Communicative Signals Promote Object Recognition Memory and Modulate the Right Posterior STS

Elizabeth Redcay, Ruth S. Ludlum, Kayla R. Velnoskey, and Simren Kanwal

Abstract

■ Detection of communicative signals is thought to facilitate knowledge acquisition early in life, but less is known about the role these signals play in adult learning or about the brain systems supporting sensitivity to communicative intent. The current study examined how ostensive gaze cues and communicative actions affect adult recognition memory and modulate neural activity as measured by fMRI. For both the behavioral and fMRI experiments, participants viewed a series of videos of an actress acting on one of two objects in front of her. Communicative context in the videos was manipulated in a 2 × 2 design in which the actress either had direct gaze (Gaze) or wore a visor (NoGaze) and either pointed at (Point) or reached for (Reach) one of the objects (target) in front of her. Participants then completed a recognition memory task with old (target and nontarget) objects and novel objects. Recognition memory for target objects in the Gaze conditions was greater than NoGaze, but no effects of gesture type were seen. Similarly, the fMRI video-viewing task revealed a significant effect of Gaze within right posterior STS (pSTS), but no significant effects of Gesture. Furthermore, pSTS sensitivity to Gaze conditions was related to greater memory for objects viewed in Gaze, as compared with NoGaze, conditions. Taken together, these results demonstrate that the ostensive, communicative signal of direct gaze preceding an object-directed action enhances recognition memory for attended items and modulates the pSTS response to object-directed actions. Thus, establishment of a communicative context through ostensive signals remains an important component of learning and memory into adulthood, and the pSTS may play a role in facilitating this type of social learning. ■

INTRODUCTION

Each day humans engage in countless instances of shared attention with others. These episodes of shared attention may be incidental, such as when two strangers at the grocery store happen to look at the same cake, or may be intentional, such as when a woman looks at her spouse who has an upcoming birthday and points at a cake. These intentional shared attention contexts, known as *joint attention*, provide a platform for transmission of information between social partners. Although many studies have demonstrated that engaging in joint attention is an important tool for learning in infancy (Mundy, Sullivan, & Mastergeorge, 2009; Tomasello, Carpenter, Call, Behne, & Moll, 2005; Kuhl, Tsao, & Liu, 2003; Morales, 2000), few studies have examined how joint attention promotes information transfer in adults.

One necessary component for joint attention is that the initiator of joint attention establishes an intent to communicate by producing communicative, ostensive signals (e.g., eye contact or calling one's name) and directs the addressee's attention toward an object of interest through communicative, referential actions toward that object (typically a point or speech act). These actions performed for the purpose of communication are distinct from noncommunicative object-directed actions that simply convey a personal goal (such as reaching for a preferred object; Csibra & Gergely, 2009; Tomasello, Carpenter, & Liszkowski, 2007; Behne, Carpenter, & Tomasello, 2005; Clark, 1996; Sperber & Wilson, 1996). However, preceding an object-directed action (i.e., reach) with an ostensive cue, such as direct gaze, can make manifest the communicative intention of that action (e.g., Note that I am picking "this" one; Csibra & Gergely, 2009; Tomasello et al., 2007; Sperber & Wilson, 1996). Thus, both ostensive signals (e.g., gaze) and communicative, referential actions (i.e., point gestures) may establish communicative context, but whether and how these signals promote information transfer between social partners are less clear.

Studies in infants and young children suggest that establishing a communicative context through both ostensive and referential signals may create a special learning context by biasing attention toward object features that generalize beyond time, person, and place, such as the identity of the object (Csibra, 2010). For example, the presence of direct gaze and pointing cues in a joint attention context results in greater memory for the identity of objects relative to incidental shared attention contexts involving no direct gaze (Kopp & Lindenberger, 2011). Furthermore, the combination of ostensive signals and communicative point actions results in greater memory for the identity of an object over the nongeneralizable

University of Maryland

features of the object, such as its location (Yoon, Johnson, & Csibra, 2008). Studies examining neural activity in infants demonstrate that joint attention contexts initiated by direct gaze cues result in neural activity representative of enhanced attention and encoding (Hoehl, Reid, Mooney, & Striano, 2008; Striano, Reid, & Hoehl, 2006; Reid & Striano, 2005; Reid, Striano, Kaufman, & Johnson, 2004). Although these studies suggest that intentional joint attention contexts alter neural processing and promote memory for object features in infants, less is known about whether joint attention alters information processing in adults.

A previous study examined how the presence of communicative signals affects information processing in adults. Marno, Davelaar, and Csibra (2014) conducted a series of experiments using a change blindness paradigm and found that, similar to infants, adults showed better immediate memory (i.e., change detection) for the identity of objects (relative to the location) in the presence of ostensive signals followed by pointing. Pointing or reaching actions without ostensive signals, on the other hand, promoted greater attention to the location of the object than the identity. This finding supports the theory that ostensive signals may promote attention toward object features (Csibra & Gergely, 2009), but studies have not examined the relative contributions of ostensive and referential signals to enhance recognition memory for objects.

Similarly, although numerous studies have examined the neural correlates of action processing, the brain systems supporting intentional, communicative actions that promote joint attention contexts remain unclear. Neuroimaging and electrophysiological studies have highlighted the posterior STS (pSTS) as sensitive to human actions and the intentions motivating those actions (Redcay & Saxe, 2013; Pelphrey & Carter, 2008; Brass, Schmitt, Spengler, & Gergely, 2007; Pelphrey, Singerman, Allison, & McCarthy, 2003). The pSTS coordinates with two distinct networks associated with action and intention processing: the mentalizing (MENT) and mirroring (MS) systems (Yang, Rosenblau, Keifer, & Pelphrey, 2015). The MENT system is engaged when participants infer and reason about an actor's mental states, even without explicit action information (Van Overwalle & Baetens, 2009).The MS system, on the other hand, is involved when observing or executing actions and is thought to play a role in understanding the goal of the action based on how the action is performed (e.g., grasping the handle of a cup to drink vs. the top of the cup to clean; Spunt, Falk, & Lieberman, 2010; Iacoboni et al., 2005). The role of the pSTS within these networks likely extends beyond simple action processing, as the pSTS is engaged during a communicative game that does not utilize human actions (Noordzij et al., 2009) and is modulated more when actions are viewed in a social-communicative context (Redcay et al., 2010; Materna, Dicke, & Thier, 2008).

