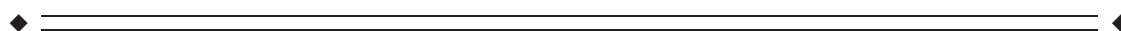


Perceived Communicative Intent in Gesture and Language Modulates the Superior Temporal Sulcus

Elizabeth Redcay,^{1*} Kayla R. Velnoskey,¹ and Meredith L. Rowe²

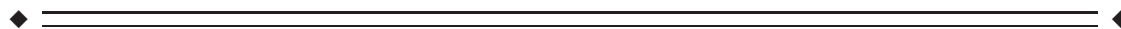
¹Department of Psychology, University of Maryland, College Park, Maryland

²Graduate School of Education, Harvard University, Cambridge, Massachusetts



Abstract: Behavioral evidence and theory suggest gesture and language processing may be part of a shared cognitive system for communication. While much research demonstrates both gesture and language recruit regions along perisylvian cortex, relatively less work has tested functional segregation within these regions on an individual level. Additionally, while most work has focused on a shared semantic network, less has examined shared regions for processing communicative intent. To address these questions, functional and structural MRI data were collected from 24 adult participants while viewing videos of an experimenter producing communicative, Participant-Directed Gestures (PDG) (e.g., “Hello, come here”), noncommunicative Self-adaptor Gestures (SG) (e.g., smoothing hair), and three written text conditions: (1) Participant-Directed Sentences (PDS), matched in content to PDG, (2) Third-person Sentences (3PS), describing a character’s actions from a third-person perspective, and (3) meaningless sentences, Jabberwocky (JW). Surface-based conjunction and individual functional region of interest analyses identified shared neural activation between gesture (PDGvsSG) and language processing using two different language contrasts. Conjunction analyses of gesture (PDGvsSG) and Third-person Sentences versus Jabberwocky revealed overlap within left anterior and posterior superior temporal sulcus (STS). Conjunction analyses of gesture and Participant-Directed Sentences to Third-person Sentences revealed regions sensitive to communicative intent, including the left middle and posterior STS and left inferior frontal gyrus. Further, parametric modulation using participants’ ratings of stimuli revealed sensitivity of left posterior STS to individual perceptions of communicative intent in gesture. These data highlight an important role of the STS in processing participant-directed communicative intent through gesture and language. *Hum Brain Mapp* 00:000–000, 2016. © 2016 Wiley Periodicals, Inc.

Key words: social perception; sentence processing; communication; second-person neuroscience; semantics; functional MRI; surface analyses



Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: NSF ADVANCE Program at the University of Maryland

*Correspondence to: Elizabeth Redcay, 2147D Biology-Psychology Building, University of Maryland, College Park, MD 20742. E-mail: redcay@umd.edu

Received for publication 25 November 2015; Revised 25 March 2016; Accepted 27 April 2016.

DOI: 10.1002/hbm.23251

Published online 00 Month 2016 in Wiley Online Library (wileyonlinelibrary.com).

INTRODUCTION

Social-cognitive abilities, such as use of gesture and engagement in episodes of joint attention, are pivotal to the acquisition and development of language [Baldwin, 1991; Brooks and Meltzoff, 2008; Goldin-Meadow, 1998; Rowe and Goldin-Meadow, 2009b]. Some have hypothesized that language emerged phylogenetically and ontogenetically from social-cognitive systems, specifically the understanding and coordinating of intentions through joint action [Clark, 1996; Tomasello et al., 2005]. This common system encompassing language and social cognition allows for the flexible use of gesture and linguistic symbols in the pursuit of communication. However, the neural systems supporting language and social cognition are often studied as separate domains. Identifying core brain systems supporting processes that cut across traditional boundaries will provide insights into the link between these behaviors as well as inform our understanding of functional brain organization.

The domains of gesture and language provide an ideal window to identify core brain mechanisms underlying social communication as evidence suggests a shared cognitive basis [McNeill, 1992] and neural basis [Andric and Small, 2012; Bates and Dick, 2000, 2002; Enrici et al., 2011; Redcay, 2008; Xu et al., 2009]. Developmental evidence provides direct behavioral links between gesture and language. Specifically, the number of meanings 1-year olds produce in gesture is predictive of their early vocabulary ability as well as their longer term vocabulary size at 54 months [Rowe and Goldin-Meadow, 2009a]. Further, this relation between gesture and language is specific to the language ability under study. That is, producing meanings in gesture relates to producing meanings in speech (i.e., vocabulary), but not to sentence complexity (i.e., syntax), whereas producing early gesture-word combinations (multimodal sentences) at 18 months predicts sentence complexity at 42 months, but not vocabulary size [Rowe and Goldin-Meadow, 2009b]. Behavioral data from adults similarly suggest that gesture and language processing share a common cognitive representation [McNeill, 1992]. Taken together this theoretical and behavioral work suggests that communication via language may be derived from understanding gesture as communicative and that these domains form an integrated communicative system [e.g., Goldin-Meadow, 1998].

Evidence from brain imaging studies suggests this cognitive link between gesture and language is reflected in shared neural functional organization [Andric and Small, 2012; Andric et al., 2013; Enrici et al., 2011; Redcay, 2008; Straube et al., 2012; Xu et al., 2009]. A large body of work has established that gesture processing relies on a bilateral neural system encompassing perisylvian regions, including inferior frontal gyrus (IFG) and posterior temporal regions. As these are also regions which are engaged for language processing, researchers have suggested that gesture and language share a common neural basis [review, Bates and

Dick, 2002; Dick and Broce, 2015]. While this work establishes a role of these language-relevant regions in gesture processing, it doesn't directly test for shared regions of gesture and language processing within the same individuals.

Most work that has examined language and gesture processing within the same individuals has focused on the integration of gestural information with co-occurring speech [Dick et al., 2009, 2012, 2014; Holle et al., 2010; Hubbard et al., 2009, 2012; Skipper et al., 2007; Straube et al., 2010, 2011; Willems et al., 2009]. These studies demonstrate that semantic information is extracted from gestures when presented with speech and that these semantic relations are processed primarily within IFG and posterior temporal regions [review, Bates and Dick, 2002; Dick and Broce, 2015]. Specifically, IFG activation is present when the gesture and speech information combine to create a novel semantic interpretation [Dick et al., 2009, 2014; Kircher et al., 2009; Straube et al., 2011; Willems et al., 2009] even if the gesture information alone does not convey meaning. In contrast, posterior temporal regions are modulated more when speech and co-speech gestures both can convey meaning independently (e.g., iconic gestures or pantomimes) and that meaning is the same across both modalities [Holle et al., 2008; Straube et al., 2011; Willems et al., 2009]. While these studies offer an important window into neural mechanisms supporting gesture and language integration, they do not test whether inferring meaning and communicative intent from gesture and language stimuli *alone* rely on common neural mechanisms. This is because, in these designs, language is not dissociated from gestural communication. This dissociation is important to identify core processes that support communication independent of modality, as individuals can and do use gestures alone as a medium for communication.

A relatively small number of studies have directly addressed the question of whether the same cortical regions are engaged in the processing of gestures and language when presented independently to the same participants [Andric et al., 2013; Enrici et al., 2011; Straube et al., 2013, 2012; Xu et al., 2009]. For example, Xu et al. [2009] demonstrated that comprehension of spoken words and iconic or pantomime gestures (e.g., "open a jar") recruit overlapping regions of activation within middle temporal gyrus, extending into the superior temporal sulcus, and inferior frontal gyrus [Xu et al., 2009], suggesting an amodal representation of symbolic communication. Similarly, using an anatomical region of interest approach, Andric et al. [2013] demonstrated significantly greater response to speech and emblems compared to grasping actions within right MTG, right anterior STG, and left IFG. Finally Straube et al. [2012] identified a supramodal network of semantic processing within left IFG and MTG for processing iconic single gestures (e.g., "huge"; "round") without speech and sentences (e.g., "The fisherman caught a huge fish") without gestures. While these studies provide

evidence of shared cortical representations for gesture and language, they are limited in two ways. First, the methods used leave open the possibility that semantic processing through gesture or language may be accomplished via nearby but distinct regions of cortex. This is because they relied only on conjunctions of *group*-averaged maps [cf. Fedorenko et al., 2012] or anatomical region of interest analyses and performed registration and smoothing within the volume, rather than on the cortical surface. These volumetric analyses can result in poor inter-subject alignment and blurring of signal across regions that do not respect cortical anatomy [Jo et al., 2007; Oosterhof et al., 2011]. Further, given large variability in individual activation patterns to these stimuli, overlap on group-averaged volumetric maps or significant modulation within large anatomical regions of interest could mask finer-grained functional segregation or overlap for gesture and language at the individual level [cf. Glezer and Riesenhuber, 2013]. An individual functional region of interest (fROI) approach, however, can test whether the peak region sensitive to gesture is the same as that sensitive to language within the same individuals [Fedorenko et al., 2012; Glezer and Riesenhuber, 2013]. In the current study, we address these past limits through the use of surface-based analyses and individual subject fROI analyses.

A second limitation is that previous studies examining overlapping regions of activation for language and gesture have focused on testing for a common semantic or symbolic system (for example through use of pantomimes or iconic gestures out of context) without focus on the critical role that detecting the communicative intention of the speaker (or communicative intent) plays in each of these domains. Gestures and speech convey both meaning and the intent of the producer to send a communicative message to the receiver [Sperber and Wilson, 1996]. Both gestures and language can be thought of as human actions produced by a social partner with the intention to communicate [Searle, 1969]. One seminal study has examined shared regions underlying communicative intentions that are conveyed either by linguistic (sentence) or gestural (pointing) communication [Enrici et al., 2011]. For example, in a picture with two people and a bottle on the table, the participant either saw “Please pass the bottle” in text in the linguistic condition or saw a person point to the bottle in the gesture condition. Both linguistic and gesture conditions recruited regions along the superior temporal sulcus, inferior frontal gyrus, and posterior cingulate. However, the baseline comparison condition did not control for critical stimulus properties that may have resulted in shared activation (i.e., presence of pictures of people vs. objects). Furthermore, in this study, the participant inferred the communicative intent between two characters from a third-person perspective. However, language and gesture in a social-interactive context require a second-person perspective—that is, a feeling of being personally addressed by one’s social partner. This second-person

perspective may fundamentally differ from a third-person stance [Schilbach et al., 2013]. Second-person communicative intent conveys relevance of the message to the participant [Sperber and Wilson, 1996].

