

## ORIGINAL ARTICLE

# Functional Organization of Social Perception and Cognition in the Superior Temporal Sulcus

Ben Deen, Kami Koldewyn, Nancy Kanwisher, and Rebecca Saxe

Department of Brain and Cognitive Sciences and McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Address correspondence to Ben Deen, Room 46-4021, 43 Vassar St., Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Email: bdeen@mit.edu

## Abstract

The superior temporal sulcus (STS) is considered a hub for social perception and cognition, including the perception of faces and human motion, as well as understanding others' actions, mental states, and language. However, the functional organization of the STS remains debated: Is this broad region composed of multiple functionally distinct modules, each specialized for a different process, or are STS subregions multifunctional, contributing to multiple processes? Is the STS spatially organized, and if so, what are the dominant features of this organization? We address these questions by measuring STS responses to a range of social and linguistic stimuli in the same set of human participants, using fMRI. We find a number of STS subregions that respond selectively to certain types of social input, organized along a posterior-to-anterior axis. We also identify regions of overlapping response to multiple contrasts, including regions responsive to both language and theory of mind, faces and voices, and faces and biological motion. Thus, the human STS contains both relatively domain-specific areas, and regions that respond to multiple types of social information.

**Key words:** social cognition, social perception, superior temporal sulcus

## Introduction

Humans are profoundly social beings, and accordingly devote considerable cortical territory to social cognition, including lower tier regions specialized for perceiving the shapes of faces and bodies (Kanwisher 2010), and high-level regions specialized for understanding the meaning of sentences (Binder et al. 1997; Fedorenko et al. 2011) and the contents of other people's thoughts (Saxe and Kanwisher 2003; Saxe and Powell 2006). Yet between these 2 extremes lies a rich space of intermediate social processes, including the ability to discern the goal of an action, the significance of a fleeting facial expression, the meaning of a tone of voice, and the nature of the relationships and interactions in a social group. How do we so quickly and effortlessly extract this multifaceted social information from "thin slices" of social stimuli (Ambady and Rosenthal 1992)? Here we investigate the functional organization of our computational machinery for

social cognition by using fMRI to target a brain region that has long been implicated as a nexus of these processes: the superior temporal sulcus (STS).

The STS is one of the longest sulci in the brain, extending from the inferior parietal lobe anteriorly along the full length of the temporal lobe. Subregions of the STS have been implicated in diverse aspects of social perception and cognition, including the perception of faces (Puce et al. 1996; Haxby et al. 2000; Pitcher et al. 2011), voices (Belin et al. 2000), and biological motion (Bonda et al. 1996; Allison et al. 2000; Grossman et al. 2000; Grossman and Blake 2002; Pelphrey, Mitchell, et al. 2003; Pelphrey et al. 2005), and understanding of others' actions (Pelphrey, Singerman et al. 2003; Pelphrey, Morris, et al. 2004; Brass et al. 2007; Vandenberg et al. 2009), and mental states (Fletcher et al. 1995; Gallagher et al. 2000; Saxe and Kanwisher 2003; Saxe and Powell 2006; Ciaramidaro et al. 2007). Regions of the STS have also been

implicated in linguistic processing (Binder et al. 1997; Vigneau et al. 2006; Fedorenko et al. 2012), as well as basic perceptual and attentional functions, such as audiovisual integration (Calvert et al. 2001; Beauchamp et al. 2004; Taylor et al. 2006), and the control of visual attention (Corbetta and Shulman 2002). But because most prior studies have investigated only a small subset of these mental processes, the relationship between the regions involved in each process remains unknown.

One possibility is that the STS is composed of a number of distinct, functionally specialized subregions, each playing a role in one of these domains of processing and not others. This would point to a modular organization, and would further point to separate streams of processing for the domains listed above—processing faces, voices, mental states, etc. Another possibility is that responses to these broad contrasts overlap. Overlap could either point to 1) STS subregions involved in multiple processes, indicating a nonmodular organization; or 2) a response driven by an underlying process shared across multiple tasks, such as integration of information from multiple modalities or domains.

Hein and Knight (2008) performed a meta-analysis assessing locations of peak coordinates of STS responses from biological motion perception, face perception, voice perception, theory of mind (ToM), and audiovisual integration tasks. They found that peak coordinates from different tasks did not fall into discrete spatial clusters, and thus argued 1) that the STS consists of multifunctional cortex, whose functional role at a given moment depends on coactivation patterns with regions outside of the STS; and 2) that there is little spatial organization to the STS response to different tasks.

However, meta-analyses cannot provide strong evidence for overlap between functional regions. Because the anatomical location of each functional region varies across individual subjects, combining data across subjects in a standard stereotactic space can lead to findings of spurious overlap (Brett et al. 2002; Saxe et al. 2006; Fedorenko and Kanwisher 2009). These issues are compounded in meta-analyses, which combine data across studies using different normalization algorithms and stereotactic coordinate systems. Furthermore, to investigate overlap between regions responding to distinct tasks, we would ideally want to study the full spatial extent of these regions, rather than simply peak coordinates.

The present study addresses these limitations by scanning the same set of subjects while they engage in face perception, biological motion perception, mental state understanding (termed ToM), linguistic processing, and voice perception. Within individual subjects, we compare STS responses across different tasks. We show that distinct input domains evoke distinct patterns of activation along the STS, pointing to different processes engaged by each type of input. In particular, we find that a dominant feature of this spatial organization consists of differences in response profile along the anterior-posterior axis of the STS. Investigating focal regions that respond maximally to each contrast within individual subjects, we are able to find strongly selective regions for most processes assessed, including biological motion perception, voice perception, ToM, and language. These selective regions are also characterized by distinct patterns of whole-brain functional connectivity, and similarity in functional connectivity profiles across regions is predictive of similarity in task responses. In addition to these selective regions, we identify a number of regions that respond reliably to multiple contrasts, including language and ToM, faces and voices, and faces and biological motion. Thus, the STS appears to contain both subregions specialized for particular domains of social processing, as well as areas responsive to information from multiple domains, potentially playing integrative roles.

