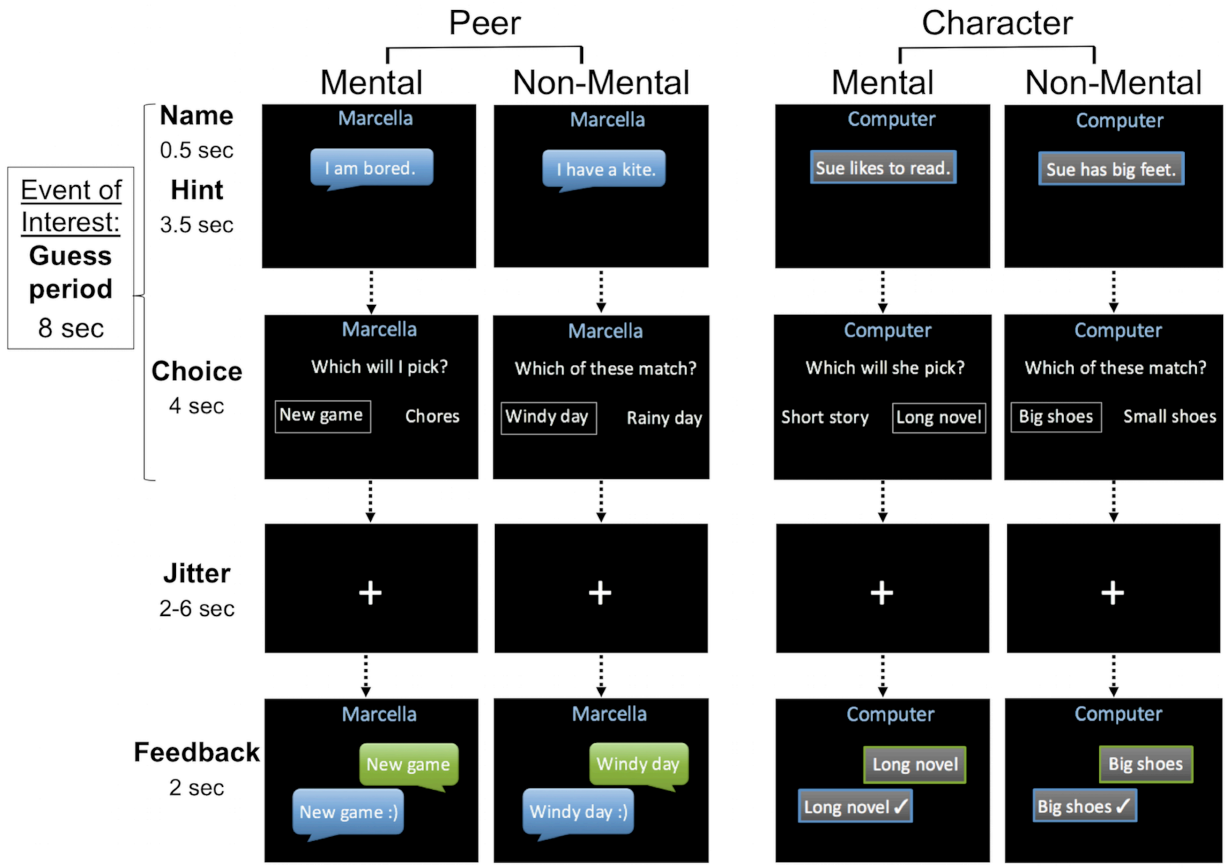


Figure 1. The interactive mentalizing task. Children completed 24 trials of each condition (Peer Mental, Character Mental, Peer Non-Mental, Character Non-Mental) in an event-related design. Mental trials required reasoning about mental states, while Non-Mental trials did not. In the Peer trials, children believed they were interacting with a child being scanned in another laboratory, whereas in Character trials, they believed they were answering questions about a fictional character provided by a computer. All trials had predetermined peer or computer responses. A smiley face (Peer) or check mark (Character) in the Feedback period indicated a match between the child's response and the peer or computer response.