Studies have begun to address the brain bases for processing gestures from a second-person versus third-person perspective and, while they generally find evidence for posterior temporal (STS and MTG) and midline (MPFC and PCC) structures, the evidence is not entirely consistent [Ciaramidaro et al., 2014; Holler et al., 2014; Nagels et al., 2015; Redcay et al., 2010; Schilbach et al., 2006]. While this literature is still nascent, some inconsistency across studies could be due to the introduction of confounds across methods used to isolate communicative intent in gestures. For example, many studies manipulate the angle at which the experimenter produces the gesture (facing forward vs. lateral) as a means to convey participant-directed communicative intent. While this manipulation nicely isolates a second vs. third person perspective while controlling for biological motion processing, it also introduces a confound by eliciting differences in spatial attention between conditions. One means to eliminate this problem is to identify neural regions supporting amodal representations of participant-directed communicative intent though both gesture and language stimuli. While each condition on its own may present a challenge to isolate participant-directed communicative intent without confound (e.g., spatial attention), common activation to both provides evidence for a shared process that supersedes confounds constrained to the modality. However, no study has yet tested for an amodal neural system supporting participant-directed communicative intent within the same group of participants.

Taken together, previous studies offer evidence for a shared network supporting gesture and language processing within posterior superior temporal sulcus, middle temporal gyrus, and inferior frontal regions. However, due to limitations discussed above, whether fine-grained functional segregation exists in the brain systems supporting participant-directed communicative intent through language and gesture remains an open question. In the current study, we overcame previous methodological limitations by using surface-based analysis methods and an individual subject functional region of interest approach to test for overlapping brain activation for gesture and language processing. With these methods we tested for amodal regions supporting semantic processing (similar to previous studies) as well as participant-directed communicative intent. In the gesture conditions participants viewed videos of an actress performing gestures that were communicative (i.e., second-person directed request, instrumental, or iconic gestures in a two gesture string) or noncommunicative (i.e., grooming gestures such as brushing her hair or rubbing her arm). Comparison of communicative (participant-directed) to noncommunicative (self-adaptor) gestures will reveal regions sensitive to

TABLE I. Stimuli examples

Sentences			Gestures	
Participant-Directed	Third-person	Jabberwocky	Participant-Directed	Self-adaptor
Hello, it's nice to meet you.	He was so tired, he overslept.	Eem tibe a pazz with his derbist.	Wave, hold out hand for shake	Smooth hair, pull shirt
I want you to listen to me.	He wore a sweater to keep warm.	Alf zopeed up in ler and glay dact.	Point forward, tap ear	Fist over mouth (cough), scratch head
I don't know. Why don't you call me?	The child bent down and smelled the rose.	Meeda saunted at the tewlaire she gwized.	Shrug, Hand like phone to ear	Crack back gesture, scratch face

both semantic processing and participant-directed communicative intent. To test for regions sensitive to language processing (including both semantics and communicative intent), we used a method similar to the previously published language localizer [Fedorenko et al., 2012], which compared sentences to jabberwocky. To test for regions sensitive to participant-directed communicative intent in language while controlling for semantic processing, we included a third language condition with the same participant-directed content as the communicative gesture strings. Comparison of these explicitly participant-directed sentences with the standard third-person-directed sentences in the language localizer identified regions sensitive to language presented in a second-person context, or with communicative intent. Thus, with this design and analyses, the current study identifies shared cortical regions involved in language and gesture processing and tests the extent to which these shared neural representations are modulated by perceived communicative intent.

METHODS

Participants

All participants were college students participating for course credit or payment through the University of Maryland SONA system. All procedures were approved by the University institutional review board. Data from the first sample of participants ($n = 13$) were used to select the gesture and sentence stimuli that would be used in the subsequent fMRI and behavioral rating tasks with the second sample of participants ($n = 28$). MRI participants were screened for any history of head injury, learning disabilities, psychoactive medication or recreational drug use as well as any contraindication for MRI scanning (e.g., metal in the body). One participant reported mild depression but was included in the study. English was the native language for all participants. Four participants were excluded from the fMRI sample because of excessive motion (see below) for a final sample of 24 (age 22(2.1) years; 14F). In the final sample 18 participants were right-handed, 3 left-handed, 1 ambidextrous, and 2 did not report handedness information. Handedness data was

determined based on self-report on a handedness scale [Chapman and Chapman, 1987].

Gesture Stimuli

We used two gesture conditions: gestures directed to the participant in a second-person context (Participant-Directed) and noncommunicative actions directed towards the experimenter (Self-adaptor gestures). For Participant-Directed gestures (PDG), two gestures from a set of 26 unique deictic (e.g., POINT), conventional (e.g., WAVE hello/goodbye), or iconic/pantomime (e.g., CALL ME gesture with hand shaped like phone put to ear) gestures were combined to make a coherent gesture string (e.g., "I'm cold. Are you?") (see Table I for examples and Supporting Information for full list of stimuli). For Self-adaptor Gestures (SG), two self-directed grooming actions were combined (e.g., brush arm, rub nose) from a set of 26 unique self-adaptor gestures such that both participant-directed, communicative gestures and self-adaptor gestures were the same length. Three of the Self-adaptor Gesture stimuli included American Sign Language (ASL) signs that were meaningless to the participants as none were familiar with ASL. To control for differences in biological motion between the two conditions, the self-adaptor gestures were chosen in order to match the amount of human motion within the PDG. Specifically, for both PDG and SG conditions, each gesture string contained two body actions. We matched these actions on proximity to the head or body (e.g., body: "I'm cold" and smooth shirt), direction of attention (e.g., lateral: point left and twist to the side), and location of the gesture in the midline or side of the screen (e.g., midline: "shhh!" and rub nose; side: wave hello and rub left shoulder). All gestures were videotaped with an experimenter wearing a visor. The visor was important for the self-adaptor gesture condition given that the cue of direct gaze combined with producing meaningless actions elicits perceptions of communicative intent [Ferri et al., 2014; Redcay et al., 2016; Tylén et al., 2012]. Because the visor was worn in the Self-adaptor condition, we also included a visor within the Participant-Directed condition, in order to ensure effects were not due simply to differences in eye contact. This

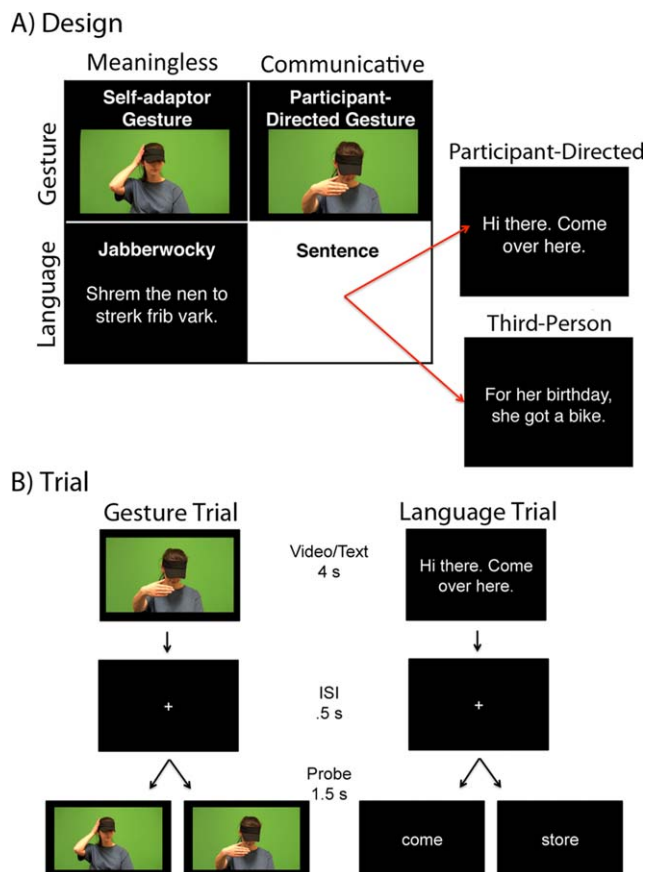


Figure 1.

fMRI task design. Participants viewed two gesture and three language conditions. Example frames are given for each stimulus in (A) Design. The timing of one trial for the Participant-Directed Gesture (left) and Standard Sentence (right) trials is given in (B). The timing was the same for all conditions. One probe image

was presented per trial that was either a match or mismatch from the previous video or text. Participants had to judge within the 1.5 s window whether it was a match or mismatch. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

inclusion of a visor may have reduced the perception of the communicative intent; however, ratings for perceived communicative intent were still high and significantly greater in the PDG than SG conditions (see Results). In total, 46 gesture strings were created for each condition, which were used in the Ratings Task described below to identify the best 39 stimuli to be used in the fMRI task.

Language Stimuli

We had three language conditions: Participant-Directed Sentences, Third-person Sentences, and Jabberwocky. All language stimuli were presented as written text (25-point Geneva font) that appeared in the center of the screen (see Fig. 1). While written text is not as inherently communicative as spoken language, written text is now widely used as a communicative, conversational medium through text-

ing, online chats, etc. (e.g., see US Department of Education, 2011). The choice of written text over spoken speech was made in order to provide the strongest test of whether overlapping activation was due to processing the communicative intent or semantic content of the message. If spoken speech were used, common activation between speech and gestures could be due to simply person perception or additional supralinguistic properties. Thus, our choice of written text biased against finding overlapping activation but also allowed for a stronger claim of the common computation underlying the overlap if found.