Studies have begun to elucidate the role of the pSTS, in concert with MENT and MS systems, in processing ostensive and referential signals beyond simple action perception. The MENT system, particularly the medial pFC

(MPFC) and pSTS, is sensitive to signals of communicative intent (Frith & Frith, 2010). For example, the MPFC responds to ostensive signals such as direct gaze, hearing one's own name, and communicative facial expressions (Kuzmanovic et al., 2009; Amodio & Frith, 2006; Schilbach et al., 2006; Kampe, Frith, & Frith, 2003). The right pSTS (RpSTS) is also sensitive to ostensive signals, particularly when they are dynamic (Pelphrey, Viola, & Mccarthy, 2004) or combined with meaningless hand actions (Ferri et al., 2014; Tylén, Allen, Hunter, & Roepstorff, 2012). Although studies have not specifically examined the neural correlates of responding to communicative, referential signals, such as communicative point actions, studies examining the perception of communicative actions more broadly suggest a role of the MENT or MS system and, in some cases, coactivation and interaction of these two systems (Committeri et al., 2015; Ciaramidaro, Becchio, Colle, Bara, & Walter, 2014; Trapp et al., 2014; Mainieri, Heim, Straube, Binkofski, & Kircher, 2013; Schippers, Gazzola, Goebel, & Keysers, 2009; Montgomery, Isenberg, & Haxby, 2007; Ferrari, Gallese, Rizzolatti, & Fogassi, 2003). Thus, the pSTS and other regions of the MENT and MS systems likely play a role in supporting the intentional, communicative actions that promote a joint attention context. However, studies have not systematically investigated how ostensive signals, communicative and noncommunicative object-directed actions (e.g., point vs. reach), and their interaction modulate MENT and MS systems and how this modulation affects object recognition memory.

In the current study, we examined the extent to which communicative actions promote memory for objects that are the target of an experimenter's actions and how the brain responds differentially to these actions when they convey communicative, as compared to personal, goals (e.g., pointing to show vs. reaching to grasp an object). Specifically, in a 2 × 2 design, we investigated how the presence of the ostensive signal of eye contact (Gaze vs. NoGaze) and presence of a communicative, referential action (Point vs. Reach) affected subsequent recognition memory for the target object and patterns of brain activation. We predicted that these communicative signals (pointing and gaze) would promote object memory and differentially engage brain systems sensitive to communicative contexts.

METHODS

Overview

We report the results of two experiments using the novel Communicative Actions task, a recognition memory paradigm, with a final total sample size of 57 participants. The behavioral experiment addressed the question of what signals are necessary to establish a communicative context that promotes object memory. The fMRI experiment examined how the communicative signals identified in the behavioral experiment modulated brain activity and

the extent to which this modulation was related to subsequent recognition memory performance. The University of Maryland institutional review board approved all procedures.

Participants

Participants were recruited from the University of Maryland student body and local community. All participants provided informed written consent and were compensated for their time with course credit or monetary payment. Behavioral data were collected from 52 adults, and 17 were excluded because of either self-reported current neurological disorders (10) or poor performance (i.e., <75%) on the 1-back task during video viewing (7) for a final sample of 35 participants (29 women). Twenty-four separate individuals participated in the scan experiment and second behavioral study, but two were excluded because of excessive motion (1) or falling asleep (1) during the scan for a final sample of 22 participants (12 women).

Communicative Actions Behavioral Task

Design

The Communicative Actions task had two parts: (1) video viewing followed by (2) a surprise recognition memory task.

Video viewing task. Participants viewed videos of an actress acting on one of two objects on a table in front of her. To ensure participants paid attention, they performed a 1-back task in which they pushed a button when the same video was repeated. There were 72 unique videos, eight of which were repeated twice for a total of 80 trials. Each video contained two unique objects, totaling 144 objects. Videos were presented on a computer screen using the Psychophysics Toolbox for MATLAB (Brainard, 1997). Each video consisted of 1500 msec of movement with a 500-msec pause on the first and last

frame each, totaling 2500 msec in duration. Each video was separated by a 500-msec gap with a crosshair fixation point in the center of the screen, which was the location where the actress's torso would appear when the next video began (Figure 1A).

Stimuli. In each video, the same female actress with a neutral facial expression sat behind a table with two unique objects resting on it. The actress then slowly reached (Reach) or pointed (Point) toward one of the objects (the target). The amount of her gaze that was visible also varied. In each video, the actress either made eye contact with the participant (Gaze) before shifting her gaze and turning her head to look down at the target object or wore a visor to obscure her eyes (NoGaze), resulting in four conditions: Point-Gaze (PG), Point-NoGaze (PNG), Reach-Gaze (RG), and Reach-NoGaze (RNG). This 2 × 2 design allowed us to identify whether eye contact (gaze), gesture (point vs. reach), or their interaction affect object memory (Figure 1B). The center of each object was consistently at one of two locations on the table (left or right). Within each video, the size, graspability, and thematic category (e.g., common medium-sized electronics) of the two objects were similar to avoid one object being inherently more memorable than the other. Objects were counterbalanced across participants so that each was presented on the left and right side of the table an equal number of times. Target objects were the same across participants, but the condition in which each target object appeared was counterbalanced across participants.

Recognition memory task. After participants completed the 1-back task, they performed a surprise recognition memory task in which they viewed pictures of objects that were either from the videos earlier—including both targets (64 items) and nontargets (30 items)—or were novel (67 items). Objects that were presented twice due to the nature of the 1-back task were not used for the recognition memory task. Participants chose between three responses: "OLD," "NEW," and "FAMILIAR." They

Figure 1. Communicative Actions task design. Still images from videos for each of the four conditions are displayed in A. The same objects are depicted for consistency, but in the actual experiment objects were counterbalanced across conditions such that each object was only viewed in one condition for each person. An example of one trial from the Point-Gaze condition is depicted in B. Trials began with the actress facing forward with either direct gaze (Gaze) or wearing a visor (NoGaze). After

500 msec, the point or reach action occurred toward the target object (i.e., the pepper in this example), and the video ended with a 500-msec pause. Each video was separated by a jittered intertrial interval with presentation of a fixation cross.
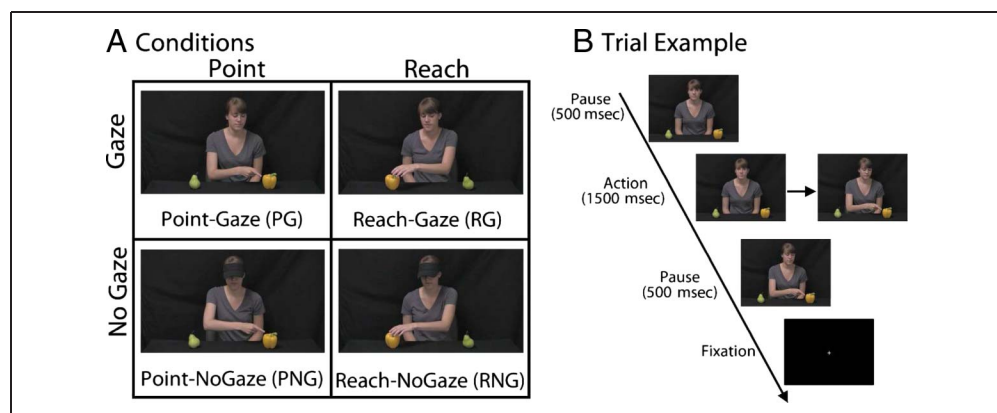
were told that if they recognize the object from earlier, choose "OLD," if they do not recognize the object from earlier, choose "NEW," and if the object seems familiar but they are unsure, choose "FAMILIAR." Thus, we had two measures of object recognition memory that varied on the confidence with which participants chose them (Old and Familiar). In only a small percentage of trials did participants choose Familiar, and thus subsequent analyses focus only on the combined Old and Familiar conditions (Total) and the Old only condition (Old). Participants were given up to 5 sec to respond. When a response was made, the item was removed from the screen, and a fixation screen appeared for 1 sec followed by the next item.