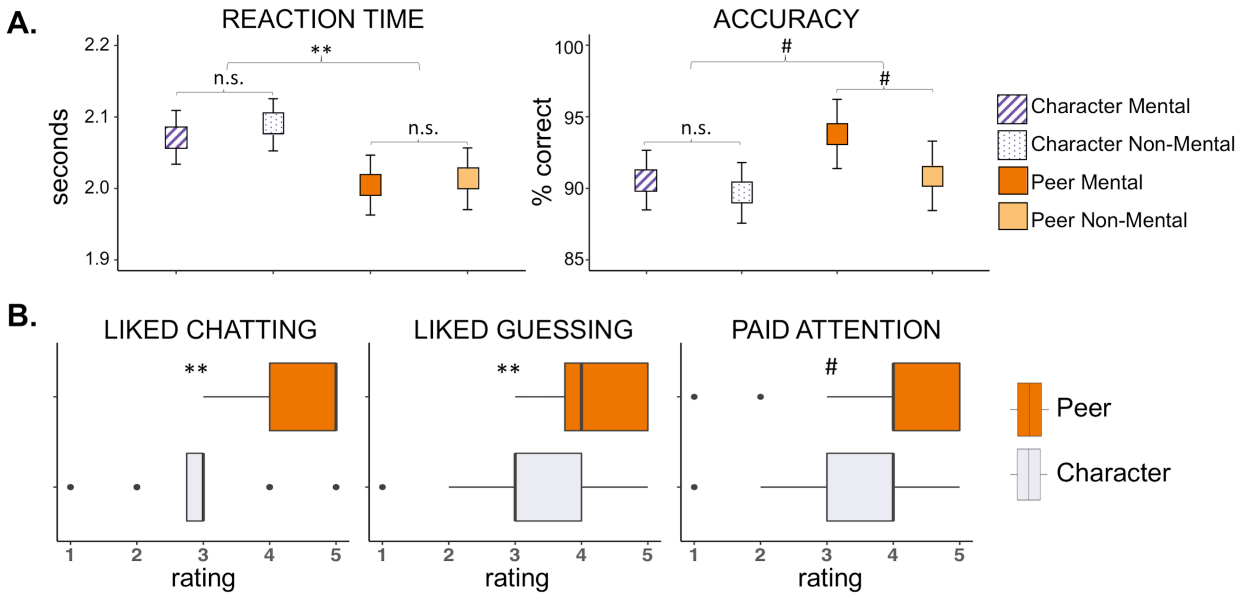


Figure 2. Behavioral results. **A.** In-scanner performance by condition. Mean values are plotted for reaction time (seconds) and accuracy (% correct) for each of the four conditions. Repeated measures ANOVA indicated a significant main effect of social interaction on reaction time such that children responded more quickly on Peer than Character trials. Error bars represent 95% confidence intervals. # $p < 0.1$; ** $p < 0.005$ **B.** Post-test questionnaire. For Peer and Character conditions separately, children rated on a Likert-type scale of 1 to 5 how much they enjoyed interacting with their partners (Peer) and answering questions from the computer (Character), how much they liked guessing what their partners would pick (Peer) and what came next in the story (Character), and how much they paid attention when interacting with their partners (Peer) and when answering questions from the computer (Character). Wilcoxon signed rank tests were used to compare ratings between Peer and Character conditions. # $p < 0.1$; ** $p < 0.005$

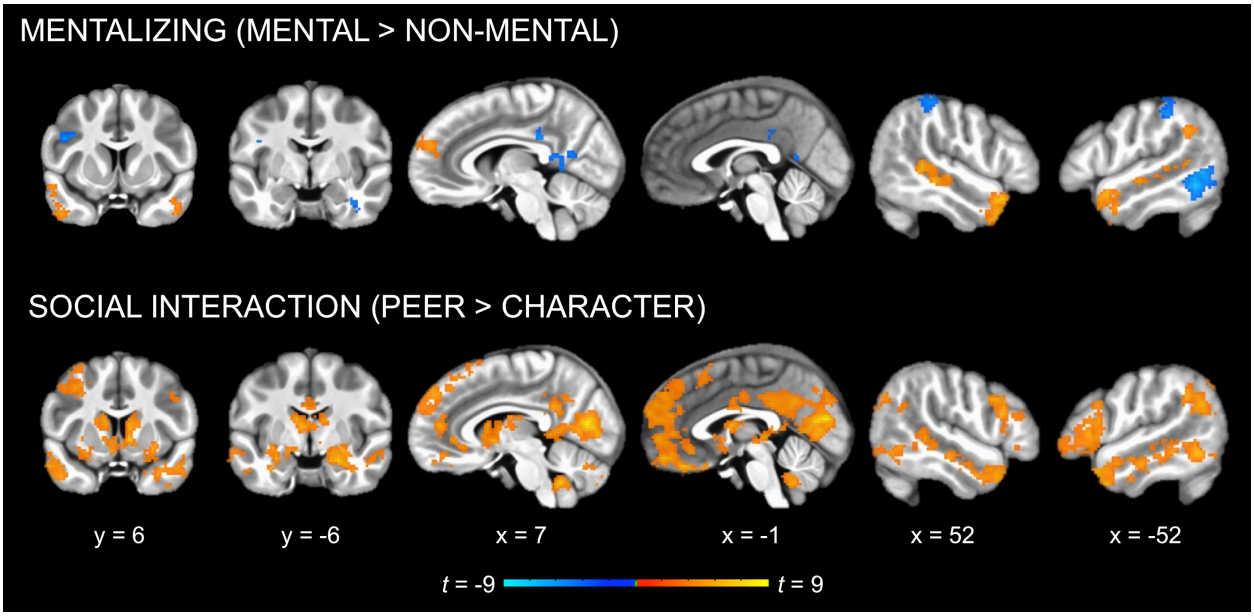


Figure 3. Whole-brain analysis of the main effects of mentalizing and social interaction (cluster corrected $p < 0.05$). Mentalizing (Mental vs. Non-Mental) activated regions previously identified in the mentalizing literature (dMPFC, TPJ, STS, and ATL). Social interaction (Peer vs. Character) activated similar regions, as well as additional cortical midline regions and subcortical structures associated with reward (e.g., amygdala, striatum).