The Participant-Directed Sentences (PDS) were the text-equivalent of the Participant-Directed Gesture strings (e.g., "I'm cold. Are you?"). The Third-person Sentences (3PS) and Jabberwocky (JW) were created by modifying stimuli from a previously validated language localizer task [Fedorenko et al., 2012]. In order to be more consistent with the gesture stimuli and to eliminate possible confounds,

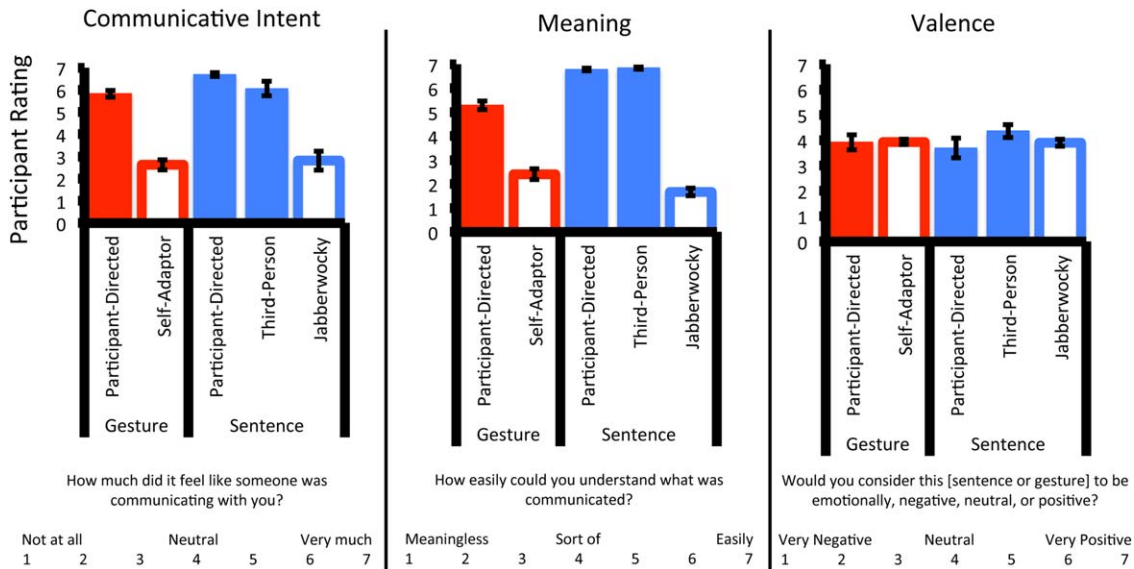


Figure 2.

Postscan stimuli ratings. Average participant ratings are given for each condition for communicative intent, meaning, and valence. Error bars represent the standard error of the mean. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

several modifications were made to these sentences. Modifications included removing proper names (e.g., Mary) to avoid possible familiarity effects, and substituting words so that the 3PS/JW sentences contained the same number of syllables on average as the Participant-Directed Sentences. To control for animacy between PDS and 3PS conditions, all Third-person Sentences contained an animate subject. The Jabberwocky sentences were created using words from the Fedorenko et al.'s (2012) jabberwocky condition, but ordered to match the average number of syllables of the other two sentence conditions. Sentences were thus matched on average number of syllables (7.25–7.75) and average mean length of utterance in words (6.1–6.7). Average word frequency was calculated using the SUBLEXTus database [Brysbaert and New, 2009], which is a corpus of American English subtitles that provides a more naturalistic, conversation-based frequency. Average word frequency did not significantly differ between PDS and 3PS sentences ($P > 0.05$). Third-person Sentences consisted primarily of transitives with modifier phrases (in some cases preposed), intransitives with modifier phrases, and compound sentences.

Stimuli Selection

To select and validate the stimuli, 13 undergraduate participants rated the stimuli on level of communicative intent (“How much did it feel like someone was communicating with you? In other words did it feel like someone was conveying something to you?”), meaning (“How easily could you understand what the person was communicating?”), and valence (“Would you consider

[this sentence or the movements] to be emotionally, negative, neutral, or positive?”). Ratings were made on a 7-point Likert scale ranging from 1 = Not at all to 7 = Very much (Communicative), 1 = Meaningless to 7 = Easily (Meaning), 1 = Very negative to 7 = Very positive (Valence). Based on these ratings, seven videos per condition were removed for a final set of 39 videos per condition to be used in the fMRI task. Because communicative intent was our primary question of interest, stimuli that scored too high on the communicative scale in noncommunicative conditions or too low in communicative conditions were removed (i.e., >3 for Jabberwocky, >5.7 for Third-person Sentences, <6 for Participant-directed Sentences, <5 for Participant-Directed Gestures >3.5 for Self-adaptor Gestures). For the remaining 39 stimuli per condition, valence ratings were in the neutral range and did not differ between conditions, with the exception of Third-person Sentences, which showed significantly more positive ratings than the other four conditions ($P < 0.05$).

fMRI Task

Stimuli were presented in an event-related design with each event lasting 6 s. On each trial, participants viewed a video or text stimulus for 4 s followed by a 0.5 s fixation cross and then a probe image for 1.5 s. The probe image was a still frame or word for the video or sentence, respectively, which was from the preceding video or text for half the trials or from a different video or text for the other half. The participant’s task was to judge whether the frame or word was the same as the video or text they just viewed (Fig. 1). This active task was chosen in order to

ensure participants were paying attention to all trials. We also chose this task as it has been used previously as a language localizer to address similar questions of neural overlap [see Fedorenko et al., 2011]. Forty trials were presented for each of the five conditions across four runs of data acquisition without repeats. Thus, each run contained 10 unique trials of each condition. Sixty seconds of null events (fixation cross) were included as baseline and the order of trials was determined using `optseq` to optimize event order (`optseq2`). All stimuli were presented using the Psychophysics Toolbox for MATLAB [Brainard, 1997].

Behavioral Ratings Task

Behavioral ratings were collected for 15 of the 24 participants following the scan session either immediately after the scan or up to 3 months for one participant. Ratings were not collected from the first six participants and ratings from three were lost due to experimenter error. The Ratings task was the same as in the Stimuli Selection section described above. Participants judged each video and text stimulus on communicative intent, meaning, and valence using the questions described above and in Figure 2.

MRI Data Collection

Structural and functional MRI data were collected at the Maryland Neuroimaging Center at the University of Maryland from 28 participants. Four participants' data were excluded due to excessive head motion or falling asleep for a final sample of 24. For two participants only three runs were included in the analysis due to excessive head motion during one run. Excessive motion was defined as greater than 3 mm total motion across any single 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 3T Siemens Tim Trio scanner using a 32-channel head coil ($n = 20$) or 12-channel head coil ($n = 4$). For the functional scan, whole brain, T2*-weighted gradient echo planar images (EPI) were collected (repetition time = 2,000 ms; echo time = 24 ms; flip angle = 90°; field of view = 19.2 cm²) with 36 interleaved oblique slices per volume (slice thickness = 3 mm) and 252 volumes per run. Structural data were acquired using a T1-weighted MPRAGE sequence (192 slices in the sagittal plane, slice thickness = 0.9 mm; repetition time = 1,900 ms; echo time = 2.32 ms).

fMRI and MRI Data Analyses

Surface-based fMRI analyses

To perform surface-based fMRI analyses, cortical surface models were created from the structural MRI data using Freesurfer's (version 5.1.0 with RedHat 6.3 Linux terminal)

automated pipeline, which has been documented extensively elsewhere [Dale et al., 1999; Desikan et al., 2006; Fischl et al., 1999a,b; Fischl and Dale, 2000; Fischl et al., 1999a,b,]. All subsequent analyses were conducted using the Analysis of Functional Neuroimages (AFNI) [Cox, 1996] and surface-mapping (SUMA) programs [Saad et al., 2004; Saad and Reynolds, 2012]. For each individual subject, SUMA was used to create standard-mesh surfaces (MNI N27) with 141,000 nodes from the surface models output from Freesurfer. These surface models were aligned with the structural volume and an aligned surface volume was created. This surface volume was used to align the functional data to the surface in subsequent processing steps.

Initial preprocessing steps for functional data were performed in volume space. These included correction for differences in slice time acquisition within each volume and registration of each functional volume to the first volume of the experiment using a rigid transformation. Functional data were transformed from oblique to cardinal orientation to match the structural scan and then co-registered with the structural volume. The surface volume described above was then aligned to the functional data before projecting the functional volume data (timeseries) to the surface. On the surface, data were then smoothed using a Gaussian smoothing kernel with full-width half maximum of 5 mm and intensity normalized. Smoothing was performed on the surface in order to avoid volumetric smoothing errors where signal from non-neighboring voxels (e.g., from two gyri that touch in volume space) would be smoothed.

For first-level analyses, Generalized Least Squares regression analyses were run using the REML estimation methods to account for temporal autocorrelation in the time series. The regression included regressors for each of the five conditions (Participant-Directed Gestures, Self-adaptor Gestures, Participant-Directed Sentences, Third-person Sentences, and Jabberwocky Sentences) as well as nuisance regressors including the baseline and linear, quadratic, and cubic trends as well as twelve motion regressors to model the residual effects of head motion. The motion regressors were the frame deviation at each volume for the six directions of translation and rotational motion (roll, pitch, yaw, x , y , z) and their derivatives. Additionally, outlier volumes were censored from the analyses. Regressors for each of the five conditions were created by convolving a gamma-variate basis function with the stimulus timing function with a duration of 6 and amplitude of 1. Contrasts were estimated for each condition of interest and comparisons of gesture type (i.e., PDGvsSG) and sentence type (i.e., PDSvs3PS, PDSvsJW, and 3PSvsJW). Additionally a main effect of communicative and semantic content was calculated (PDS + PDG + 3PS vs. SG + JW).

Second-level analyses

Coefficients and t -statistics for each contrast were brought to second-level analyses using mixed effect models (3dMEMA) [Chen et al., 2013] to model both within-

and between-subject variance. Specifically, for each contrast, we calculated an effect of group across all participants for each node using mixed effect models. All second-level analyses were corrected for multiple comparisons using a cluster-correction of 250 mm² which maintained an overall alpha of $P < 0.05$ with a voxel threshold of $P < 0.001$. The minimum cluster-volume needed was estimated using Monte Carlo simulations (10,000 iterations) on the cortical surface representation.

Overlap Analyses

Conjunction analyses

To test for overlapping activation at the group level for gesture and language stimuli, conjunction analyses were conducted by multiplying group-level contrast maps to identify voxels significant at $P < 0.05$, corrected for both contrasts [Nichols et al., 2005].