Recognition Memory Data Analysis

Each participant's performance on the recognition memory task was measured by calculating the participant's sensitivity index (d'). This was done by first determining the proportion of old objects a participant correctly identified as old (hits-old) or familiar (hits-familiar) and the proportion of novel objects incorrectly identified as old (false alarms-old) or familiar (false alarms-familiar). To calculate Total d' scores, we summed across Old and Familiar proportions for hits and false alarms (e.g., hits-old + hits-familiar). Old d' only included "Old" responses. Proportions of false alarms and hits were converted to z scores, and d' was calculated as the z-transformed hit rate minus false alarms. Total and Old d' scores were calculated for all target objects across the four conditions and all nontarget objects. To determine whether objects that were the target of the actress's action were remembered more than nontarget objects, a paired t test compared d' for all target to d' for all nontarget items. To examine the effects of gaze and gesture on object recognition memory, a 2 × 2 repeated-measures ANOVA with factors Gesture (Point and Reach) and Gaze (Gaze and NoGaze) was used to test for main effects and interactions of condition on d' scores. To test whether d' differed for target objects within each of the four conditions and all nontarget objects (collapsed across condition), pairwise contrasts on d' scores (Point-Gaze, Point-NoGaze, Reach-Gaze, Reach-NoGaze, and all nontarget) were calculated using Tukey's honest significant difference (HSD) to control for multiple comparisons. These analyses were conducted for both Total and Old d' scores separately. Condition effects were not examined for nontarget items.

Pilot study. A pilot study was first conducted with a total of 80 unique videos ($n = 30$ adults). Because of relatively low overall d' scores from these data, several changes were made to the design. First, eight videos containing target items that consistently demonstrated very high or low d' scores across conditions were removed from the video viewing. The rationale for removing these items was that those items with consistently high or low

d' scores across conditions may be either highly salient or more difficult to remember, respectively, due to specific aspects of the item and thus could weaken effects of condition. Second, the number of nontarget items in the recognition memory task was reduced (from 72 to 30) to reduce the overall number of trials and better balance the proportion of old and new items.

Communicative Actions fMRI Scan

Design

The design and stimuli for the Communicative Actions fMRI task were similar to those described above with several exceptions. First, to increase power to detect signal, the full video-viewing task was presented twice within the scanner such that all videos were seen twice (once per run). Second, eight videos were added such that each condition contained 20 unique videos for a total of 40 trials per each of the four conditions (PG, PNG, RG, RNG). Within each condition, two videos repeated for the 1-back task and thus were not analyzed. Third, 160 sec of fixation were added into each run to introduce jitter between video events. All data were presented in a rapid event-related design with the trial order and jitter intervals optimized for main effects of condition using OptSeq2 (surfer.nmr. mgh.harvard.edu/optseq/).

Recognition memory task. Similar to the behavioral study, a recognition memory task followed the fMRI scan (outside the scanner). Participants viewed 72 target (old), 72 nontarget (old), and 59 novel (new) photos and responded with "Old," "Familiar," or "New" judgments with up to 5 sec to respond to each item. Data analyses with d' values were conducted as described above for the separate behavioral study. Seven participants' memory data were lost due to experimenter error.

fMRI Data Acquisition and Analyses

All fMRI data were acquired at the Maryland Neuroimaging Center at the University of Maryland. fMRI and structural MRI data were collected from 24 participants. One was excluded from further analyses because of excessive motion and one for falling asleep during the scan. Excessive motion was defined as greater than 3 mm total motion across the run or greater than 10% of outlier timepoints (with outliers defined as greater than 1 mm scan-to-scan deviation or 3 SD global signal). Data were collected on a Siemens (Malvern, PA) 3T Tim Trio scanner using a 32-channel head coil ($n = 17$) or 12-channel head coil ($n = 5$). Whole-brain, T2*-weighted gradient EPIs were collected (repetition time = 2000 msec; echo time = 24 msec; flip angle = 90; field of view = 19.2 cm^2) with 36 interleaved oblique slices per volume (slice thickness = 3 mm). During the Communicative Actions task, 203 volumes were collected per run (two runs in total). A structural scan and

second functional task were collected during the same scanning session, but not included in the current analyses. The Communicative Actions task was always collected after these other scans to keep the length of time between the imaging task and the recognition memory task as short as possible and consistent across participants.

SPM8 (www.fil.ion.ucl.ac.uk/spm) and in-house MATLAB scripts were used for imaging data analyses. Data were first adjusted for timing differences in slice acquisition across each volume acquisition. Data from all functional runs were realigned to the first volume of the first run using a 6-degree rigid spatial transformation. Images were then spatially normalized to Montreal Neurological Institute (MNI) space using a 12-parameter affine transformation to match to reference EPI template and spatially smoothed (FWHM = 5). A high-pass filter (128 sec) was then applied to the data. We used the artifact detection toolbox (ART; www.nitrc.org/projects/artifact_detect/) to identify volumes in which motion between two neighboring timepoints exceeded 1 mm or in which global signal deviation exceeded 3 SDs. These timepoints were used as regressors of no interest in subsequent regression analyses.

First-level analyses were conducted within each participant using the general linear model with each of the four conditions (Point-Gaze, Point-NoGaze, Reach-Gaze, Reach-NoGaze) modeled as regressors of interest and motion parameters from the realignment step and outlier timepoints were modeled as regressors of no interest. Regressors for each condition were calculated by convolving each video event per condition with a canonical hemodynamic response function (HRF) with the onset of the event at the beginning of the video and with a duration of 2500 msec. Regressors were temporally high-pass filtered (128 sec). Contrasts were estimated by averaging parameter estimates (Beta values) for each condition across each run. These contrast values for each condition were used for second-level analyses and correlation analyses. To examine how communicative signals of gaze and gesture modulate brain activity, a second-level 2 × 2 repeated-measures ANOVA was conducted using the flexible factorial design within SPM8 with the contrast estimates for each individual condition with Gaze (Gaze, NoGaze) and Gesture (Point, Reach) as factors. All analyses were thresholded at $p < .001$ and corrected for multiple comparisons at the cluster level ($p < .05$, FDR-corrected). Second-level analyses were masked using the intersection of all individual level masks identified using an implicit mask with threshold of .8. Use of this mask led to the exclusion of portions of the cerebellum, inferior temporal gyrus, OFC, and anterior MPFC from the analyses.