## Materials and Methods

### Participants

Twenty adult subjects (age 19–31 years, 11 females, all right-handed) participated in the study. Participants had no history of neurological or psychiatric impairment, had normal or corrected vision, and were native English speakers. All participants provided written, informed consent.

### Paradigm

Each participant performed 5 tasks over the course of 1–3 scan sessions. These included a ToM task, biological motion perception task, face perception task, voice perception task, and an auditory story task that yielded multiple contrasts of interest. The paradigms were designed such that roughly 5 min of data was collected for each condition within each experiment.

In the ToM task, participants read brief stories describing either false beliefs (ToM condition) or false physical representations (control condition), and then answered true/false questions about these stories. Stories were chosen based on a prior study that identified false belief stories that elicited the largest response in the right temporo-parietal junction (TPJ) (Dodel Feder et al. 2011). Stories were presented for 10 s, followed by a 4-s question phase, and 12-s fixation period, with an additional 12 s fixation at the start of the run. Twenty stories (10 per condition) were presented over 2 runs, each lasting 4:32 min. Conditions were presented in a palindromic order (e.g., 1, 2, 2, 1) counterbalanced across runs and subjects. The stimuli and experimental scripts are available on our laboratory's website (<http://saxelab.mit.edu/superloc.php>, Last accessed 16/05/2015).

In the biological motion task, participants watched brief point-light-display (PLD) animations that either depicted various human movements (walking, jumping, waving, etc.) or rotating rigid 3D objects with point-lights at vertices (Vanrie and Verfaillie 2004). Animations consisted of white dots moving on a black background. Individual animations lasted 2 s, and were presented in blocks of 9, with a 0.25-s gap between animations. Participants performed a one-back task on individual animations by pressing a button for repeated stimuli, which occurred once per block. Additionally, 4 other conditions were included which are not reported here; these included spatially scrambled versions of human and object PLDs, linearly moving dots, and static dots. In each of 6 runs, 2 blocks per condition were presented in palindromic order, with condition order counterbalanced across runs and subjects. Runs consisted of 12 20.25-s blocks as well as 18-s fixation blocks at the start, middle, and end, for a total run time of 4:57 min. Due to timing constraints, one subject did not complete the biological motion task.

In the face perception task, participants passively viewed 3-s movie clips of faces or of moving objects, using stimuli that have been previously described (Pitcher et al. 2011). We chose to use dynamic as opposed to static face stimuli as dynamic stimuli have been shown to yield a substantially stronger response in the face area of posterior STS (pSTS) (Pitcher et al. 2011), facilitating the ability to identify face-responsive regions of the STS in individual subjects. Stimuli were presented in blocks of 6 clips with no interval between clips. Additionally, a third condition presenting movies of bodies was included, but not assessed in this report. In each of 3 runs, 4 blocks per condition were presented in palindromic order, with condition order counterbalanced across runs and subjects. Runs consisted of 12 18-s stimulus blocks as well as 18-s rest blocks at the start, middle, and end, for a total run time of 4:30 min.

Downloaded from https://academic.oup.com/jn/article-abstract/202/4/4590/2307585 by Universitat de València user on 08 March 2019

In the voice perception task, participants passively listened to audio clips consisting either of human vocal sounds (e.g., coughing, laughing, humming, sighing, speech sounds), or nonvocal environmental sounds (e.g., sirens, doorbells, ocean sounds, instrumental music). Stimuli were taken from a previous experiment that identified a voice-responsive region of STS (Belin et al. 2000) (<http://vnl.psy.gla.ac.uk/resources.php>, Last accessed 16/05/2015). Clips were presented in 16-s blocks that alternated between the 2 conditions, with a 12-s fixation period between blocks and 8 s of fixation at the start of the experiment. Condition order was counterbalanced across subjects. A single run was given, with 10 blocks per condition, lasting 9:28. Due to timing constraints, 4 subjects did not receive the voice perception task.

During the auditory story task, participants listened to either stories or music. Four conditions were included: ToM stories, physical stories, jabberwocky, and music. This task provides a language contrast (physical stories vs. jabberwocky), a second ToM contrast (ToM vs. physical stories), and a second voice contrast (jabberwocky vs. music). Two additional conditions, stories depicting physical and biological movements, were included for separate purposes, and are not analyzed in the present report. ToM stories consisted of stories describing the false belief of a human character, with no explicit descriptions of human motion. Physical stories described physical events involving no object motion (e.g., streetlights turning on at night) and no human characters. All stories consisted of 3 sentences, and stories were matched across conditions on number of words, mean syllables per word, Flesch reading ease, number of noun phrases, number of modifiers, number of higher level constituents, number of words before the first verb, number of negations, and mean semantic frequency (log Celex frequency). Jabberwocky stimuli consisted of English sentences with content words replaced by pronounceable nonsense words, and with words temporally reordered. This condition has minimal semantic and syntactic content, but preserves prosody, phonology, and vocal content. Music stimuli consisted of clips from instrumental classical and jazz pieces, with no linguistic content. Each auditory stimulus lasted 9 s. After a 1-s delay, participants performed a delayed-match-to-sample task, judging whether a word (or music clip) came from the prior stimulus. Each run consisted of 4 trials per condition, for a total of 24 trials. Four runs of 8:08 min were given. Stimuli were presented in a jittered, slow event-related design, with stimulus timing determined using FreeSurfer's optseq2 to optimize power in comparing conditions.

Additionally, resting-state data were acquired to investigate functional connectivity of STS subregions. For these scans, participants were asked to keep their eyes open, avoid falling asleep, and stay as still as possible. These scans lasted 10 min.