Individual functional region of interest analyses

To determine whether overlap exists at the individual subject level, we conducted a within-subjects functional region of interest analysis. Note this method differs from studies using anatomical regions of interest [Dick et al., 2009, 2012, 2014; Skipper et al., 2007] or group-level functional regions of interest [Holle et al., 2008] that aim to provide greater anatomical or functional precision in relation to previous studies. Rather the goal of this approach is to directly test for overlapping regions of activation to gesture and language stimuli within the same individual with greater spatial resolution than group-averaged methods [see Fedorenko et al., 2012 for discussion]. Specifically, ROIs were created for each individual participant for the contrast of Participant-Directed vs. Self-adaptor Gesture (PDG vs. SG). Based on the group overlap maps, these regions of interest were created within bilateral superior temporal sulcus, including an anterior, mid, and posterior region. Peak nodes were selected if they fell within either of these three regions of the STS at $P < 0.05$, 20 node minimum, at the individual level. The freesurfer parcellation was used as a guide to choose clusters within the superior temporal sulcus. Additionally, the Yeo functional parcellation scheme [Yeo et al., 2011] was used in conjunction with freesurfer parcellations as guides to determine clusters within the pSTS/TPJ. A 6-mm sphere was centered around the peak node to create the region of interest within each individual. Coefficient values from each of the three language conditions (PDS, 3PS, and JW) were extracted from within each of these ROIs for each subject. Coefficients were entered into a one-way repeated-measures ANOVA using JMP statistical software (JMP Pro 11. SAS Institute Inc., Cary, NC) to identify a main effect of condition for each region of interest. Pairwise t -tests were conducted for regions showing a significant effect of condition.

Parametric Modulation of Communicative Rating

To examine the extent to which individual perceptions of communicative intent for each gesture stimulus modulated brain activation, participant ratings of communicative intent (Fig. 2) for all gesture stimuli were entered into a parametric modulation analysis for each individual participant who provided ratings ($n = 15$). Only gestures were examined parametrically as the ratings of the sentence stimuli did not provide sufficient variability in many of the subjects. Two parametric analyses were run. First, all movie stimuli (i.e., PDG and SG) were combined into a single condition with a single regressor of communicative rating. All text stimuli were modeled as a separate condition with no parametric regressor so that text condition effects were not modeled as baseline. In the second analysis, each condition was modeled separately and both gesture types had a regressor for the communicative rating of each stimulus. Text stimuli were included as before. Regressors were mean-centered. For both models, the same nuisance regressors were included in the model as in the nonparametric first-level analyses. As before, Generalized Least Squares regression analyses were used with the REML estimation methods. First-level contrasts of the ratings for the gesture stimuli were entered into a second-level mixed effect analysis and corrected for multiple comparisons. To determine whether parametric modulation of communicative intent was seen in the regions demonstrating overlap between participant-directed language and gesture, an ROI analysis was conducted in which contrast values from the parametric analysis were extracted from the conjunction map of Participant-Directed vs. Third-person Sentences and Participant-Directed vs. Self-adaptor Gestures. A one-way t -test was used to determine whether contrast values differed significantly from zero.

RESULTS

Rating of Gesture and Language Stimuli

Behavioral ratings collected following the scan session confirmed that participants perceived the Participant-Directed Gestures as significantly more communicative than the Self-adaptor Gestures and the Participant-Directed Sentences as more communicative than the Third-person Sentences or Jabberwocky. A one-way ANOVA revealed a significant effect of condition on communicative ratings ($F_{(4,56)} = 53.77$, $P < 0.0001$). Pairwise contrasts corrected at $P < 0.05$ showed significantly higher ratings for Participant-Directed Sentences [mean 6.68 (0.77)] than any other condition, including Third-person Sentences [mean 6.05 (1.4)]. Third-person Sentences were rated as more communicative than Jabberwocky [mean 2.81 (1.9)], and Participant-Directed Gestures [mean 5.82 (1.3)] were rated as significantly more communicative than Self-adaptor Gestures [mean 2.63 (1.5)] ($P < 0.05$, corrected). Similarly, participants' ability to understand what

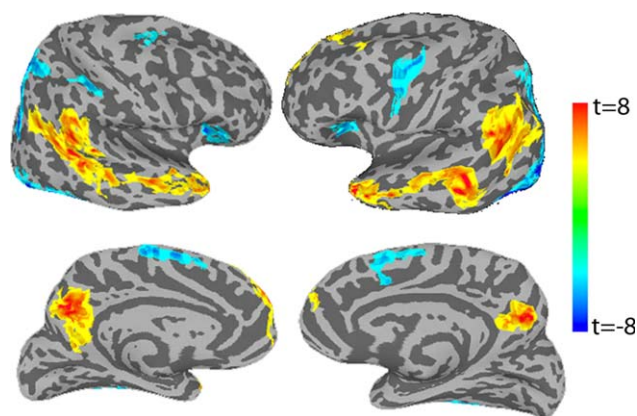


Figure 3.

Main effect of communicative, meaningful signals. Regions showing a greater response to communicative, meaningful conditions (PDG, PDS, 3PS) than meaningless, noncommunicative conditions (SG, JW) are displayed in warm colors, where colors represent t -values. The reverse contrast is shown in cool colors. Colors reflect t -values. Maps are projected onto a standardized inflated surface (MNI_N27) and thresholded at $P < 0.05$, corrected.

was being communicated (i.e., the meaning) showed a main effect of condition ($F_{(4,56)} = 270.15$, $P < 0.0001$). Both Participant-Directed Sentences [mean 6.78 (0.68)] and Third-person Sentences [mean 6.85 (0.58)] showed significantly greater meaning ratings than any other condition, but there was no difference between those two conditions. Participant-Directed Gestures [mean 5.28 (1.5)] were rated as more meaningful than Self-adaptor Gestures [mean 2.41 (1.39)] ($P < 0.05$, corrected). In these participant ratings, ratings of valence did unexpectedly differ by condition ($F_{(4,56)} = 22.9$, $P < 0.0001$). While all ratings were in the range of neutral (Fig. 2), Third-person Sentences [mean 4.36 (0.26)] were consistently rated as more positive than the other conditions while Participant-Directed Sentences [mean 3.68 (0.39)] were rated as slightly more negative than the other conditions [PDG mean 3.92 (0.30); SG mean 3.92 (0.10); JW mean 4.36 (0.14)] ($P < 0.05$, corrected) (Fig. 2).

Shared Regions for Processing Gesture and Language

We investigated the shared neural systems supporting gesture and language processing in several ways. First, a main effect of communicative and meaningful signals (i.e., sentences and gesture vs self-adaptor gestures and jabberwocky) was seen along the anterior to posterior extent of bilateral STS and MTG, as well as left precuneus and bilateral dorsomedial prefrontal cortex (dMPFC) (Fig. 3, Table II).

Conjunction analyses

Second, in a direct test of shared neural systems, a conjunction analysis was conducted to identify voxels that showed significant activation to both gesture and language stimuli when each contrast was examined independently. The contrast of participant-directed gestures to self-adaptor gestures elicited significant activation within left middle and posterior STS, extending posteriorly into the temporoparietal junction (TPJ), right middle and anterior STS, and left inferior frontal gyrus (LIFG). Comparison of Third-person Sentences versus Jabberwocky revealed activation within bilateral TPJ, middle temporal gyrus (MTG), anterior STS, and midline structures including PCC and dMPFC (Table III, Fig. 4). The TPJ region encompassed the Inferior Parietal Gyrus, Supramarginal Gyrus, and posterior Superior Temporal Sulcus. For brevity and consistency with previous literature we refer to this region as the temporoparietal junction. Conjunction analyses revealed significant overlapping activation for gestures and language within two regions of superior temporal sulcus: one in mid-STS and one in posterior STS extending into the TPJ.

Individual Functional Region of Interest Analyses

Group analyses can identify candidate regions of shared activation for gesture and language stimuli, but because the anatomical region sensitive to each stimulus type varies significantly between individuals, nearby regions that are responsive to two stimuli can appear to be overlapping at the group level. To demonstrate overlapping activation at the *individual* level, we conducted functional region of interest (ROI) analyses in which we identified the peak node discriminating Participant-Directed versus Self-adaptor Gestures within the superior temporal sulcus for each individual person (Fig. 5). We then examined whether these regions maximally sensitive to differences in gesture types on an individual level also revealed significant differences between Sentence and Jabberwocky stimuli. All six regions demonstrated a main effect of language condition. Paired t -tests ($P < 0.05$, corrected) revealed that the bilateral pSTS/TPJ and anterior STS/STG regions showed significantly greater activation to both sentence conditions (PDS and 3PS) than to Jabberwocky, consistent with the role of these regions in sentence comprehension [e.g., Fedorenko et al., 2011; Price, 2010]. See Table IV for details on ROI results.

A Modal Representation of Communicative Intent in pSTS

Perceived communicative intent in gesture and language

Communicative intent in text. Participant ratings demonstrated that the Participant-Directed Sentences were perceived as more communicative than the Third-Person

TABLE II. Main effect of meaningful, communicative signals

Contrast Region	Hemi	# nodes	Area, mm ²	Peak node	x	y	z	Peak t-value
Meaningful/communicative > meaningless/noncommunicative								
Superior Temporal Sulcus	RH	8,863	5,948.8	154,937	43	-50	14	8.612
Inferior parietal lobe/supramarginal gyrus (temporoparietal junction)	LH	4,788	4,371.6	146,394	-60	-47	26	8.26
Middle temporal gyrus/superior temporal sulcus	LH	4,743	3,375.4	20,665	-48	16	-32	9.31
Subparietal sulcus (precuneus)	LH	3,651	2,577.7	100,210	-6	-56	41	8.79
Dorsal medial prefrontal cortex	LH	2,483	2,450.9	61,416	-8	66	12	7.87
Subparietal sulcus (precuneus)	RH	1,469	980.88	146,838	14	-53	37	8.69
Superior frontal gyrus (dorsal medial prefrontal cortex)	RH	231	282.38	51,390	8	59	5	6.59
Meaningless/noncommunicative > meaningful/communicative								
Middle occipital gyrus	LH	8,407	7,421.8	161,414	-35	-92	8	-11.92
Inferior occipital sulcus	RH	3,582	2,813.5	161,886	36	-85	-5	-8.77
Superior parietal lobule	RH	3,517	1,891.7	118,001	19	-65	59	-8.88
Precentral gyrus	LH	1,652	1,136.3	85,457	-51	5	45	-6.78
Superior frontal gyrus (supplementary motor area)	LH	1,177	834.39	69,495	-41	-23	82	-6.75
Superior frontal gyrus (supplementary motor area)	RH	761	484.51	94,566	9	28	41	-5.98
Insula	RH	1,422	464.12	48,002	32	24	1	-7.66
Postcentral sulcus	RH	674	400.98	106,638	44	-33	35	-5.9
Superior parietal lobule	LH	469	386.5	113,354	-31	-63	59	-5.44
Anterior insula	LH	928	329.15	8,339	-29	29	-3	-6.3
Superior frontal sulcus	LH	346	318.53	60,381	-28	30	40	-5.48
Superior frontal gyrus	RH	614	274.67	80,972	25	-4	55	-4.78

Significant clusters are given for contrast of gesture and sentence stimuli (PDG, PDS, 3PS) compared to Jabberwocky and Self-adaptor gestures. Regions names are based on the Freesurfer parcellation. In some cases, a name in parentheses is given for reference to previous literature. x, y, z coordinates are the MNI coordinates for the peak node within each cluster.