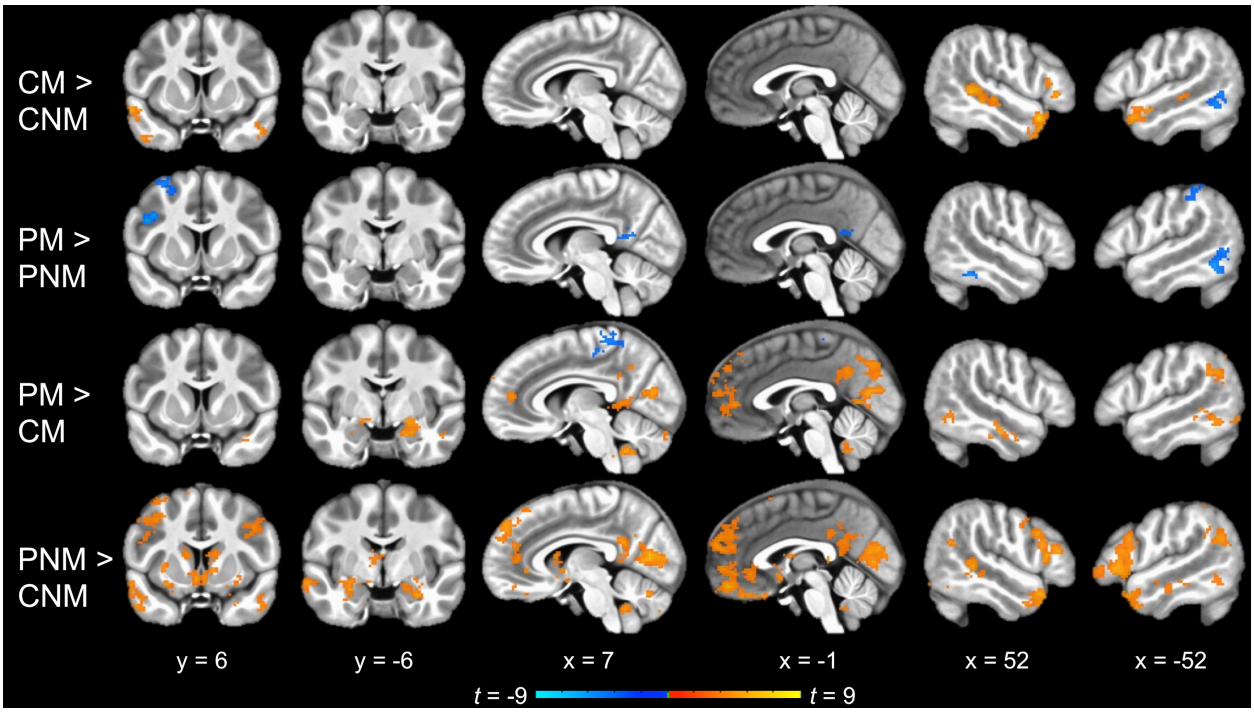


Figure 4. Whole-brain pairwise comparisons between the four conditions (cluster corrected $p < 0.05$). Offline mentalizing (Character Mental vs. Character Non-Mental) elicited a pattern of activation similar to the main effect of mentalizing (Figure 3). In contrast, no regions were significantly more active for Peer Mental than Peer Non-Mental. Social interaction without explicit mentalizing demands (Peer Non-Mental vs. Character Non-Mental) recruited similar regions as in the main effect of social interaction (Figure 3), whereas a smaller subset of these regions was more active for mentalizing within social interaction than offline mentalizing (Peer Mental vs. Character Mental). CM = Character Mental, CNM = Character Non-Mental, PM = Peer Mental, PNM = Peer Non-Mental

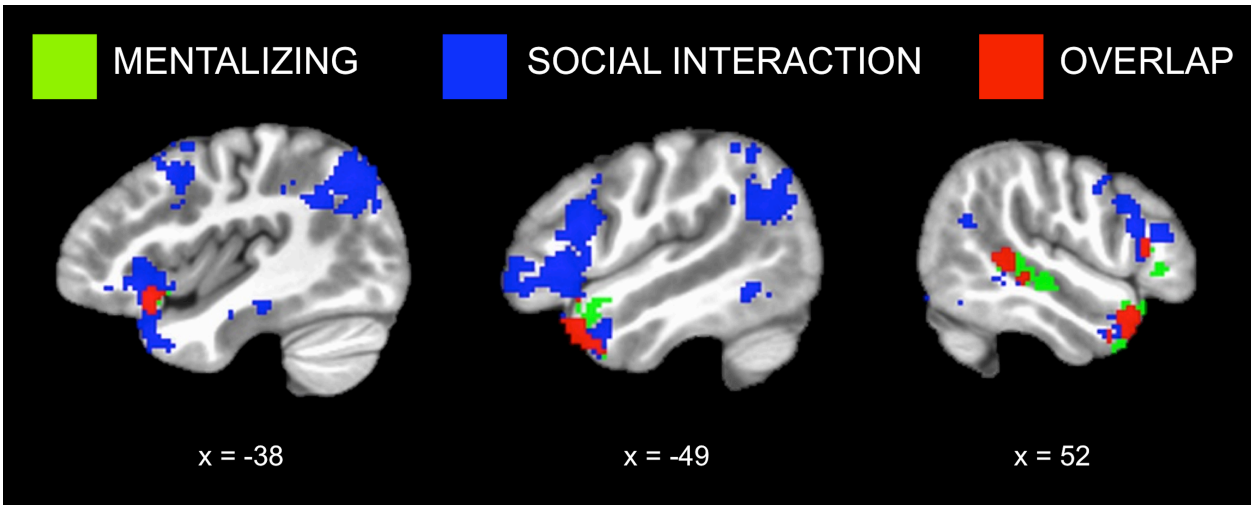


Figure 5. Shared regions for mentalizing and social interaction. Binarized, cluster-corrected maps for offline mentalizing (Character Mental > Character Non-Mental; green) and social interaction without mentalizing demands (Peer Non-Mental > Character Non-Mental; blue) are shown along with their conjunction (red), which reveals both overlapping and distinct regions of activation. CM = Character Mental, CNM = Character Non-Mental, PNM = Peer Non-Mental