## Data Acquisition

Data were acquired using a Siemens 3T MAGNETOM Tim Trio scanner (Siemens AG, Healthcare, Erlangen, Germany). High-resolution T<sub>1</sub>-weighted anatomical images were collected using a multi-echo MPRAGE pulse sequence (repetition time [TR] = 2.53 s; echo time [TE] = 1.64, 3.5, 5.36, 7.22 ms, flip angle  $\alpha$  = 7°, field of view [FOV] = 256 mm, matrix = 256 × 256, slice thickness = 1 mm, 176 near-axial slices, acceleration factor = 3, 32 reference lines). Task-based functional data were collected using a T<sub>2</sub><sup>\*</sup>-weighted echo planar imaging (EPI) pulse sequence sensitive to blood oxygen level-dependent (BOLD) contrast (TR = 2 s, TE = 30 ms,  $\alpha$  = 90°, FOV = 192 mm, matrix = 64 × 64, slice thickness = 3 mm, slice gap = 0.6 mm, 32 near-axial slices). Resting-state functional data were also collected using a T<sub>2</sub><sup>\*</sup>-weighted EPI sequence (TR = 6 s,

TE = 30 ms,  $\alpha$  = 90°, FOV = 256 mm, matrix = 128 × 128, slice thickness = 2 mm, 67 near-axial slices). Resting data were acquired at higher resolution (2 mm isotropic) to reduce the relative influence of physiological noise (Triantafyllou et al. 2005, 2006).

## Data Preprocessing and Modeling

Data were processed using the FMRIB Software Library (FSL), version 4.1.8, supplemented by custom MATLAB scripts. Anatomical and functional images were skull-stripped using FSL's brain extraction tool. Functional data were motion corrected using rigid-body transformations to the middle image of each run, corrected for interleaved slice acquisition using sinc interpolation, spatially smoothed using an isotropic Gaussian kernel (5-mm FWHM unless otherwise specified), and high-pass filtered (Gaussian-weighted least-squares fit straight line subtraction with  $\sigma$  = 50 s (Marchini and Ripley 2000)). Functional images were registered to anatomical images using a rigid-body transformation determined by FreeSurfer's bbregister (Greve and Fischl 2009). Anatomical images were in turn normalized to the Montreal Neurological Institute-152 template brain (MNI space), using FMRIB's nonlinear registration tool (FNIRT). Further details on the preprocessing and modeling of resting-state data are provided below (see Resting-state functional connectivity analysis).

Whole-brain general linear model-based analyses were performed for each subject, run, and task. Regressors were defined as boxcar functions convolved with a canonical double-gamma hemodynamic response function. For the ToM task, the story and response periods for each trial were modeled as a single event, lasting 14 s. For the auditory story task, 9-s-long stories were modeled as single events; these did not include the response period, as the response was unrelated to the processes of interest for this task. For the face, biological motion, and voice perception tasks, the regressor for a given condition included each block from that condition. Temporal derivatives of each regressor were included in all models, and all regressors were temporally high-pass filtered. FMRIB's improved linear model was used to correct for residual autocorrelation, to provide valid statistics at the individual-subject level (Woolrich et al. 2001).

Subsequently, data were combined across runs for each subject using second-level fixed-effects analyses, after transforming beta maps to MNI space. For split-half analyses (further described below), data were combined across even and odd runs separately. For the voice localizer, which only had a single run, the data were temporally split into first and second halves, each with 5 blocks per condition, and these were analyzed as if they were separate runs.

The contrasts analyzed were as follows: from the ToM task, false belief versus false physical representation stories (termed ToM 1); from the face perception task, faces versus objects (Faces); from the biological motion task, biological motion versus rigid object motion (Biological Motion); from the voice perception task, vocal versus nonvocal sounds (Voice 1); and from the auditory story task, false belief versus physical stories (ToM 2), physical stories versus jabberwocky (Language), and jabberwocky versus music (Voice 2). Note that the ToM 2 and Language contrasts are nonorthogonal and thus statistically dependent, as are the Language and Voice 2 contrasts. As a result, these pairs of contrasts are biased toward finding nonoverlapping sets of voxels: for instance, voxels with high responses in the physical condition are less likely have significant effects of ToM 2, and more likely to have significant effects of Language. However, in

both of these cases we have contrasts from separate datasets (ToM 1 and Voice 1 contrasts) to validate the results.

Because we were specifically interested in responses within the STS, second-level analyses were restricted to voxels within a bilateral STS mask, defined by drawing STS gray matter on the MNI template brain. Posteriorly, the STS splits into 2 sulci surrounding the angular gyrus. Our mask included both of these sulci as well as gray matter in the angular gyrus, because responses to ToM contrasts have previously been observed on the angular gyrus. Statistical maps were thresholded using a false discovery rate of  $q < 0.01$ , which controls the proportion of positive results that are expected to be false positives, to correct for multiple comparisons; supplementary analyses also used different thresholds to determine the effect on overlap estimates.

### Anterior–Posterior Organization

We first investigated the large-scale spatial organization of STS responses to different contrasts, by assessing how responses to each contrast vary as a function of position along the length of the sulcus. We sought to define a series of regions of interest (ROIs) that carved the STS into slices along its length. Prior studies have analyzed responses in coronal slices of the STS, assessing how responses vary as a function of the y-coordinate in MNI space (Pelphrey, Mitchell, et al. 2003; Pelphrey, Singerman, et al. 2003; Pelphrey, Viola, et al. 2004; Morris et al. 2005). However, the STS has an oblique orientation in the y–z plane of MNI space, and we wished to define ROIs that extended perpendicularly to the local direction of the STS in the y–z plane.