ROI Analyses

To examine the extent to which regions of the MENT or MS systems were modulated during action observation, average contrast values from each of the four conditions compared to baseline were extracted from regions of the MENT and MS systems identified in a meta-analysis (Van Overwalle & Baetens, 2009). To create ROIs, a 6-mm radius sphere was created at the peak coordinate for each region identified in the meta-analysis. Mentalizing regions included right and left TPJ (±50, −55, 25), posterior cingulate (0, −60, 40), and MPFC (0, 50, 20). The MS regions included bilateral anterior intraparietal sulcus (±40, −40, 45) and bilateral premotor cortex (±40, 5, 40). Given its role in both MENT and MS systems (Yang et al., 2015), we considered the bilateral pSTS (±50, −55, 10) as part of both systems. Separate 2 × 2 repeated-measures ANOVAs were conducted for each ROI to test for main effects of Gaze and Gesture as well as their interaction on average contrast values within each ROI.

To examine whether activation within MS and MENT systems were related to recognition memory, d' scores were regressed on average contrast values for each region in separate linear regression analyses. Separate regressions were run for d' scores for all target versus nontarget objects as well as for difference scores between Gaze and NoGaze conditions (i.e., (Point-Gaze + Reach-Gaze) − (Point-NoGaze + Reach-NoGaze)). These comparisons were chosen given the significant behavioral effects in Study 1. Coil type (12 vs. 32 channel) was included as a covariate in all regression analyses.

Functional Connectivity Analyses

Psychophysiological interaction (PPI) analyses were used to examine whether and how communicative context modulated connectivity between the pSTS and regions of the MENT and MS systems. PPI analyses can identify brain regions that show a change in correlation (or functional connectivity) depending on the task condition (or psychological state). All PPI analyses were conducted using the generalized PPI toolbox (McLaren, Ries, Xu, & Johnson, 2012). Whole-brain voxelwise PPI models were run for each participant and included four PPI regressors (one per condition). To create the PPI regressor for each task condition (Point-Gaze, Point-NoGaze, Reach-Gaze, Reach-NoGaze), the BOLD time series was extracted from the RpSTS seed regions for each participant. The RpSTS was chosen because that was the only region identified by the main effect of gaze. The RpSTS seed region used was the same as in the ROI analyses (±50, −55, 10). This time series was then deconvolved with the canonical HRF to estimate the neural response. This deconvolved time series was multiplied by the model for each condition (i.e., the onset and duration of each event for each condition) and then convolved with the canonical HRF, resulting in a PPI term for each condition. The seed region time series (physiological) and task timing for each condition (psychological) were included as regressors in addition to the six motion parameters and outlier timepoints as described in the task-based analyses. Because our behavioral results suggested that gaze (not point)

was the communicative signal which enhanced recognition memory, PPI analyses examined the effect of Gaze (i.e., Point-Gaze vs. Point-NoGaze and Reach-Gaze vs. Reach-NoGaze) on connectivity between RpSTS and MENT and MS systems within the point and reach conditions separately. Parameter estimates for the PPI were averaged across runs to create contrast values and submitted to second level analyses. Whole-brain t tests were conducted on contrast values to identify voxels showing a significant effect within group ($p < .05$, FDR-corrected).

To directly examine integration between RpSTS and MENT and MS systems, contrast values for the PPI interaction term were extracted from the same MENT and MS ROIs described above from each individual participant. Contrast values were extracted separately for the contrasts of Point-Gaze vs. Point-NoGaze and Reach-Gaze vs. Reach-NoGaze conditions. Four t tests were used to test whether average connectivity within MENT and MS regions, separately, differed in the Point-Gaze vs. Point-NoGaze and Reach-Gaze vs. Reach-NoGaze contrasts. Further, in exploratory analyses we examined whether pSTS connectivity with MENT and MS systems predicted subsequent memory performance for target objects specifically within the Gaze condition. To test this, we regressed the same d' differences scores for Gaze vs NoGaze conditions described above on gaze-modulated connectivity between MENT and MS ROIs and the RpSTS. These regressions were run separately for Point and Reach conditions (Point-Gaze vs. Point-NoGaze PPIs and Reach-Gaze vs. Reach-NoGaze PPIs).

RESULTS

Gaze Cues Enhance Object Recognition Memory

Behavioral Sample (Study 1)

Accuracy (hit rates) and d' values are given in Table 1 for each condition.

Total d' (combined old and familiar responses). A paired t test comparing d' scores between all target and nontarget objects collapsed across the four conditions revealed significantly greater d' to target objects than to nontarget objects ($t(34) = 4.28, p < .0001$), demonstrating that the experimenter's actions, regardless of action type, attracted participants' attention to the object and facilitated object recognition memory. To determine whether communicative context (e.g., the presence of eye contact or a communicative action) modulated recognition memory, we conducted a 1-way repeated-measures ANOVA with Condition as the repeated-measure, where condition included d' scores for target objects viewed within each of the four conditions (Point-Gaze, Point-NoGaze, Reach-Gaze, Reach-NoGaze) as well as all nontarget objects. The ANOVA revealed a main effect of Condition ($F(4, 136) = 7.43, p < .0001$). Follow-up pairwise comparisons demonstrated significantly greater d' within the Point-Gaze, Point-NoGaze, and Reach-Gaze conditions as compared to Reach-NoGaze and nontarget objects ($p < .05$, Tukey's HSD). To examine how the factors gaze and gesture differentially contributed to recognition memory, a 2 × 2 repeated-measures ANOVA was run with Gaze

Table 1. Recognition Memory Performance

	Point-Gaze	Point-NoGaze	Reach-Gaze	Reach-NoGaze	Target	Nontarget
Study 1. Behavioral Only (n = 35)						
d'						
Total	1.51 (.11)	1.40 (.10)	1.53 (.11)	1.32 (.08)	1.36 (.11)	1.06 (.06)
Old only	1.94 (.10)	1.80 (.09)	1.87 (.11)	1.72 (.10)	1.81 (.09)	1.40 (.09)
Accuracy (% hits)						
Total	75.4 (2.0)	72.7 (2.5)	74.5 (2.8)	70.4 (2.2)	73.3 (1.9)	61.6 (2.4)
Old only	66.0 (2.6)	62.9 (2.8)	64.6 (3.2)	60.0 (2.4)	63.4 (2.3)	48.1 (2.2)
Study 2. fMRI + Behavioral (n = 15)						
d'						
Total	1.34 (.193)	1.15 (.187)	1.20 (.150)	1.35 (.189)	1.18 (.141)	.563 (.114)
Old only	1.48 (.16)	1.36 (.17)	1.46 (.14)	1.33 (.14)	1.39 (.14)	.72 (.11)
Accuracy (% hits)						
Total	79.0 (3.5)	74.4 (4.4)	78.4 (3.0)	79.0 (3.1)	77.8 (3.0)	58.1 (4.1)
Old only	68.3 (4.4)	64.1 (4.9)	67.3 (4.5)	63.1 (4.1)	65.7 (4.0)	42.2 (4.4)