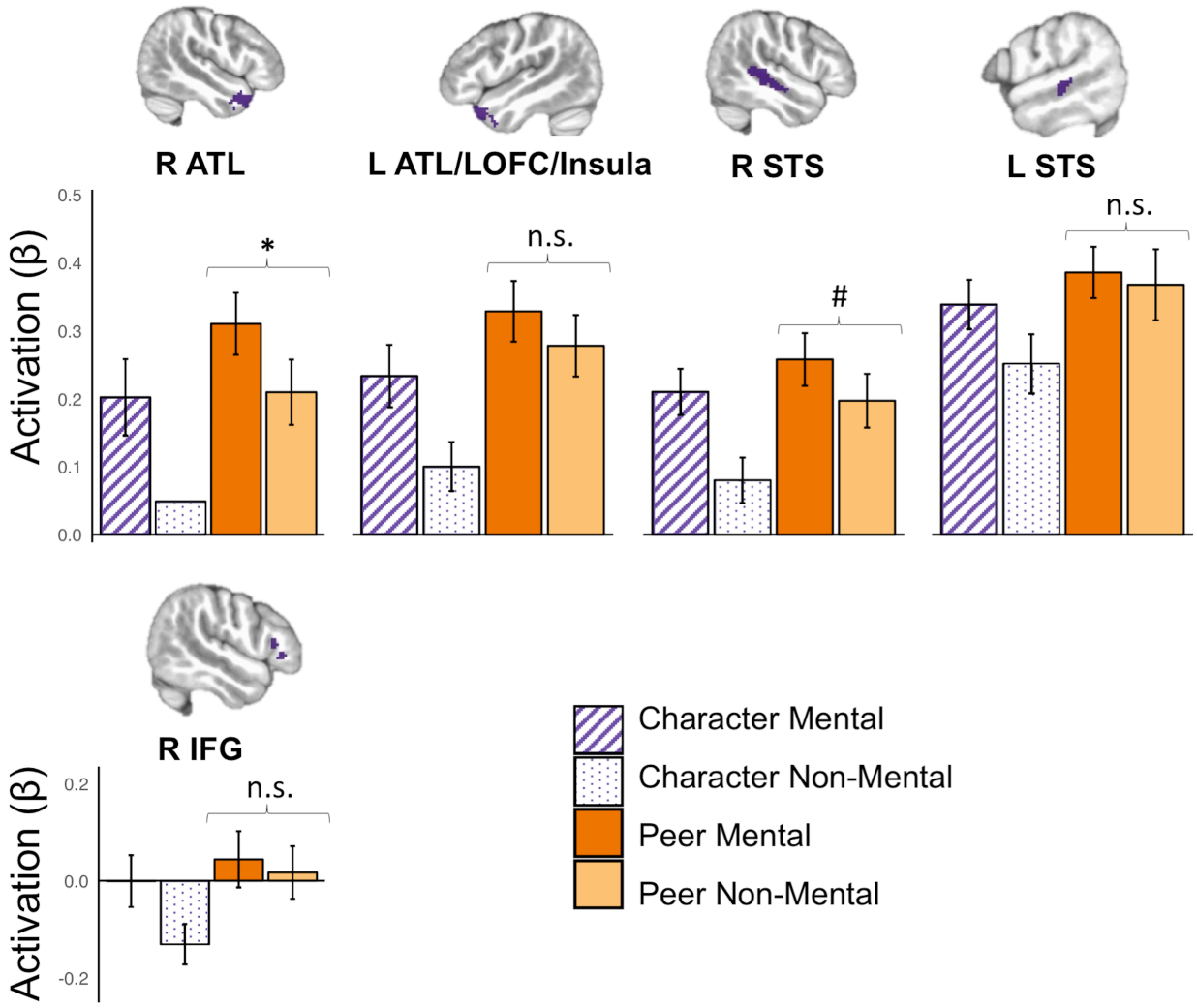


Figure 6. ROI analysis of mentalizing regions during social interaction. Regions of interest (ROIs) were defined by the Character Mental > Character Non-Mental contrast. Individual beta values for each condition within each ROI were extracted; average values are plotted with error bars representing 95% confidence intervals. Paired t-tests indicated non-significant differences between Peer Mental and Peer Non-Mental in all ROIs except R ATL, as well as non-significant differences between Character Mental and Peer Non-Mental in all ROIs. ATL = anterior temporal lobe, IFG = inferior frontal gyrus, LOFC = lateral orbitofrontal cortex, STS = superior temporal sulcus. * $p < 0.05$ corrected; # $p < 0.1$ corrected