To this end, we used our STS mask to estimate the local orientation of the sulcus at different y-coordinates. Mask coordinates were averaged across the x- and z-dimensions, to obtain a function specifying the mean z-coordinate of the STS for a given y-coordinate. Next, for each y-coordinate, the local slope of the STS was determined by fitting a linear regression to z-coordinates in a 1-cm window along the y-dimension. This slope was used to define “slice” ROIs along the length of the STS, by constructing an anisotropic Gaussian ROI and intersecting this with the STS mask (sample ROIs are shown in Fig. 2). Note that, for the posterior segment of the STS, where it splits into 2 sulci, our approach does not treat these sulci separately, instead computing the local slope of a mask including both sulci and the angular gyrus. For each ROI, hemisphere, subject, and contrast, percent signal change values were extracted, and plotted as a function of y-coordinate.

Additionally, we asked whether these patterns differed across the upper and lower banks of the STS. The upper and lower banks of the STS were drawn on individual subjects' cortical surface representations, and intersected with the slice ROIs defined above. Percent signal change values were extracted from the resulting “upper and lower slice” ROIs. For this analysis, we only considered portions of the STS that were anterior to the point at which the STS splits into 2 sulci posteriorly, and data were only smoothed at 3-mm FWHM to minimize bleeding across upper and lower banks.

These analyses revealed that the positions of regions with the strongest response to each contrast were ordered as follows, from posterior to anterior: ToM, biological motion, faces, voices, language. We next aimed to statistically assess these differences in spatial position. Although some contrasts elicited responses in multiple regions along the STS, we aimed to compare responses specifically within the region of maximal response to each task, and thus assessed active regions in individual subjects within spatially constrained group-level search spaces. Search spaces for each contrast were defined from group-level activation

maps, computed using a mixed-effects analysis (Woolrich et al. 2004), as the set of active voxels within a 15-mm sphere around a peak coordinate (shown in Supplementary Fig. 1). For each spatially adjacent pair of search spaces (e.g., ToM and biological motion, biological motion and faces, etc.), we combined the 2 search spaces, and identified regions of activation in individual subjects to each of the 2 contrasts within this combined search space. We then computed the center of mass of these regions, and compared their y-coordinates across the 2 contrasts using a paired, two-sample, two-tailed t-test.

### Responses in Maximally Sensitive Regions

We next asked whether the STS contains selective regions responding to one contrast and not others. We focused on small regions that were maximally responsive to each contrast in a given subject, to increase the likelihood of finding selective responses, and extracted responses in these regions across all conditions. ROIs were defined using data from odd runs of each task. For the face localizer, which had 3 runs, runs 1 and 3 were used to define ROIs. The group-level search spaces defined above (Anterior–Posterior Organization section) were used to spatially constrain ROI definition. For each contrast, hemisphere, and subject, we identified the coordinate with the global maximum response across a given search space, placed a 5-mm radius sphere around this coordinate, and intersected this sphere with the individual subject's activation map for that contrast.

Responses to each condition in each task were then extracted from these ROIs. For the task used to define the ROI, data were extracted from even runs while, for other tasks, the full dataset was used, such that the extracted responses were always independent from data used to define the ROI. Percent signal change was extracted by averaging beta values across each ROI and dividing by mean BOLD signal in the ROI. t-Tests were used to test for an effect of each of the seven contrasts of interest in each ROI, with a threshold of  $P < 0.01$  (one-tailed). Participants who lacked a certain ROI were not included in the statistical analysis for that ROI.

### Resting-State Functional Connectivity Analysis

We next probed another aspect of spatial organization in the STS: do subregions of the STS have different patterns of functional connectivity, and do these patterns relate to the task response profile of that region? Specifically, we assessed resting-state functional connectivity of functionally defined STS subregions.

For resting-state data, several additional preprocessing steps were performed to diminish the influence of physiological and motion-related noise. Time series of 6 motion parameter estimates, computed during motion correction, were removed from the data via linear regression. Additionally, time series from white matter and cerebrospinal fluid (CSF) were removed using the CompCorr algorithm (Behzadi et al. 2007; Chai et al. 2012). White matter and CSF ROIs were defined using FSL's Automated Segmentation Tool and eroded by one voxel. The mean and first principal components of time series from these masks were computed and removed from the data via linear regression.

We again focused on regions with maximum responses to a given contrast, to isolate spatially distinct subregions of the STS. ROIs were defined in the same way as above (see Responses in maximally sensitive regions): as the set of active voxels within a 5-mm-radius sphere around the peak coordinate of response from a given task and participant. Although these ROIs were defined using the same procedure as described above, they were defined using the full dataset from each task, rather than half of

the data, and thus differed slightly from the ROIs used above. Time series were extracted from each ROI, and correlations were computed with time series from every voxel in a Freesurfer-derived gray matter mask (excluding within-hemisphere STS voxels), to derive a whole-brain functional connectivity map for each region. We then computed correlations between whole-brain functional connectivity maps from different regions, to determine the degree of similarity of functional connectivity maps from different STS subregions (functional connectivity similarity). We also computed task responses of these ROIs across the 12 conditions assessed in this study, and computed correlations between these task response vectors across ROIs, to assess similarity of response profiles (response similarity).

Lastly, we assessed the relationship between functional connectivity similarity and response similarity, after accounting for effects of spatial proximity of ROIs. Within each hemisphere, a linear mixed model was performed with functional connectivity similarity values (Fisher-transformed) across ROI pairs and subjects as the dependent variable. To avoid pairs of ROIs with trivially similar response profiles due to similarity in physical location, we excluded ROIs from the ToM 2 and Voice 2 contrasts, as well as any pair of ROIs whose centers were closer than 1 cm, which prevents any overlap between pairs of ROIs. Response similarity (Fisher-transformed) was used as the explanatory variable of interest. Additionally, physical distance between ROIs, as well as the square and cube roots thereof, were included as nuisance regressors to account for effects of spatial proximity on similarity of functional connectivity maps. These specific nonlinear functions of physical distance were found to accurately model the relation between physical distance and functional connectivity similarity values. A mixed model with random-effect terms for the intercept and the effect of response similarity was used. Parameters were estimated using an approximate maximum likelihood method (Lindstrom and Bates 1990), implemented using MATLAB's nlmefit function. A Wald test was used to assess the relationship between response similarity and functional connectivity similarity in each hemisphere; the use of a normal approximation is justified by the large number of data points in each model ( $N = 291$ , right hemisphere;  $N = 233$ , left hemisphere).