Means (*SE*s) are given for condition for both d' and accuracy for Study 1 and Study 2. Total is d' calculated by combining hits-familiar and hits-old and false alarm-familiar and false alarm-old. False alarm rates: Study 1: Total 24.1 (1.7), Old 9.1 (1.2); Study 2: Total 37.4 (3.9), Old 19.1 (3.0).
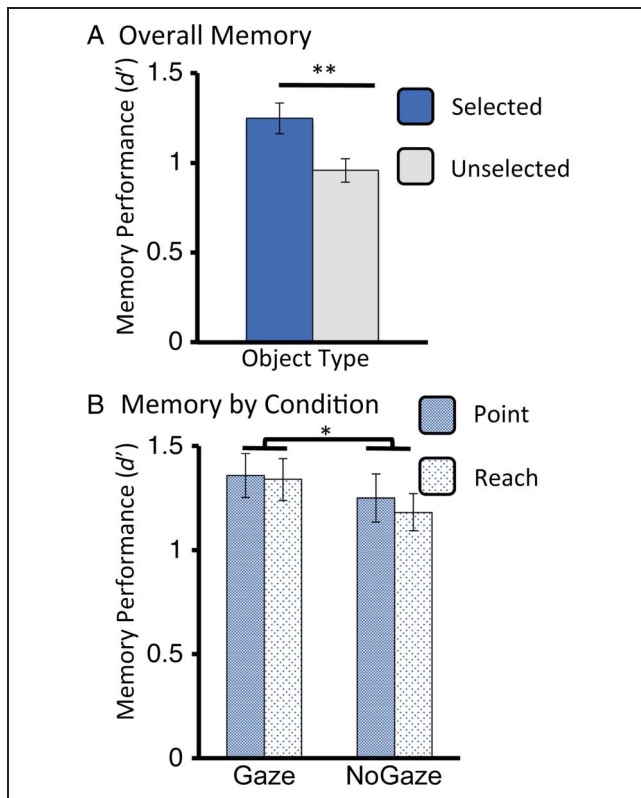
Figure 2. Effects of communicative signals on recognition memory. (A) Recognition memory (d') was significantly greater for objects that were the target of the experimenter's actions than for those that were also present in the video but not the target. (B) Objects viewed in Gaze conditions elicited significantly greater d' scores than in the NoGaze conditions.

(Gaze, NoGaze) and Gesture (Point, Reach) as the within-subject factors on d' values. This analysis demonstrated a significant effect of Gaze on object recognition memory ($F(1, 34) = 7.37, p < .01$), but neither the main effect of Gesture nor interactions between Gaze and Gesture were significant (Figure 2).

Old d'. For Old responses, d' for target items was significantly greater than for nontarget items ($t(34) = 6.21, p < .0001$). A 1-way ANOVA revealed a significant effect of Condition on Old d' scores ($F(4, 136) = 12.00, p < .0001$). Follow-up pairwise contrasts revealed that, as with Total d', Old d' for Point-Gaze, Point-NoGaze, and Reach-Gaze conditions were not significantly different from each other but were significantly different from Reach-NoGaze and nontarget items. Finally, a 2 × 2 repeated-measures ANOVA revealed a significant effect of Gaze ($F(1, 34) = 5.80, p < .022$) on Old d' scores but no effect of Gesture or interaction.

fMRI Sample (Study 2)

Behavioral analyses of the memory performance following the fMRI scan revealed a significant effect of Target object on Total d' scores ($t(15) = 8.37, p < .0001$) but

no significant effects of Condition. Recognition memory based on the Old answers alone similarly revealed significantly greater d' scores for target than nontarget objects ($t(15) = 9.48, p < .0001$). Although not significant ($F(1, 16) = 2.29, p = .15$), Old d' scores showed a consistent pattern of effects as in Study 1 (e.g., higher for Gaze than NoGaze trials). See Table 1.

fMRI Sample 1-Back Task

Accuracy on the 1-back task during the fMRI scan was high (89% ± 3%) with only two participants scoring below 75% (50% and 62.5%). Analyses conducted without these two participants revealed similar patterns to the full group and thus these participants were included in the data reported here.

Gaze Cues Modulate the STS

A whole-brain repeated-measures ANOVA demonstrated a significant effect of Gaze but no main effect of Gesture or interaction between Gaze and Gesture (Figure 3). The effect of Gaze was only seen within RpSTS [(56, −42, 10), $t = 5.08, k = 185$]. Paired t tests were conducted to compare the effect of gaze within each gesture condition. Although Point-Gaze versus Point-NoGaze elicited a significant cluster within RpSTS, no significant effect of Gaze was seen within the Reach condition (Reach-Gaze vs. Reach-NoGaze). However, the reverse contrast of Reach-NoGaze versus Reach-Gaze revealed significant activation within several prefrontal and parietal regions including left middle frontal gyrus [(−40, 18, 34), $t = 3.53, k = 213$], right inferior parietal lobe [(48, −52, 46), $t = −5.77, k = 306$], and SMA [(2, 28, 54), $t = −4.74, k = 141$]. No significant activation was found for the Point-NoGaze versus Point-Gaze contrast (Figure 4).

MENT and MS ROI Analyses

Consistent with the whole-brain analyses, the RpSTS ROI showed a significant main effect of Gaze [$F(1, 21) = 15.2, p < .0008$], and this was also significant in the left
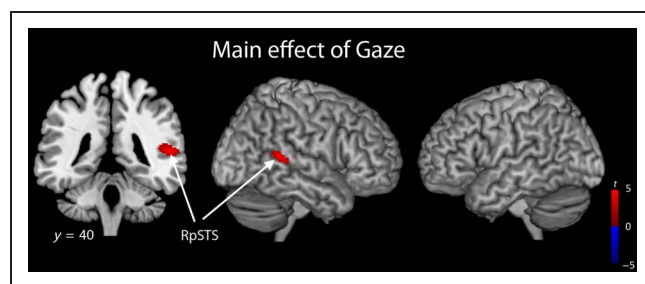

Figure 3. Neural correlates of ostensive gaze cues. The whole-brain statistical map of the significant main effect of Gaze is overlaid on a template brain in MNI space ($p < .001$, cluster-corrected $p < .05$ FDR-corrected).
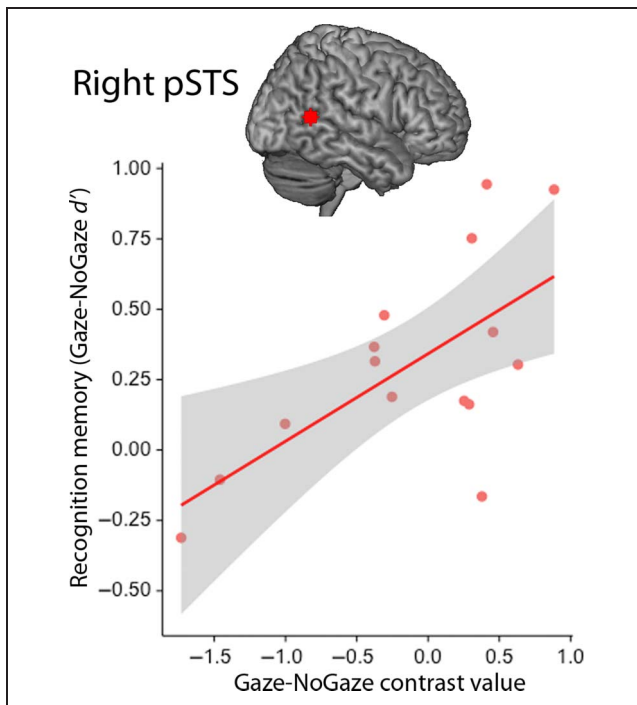
Figure 4. RpSTS activity predicts recognition memory. Greater activation to Gaze versus NoGaze conditions in the RpSTS predicts greater subsequent memory for objects viewed in Gaze compared to NoGaze conditions. Gaze-NoGaze contrast values extracted from the RpSTS ROI (top) are plotted on the y axis, and Gaze-NoGaze d' difference scores are plotted on the x axis. Regression lines with 95% confidence intervals are depicted.