LH = Left hemisphere; RH = Right hemisphere.

Sentences. Comparison of PDS to 3PS revealed activation in bilateral mid- and anterior-STS (and STG), left posterior STS extending into the middle temporal gyrus and TPJ (inferior parietal and supramarginal gyrus) and left inferior frontal gyrus—a pattern strikingly consistent with Participant-Directed Gestures compared to Self-adaptor Gestures (Fig. 6).

Conjunction. Conjunction analyses of the contrasts PDSvs3PS and PDGvsSG revealed significant overlap within left IFG, left middle and posterior STS, and right middle and anterior STS (Fig. 6, Table III).

Individual Functional Region of interest analyses. As described above, a stronger test for overlapping activation is to use ROIs from the individual level. Of the six ROIs examined, bilateral mid-STS and anterior STS were significantly more active to Participant-Directed Sentences than Third-Person Sentences ($P < 0.05$, corrected). See Fig. 5, Table IV.

Perceived communicative intent in gesture. Conjunction and ROI analyses suggested the bilateral STS and IFG may be modulated by the perceived communicative value of

language stimuli and that these same regions are sensitive to Participant-Directed Gestures. To determine whether the perceived communicative intent within gesture stimuli modulated activation within these same overlapping regions we conducted a parametric analysis in which participant's own ratings of communicative intent within each gesture stimulus were used as a parametric regressor. Combining ratings for Participant-directed and Self-adaptor gestures into one regressor revealed a cluster of activation with left middle and posterior STS [Nodes = 607, Area = 350 mm² Peak = Node 143351, (-57, -37, 11), $t = 6.29$] that showed positive modulation with increasing perception of communicativeness. Analyses within each gesture condition (i.e., PDG and MG) did not reveal significant effects at the whole-brain level when corrected for multiple comparisons. However, ROI analyses using the regions of overlap identified in Figure 5 demonstrated a significant effect of communicative ratings within individuals for the Self-adaptor Gestures ($t_{(14)} = 2.1, P < 0.028$) but not the Participant-Directed Gestures ($t_{(14)} = 0.24, P < 0.4$). This effect was only significant in the left hemisphere regions of overlap.

TABLE III. Contrasts within gesture and language conditions

Contrast region	Hemi	No. nodes	Area (mm ²)	Peak node	x	y	z	Peak t-value
Participant-Directed gestures > Self-adaptor gestures								
Superior temporal sulcus	LH	2,301	1,073.66	142,556	-56	-29	6	7.88
Superior temporal sulcus	RH	1,401	579.84	158,854	56	-22	1	7.23
Inferior parietal/supramarginal gyrus (temporoparietal junction)	LH	481	429.22	145,090	-50	-49	13	7.08
Inferior frontal gyrus/operculum	LH	263	308.31	99,111	-53	27	21	6.79
Anterior superior temporal gyrus	RH	269	266.66	38,177	55	11	-16	6.09
Self-adaptor gestures > Participant-Directed gestures								
Superior parietal gyrus	LH	4,362	2,155.85	113,225	-30	-57	61	-7.521
Superior occipital sulcus	LH	3,394	2,049.12	118,711	-28	-74	30	-10.96
Superior parietal lobule	RH	3,168	1,956.66	115,734	21	-59	68	-8.9
Inferior parietal lobe/supramarginal gyrus	RH	1,811	1,168.02	179,313	59	-17	13	-6.5
Middle occipital gyrus	RH	791	781.14	166,515	43	-78	20	-9.47
Inferior temporal sulcus	RH	757	608.63	152,970	47	-57	-5	-6.45
Postcentral gyrus	LH	1,533	495.29	129,308	-50	-27	24	-7.47
Superior occipital gyrus	RH	508	298.04	119,479	12	110	59	-8.1
Occipital pole	LH	346	279.56	166,271	-3	-102	-4	-7.56
Participant-Directed sentences > Third-person sentences								
Posterior superior temporal gyrus (temporoparietal junction)	LH	1,686	1,783.08	146,396	-60	-48	25	7.28
Superior temporal sulcus	LH	2,576	946.18	22,303	-52	-18	-13	8.75
inferior frontal gyrus/operculum	LH	691	791.96	98,754	-52	19	25	6.8
Anterior superior temporal gyrus	LH	730	702.77	17,841	-51	15	-20	8.7
Superior temporal sulcus	RH	1,518	666.37	158,958	53	-19	-3	6.66
Anterior superior temporal gyrus	RH	417	386.23	38,047	55	12	-17	6.62
Superior frontal gyrus	LH	294	300.51	65,680	-4	47	44	5.34
Third-person sentences > Participant-Directed sentences								
None								
Third-person sentences > Jabberwocky								
Middle temporal gyrus/superior temporal sulcus	RH	7,576	5,332.63	154,715	58	-53	6	9.97
Inferior parietal/angular gyrus	LH	5,242	4,428.75	119,679	-38	-71	35	8.45
Precuneus	LH	5,967	3,920.28	192,967	-12	-42	34	8.49
Anterior inferior temporal sulcus/superior temporal sulcus	LH	1,183	1,397.2	24,368	-49	-9	-23	7.73
Superior frontal gyrus (dorsal medial prefrontal cortex)	LH	1,163	1,248.49	61,282	-9	65	12	6.78
Middle temporal gyrus/superior temporal sulcus	LH	804	981.7	174,037	-67	-38	-1	9.86
Precuneus	RH	1,397	884.57	146,763	14	-54	37	8.64
Cuneus	LH	726	786.83	166,732	-3	-99	3	9.07
Superior frontal gyrus (dorsal medial prefrontal cortex)	RH	701	567.5	53,389	10	59	-6	6.27
Medial occipital-temporal sulcus	LH	1,100	508.23	33,358	-28	-30	-16	8.82
Superior temporal sulcus	RH	538	477	39,318	49	-19	-8	7.29
Occipital pole	RH	446	405.98	177,886	14	-94	10	10.07
Superior frontal sulcus	LH	466	357.86	62,388	-24	37	35	5.73
Middle frontal gyrus	LH	368	290.3	50,090	-29	31	40	5.12
Posterior cingulate	RH	1,017	287.04	129,561	11	-31	42	7.36
Jabberwocky > Third-person sentences								
Middle occipital sulcus	LH	5,060	4,878.49	161,546	-34	-88	8	-10.22
Inferior occipital sulcus	RH	2,686	2,060.64	162,131	35	-84	-4	-9.5
Intraparietal sulcus	RH	2,536	1,327.1	118,645	21	-66	45	-7.47
Precentral gyrus	LH	1,353	931.59	85,263	-51	3	44	-6.91
Superior frontal gyrus (supplementary motor area)	LH	1,224	923.45	69,304	-6	5	70	-8.3

TABLE III. (continued).

Contrast region	Hemi	No. nodes	Area (mm ²)	Peak node	<i>x</i>	<i>y</i>	<i>z</i>	Peak <i>t</i> -value
Inferior frontal gyrus/operculum	LH	1,004	502.15	98,772	-55	9	16	-6.9
Anterior insula	LH	1,247	488.38	10,799	-28	23	0	-8.5
Insula	RH	1,196	358.48	48,002	32	24	1	-7.9
Precentral gyrus	RH	486	299.09	67,142	41	10	23	-5.4
Inferior frontal sulcus	LH	346	282.32	51,747	-42	32	25	-5.92
Postcentral sulcus	RH	410	277.83	107,255	40	-34	42	-6.32
Superior frontal gyrus (supplementary motor area)	RH	394	276.34	94,772	10	29	40	-5.85

Significant clusters are listed for each contrast. *x*, *y*, *z* coordinates are the MNI coordinates for the peak node within each cluster. Regions names are based on the Freesurfer parcellation. In some cases, a name in parentheses is given for reference to previous literature.

LH = left hemisphere; RH = right hemisphere.

DISCUSSION

The current study extended past work demonstrating shared processing of gesture and language by using a combination of surface-based analyses and individually-localized functional region of interest approaches to test for shared neural processing at the group and individual level. Specifically, overlapping regions within bilateral superior temporal sulcus (STS) were significantly more engaged for both gestures and sentences that were rated as more communicative and meaningful. Further, we extended previous work on gesture and language by demonstrating that one common process that underlies this shared brain activation within the STS (mid-STS bilaterally and left posterior STS) is the perception of participant-directed communicative intent, specifically a feeling of being personally addressed in a second-person context.

Language and gesture are inherently social processes that both typically unfold and are learned during the context of social interactions. However, the study of these processes is often divorced from this social-interactive context. Growing work suggests the importance of considering this social-interactive context and taking a “second-person” perspective [Redcay et al., 2010; Schilbach et al., 2013]. The motivation for using participant-directed gesture strings and written sentences in the current design was to elicit a sense of being personally addressed—which was confirmed based on participant’s greater communicative ratings of the Participant-Directed compared to Third-Person Sentences. Thus, comparison of Participant-Directed Sentences to Third-person Sentences allowed for identification of regions sensitive to communicative intent. This comparison revealed overlapping activation with the PDGvsSG contrast within left IFG, left middle and

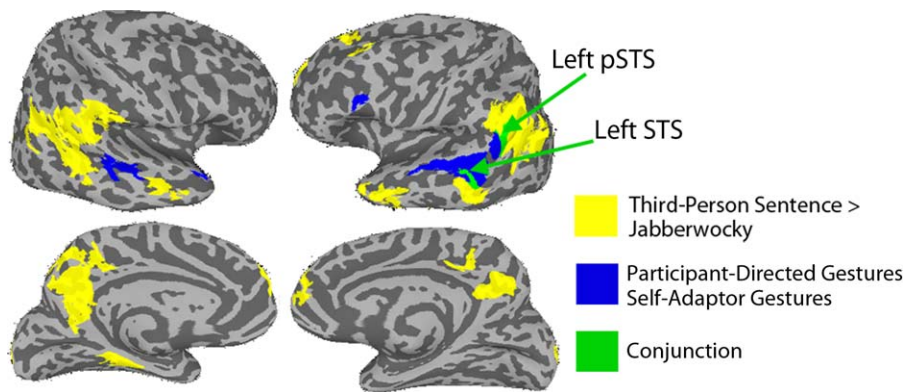


Figure 4.