## Overlap Analysis

Having probed the response profile of focal, maximally responsive ROIs for each contrast, we next investigated the full spatial extent of responses to each contrast. Specifically, we assessed the amount of overlap between significantly active STS voxels in each hemisphere across contrasts. To illustrate our method for quantifying overlap, suppose we have 2 regions, called A and B, defined by 2 different contrasts, and let AB denote the region of overlap between A and B. We compute 2 quantities to assess the overlap between A and B: the size of AB divided by the size of A, and the size of AB divided by the size of B. In addition to the amount of overlap, these measures provide some insight into the type of overlap occurring. For instance, if region A encompasses and extends beyond region B, then  $\text{size}(AB)/\text{size}(B) = 1$ , while  $\text{size}(AB)/\text{size}(A) < 1$ . In contrast, if the regions are of equal size, then  $\text{size}(AB)/\text{size}(B) = \text{size}(AB)/\text{size}(A)$ . Furthermore, these quantities have an intuitive and straightforward interpretation, as the proportion of one region (A or B) that overlaps with the other. Overlap values were averaged across subjects. For each pair of contrasts and each hemisphere, subjects who lacked any STS response to one or both contrast were excluded from this average.

fMRI overlap values depend on both the extent of spatial smoothing applied, as well as the statistical threshold used to

define the extent of active regions. For this reason, we additionally computed overlap values at a range of different thresholds ( $q < 0.05$ ,  $q < 0.01$ , and  $q < 0.005$ ) as well as smoothing kernels (5-, 3-, and 0-mm FWHM) to ask whether overlap could be consistently observed across these different parameters.

Lastly, we investigated the response profiles of overlapping regions, focusing on pairs of contrasts for which substantial overlap was observed. Specifically, we focused on regions responsive to language and ToM, faces and voices, and faces and biological motion. We used a split-half analysis approach. Regions were defined in the first half of the dataset as the set of all voxels that responded to 2 given contrasts. Responses were then extracted in the second half of the data for the 2 tasks used to define the region, or the full dataset for other tasks, such that responses were always extracted from data that was independent of those used to define the ROI. Unsmoothed data were used to extract responses, such that overlapping responses could not be introduced by spatial smoothing. t-Tests were used to test for an effect of each of the 7 contrasts of interest in each ROI, with a threshold of  $P < 0.01$  (one-tailed).

## Results

### Individual Subject Activations

Individual subject activations are shown in Figure 1, and mean peak coordinates of response (within search spaces for each contrast) are given in Table 1. For the ToM contrasts, the most commonly observed response bilaterally was in the angular gyrus or one of the 2 branching sulci of the STS, a region previously termed the TPJ (Saxe and Kanwisher 2003). Additionally, responses in middle and anterior STS region were often observed. In some subjects, these responses were relatively focal (e.g., Subject 2 in Fig. 1), while in others this response encompassed a large portion of middle to anterior STS (e.g., Subject 1 in Fig. 1).

Responses to the language contrast were generally stronger in the left than right hemisphere. In the left hemisphere, most subjects had several distinct regions of activation along the STS, with variable positions across subjects, ranging from angular gyrus to middle and anterior STS. In the right hemisphere, the most commonly observed response was a single region of far-anterior STS as seen in both subjects in Figure 1.

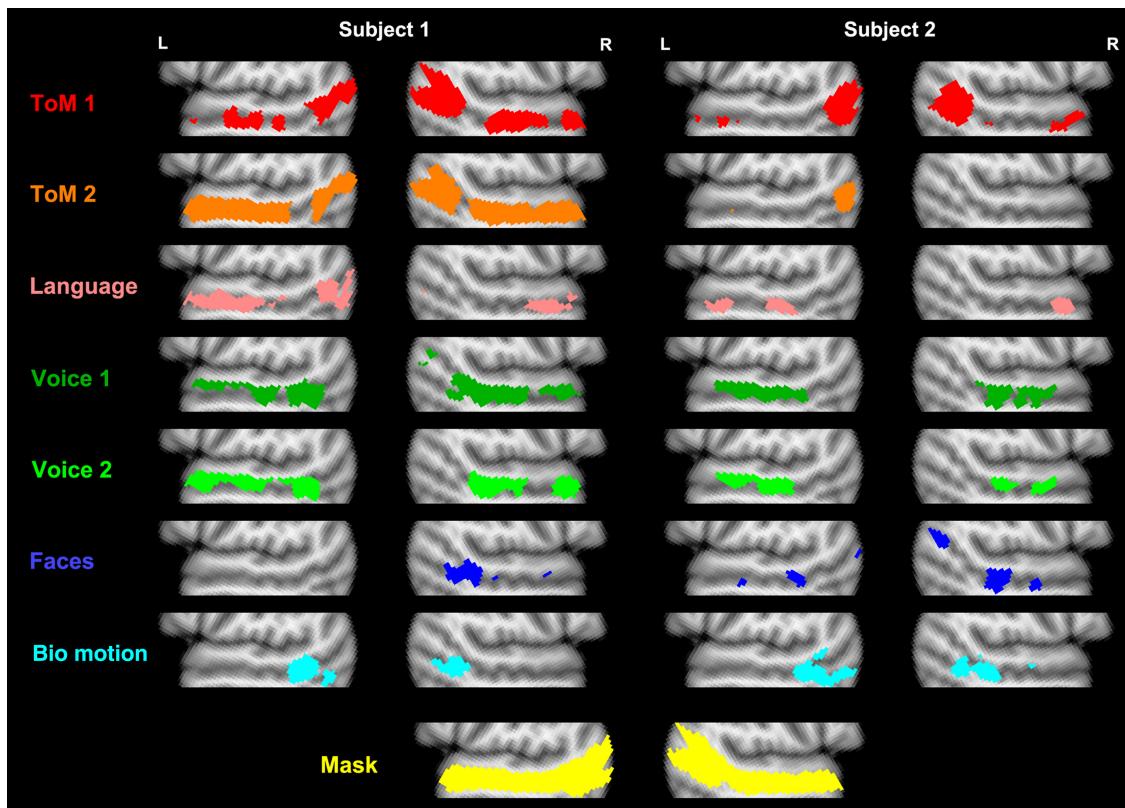
For the voice contrast, activations were generally centered in the middle STS, and were typically stronger in the upper than lower bank of the STS. These responses varied substantially in extent across subjects, with some being relatively focal (e.g., Subject 2 in Fig. 1), and some extending along nearly the full length of the STS anterior to the angular gyrus (e.g., Subject 1 in Fig. 1).