hemisphere [$F(1, 21) = 9.58, p < .006$]. The RpSTS also showed a significant main effect of Gesture [$F(1, 21) = 19.7, p < .0002$], with a greater response to Reach than Point conditions. No significant main effects or interactions of Gaze or Gesture were found within any other regions of the MENT system (right TPJ, left TPJ, MPFC, posterior cingulate). The MS regions, however, revealed a significant interaction between Gaze and Gesture but no main effect of either [left anterior intraparietal sulcus, $F(1, 21) = 5.04, p < .036$; right anterior intraparietal sulcus, $F(1, 21) = 5.03, p < .036$]. In both regions, the interaction was driven by increased activation during the Reach condition in NoGaze conditions, as compared to Gaze conditions. However, follow-up pairwise comparisons did not reveal significant differences between conditions.

STS Activation to Gaze Cues Is Related to Subsequent Recognition Memory

To determine whether differential activation for Gaze versus NoGaze conditions was related to recognition memory performance, a Gaze versus NoGaze difference score was calculated with contrast values (i.e., (Point-Gaze + Reach-Gaze) − (Point-NoGaze + Reach-NoGaze)) for each ROI. d' values for memory for objects viewed in Gaze versus NoGaze conditions and Target versus Nontarget objects were regressed on these Gaze versus NoGaze

values in separate regression models. Only the RpSTS demonstrated a significant relation between differential activation to Gaze cues and recognition memory performance. Specifically, greater RpSTS activation for Gaze, as compared with NoGaze, conditions was related to greater recognition memory (d') for objects viewed in Gaze compared to NoGaze conditions [$F(1, 14) = 8.86, p < .01$]. There were no significant relations between activation within other regions of the MENT and MS systems and d' scores.

Functional Connectivity Analyses

PPI analyses were carried out with the RpSTS seed region given that this region showed an effect of communicative context, specifically sensitivity to gaze conditions, in both whole-brain and ROI analyses. We examined both whole-brain effects of gaze conditions on RpSTS functional connectivity and explored individual differences in the extent to which gaze modulated RpSTS connectivity. Whole-brain PPI analyses with the RpSTS seed did not reveal any significant clusters in which connectivity differed as a function of Gaze condition (i.e., Gaze vs. NoGaze) when corrected for multiple comparisons at $p < .05$ (FDR-corrected cluster). Despite no significant effect at the group level, we conducted exploratory analyses to determine whether the extent to which the presence of gaze-modulated RpSTS connectivity to MENT and MS regions was related to recognition memory performance, similar to the regressions above. Modulation of RpSTS connectivity with the MENT system for Gaze versus NoGaze conditions was not related to d' for either Target versus Nontarget objects nor Gaze versus NoGaze objects.

DISCUSSION

In the current study, we examined the effect of communicative context on object recognition memory and brain activation by manipulating the presence of ostensive signals (Gaze vs. NoGaze) and communicative actions (Point vs. Reach). In the behavioral study, all conditions except Reach-NoGaze led to significantly greater recognition memory for the target objects compared to the nontarget objects. This suggests that the presence of any communicative signal, whether direct gaze or point, has the ability to enhance recognition memory. However, gaze preceding an action was the only communicative signal that reliably increased recognition memory when evaluating the main effects of gaze and gesture across conditions. This modulation by gaze but not point is consistent with the theoretical framework of previous research, which suggests that, although pointing is inherently communicative, the specific intention behind the action cannot be interpreted without a broader communicative context (Liebal, Behne, Carpenter, & Tomasello, 2009; Tomasello et al., 2007; Behne et al., 2005; Clark, 1996; Sperber & Wilson, 1996). The ostensive signal of direct gaze indicates

a deliberate communicative act for both referential point and reach actions. Thus, a previously nonreferential action (i.e., reach) becomes a communicative, referential action with the presence of an ostensive signal (cf. Csibra & Gergely, 2009; Sperber & Wilson, 1996) and the presence of this communicative, referential action toward an object enhances recognition memory for that object. Similarly, in the fMRI experiment, gaze preceding an action differentially modulated the RpSTS, and the extent to which the RpSTS was modulated by gaze was related to subsequent recognition memory performance. These findings highlight the important role of the ostensive signal of gaze in creating and the pSTS in detecting communicative, referential acts that promote memory for objects of shared attention.

Previous research with infants and toddlers has demonstrated effects of joint attention or communicative actions on object processing (e.g., Hoehl et al., 2008; Yoon et al., 2008; Reid & Striano, 2005). Our findings are consistent with this work and demonstrate that communicative signals continue to alter object processing into adulthood. Using a change blindness paradigm, a previous study (Marno et al., 2014) also demonstrated effects of communicative signals on object processing in adults. Marno and colleagues demonstrated that the presence of ostensive, communicative signals (e.g., direct gaze, smiling, waving) before an object-directed action (point or reach) led to worse detection of a change in object location and better (though nonsignificant) detection of a change in object identity. Our findings complement and extend these to demonstrate significant positive effects of ostensive signals before an object-directed action (point or reach) specifically on subsequent recognition memory for cued objects in adults. Although this pattern of behavioral findings did not reach significance in our fMRI sample, d' scores and accuracy for Old responses did show a numerical pattern consistent with greater recognition memory in the Gaze conditions (Table 1). The lack of statistical significance may be due to the differences in fMRI testing conditions including a change in context for the recognition memory test, smaller sample, and repetition of each video to increase statistical power for fMRI analyses.

Consistent with our behavioral findings, the fMRI study found a main effect of Gaze in the RpSTS. Because all conditions contained intentional human actions, the modulation of the RpSTS in cases where direct gaze preceded action suggests that the RpSTS is sensitive to ostensive signals indicating a deliberate communicative act and not simply action processing or intention processing alone. These findings are consistent with previous studies demonstrating greater pSTS activity during joint attention tasks (Redcay, Kleiner, & Saxe, 2012; Redcay et al., 2010; Materna et al., 2008), communicative games without human action processing (Noordzij et al., 2009), and in response to ostensive (compared to non-ostensive) signals before noncommunicative actions (Ferri et al., 2014; Tylen et al., 2012; Schilbach et al., 2006).