Shared network for gesture and language. Thresholded ($P < 0.05$, corrected), binary maps are projected on a standardized inflated surface for the gesture contrast (Participant-Directed vs. Self-adaptor Gesture) in blue, language contrast (Third-Person Sentence vs. Jabberwocky) in yellow, and overlap of the two in green. Overlapping activation is seen within mid Superior Temporal Sulcus (STS) and posterior STS/Temporoparietal Junction.

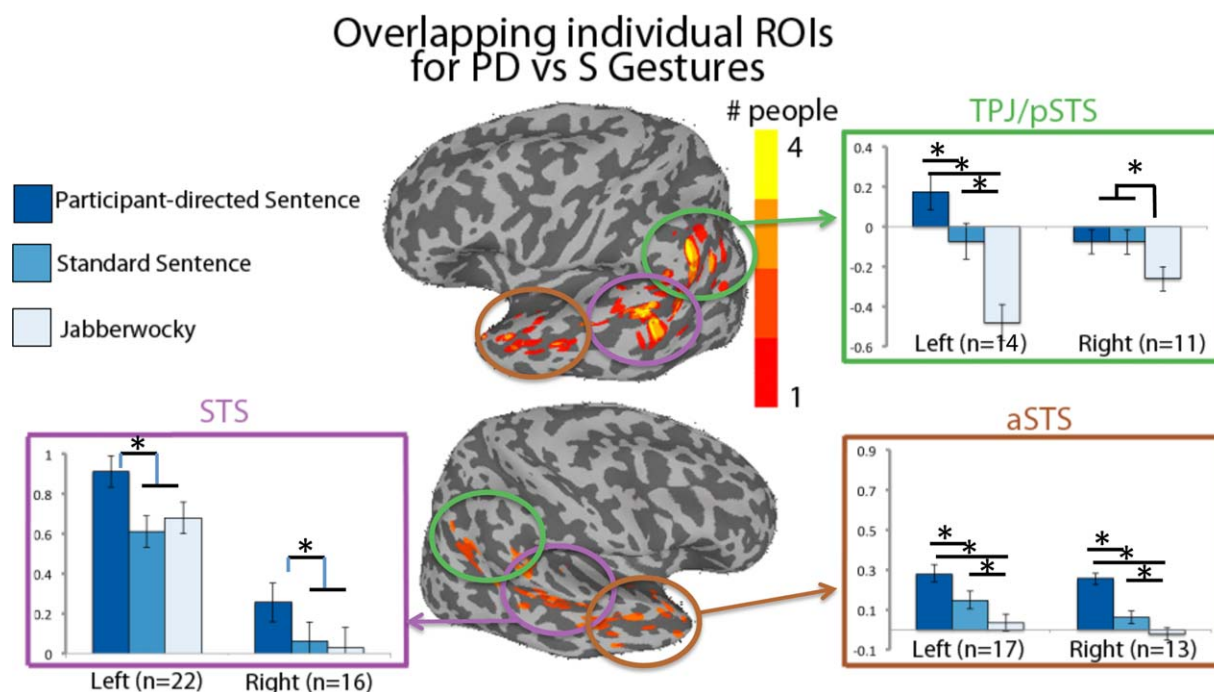


Figure 5.

Overlap on an individual level. Individually-defined regions of interest (ROIs) were created from identifying the peak region of activation for Participant-Directed Gestures vs. Self-adaptor Gestures within left and right anterior, middle, and posterior superior temporal sulcus (STS) from each individual's activation map. These individual ROIs were summed across participants such that the colors on the map represent the number of

participants for whom their individual peak sensitivity to gesture overlapped with other participants' peak region of sensitivity. Contrast values for each of the language conditions are presented for each region in bar graphs. Error bars reflect the standard error of the estimate. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

posterior STS, and right middle and anterior STS/STG at the group level and significant modulation within all gesture regions of STS, except right posterior STS, at the individual fROI level. This left-dominant network is composed of regions typically associated with language-specific processing. For example, Pallier et al. (2011) argued that the left IFG and STS (similar regions to those in the current study) form part of a network supporting processing of syntactic structure. They demonstrated that greater nested linguistic structure (i.e., comparing a sentence that is 12 words long with four sentences that are three words long) modulates left IFG and the full extent of the STS into to the TPJ. While anterior STS and TPJ regions were only sensitive to syntactic structure when words (as opposed to pseudowords) were used, left IFG and posterior STS responded in both cases, suggesting a role of these regions in processing increasing syntactic complexity independent of semantic context. By this metric, however, our Participant-Directed Sentences were actually less syntactically complex (two short sentences) than our Third-person Sentences (one long sentence). Thus, syntactic complexity alone is unlikely to drive the greater response to Participant-Directed Sentences. Similarly, neuroimaging

work reveals a role of the left STS in phonological processing [Emmorey et al., 2011; Hickok and Poeppel, 2007]. However, in the current study, we do not expect differences in phonological processing for these two visually presented sentence conditions. Nonetheless, the current sentences, as designed, were quite different across conditions and so we cannot rule out the existence of some other linguistic difference between conditions that could contribute to the left IFG and left STS activation. Future studies should strive to create sentences tightly matched on linguistic parameters that differ only in their second-person context [e.g., Rice and Redcay, 2016].

Converging evidence for a role of participant-directed, or "second-person," communicative intent (rather than simply differing linguistic demands) in response to our participant-directed stimuli is seen in the parametric analysis on the gesture stimuli. The extent to which participants felt as though the actress was communicating with them related to activation within left middle and posterior STS. Furthermore, this parametric modulation was also significant within the same regions of the communicative intent conjunction analysis in Figure 6 (left IFG, STS, and TPJ) when looking within the Self-adaptor gestures alone.

TABLE IV. Effects of language conditions within gesture ROIs

Participant-Directed gesture > Self-adaptor gesture ROI				
Region	No. participants	PDSvsJW	3PSvsJW	PDSvs3PS
pSTS/TPJ				
LH	14	0.66 [0.47 to 0.85]	0.41 [0.22 to 0.59]	0.25 [0.06 to 0.44]
RH	11	0.19 [0.06 to 0.31]	0.19 [0.06 to 0.31]	0 [-0.13 to 0.13]
STS				
LH	22	0.23 [0.06 to 0.41]	0.07 [-0.11 to 0.24]	0.30 [0.12 to 0.48]
RH	16	0.48 [0.28 to 0.67]	0.16 [-0.36 to 0.03]	0.31 [0.11 to 0.51]
aSTS				
LH	17	0.25 [0.16 to 0.34]	0.11 [0.03 to 0.20]	0.13 [0.04 to 0.22]
RH	13	0.28 [0.21 to 0.34]	0.08 [0.02 to 0.15]	0.19 [0.13 to 0.26]

Mean difference and 95% confidence intervals for contrast values for each language contrast are given. Number of participants indicates the number of participants with a significant cluster for the PDGvsSG contrast within the region. Bold indicates a significant effect at $P < 0.05$, corrected.

That is, the more communicative a gesture felt, the greater the modulation of regions identified as sensitive to communicative intent. These two lines of converging evidence suggest that these differences in neural activity are not due to differences in linguistic demands but rather perception of communicative intent from a second-person perspective. This finding is consistent with work demonstrating that pSTS and TPJ are modulated when hearing speech or viewing actions that are believed to be directed at the participant in the context of a social interaction [Redcay et al., 2010; Rice and Redcay, 2016] or when engaging in joint attention with a social partner [Caruana et al., 2015; Redcay et al., 2012]. This converging evidence across multiple different paradigms provides a strong case for the important role of the left superior temporal sulcus and temporo-parietal junction in processing self-relevant communicative intent directed at the participant [cf. Noordzij et al., 2009].

Whether the STS is involved in the semantic processing of gestures is debated [e.g., Andric and Small, 2012]. We found evidence for bilateral STS activation in both language contrasts as well as in our participant-directed (vs. self-adaptor) gestures. Two previous studies similarly identified a role of the STS in semantic processing when iconic gestures are combined with speech to disambiguate meaning [Holle et al., 2008] or when pantomimes, which convey meaning independent of speech context, were presented with speech that was conceptually matched with the gesture [Willems et al., 2009]. Willems et al. [2009] suggest the role of the STS in gesture processing is in matching two input streams with a common object representation. Other studies, however, have not identified STS for meaningful compared to meaningless gestures. These include studies examining iconic, pantomime, and emblem gestures presented independent of speech [Andric et al., 2013; Straube et al., 2012; Xu et al., 2009] and

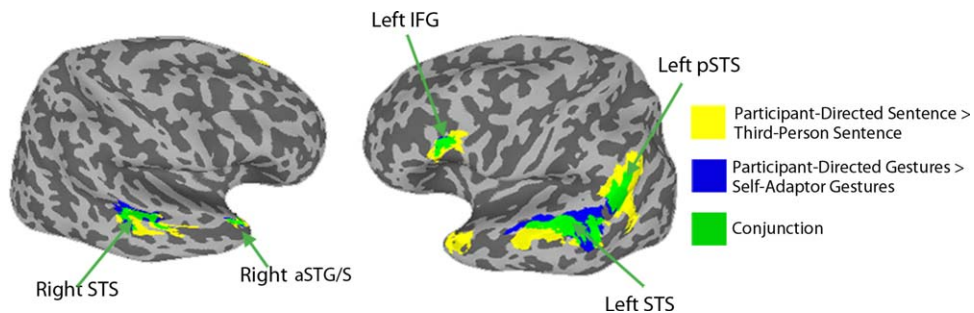


Figure 6.