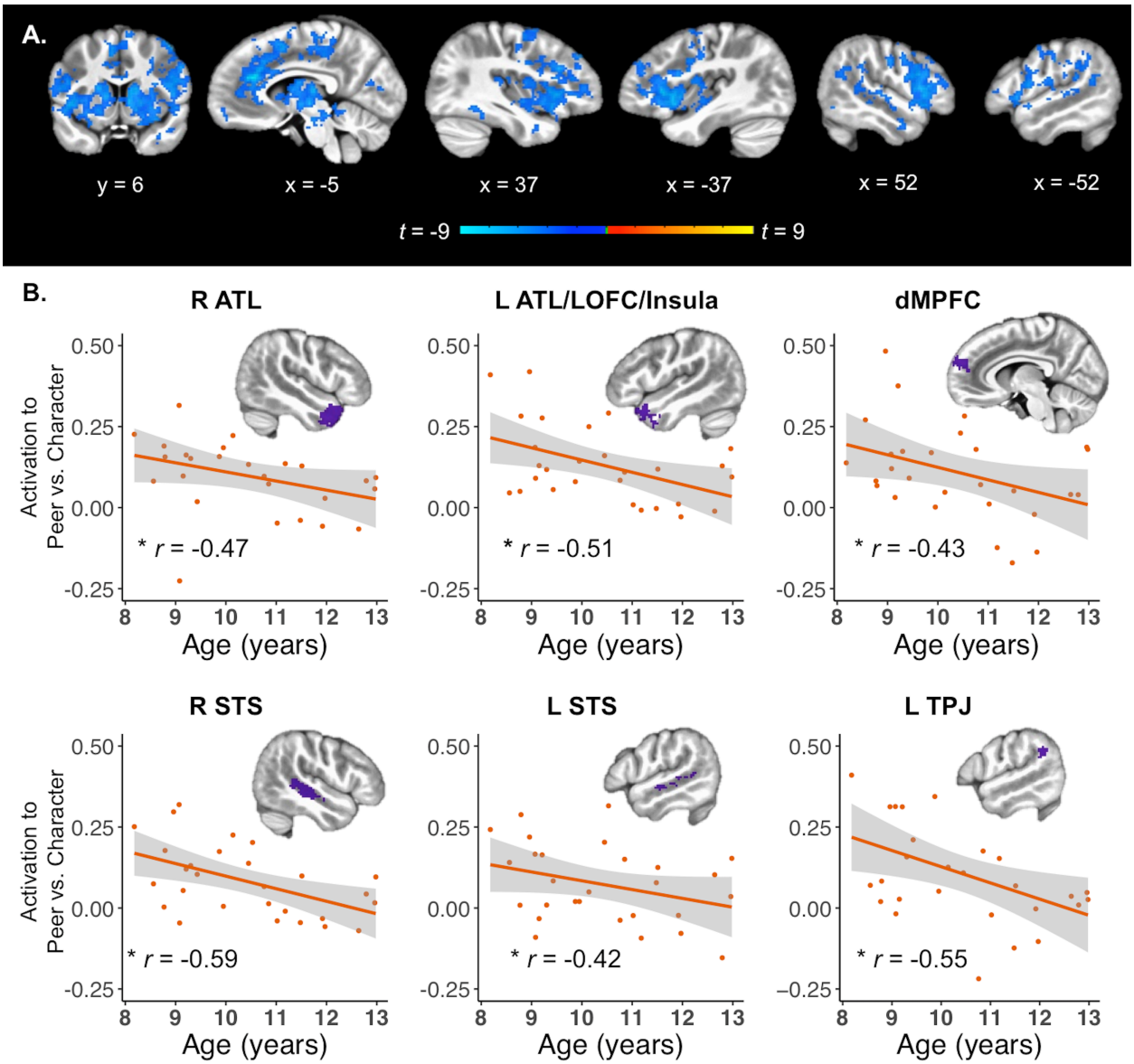


Figure 7. Effects of age on neural response to social interaction. **A.** Whole-brain effects of age on social interaction (Peer vs. Character; cluster corrected $p < 0.05$). Differences in activation to Peer versus Character conditions decreased with age in several frontal, temporal, insular, and subcortical areas. **B.** Effect of age on mentalizing ROIs. Regions of interest were defined by the main effect of mentalizing (Mental > Non-Mental). All regions showed a significant negative correlation between age and difference in activation to Peer versus Character conditions. ATL =

anterior temporal lobe, dMPFC = dorsomedial prefrontal cortex, LOFC = lateral orbitofrontal cortex, STS = superior temporal sulcus, TPJ = temporoparietal junction. ***** $p < 0.05$ corrected

Table 1. Whole-brain results for main effects of mentalizing, social interaction, the interaction term, and pairwise comparisons between the four conditions.

Region	Side	Peak		Cluster	MNI Coordinates		
		t	β†	*k*	*x*	*y*	*z*
Main Effect of Mentalizing							
Mental > Non-Mental							
Anterior temporal lobe	R	6.25	0.37	414	52	18	-20
Superior temporal sulcus	R	6.96	0.27	407	54	-40	2
Anterior temporal lobe	L	5.61	0.24	375	-50	10	-18
Lateral OFC/insula*	L	4.90	0.23		-39	18	-18
Superior temporal sulcus	L	5.19	0.15	147	-58	-23	-2
dMPFC	R	4.81	0.23	140	5	56	21
TPJ	L	5.29	0.18	109	-56	-48	29
Non-Mental > Mental							
Inferior parietal cortex	L	6.97	0.25	747	-24	-71	47
Inferior temporal gyrus	L	7.63	0.28	585	-54	-54	-9
Fusiform gyrus	L	7.26	0.33	390	-28	-30	-20
Inferior parietal cortex	R	5.04	0.22	295	33	-71	45
PCC	R	5.31	0.27	273	12	-52	9
Fusiform gyrus	R	5.66	0.22	256	35	-24	-23
Lateral OFC	L	6.73	0.22	191	-33	37	-11
PCC	L	6.67	0.20	167	-5	-54	11
SMG/inferior parietal cortex	R	5.53	0.13	157	46	-38	45
SMG/postcentral gyrus	L	4.86	0.17	114	-54	-33	45

PCC	L	4.75	0.15	91	8	-33	29
IFG_{tri}	R	5.33	0.30	87	50	43	14
IFG_{oper}	L	5.43	0.19	86	-44	6	27
Main Effect of Social Interaction							
Peer > Character							
Pericalcarine/cuneus	R	9.14	0.42	19,206	10	-73	11
Putamen/AMY/hippocampus*	L	8.82	0.19		-31	-19	-9
dMPFC/ACC*	L	7.83	0.20		-12	39	18
Anterior temporal lobe*	R	7.32	0.34		48	16	-29
Caudate/putamen*	R	6.96	0.21		10	5	-1
Lateral OFC/insula*	L	6.84	0.31		-41	17	-16
Precuneus*	L	6.80	0.26		-12	-48	25
AMY/putamen/insula*	R	6.76	0.19		25	-7	-20
Caudate*	L	6.74	0.15		-12	17	10
Caudate*	R	6.71	0.16		10	9	12
PCC/LG*	R	6.52	0.30		18	-46	-2
Thalamus*	L	6.51	0.35		-6	-4	9
Inferior temporal gyrus*	L	6.37	0.21		-50	-56	-9
Inferior parietal cortex*	R	6.29	0.24		33	-73	38
IFG_{oper}*	R	6.24	0.25		52	12	32
PHG*	R	6.20	0.22		18	-37	-13
dMPFC*	R	6.01	0.30		4	45	11
Caudate/ventral striatum*	L	5.90	0.21		-5	10	0