These data extend past studies on the neural correlates of gaze and joint attention and provide an important novel contribution to this literature by demonstrating that this pSTS sensitivity to communicative contexts is related to subsequent recognition memory for objects viewed in a communicative context. This brain–behavior relation provides a neural link for theories of joint attention that suggest these communicative signals are important in constraining information processing and promoting learning about one's shared environment—such as the identity of an object (Csibra & Gergely, 2009; Mundy et al., 2009). Specifically these findings highlight an important role of the STS in gaze-modulated object memory. Although the current study focused on how communicative context modulates object memory, these data mirror findings in other domains examining the role of context in modulating memory. For example, pictures viewed in a high reward context are remembered better than in low reward and the extent of activation within reward-relevant brain regions differentiated remembered and forgotten pictures (Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006). Similarly, stimuli eliciting high arousal are remembered better than low arousal, and this effect is mediated by the amygdala (e.g., Kensinger & Corkin, 2004). In the current study, the ostensive signal of gaze preceding an object-directed action recruits pSTS, which may promote greater attention to or deeper encoding of objects of another's interest, leading to greater recognition memory for those objects. Future studies should examine whether these pSTS effects on object memory are specific to gaze or whether other communicative signals such as verbal requests, which also rely on pSTS (e.g., Redcay, 2008) would have similar effects.

Although the findings of this study and past literature emphasize the role of the pSTS in the reception of communicative acts, the pSTS has also been shown to be involved in the initiation of communicative acts (Redcay et al., 2012; Cleret de Langavant et al., 2011; Noordzij et al., 2009). Initiating joint attention can also improve recognition memory for the target object, with even greater improvements than when responding to joint attention (Kim & Mundy, 2012). An important future direction will be to investigate how brain activation during both initiating and responding to joint attention is related to memory for objects of shared attention.

Although we predicted an effect of communicative signals within the pSTS, the fact that only the pSTS was sensitive to communicative signals was surprising. We had predicted greater involvement of regions of the MENT system, particularly the MPFC, given its role in processing communicative intent (Kampe et al., 2003). For example, the MPFC and other regions of the MENT system are engaged when participants hear someone call their own name (Kampe et al., 2003), perceive an extended period of eye contact (Kuzmanovic et al., 2009), view social facial gestures (Schilbach et al., 2006), and jointly attend to an object with a partner (Redcay et al., 2012; Schilbach et al.,

2010). These communicative signals may establish a "meeting of the minds" (Amodio & Frith, 2006) in which participants represent their partner's mental states in the context of a dyadic interaction. One reason why we may not have seen MENT activation in the current study is that the relatively brief eye contact may not have been sufficient to create a perceived meeting of the minds. In fact, MPFC activation increases with increasing duration of direct gaze (Kuzmanovic et al., 2009).

On the basis of previous findings (Trapp et al., 2014; Schippers et al., 2009; Ferrari et al., 2003), we had hypothesized that increasing social-interactive context (through gaze cues) could modulate the MS system but did not find support for this hypothesis, either in ROI analyses or functional connectivity analyses. Studies finding effects of communicative context on the MS system have differed from ours in several ways. In one study, participants played a game of charades with a social partner (Schippers et al., 2009). This task required careful attention to hand actions to infer their communicative intent, which may have more robustly driven MENT and MS systems. This interpretation would be consistent with previous studies suggesting that the MS system is involved in analyzing the "how" of an action rather than the "why" (Spunt, Satpute, & Lieberman, 2011). In a second study (Ciaramidaro et al., 2014), the communicative actions involved an experimenter holding an object in front of her (toward the participant), which is a hand action that may more strongly prime a reciprocal motor response from a social partner (even if not acted upon). Contrary to expectations, in the current study, regions of the MS system were recruited to a greater extent during the least communicative condition—the reach conditions without gaze—as compared to reach with gaze. The reach condition may have facilitated greater attention to the specific hand actions, as these differed slightly depending on the object. Furthermore, without gaze cues, more attention may have been paid to the hands than the eyes. This post hoc conclusion remains speculative, given that eye-tracking data were not collected during fMRI scanning and identification of MS system was based on coordinates from other papers rather than an action execution localizer task. Thus, communicative context may modulate the MS system when detection of the communicative intent demands attention to the "how" of an action or when an action affords a reciprocal response but the current data cannot directly speak to these hypotheses.

Taken together, the current study demonstrates that the presence of ostensive signals of gaze coupled with an intentional action (a point or a reach) on an object modulates the pSTS and promotes subsequent recognition memory for that object. These findings add behavioral and neural support to theories that detection of ostensive, communicative signals may promote learning about generalizable properties of one's environment such as the identity of an object (Csibra & Gergely, 2009). Importantly, these findings do not imply that communicative

context is necessary for or the only means by which learning from others can occur. Attention to others' intentional actions and goals can provide important learning opportunities even without ostensive signals (e.g., Shneidman, Todd, & Woodward, 2014; Yu & Smith, 2013). Rather these findings highlight that communicative signals promote learning above and beyond the learning seen from intentional actions alone and that the STS plays an important role in processing these communicative signals and promoting subsequent memory. These findings have important implications for autism—a disorder that is characterized by reduced engagement in joint attention (Mundy & Newell, 2007; Kasari, Sigman, Mundy, & Yirmiya, 1990) and reduced attention to social cues (Pierce, Conant, Hazin, Stoner, & Desmond, 2011; Dawson, 2008; Klin, Jones, Schultz, & Volkmar, 2003) as well as atypical structural and functional development of the pSTS (for reviews, see Pelphrey, Shultz, Hudac, & Vander Wyk, 2011; Redcay, 2008; Zilbovicius et al., 2006).

REFERENCES

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. E. (2006). Reward-motivated learning: Mesolimbic activation precedes memory formation. *Neuron, 50,* 507–517.

Amodio, D., & Frith, C. (2006). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience, 7,* 268–277.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science, 8,* 492–499.

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

Brass, M., Schmitt, R. M., Spengler, S., & Gergely, G. (2007). Investigating action understanding: Inferential processes versus action simulation. *Current Biology, 17,* 2117–2121.

Ciaramidaro, A., Becchio, C., Colle, L., Bara, B., & Walter, H. (2014). Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. *Social Cognitive and Affective Neuroscience, 9,* 909–916.

Clark, H. (1996). *Using language.* Cambridge: Cambridge University Press.

Cleret de Langavant, L., Remy, P., Trinkler, I., McIntyre, J., Dupoux, E., Berthoz, A., et al. (2011). Behavioral and neural correlates of communication via pointing. *PloS One, 6,* e17719.

Committeri, G., Cirillo, S., Costantini, M., Galati, G., Romani, G. L., & Aureli, T. (2015). Brain activity modulation during

the production of imperative and declarative pointing. *Neuroimage, 109,* 449–457.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language, 25,* 141–168.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13,* 148–153.

Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology, 20,* 775–803.

Ferrari, P. F., Gallese, V., Rizzolatti, G., & Fogassi, L. (2003). Mirror neurons responding to the observation of ingestive and communicative mouth actions in the monkey ventral premotor cortex. *European Journal of Neuroscience, 17,* 1703–1714.

Ferri, F., Busiello, M., Campione, G. C., De Stefani, E., Innocenti, A., Romani, G. L., et al. (2014). The eye contact effect in request and emblematic hand gestures. *European Journal of Neuroscience, 39,* 841–851.