Shared network for communicative intent. Thresholded ($P < 0.05$, corrected), binary maps are projected on a standardized inflated surface. Regions showing a significantly greater response to Participant-Directed Sentences compared to Third-Person Sentences are displayed in yellow. Those showing a greater

response to Participant-Directed Gestures compared to Self-adaptor Gestures are shown in blue and their conjunction is shown in green. STS = superior temporal sulcus. STG = superior temporal gyrus. IFG = inferior frontal gyrus.

metaphoric or iconic gestures presented within speech context [Dick et al., 2009]. Because the participant-directed vs. self-adaptor gesture contrast cannot independently isolate semantic versus communicative processing, recruitment of the STS in the current study may indeed be due to the STS' role in processing participant-directed communicative intent, rather than semantics per se.

The study of signed languages used by deaf and hearing populations provides an opportunity to dissociate neural activation related to communicative, semantic, and linguistic processing conveyed by gesture. To deaf signers, signed languages are perceived as communicative, meaningful, and linguistic. However, to nonsigners signed languages convey no linguistic information though can be perceived as communicative. Previous work exploiting this distinction has demonstrated engagement of left lateral temporal regions, including STS, when deaf signers who are fluent in American Sign Language (ASL) view ASL signs compared to self-adaptor gestures [Corina et al., 2007]. However, hearing non-signers (like the participants in our study) showed greater activation for ASL than self-adaptor gestures within left middle temporal gyrus, a cluster which did not appear to extend to the STS region [Corina et al., 2007]. One explanation for lack of differential STS response to “communicative” ASL signs in hearing but not deaf participants is that semantic content is necessary to drive STS activation. However, studies described in the above paragraph report inconsistent engagement of the STS for semantic processing in gestures presented without speech. An alternative is that individuals who are not ASL signers may not perceive ASL signs to be as communicative as the communicative, participant-directed gestures used in the current study. In fact pilot work from our lab using the same communicative rating scale described in the current paper found that ASL signs presented to non-signers were only rated slightly more communicative than self-adaptor gestures (mean communicative rating: 4.53 ± 0.42 and 4.22 ± 0.94 , respectively, whereas participant-directed gestures were rated as 6.2 ± 0.54). Taken together, these data highlight a role for the STS in processing communicative intent through gesture, though the extent to which communicative and semantic information in gestures are necessary to engage the STS requires further testing.

While we focused on participant-directed communicative intent, the STS may represent amodal communicative intent for both participant-directed and third-person-directed intent. In a seminal study, Enrici et al. [2011] found an effect of third-person communicative intent across both gesture and language modalities within pSTS/TPJ, as well as anterior and posterior midline regions and IFG. This was an important first contribution to our understanding of amodal processing of communicative intent. The current study extends these findings in two ways. First, while Enrici et al. relied on a main effect to demonstrate overlap the present study used conjunction analyses [Nichols et al.,

2005] and individual region of interest approach, thus providing stronger evidence for shared neural activation. Second, the current study employed comparison conditions that controlled for the presence of people, which allows for stronger claims that these regions support communicative intent, rather than person perception more broadly. Taken together, these studies suggest that the STS and TPJ support amodal processing of communicative intent—both participant-directed and third-person directed.

The current study revealed two clusters within the STS, one within mid-STS and one within the posterior STS extending into the TPJ. This latter region demonstrated deactivation during the self-adaptor gestures suggesting that this region is part of the “default mode” network. This default mode network is associated with diverse functions including semantic processing [Binder and Desai, 2011] and social cognition [Gusnard and Raichle, 2001; Spreng and Mar, 2012], which are consistent with the current study demonstrating a role of this region in both processes [cf. Mar, 2011]. Future studies should aim to dissociate communicative from semantic processing in gesture and language processing to test for functional specialization along the anterior to posterior extent of the STS/TPJ.

The finding that overlapping regions within the STS were engaged for participant-directed gesture and language processing (both participant-directed and third-person language) is consistent with previous hypotheses suggesting the STS may perform a common function in language and social processing [Redcay, 2008]. Additional studies have demonstrated through meta-analyses, functional connectivity analyses, or single subject overlap studies that the STS is involved in diverse tasks, including theory of mind, action perception, biological motion processing, voice processing, and narrative processing [Deen et al., 2015; Hein and Knight, 2008; Lee and McCarthy, 2014; Yang et al., 2015]. Only one previous study, however, has used a single subject ROI approach similar to the one used here to address this question [Deen et al., 2015]. This study examined overlapping activation along the STS for theory of mind, narrative, vocal sounds, music, faces, and biological motion (but not gestures). Regions along the STS that were maximally sensitive to narrative (stories vs. jabberwocky) were also sensitive to theory of mind (stories with false beliefs vs. stories with false representations of reality) but not face, voices, music, or biological motion. While these data are suggestive, they do not address a common process underlying patterns of neural overlap. In the current study, we find shared neural regions that are modulated by communicative intent in both language and gesture stimuli. One explanation for this increased STS activation is that communicative intent may simply amplify these language- or gesture-relevant processes, independently, as the sentences or gestures contain greater self-relevance. While this is possible, this explanation does not account for why communicative

intent modulates overlapping regions for both gesture and language. In the current study, we provide evidence that a common shared process across gesture and language is detection of participant-directed (“second-person”) communicative intent. The STS is a prime location for this process given that communicative intent is at the intersection of STS-associated processes (e.g., human action processing, language, and theory of mind).

To ensure attention to all stimuli, we chose an active task in which participants had to match a frame or word from the preceding trial in order to ensure participants maintained attention on each trial. However, use of an active compared to passive task has the potential to alter BOLD response to gesture and language stimuli [see Caplan and Gow, 2012 for careful discussion]. This inclusion of an active task is problematic when the task induces differential processing between stimulus types, which could account for activation differences due to the ancillary task rather than the construct of interest. In the current study, short-term memory for simple sentences may be easier than that for nonsense words (jabberwocky). This difference in level of difficulty could result in greater activation of the default mode network for sentence processing [Mckiernan et al., 2003; but see Seghier and Price, 2012]. However, we think this explanation is unlikely given meta-analyses on semantic processing across a wide array of task conditions highlight a role of the default mode network in semantic processing [Binder et al., 2009]. A second limitation of including a memory judgment task is that it may make participants more likely to engage in subvocalization during the gestures to enhance memory performance. Future studies could compare whether a passive viewing version of this task results in similar patterns of activation to gesture and text.

CONCLUSION

The current study extends previous work demonstrating a common neural network for semantic processing in gestures and language [Andric et al., 2013; Straube et al., 2012; Xu et al., 2009] by demonstrating overlap to gesture strings and written text in anterior and posterior temporal lobe regions using surface-based and individual functional ROI approaches. Further, by demonstrating neural overlap within the STS for participant-directed human actions and written sentences, the current study extends previous work by suggesting one candidate unifying process underlying this multifaceted STS region—that is, detecting communicative intent from human actions (i.e., gestures and language). This fine-grained overlap at the neural level suggests a common mechanism underlying the extraction of communicative relevance across diverse communicative modalities. Further studies examining overlap within regions and networks will be critical to uncover the core processes that subservise the intertwined domains of lan-

guage, gesture, and social cognition in the service of communication.

ACKNOWLEDGMENTS

The authors thank Briana Viscomi for assistance with stimulus creation, stimulus presentation, and data collection and Dustin Moraczewski for assistance with data analyses. The authors also thank the Institute for Disabilities Research and Training and Dana Breakstone for assistance with stimulus development and creation.

REFERENCES

- Andric M, Small SL (2012): Gesture’s Neural Language. *Front Psychol* 3:99.
- Andric M, Solodkin A, Buccino G, Goldin-Meadow S, Rizzolatti G, Small SL (2013): Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia* 51: 1619–1629.
- Baldwin DA (1991): Infants’ contribution to the achievement of joint reference. *Child Dev* 62(5), 875–890.
- Bates E, Dick F (2000): Beyond phrenology: brain and language in the next millennium. *Brain Lang* 71:18–21.
- Bates E, Dick F (2002): Language, gesture, and the developing brain. *Dev Psychobiol* 40:293–310.
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex* (New York, N.Y.: 1991) 9(12), 2767–2796.
- Binder JR, Desai RH (2011): The neurobiology of semantic memory. *Trends Cognit Sci* 15:527–536.
- Brainard DH (1997): The psychophysics toolbox. *Spat Vis* 10:433–436.
- Brooks R, Meltzoff AN (2008): Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: a longitudinal, growth curve modeling study. *Journal of Child Language* 35:207–220.
- Brysbaert M, New B (2009): Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behav Res Methods* 41:977–990.
- Caplan, D., & Gow, D. (2012). Brain & Language Effects of tasks on BOLD signal responses to sentence contrasts: Review and commentary. *Brain and Language* 120(2), 174–186.
- Caruana N, Brock J, Woolgar A (2015): A frontotemporal network common to initiating and responding to joint attention bids. *Neuroimage* 108:34–46.
- Chapman L, Chapman J (1987): The measurement of handedness. *Brain Cognit* 6:175–183.
- Chen G, Saad ZS, Britton JC, Pine DS, Cox RW (2013): Linear mixed-effects modeling approach to fMRI group analysis. *Neuroimage* 73:176–190.
- Ciaramidaro A, Becchio C, Colle L, Bara B, Walter H (2014): Do you mean me? Communicative intentions recruit the mirror and the mentalizing system. *Soc Cognit Affect Neurosci* 9:909–916.
- Clark, H. (1996): *Using Language*. Cambridge: Cambridge University Press.
- Corina, D., Chiu, Y., Knapp, H., Greenwald, R., Jose-robertson, L. S., & Braun, A. (2007). Neural correlates of human action observation in hearing and deaf subjects, 1152, 111–129.