Face responses were most commonly observed in a region anterior and/or just anterior to the point at which the STS breaks into 2 sulci, previously termed the pSTS (Pelphrey, Mitchell, et al. 2003; Shultz et al. 2011). Additionally, many subjects had several other discrete face-sensitive regions both anterior and posterior to this pSTS region, with locations varying across individuals.

Biological motion responses were typically observed in a similar region of pSTS. This region was typically overlapping with the face-responsive pSTS region, but was centered slightly posteriorly in most subjects (e.g., Subject 2 in Fig. 1; also see Anterior–Posterior Organization section).

### Anterior–Posterior Organization

To summarize the large-scale organization of responses to each contrast, we next analyzed the strength of BOLD responses to



**Figure 1.** Individual-subject activations to 7 different contrasts in 2 representative example subjects. Analyses were restricted to the bilateral STS mask shown in yellow at the bottom, and were thresholded at a false discovery rate of  $q < 0.01$ . The slices displayed are at MNI x-coordinate  $\pm 52$ .

**Table 1** Peak coordinates (in MNI space) of response to each task

ROI	x	y	z
<b>LH</b>			
ToM 1	47.4	-58.7	22.9
ToM 2	45.6	-59.6	24.2
Bio motion	49.9	-59.7	11.4
Faces	55.6	-41.1	7
Voice 1	60.6	-17.4	-1.6
Voice 2	61.5	-15.1	-1.5
Language	53	-3.5	-15.1
<b>RH</b>			
ToM 1	-52.5	-52.2	21.1
ToM 2	-53.3	-55.2	23.2
Bio motion	-57.1	-44.1	14.9
Faces	-54.9	-36.2	7.2
Voice 1	-60.8	-16.3	-1.4
Voice 2	-61	-13.7	-2
Language	-52.5	-0.9	-17.3

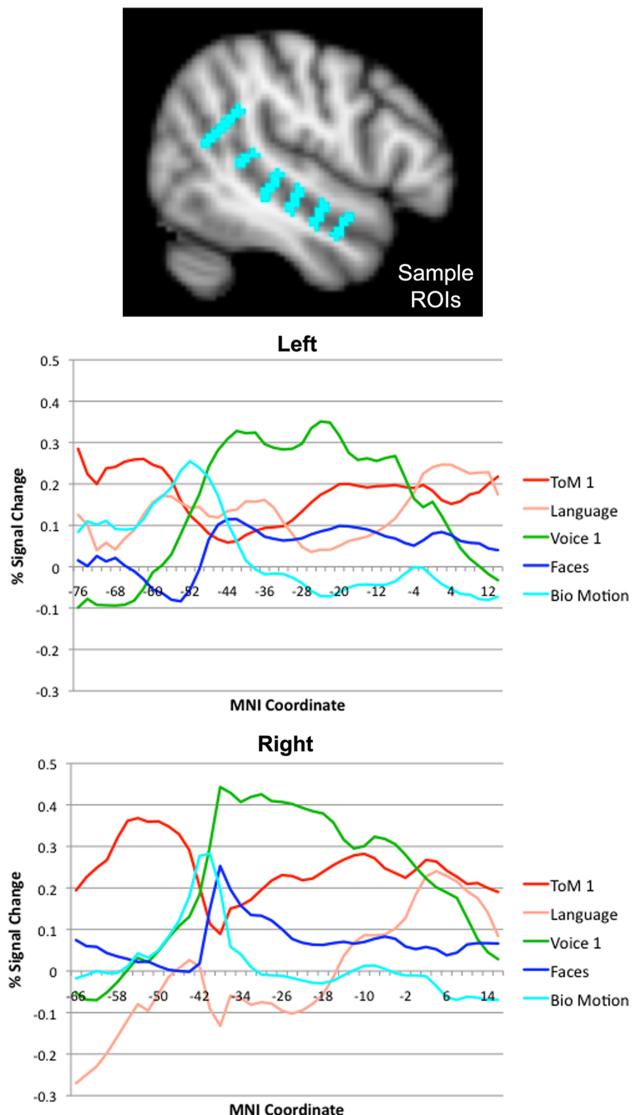
Coordinates were defined in individual participants (within search spaces for each task) and then averaged across participants.

each contrast as a function of position along the length of the STS.