Frith, U., & Frith, C. (2010). The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 365,* 165–176.

Hoehl, S., Reid, V., Mooney, J., & Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. *Developmental Science, 11,* 10–16.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology, 3,* e79.

Kampe, K., Frith, C., & 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,* 5258–5263.

Kasari, C., Sigman, M., Mundy, P., & Yirmiya, N. (1990). Affective sharing in the context of joint attention interactions of normal, autistic, and mentally retarded children. *Journal of Autism and Developmental Disorders, 20,* 87–100.

Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural processes for valence and arousal. *Proceedings of the National Academy of Sciences, U.S.A., 17,* 63–645.

Kim, K., & Mundy, P. (2012). Joint attention, social-cognition, and recognition memory in adults. *Frontiers in Human Neuroscience, 6,* 1–11.

Klin, A., Jones, W., Schultz, R., & Volkmar, F. (2003). The enactive mind, or from actions to cognition: Lessons from autism. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 358,* 345–360.

Kopp, F., & Lindenberger, U. (2011). Effects of joint attention on long-term memory in 9-month-old infants: An event-related potentials study. *Developmental Science, 14,* 660–672.

Kuhl, P. K., Tsao, F.-M., & Liu, H.-M. (2003). Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences, U.S.A., 100,* 9096–9101.

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

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science, 12,* 264–271.

Mainieri, A. G., Heim, S., Straube, B., Binkofski, F., & Kircher, T. (2013). Differential role of the mentalizing and the mirror neuron system in the imitation of communicative gestures. *Neuroimage, 81,* 294–305.

Marno, H., Davelaar, E. J., & Csibra, G. (2014). Nonverbal communicative signals modulate attention to object properties. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 752–762.

Materna, S., Dicke, P. W., & Thier, P. (2008). Dissociable roles of the superior temporal sulcus and the intraparietal sulcus in joint attention: A functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience, 20,* 108–119.

McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *Neuroimage, 61,* 1277–1286.

Montgomery, K. J., Isenberg, N., & Haxby, J. V. (2007). Communicative hand gestures and object-directed hand movements activated the mirror neuron system. *Social Cognitive and Affective Neuroscience, 2,* 114–122.

Morales, M. (2000). Responding to joint attention across the 6- through 24-month age period and early language acquisition. *Journal of Applied Developmental Psychology, 21,* 283–298.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science, 16,* 269–274.

Mundy, P., Sullivan, L., & Mastergeorge, A. (2009). A parallel and distributed-processing model of joint attention, social cognition and autism. *Autism Research, 2,* 2–21.

Noordzij, M., Newman-Norlund, S., DeRuiter, J., Hagoort, P., Levinson, S., & Toni, I. (2009). Brain mechanisms underlying human communication. *Frontiers in Human Neuroscience, 3,* 14.

Pelphrey, K. A., & Carter, E. J. (2008). Brain mechanisms for social perception: Lessons from autism and typical development. *Annals of the New York Academy of Sciences, 1145,* 283–299.

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

Pelphrey, K. A., Shultz, S., Hudac, C. M., & Vander Wyk, B. C. (2011). Research review: Constraining heterogeneity: The social brain and its development in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 52,* 631–644.

Pelphrey, K. A., Viola, R. J., & Mccarthy, G. (2004). When strangers pass: Processing of mutual and averted social gaze in the superior temporal sulcus. *Psychological Science, 15,* 598–603.

Pierce, K., Conant, D., Hazin, R., Stoner, R., & Desmond, J. (2011). Preference for geometric patterns early in life as a risk factor for autism. *Archives of General Psychiatry, 68,* 101–109.

Redcay, E. (2008). The superior temporal sulcus performs a common function for social and speech perception: Implications for the emergence of autism. *Neuroscience and Biobehavioral Reviews, 32,* 123–142.

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., et al. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *Neuroimage, 50,* 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,* 1–14.

Redcay, E., & Saxe, R. (2013). Do you see what I see? The neural bases of joint attention. In H. S. Terrace & J. Metcalfe (Eds.), *Agency and joint attention* (pp. 216–237). New York: Oxford University Press.

Reid, V. M., & Striano, T. (2005). Adult gaze influences infant attention and object processing: Implications for cognitive

neuroscience. *European Journal of Neuroscience, 21,* 1763–1766.

Reid, V. M., Striano, T., Kaufman, J., & Johnson, M. H. (2004). Eye gaze cueing facilitates neural processing of objects in 4-month-old infants. *NeuroReport, 15,* 15–17.

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., et al. (2010). Minds made for sharing: Initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience, 22,* 2702–2715.

Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., et al. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia, 44,* 718–730.

Schippers, M. B., Gazzola, V., Goebel, R., & Keysers, C. (2009). Playing charades in the fMRI: Are mirror and/or mentalizing areas involved in gestural communication? *PloS One, 4,* e6801.

Shneidman, L., Todd, R., & Woodward, A. (2014). Why do child-directed interactions support imitative learning in young children? *PloS One, 9,* e110891.

Sperber, D., & Wilson, D. (1996). *Relevance: Communication and cognition*. Oxford: Blackwell Publishers.

Spunt, R. P., Falk, E. B., & Lieberman, M. D. (2010). Dissociable neural systems support retrieval of how and why action knowledge. *Psychological Science, 21,* 1593–1598.

Spunt, R. P., Satpute, A. B., & Lieberman, M. D. (2011). Identifying the what, why, and how of an observed action: An fMRI study of mentalizing and mechanizing during action observation. *Journal of Cognitive Neuroscience, 23,* 63–74.

Striano, T., Reid, V. M., & Hoehl, S. (2006). Neural mechanisms of joint attention in infancy. *The European Journal of Neuroscience, 23,* 2819–2823.

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,* 675–691; discussion 691–735.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78,* 705–722.

Trapp, K., Spengler, S., Wüstenberg, T., Wiers, C. E., Busch, N. A., & Bermpohl, F. (2014). Imagining triadic interactions simultaneously activates mirror and mentalizing systems. *Neuroimage, 98,* 314–323.

Tylén, K., Allen, M., Hunter, B. K., & Roepstorff, A. (2012). Interaction vs. observation: Distinctive modes of social cognition in human brain and behavior? A combined fMRI and eye-tracking study. *Frontiers in Human Neuroscience, 6,* 331.

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

Yang, D. Y.-J., Rosenblau, G., Keifer, C., & Pelphrey, K. A. (2015). An integrative neural model of social perception, action observation, and theory of mind. *Neuroscience and Biobehavioral Reviews, 51,* 263–275.

Yoon, J. M. D., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences, U.S.A., 105,* 13690–13695.

Yu, C., & Smith, L. B. (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PloS One, 8,* e79659.

Zilbovicius, M., Meresse, I., Chabane, N., Brunelle, F., Samson, Y., & Boddaert, N. (2006). Autism, the superior temporal sulcus and social perception. *Trends in Neurosciences, 29,* 359–366.