Lateral occipital cortex*	R	5.70	0.19		40	-77	-3
Middle temporal gyrus*	R	5.65	0.19		59	-17	-18
Medial OFC*	R	5.45	0.32		4	44	-15
Superior temporal sulcus*	R	5.32	0.25		46	-38	2
Superior frontal gyrus*	R	5.11	0.15		14	35	50
Inferior parietal cortex*	R	4.89	0.23		50	-67	34
Cuneus*	L	4.82	0.64		-1	-79	38
Lateral OFC *	R	4.48	0.19		42	26	-4
Inferior parietal cortex/TPJ*	R	4.36	0.27		61	-56	18
Inferior parietal cortex/TPJ	L	8.16	0.38	1,572	-37	-71	40
Cerebellum	R	6.49	0.25	915	20	-73	-32
Cerebellum	L	5.26	0.24	368	-16	-92	-30
IFG_{oper}	R	6.24	0.25	310	52	12	32
Cerebellum	R	6.93	0.26	260	3	-51	-39
Lateral OFC	R	4.48	0.19	148	42	26	-4
Character > Peer							
Lateral occipital cortex	L	5.44	0.19	123	-14	-98	4
Interaction Effect (Mentalizing x Social Interaction)							
(Peer Mental > Peer Non-Mental) > (Character Mental > Character Non-Mental)							
None							
Effect of Character Mental vs. Character Non-Mental							
Character Mental > Character Non-Mental							
Superior temporal sulcus	R	7.11	0.18	368	54	-40	2

Anterior temporal lobe	L	5.79	0.17	357	-56	4	-13
Lateral OFC/insula*	L	5.27	0.18		-41	20	-20
Anterior temporal lobe	R	7.49	0.21	241	52	14	-20
Superior temporal sulcus	L	5.27	0.10	99	-58	-23	-2
IFG_{tri}	R	5.21	0.16	93	52	28	0
Character Non-Mental > Character Mental							
Inferior temporal gyrus	L	4.58	0.14	159	-54	-54	-9
Effect of Peer Mental vs. Peer Non-Mental							
Peer Mental > Peer Non-Mental							
None							
Peer Non-Mental > Peer Mental							
Inferior parietal cortex	L	5.77	0.14	561	-29	-62	45
Inferior temporal gyrus	L	5.60	0.17	421	-56	-56	-9
Fusiform gyrus	L	7.12	0.17	379	-28	-38	-14
PCC	L	4.42	0.09	180	-9	-48	13
Middle frontal gyrus	L	6.38	0.14	160	-27	17	52
Postcentral gyrus	L	4.77	0.12	143	-29	-29	65
IFG_{oper}	L	5.93	0.15	137	-44	8	27
Inferior temporal gyrus	R	5.41	0.10	128	59	-34	-20
Fusiform gyrus	R	5.50	0.16	124	33	-24	-23
Lateral OFC	L	4.78	0.16	101	-31	33	-11
Inferior parietal cortex	L	4.46	0.08	96	-44	-50	50
IFG_{tri}	L	6.10	0.13	95	-41	37	14

Effect of Peer Non-Mental vs. Character Non-Mental							
Peer Non-Mental > Character Non-Mental							
dMPFC	L	8.40	0.16	2,946	-12	56	30
dMPFC/ACC*	L	5.59	0.12		-12	39	21
ACC/medial OFC*	L	7.49	0.14		-5	33	-2
PHG/Fusiform gyrus /PCC/LG	L	7.43	0.11	2,808	-20	-34	-14
LG/cuneus/pericalcarine*	R	6.89	0.25		8	-71	11
LG*	R	4.81	0.13		20	-48	-2
PCC*	L	4.74	0.14		-9	-48	25
Lateral OFC	L	6.55	0.31	2,212	-35	22	-25
IFG_{tri}*	L	6.20	0.14		-39	25	3
Anterior temporal lobe*	L	5.94	0.27		-50	16	-24
Insula/medial OFC/AMY*	L	5.20	0.13		-26	5	-13
Inferior parietal cortex/TPJ	L	6.00	0.17	1,123	-41	-64	27
Inferior parietal cortex/TPJ*	L	5.34	0.09		-39	-50	31
Anterior temporal lobe	R	7.12	0.26	713	50	18	-29
Lateral OFC*	R	6.15	0.13		25	12	-16
AMY*	R	5.66	0.11		29	-5	-22
Caudate	L	5.44	0.12	575	-12	16	9
Thalamus*	L	4.75	0.13		-6	-3	6
Caudate*	R	4.53	0.08		8	8	13
Ventral striatum*	L	4.24	0.11		-7	8	-11
Ventral striatum*	R	3.91	0.13		5	8	-9

IFG_{oper}	R	5.57	0.13	487	46	14	23
Cerebellum	R	5.28	0.14	394	20	-75	-32
Lateral occipital cortex	R	5.60	0.10	367	31	-77	9
Middle frontal gyrus	L	5.03	0.10	285	-29	4	45
Middle temporal gyrus	L	5.23	0.17	273	-58	-23	-16
Superior temporal sulcus	R	5.46	0.14	229	54	-38	2
Cerebellum	L	6.19	0.11	216	-11	-71	-30
Middle temporal gyrus	L	4.64	0.13	208	-63	-50	2
Inferior temporal gyrus*	L	3.74	0.10		−47	−53	−8
IFG_{orb}	R	5.96	0.17	205	44	24	-2
Inferior parietal cortex/TPJ	R	5.16	0.14	183	44	-56	27
PHG	R	5.35	0.14	154	18	-38	-14
Fusiform gyrus*	R	4.27	0.09		35	-38	-18
Cerebellum	R	4.58	0.15	87	8	-51	-39
Character Non-Mental > Peer Non-Mental							
None							
Effect of Peer Mental vs. Character Mental							
Peer Mental > Character Mental							
Pericalcarine/cuneus/LG	R	6.87	0.13	1,654	16	-71	11
PCC*	R	5.04	0.13		14	-38	-2
Hippocampus*	L	5.76	0.16		-14	-36	-5
dMPFC	L	5.70	0.15	700	-7	62	14
Cerebellum	R	6.15	0.10	444	20	-73	-30