Results from this analysis are shown in Figure 2 (ToM 2 and Voice 2 contrasts are omitted for visualization purposes); results separated across the upper and lower banks are shown in *Supplementary Figure 2*. These results are consistent with the qualitative descriptions of individual-subject activation patterns

described above, and provide a visualization of these effects at the group level. ToM responses are strongest in the posterior-most part of the STS (in the angular gyrus and surrounding sulci), and were also observed in middle-to-anterior STS. Biological motion responses peaked in a pSTS region just anterior to the angular gyrus ToM response. Face responses peaked in a further-anterior pSTS region, with weaker responses also observed in middle-to-anterior STS. The voice response was very broad, encompassing much of the STS, and centered on middle STS. Lastly, language responses in the left hemisphere were observed along the extent of the STS, with peaks in posterior and anterior STS regions. In the right STS, only an anterior region of language activation was observed. Voice responses were substantially stronger in the upper bank than the lower bank of the STS, while responses to ToM, biological motion, faces, and language were largely symmetric across the 2 banks (with a slightly stronger right anterior ToM response in the lower bank).

To determine the reliability of these anterior-posterior spatial relations across individual subjects, we statistically assessed differences in center of mass along the y-axis of spatially adjacent regions. This difference was significant for all pairs of regions, including ToM 1 and biological motion regions (LH:  $t = 2.34$ ,  $P < 0.05$ ; RH:  $t = 5.65$ ,  $P < 10^{-4}$ ), biological motion and face regions (LH:  $t = 7.30$ ,  $P < 10^{-5}$ , RH:  $t = 3.17$ ,  $P < 0.01$ ), face and Voice 1 regions (LH:  $t = 2.54$ ,  $P < 0.05$ , RH:  $t = 6.10$ ,  $P < 10^{-4}$ ), and Voice 1 and language regions (LH:  $t = 7.46$ ,  $P < 10^{-5}$ , RH:  $t = 6.91$ ,  $P < 10^{-4}$ ). This result demonstrates a reliable, bilateral anterior-to-posterior ordering of responses: the TPJ response to ToM, pSTS response to biological motion, pSTS response to faces, middle STS response to voices, and anterior STS response to language.



**Figure 2.** Responses to each task as a function of position along the length of the STS. The upper figure shows the ROIs that were used to extract responses at each position. The lower 2 graphs show left and right STS responses (percent signal change) for each task, as a function of y-coordinate in MNI space.

### Responses in Maximally Sensitive Regions

Do STS subregions of maximal sensitivity to a given contrast also exhibit selectivity—a response to one contrast but not others? We tested this by extracting responses across all conditions from small ROIs surrounding peak coordinates for the response to each contrast (Fig. 3), using data independent of those used to define the ROI. Responses in regions defined by ToM 2 and Voice 2 contrasts are omitted for brevity, as the locations of these regions and their response profiles were highly similar to the ToM 1 and Voice 1 ROIs.

Among regions defined by the ToM (ToM 1) contrast, the left hemisphere region responded strongly to the ToM 1 ( $t_{(18)} = 7.13$ ,  $P < 10^{-6}$ ) and ToM 2 ( $t_{(18)} = 7.45$ ,  $P < 10^{-6}$ ) contrasts, but also responded significantly to the language contrast ( $t_{(18)} = 4.88$ ,  $P < 10^{-4}$ ). The right hemisphere region, in contrast, exhibited a selective response, with significant effects only for the ToM 1 ( $t_{(18)} = 6.84$ ,  $P < 10^{-6}$ ) and ToM 2 ( $t_{(18)} = 4.00$ ,  $P < 10^{-3}$ ) contrasts.

Language regions were defined in a relatively small number of subjects (4 for the right hemisphere and 8 for the left, in split-half data). This likely reflects the fact that this contrast was generally somewhat weaker than the others, as well as more spatially variable, leading to a substantial portion of subjects with no significant response in the anterior STS language search space when only odd runs were analyzed. Nevertheless, significant effects of the language contrast were observed in both the left ( $t_{(7)} = 6.03$ ,  $P < 0.001$ ) and right ( $t_{(3)} = 5.74$ ,  $P < 0.01$ ) hemispheres. An effect of the Voice 2 contrast was observed in the left hemisphere ( $t_{(7)} = 3.65$ ,  $P < 0.01$ ). Note that the jabberwocky condition used to define this contrast involves phonemic and prosodic information; this difference could thus still reflect linguistic processing. However, an effect of the ToM 2 contrast was also observed in the left hemisphere ( $t_{(7)} = 3.73$ ,  $P < 0.01$ ), in addition to a marginal effect of ToM 2 in the right hemisphere ( $t_{(3)} = 2.16$ ,  $P < 0.05$ ) and marginal effects of ToM 1 in the left ( $t_{(7)} = 2.63$ ,  $P < 0.05$ ) and right ( $t_{(3)} = 3.73$ ,  $P < 0.05$ ) hemispheres. These effects reflected moderately sized differences in percent signal change (0.2–0.4%), but were nevertheless statistically marginal due to the small number of subjects with a defined region. Thus, although this region appears largely language selective—with a stronger response to language conditions relative to a range of nonlinguistic visual and auditory conditions—it appears to also be modulated by mental state content.

Voice-sensitive regions in middle STS showed a clearly selective profile of responses bilaterally. These regions responded to the Voice 1 (left:  $t_{(13)} = 6.78$ ,  $P < 10^{-5}$ ; right:  $t_{(14)} = 11.12$ ,  $P < 10^{-5}$ ) and Voice 2 (left:  $t_{(13)} = 7.02$ ,  $P < 10^{-5}$ ; right:  $t_{(14)} = 6.95$ ,  $P < 10^{-5}$ ) contrasts, and did not respond significantly to any other contrast.

Strikingly, the face-sensitive region of posterior STS responded strongly to both face and voice contrasts, bilaterally. The expected response to faces was found in the left ( $t_{(6)} = 6.51$ ,  $P < 0.001$ ) and right ( $t_{(13)} = 7.50$ ,  $P < 10^{-5}$ ) hemispheres. Additionally, there was a bilateral effect of the Voice 1 (left:  $t_{(5)} = 7.42$ ,  $P < 0.001$ ; right:  $t_{(11)} = 5.13$ ,  $P < 10^{-6}$ ) and Voice 2 (left:  $t_{(6)} = 4.13$ ,  $P < 0.01$ ; right:  $t_{(13)} = 3.86$ ,  $P < 0.01$ ) contrasts. Weaker effects of the ToM 2 contrast ( $t_{(14)} = 2.80$ ,  $P < 0.01$ ) and the biological motion contrast ( $t_{(12)} = 2.90$ ,  $P < 0.01$ ) were also observed in the right hemisphere.