- Cox RW (1996): AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res* 29:162–173.
- Dale AM, Fischl B, Sereno MI (1999): Cortical surface-based analysis I: Segmentation and surface reconstruction. *Neuroimage* 194, 179–194.
- Deen B, Koldewyn K, Kanwisher N, Saxe R (2015): Functional organization of social perception and cognition in the superior temporal sulcus. *Cereb Cortex* 25(11):4596–4609.
- Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Killiany RJ (2006): An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31:968–980.
- Dick AS, Broce I (2015): The neurobiology of gesture and its development. In: Hickock G, Small SL, editors. *Neurobiology of language*. San Diego, CA: Elsevier. pp 389–398.
- Dick AS, Goldin-Meadow S, Hasson U, Skipper JL, Small SL (2009): Co-speech gestures influence neural activity in brain regions associated with processing semantic information. *Hum Brain Mapp* 30:3509–3526.
- Dick AS, Goldin-Meadow S, Solodkin A, Small SL (2012): Gesture in the developing brain. *Dev Sci* 15:165–180.
- Dick AS, Mok EH, Beharelle AR, Goldin-Meadow S, Small SL (2014): Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. *Hum Brain Mapp* 35:900–917.
- US Department of Education (2011): *Writing Framework for the 2011 National Assessment of Educational Progress*.
- Emmorey K, Xu J, Braun A (2011): Neural responses to meaningless pseudosigns: evidence for sign-based phonetic processing in superior temporal cortex. *Brain Lang* 117:34–38.
- Enrici I, Adenzato M, Cappa S, Bara BG, Tettamanti M (2011): Intention processing in communication: a common brain network for language and gestures. *J Cognitive Neuroscience* 23(9):1–17.
- Fedorenko E, Behr MK, Kanwisher N (2011): Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences* 108(39):16428–16433.
- Fedorenko E, Nieto-Castañón A, Kanwisher N (2012): Syntactic processing in the human brain: What we know, what we don't know, and a suggestion for how to proceed. *Brain Lang* 120: 187–207.
- Ferri F, Busiello M, Campione GC, De Stefani E, Innocenti A, Romani GL, Gentilucci M (2014): The eye contact effect in request and emblematic hand gestures. *Eur J Neurosci* 39:841–851.
- Fischl B, Dale AM (2000): Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences* 97:11050–11055.
- Fischl B, Sereno MI, Dale AM (1999a): Cortical surface-based analysis II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* 207, 195–207.
- Fischl B, Sereno MI, Tootell RB, Dale AM (1999b): High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp* 8:272–284.
- Fischl B, Salat DH, Busa E, Albert M, Dieterich M, Haselgrove C, Dale AM (2002): Whole brain segmentation: Neurotechnique automated labeling of neuroanatomical structures in the human brain. *Neurotechnique* 33, 341–355.
- Glezer LS, Riesenhuber M (2013): Individual variability in location impacts orthographic selectivity in the “visual word form area.” *Journal of Neuroscience* 33(27):11221–11226.
- Goldin-Meadow S. (1998): The development of gesture and speech as an integrated system. In: *The Nature and Function in Children's Communication*. San Francisco: Jossey-Bass. pp 29–43
- Gusnard DA, Raichle ME (2001): Searching for a baseline: Functional imaging and the resting human brain. *Nat Rev Neurosci* 2:685–694.
- Hein G, Knight RT (2008): Superior temporal sulcus—It's my area: or is it? *J Cognit Neurosci* 20:2125–2136.
- Hickok G, Poeppel D (2007): The cortical organization of speech processing. *Nat Rev Neurosci* 8:393–402.
- Holle H, Gunter TC, Rüschemeyer SA, Hennenlotter A, Iacoboni M (2008): Neural correlates of the processing of co-speech gestures. *Neuroimage* 39:2010–2024.
- Holle H, Obleser J, Rueschemeyer SA, Gunter TC (2010): Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage* 49:875–884.
- Holler J, Kokal I, Toni I, Hagoort P, Kelly SD, Ozyürek A (2014): Eye'm talking to you: Speakers' gaze direction modulates co-speech gesture processing in the right MTG. *Soc Cognit Affect Neurosci* 10(2):255–61.
- Hubbard AL, McNealy K, Scott-Van Zeeland A. a, Callan DE, Bookheimer SY, Dapretto M (2012): Altered integration of speech and gesture in children with autism spectrum disorders. *Brain Behav* 2:606–619.
- Hubbard AL, Wilson SM, Callan DE, Dapretto M (2009): Giving speech a hand: Gestures modulates activity in auditory cortex during speech perception. *Hum Brain Mapp* 30:1028–1037.
- Jo HJ, Lee JM, Kim JH, Shin YW, Kim IY, Kwon JS, Kim SI (2007): Spatial accuracy of fMRI activation influenced by volume- and surface-based spatial smoothing techniques. *Neuroimage* 34: 550–564.
- Kircher T, Straube B, Leube DT, Weis S, Sachs O, Willmes K, Green A (2009): Neural interaction of speech and gesture: differential activations of metaphoric co-verbal gestures. *Neuropsychologia* 47:169–179.
- Lee SM, McCarthy G (2014): Functional heterogeneity and convergence in the right temporoparietal junction. *Cereb Cortex* 26(3):1108–1116.
- Mar RA (2011): The neural bases of social cognition and story comprehension. *Annu Rev Psychol* 62:103–134.
- Mckiernan KA, Kaufman JN, Kucera-Thompson J, Binder JR (2003): A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *Neuroimage* 15(3):394–408.
- McNeill D (1992): *Hand and Mind. What Gestures Reveal About Thought*. Chicago: Chicago University Press.
- Nagels A, Kircher T, Steines M, Straube B (2015): Feeling addressed! The role of body orientation and co-speech gesture in social communication. *Hum Brain Mapp* 36:1925–1936.
- Nichols TE, Brett M, Andersson J, Poline JB, Wager T (2005): Valid conjunction inference with the minimum statistic. *Neuroimage* 25:653–660.
- Noordzij M, Newman-Norlund S, DeRuiter J, Hagoort P, Levinson S, Toni I (2009): Brain mechanisms underlying human communication. *Front Hum Neurosci* 3:14.
- Oosterhof NN, Wiestler T, Downing PE, Diedrichsen J (2011): A comparison of volume-based and surface-based multi-voxel pattern analysis. *Neuroimage* 56:593–600.
- Pallier, C., Devauchelle, A., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences, *108(6):2522–2527*.

- Price CJ (2010): The anatomy of language: a review of 100 fMRI studies published in 2009. *Ann NY Acad Sci* 1191:62–88.
- Redcay E (2008): The superior temporal sulcus performs a common function for social and speech perception: implications for the emergence of autism. *Neurosci Biobehav Rev* 32:123–142.
- Redcay E, Dodell-Feder D, Pearrow MJ, Mavros PL, Kleiner M, Gabrieli JDE, Saxe R (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. *Front Hum Neurosci* 6:1–14.
- Redcay E, Ludlum RS, Velnoskey K, Kanwal S (2016): Communicative signals promote object recognition memory and modulate the posterior superior temporal sulcus. *J Cognit Neurosci* 28:8–19.
- Rice K, Redcay E (2016): Interaction matters: A perceived social partner alters the neural response to human speech. *Neuroimage* 129:480–488.
- Rowe ML, Goldin-Meadow S (2009a): Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science* (New York, NY) 323:951–953.
- Rowe ML, Goldin-Meadow S (2009b): Early gesture selectively predicts later language learning. *Dev Sci* 12:182–187.
- Saad ZS, Reynolds RC (2012): Suma. *Neuroimage* 62:768–773.
- Saad ZS, Reynolds RC, Argall B, Japee S, Cox RW (2004): SUMA: An interface for surface-based intra- and inter-subject analysis with AFNI. Presented at: 2004 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano (IEEE Cat No. 04EX821), July 2, 2015. pp 1510–1513.
- Schilbach L, Timmermans, Reddy V, Costall A, Bente G, Schlicht T, Vogeley K (2013): Toward a second-person neuroscience. *Behav Brain* 36:441–462.
- Schilbach L, Wohlschlaeger AM, Kraemer NC, Newen A, Shah NJ, Fink GR, Vogeley K (2006): Being with virtual others: Neural correlates of social interaction. *Neuropsychologia* 44:718–730.
- Searle JR (1969): *Speech Acts: An Essay in the Philosophy of Language*. London: Cambridge University Press.
- Seghier ML, Price CJ (2012): Functional heterogeneity within the default network during semantic processing and speech production. 3:1–16.
- Skipper JI, Goldin-Meadow S, Nusbaum HC, Small SL (2007): Speech-associated gestures, Broca’s area, and the human mirror system. *Brain Lang* 101:260–277.
- Sperber D, Wilson D (1996): *Relevance: Communication and Cognition*. Relevance: Communication and Cognition. Oxford, UK: Blackwell Publishers.
- Spreng RN, Mar RA (2012): I remember you: A role for memory in social cognition and the functional neuroanatomy of their interaction. *Brain Res* 1428:43–50.
- Straube B, Green A, Bromberger B, Kircher T (2011): The differentiation of iconic and metaphoric gestures: Common and unique integration processes. *Hum Brain Mapp* 32:520–533.
- Straube B, Green A, Jansen A, Chatterjee A, Kircher T (2010): Social cues, mentalizing and the neural processing of speech accompanied by gestures. *Neuropsychologia* 48:382–393.
- Straube B, Green A, Weis S, Kircher T (2012): A supramodal neural network for speech and gesture semantics: An fMRI study. *PLoS One* 7:e51207.
- Straube B, He Y, Steines M, Gebhardt H, Kircher T, Sammer G, Nagels A (2013): Supramodal neural processing of abstract information conveyed by speech and gesture. *Front Behav Neurosci* 7:1–14.
- Tomasello M, Carpenter M, Call J, Behne T, Moll H (2005): Understanding and sharing intentions: the origins of cultural cognition. *Behav Brain Sci* 28:675–691. discussion 691–735.
- Tylén K, Allen M, Hunter BK, Roepstorff A (2012): Interaction vs. observation: Distinctive modes of social cognition in human brain and behavior? A combined fMRI and eye-tracking study. *Front Hum Neurosci* 6:331.
- Willems RM, Ozyürek A, Hagoort P (2009): Differential roles for left inferior frontal and superior temporal cortex in multimodal integration of action and language. *Neuroimage* 47:1992–2004.
- Xu J, Gannon PJ, Emmorey K, Smith JF, Braun AR (2009): Symbolic gestures and spoken language are processed by a common neural system. *Proc Natl Acad Sci USA* 106:20664–20669.
- Yang DYJ, Rosenblau G, Keifer C, Pelphrey KA (2015): An integrative neural model of social perception, action observation, and theory of mind. *Neurosci Biobehav Rev* 51:263–275.
- Yeo BTT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Buckner RL (2011): The organization of the human cerebral cortex estimated by functional connectivity. *J Neurophysiol* 106(3):1125–1165.