Inferior parietal cortex/TPJ	L	5.41	0.15	426	-37	-71	40
PCC/precuneus	L	5.80	0.15	362	-12	-46	25
Inferior/middle temporal gyrus	L	4.59	0.11	272	-50	-54	-9
AMY/putamen/hippocampus	R	5.87	0.14	231	20	-9	-9
Hippocampus/AMY/ventral DC	L	5.34	0.15	168	-26	-19	-9
Cerebellum	R	7.13	0.13	147	3	-55	-41
Cuneus/precuneus	R	4.41	0.29	138	1	-77	31
Cerebellum	L	4.67	0.13	119	-26	-75	-37
Character Mental > Peer Mental							
Paracentral gyrus	R	4.87	0.07	150	5	-31	63
Lateral occipital cortex	L	5.95	0.16	146	-9	-104	-3

† β coefficient at the peak *t* value

* sub-peaks within clusters

ACC = anterior cingulate cortex, AMY = amygdala, dMPFC = dorsomedial prefrontal cortex, IFG_{oper} = inferior frontal gyrus (pars opercularis), IFG_{orb} = inferior frontal gyrus (pars orbitalis), IFG_{tri} = inferior frontal gyrus (pars triangularis), LG = lingual gyrus, OFC = orbitofrontal cortex, PCC = posterior cingulate cortex, PHG = parahippocampal gyrus, SMG = supramarginal gyrus, TPJ = temporoparietal junction, Ventral DC = Ventral diencephalon

Table 2. Conjunction analysis: Overlapping activation between Character Mental > Character Non-Mental and Peer Non-Mental > Character Non-Mental contrasts. Coordinates are reported for the center of mass of each cluster. Clusters of fewer than 20 voxels are not reported.

Region	Side	Cluster	MNI Coordinates		
		k	***x***	***y***	***z***
Anterior temporal lobe/lateral OFC/insula	L	233	-43	15	-19
Anterior temporal lobe	R	174	48	15	-27
Posterior superior temporal sulcus	R	144	52	-37	3
IFG_{oper}	R	29	53	23	13

IFG_{oper} = inferior frontal gyrus (pars opercularis), OFC = orbitofrontal cortex

Supplementary Material

Piloting of task items. To ensure that all items were easily answerable by children aged 8–12, we conducted a pilot behavioral study on a separate sample of 10 children in this age range. After completing 96 trials, all presented in the Character condition to avoid unnecessary deception, children indicated items they found difficult, and some items were revised based on this feedback. Accuracy on the task was high (mean = 90% correct, SD = 5%) and children were able to respond within the 4-s limit (mean reaction time = 2.15 s, SD = 0.30 s). Mental and Non-Mental items did not significantly differ on accuracy (Mental: mean = 92%, SD = 7%; Non-Mental: mean = 89%, SD = 5%; $t(9) = 1.36$, $p = 0.21$) or reaction time (Mental: mean = 2.12 s, SD = 0.32 s; Non-Mental: mean = 2.17 s, SD = 0.29 s; $t(9) = -1.04$, $p = 0.32$). The final set of items was balanced across the four conditions for number of syllables and number of negations (which take longer to evaluate than affirmative statements[1]).

Demonstration of the interactive mentalizing task. Crucial to the illusion of peer interaction was the participant's understanding of not only how to perform his or her role in the "game," but also the chat partner's role. To this end, children viewed a demonstration of a chat between two fictional people. The "hint-giver" (i.e., the chat partner in the real task; female in the demonstration) is first shown a sentence with a word or phrase missing, which she completes by choosing between two words or phrases. The "guesser" (i.e., the participant in the real task; male in the demonstration) then sees this "hint" followed by either "Which will I pick?" or "Which of these match?" with two answer choices below, and his task is to guess what the hint-giver will choose. Meanwhile, the hint-giver sees the same question and answer choices and makes her own choice. The hint-giver then sees the guesser's choice and can either send a smiley

[1] Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content*. Cambridge, MA: Harvard University Press.

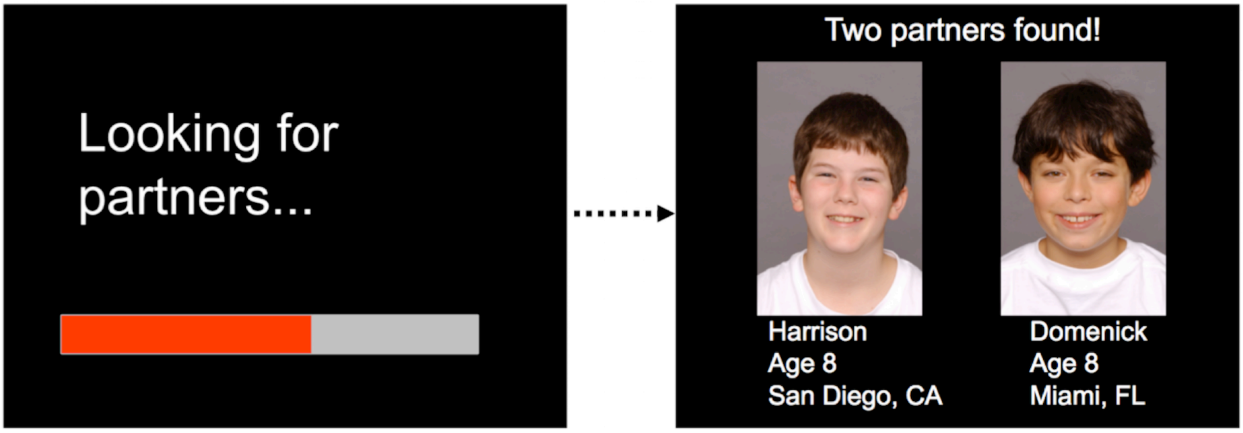
face (if the guesser's choice matches her own) or not (if the guesser's choice does not match). Lastly, the guesser is shown his own choice alongside the hint-giver's choice and smiley face (if the choices match). A similar demonstration followed in which the guesser receives a hint "from the computer," answers either "Which will she/he pick?" or "Which of these match?" and learns whether his choice matches the answer from the computer, with a check mark indicating a match. After the demonstration, participants answered a set of questions to confirm their comprehension of the task. Finally, the experimenter explained that the hints given by the chat partner would be based on questionnaires he or she filled out ahead of time about likes, dislikes, and what he or she would do in different situations. Thus, despite the peer interactions within the task being highly structured, participants had a basis for believing that their partners' hints and choices were generated by their authentic beliefs and personalities.

Example stimuli. In the Mental conditions, hints contained information about the chat partner or story character's knowledge (e.g., "I know my brother is hiding in the closet"; answer choices: "Keep looking" or "Open closet"), belief (e.g., "Sue does not think the teacher heard her"; answer choices: "Repeat question" or "Wait for answer"), desire (e.g., "I want to get a good grade"; answer choices: "Watch TV" or "Study for test"), preference ("Sue really likes to laugh"; answer choices: "Funny movie" or "Scary movie"), or emotion (e.g., "I am angry at my mom" ; answer choices: "Smile at mom" or "Glare at mom"). Non-Mental hints conveyed factual information about the chat partner or story character such as location ("I live far away from school"; answer choices: "Bus" or "Bike"), activity ("Sue is going on a hike"; answer choices: "Soup" or "Trail mix"), possession ("I have a lot of clothes"; answer choices: "Big closet" or "Small closet"), ability ("Tim can speak Spanish"; answer choices: "Mexico" or "Japan"), or

physical characteristic ("I am very tall for my age"; answer choices: "Gymnastics" or "Basketball").

To avoid the suggestion of intentionality, Non-Mental hints were followed by "Which of these match?" as opposed to "Which will I/she/he pick?" for Mental hints.

Post-test questionnaire. Belief in the live illusion was assessed in the majority of children by asking whether the peer and character, respectively, were real people. Six children included in the final sample also participated in a second MRI session that used a similar peer deception (see Warnell et al., 2018). To avoid giving away the deception after the first scan (reported here), these children were not asked whether the peer or character were real and were not debriefed until after the second scan. All participants were asked after each session whether they felt there was more to the task than they were told.



Supplementary Figure S1. Partner selection for the interactive mentalizing task. Children were given a choice between two age- and gender-matched children. Photos came from either the NIMH Child Emotional Faces Pictures Set (Egger et al., 2011), stock photography website Getty Images (www.gettyimages.com), or a Google Images search. All photos featured a headshot of a child against a solid background, smiling, and gazing directly at the camera. During scanning, the chosen partner's photograph was displayed at the end of each run to maintain the live illusion.

Supplementary Table S1. Whole-brain effects of age on mentalizing (Mental vs. Non-Mental) and social interaction (Peer vs. Character).

Region	Side	Peak *t*	β†	Cluster *k*	MNI Coordinates *x*	*y*	*z*
Mental vs. Non-Mental							
None							
Peer vs. Character							
ACC	L	-9.21	-0.19	15,740	-5	31	21
Caudate/putamen/thalamus*	R	-8.51	-0.25		12	0	11
Paracentral lobule*	R	-8.43	-0.11		12	-23	47
Caudate/putamen/thalamus*	L	-8.23	-0.20		-14	-2	14
Insula/ IFG_{oper} /lateral OFC*	L	-7.96	-0.31		-41	12	-9
IFG_{oper}*	R	-7.10	-0.20		56	14	7
Middle frontal gyrus*	R	-6.97	-0.60		44	8	57
IFG_{orb}/lateral OFC/insula*	R	-6.50	-0.17		42	26	0
Insula*	L	-6.50	-0.10		-31	-6	16
ACC*	R	-6.48	-0.18		5	23	14
Paracentral lobule*	R	-6.36	-0.13		3	-23	63
Middle frontal gyrus*	R	-6.00	-0.14		37	25	25
IFG_{tri}*	R	-5.62	-0.20		46	47	10
Superior frontal gyrus*	L	-5.44	-0.15		-5	4	48
Middle frontal gyrus*	R	-5.32	-0.21		27	50	37
Paracentral lobule*	L	-5.32	-0.09		-16	-29	41
TPJ*	L	-5.23	-0.15		-63	-46	27

TPJ*	L	-5.03	-0.18		-54	-52	18
TPJ	R	-6.35	-0.21	1,111	65	-44	16
Posterior STS*	R	-3.64	-0.13		49	-42	8
Brain stem/cerebellum	L	-6.34	-0.15	471	-1	-30	-18
Cerebellum*	R	-4.42	-0.10		3	-44	-16
Fusiform gyrus	R	-5.33	-0.14	407	44	-55	-21
Brain stem*	R	-4.60	-0.12		9	-30	-20
Middle frontal gyrus	R	-5.86	-0.31	184	25	55	-11
Anterior temporal lobe	R	-4.51	-0.12	139	44	1	-34
Cuneus	L	-5.29	-0.18	99	-1	-81	15

† β coefficient at the peak *t* value

* sub-peaks within clusters

ACC = anterior cingulate cortex, IFG_{oper} = inferior frontal gyrus (pars opercularis), IFG_{orb} = inferior frontal gyrus (pars orbitalis), IFG_{tri} = inferior frontal gyrus (pars triangularis), OFC = orbitofrontal cortex, STS = superior temporal sulcus, TPJ = temporoparietal junction