Lastly, the posterior STS regions defined by the biological motion contrast responded to biological motion bilaterally (left:  $t_{(16)} = 15.20$ ,  $P < 10^{-10}$ ; right:  $t_{(16)} = 13.76$ ,  $P < 10^{-10}$ ). There was also an effect of the Voice 1 contrast in the right hemisphere ( $t_{(13)} = 2.78$ ,  $P < 0.01$ ) and marginally in the left hemisphere ( $t_{(13)} = 2.43$ ,  $P < 0.02$ ), although neither region showed an effect of the Voice 2 contrast. This indicates the presence of an STS subregion that is quite selective for processing biological motion.

Could differences in effect sizes across regions in this analysis reflect differences in signal quality from different subregions of the STS? To address this, we computed temporal signal-to-noise ratios (tSNR) for the ROIs assessed here (Supplementary Fig. 3). These results indicate that tSNR values are largely similar across ROIs, with a slight (~25%) decrease in values for voice ROIs. This suggests against the possibility that these functional dissociations result from signal quality differences.

### Resting-State Functional Connectivity Analysis

We next asked about another dimension of spatial organization: do functionally and anatomically distinct subregions of the STS have differing patterns of functional connectivity with the rest of the brain? We identified regions of maximum response to each task within individual subjects, and assessed similarity between their functional connectivity maps, as well as their task response profiles.

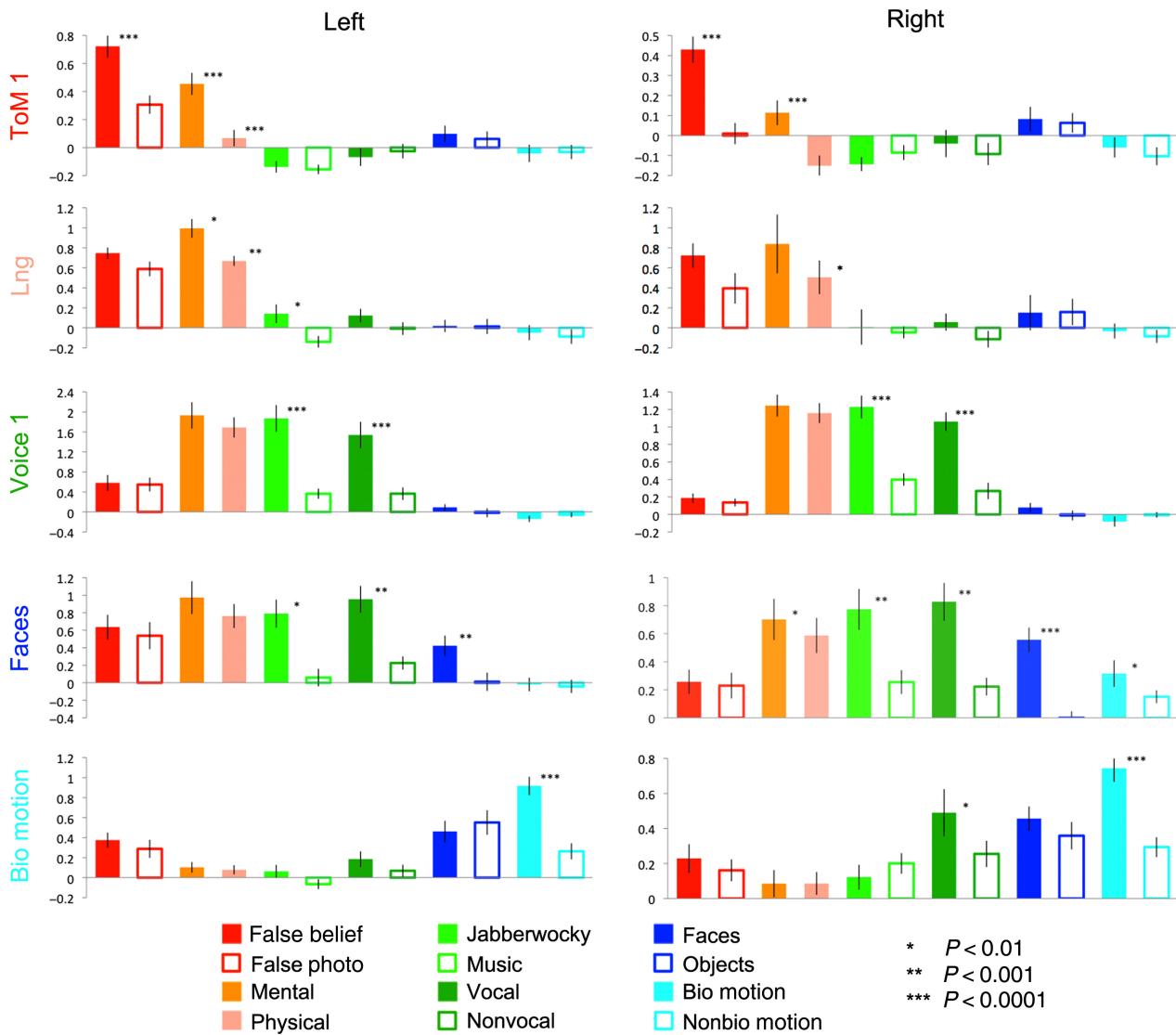


Figure 3. Responses (in percent signal change) of maximally sensitive regions for each contrast, across all conditions. Responses were measured in data independent of those used to define the ROIs.

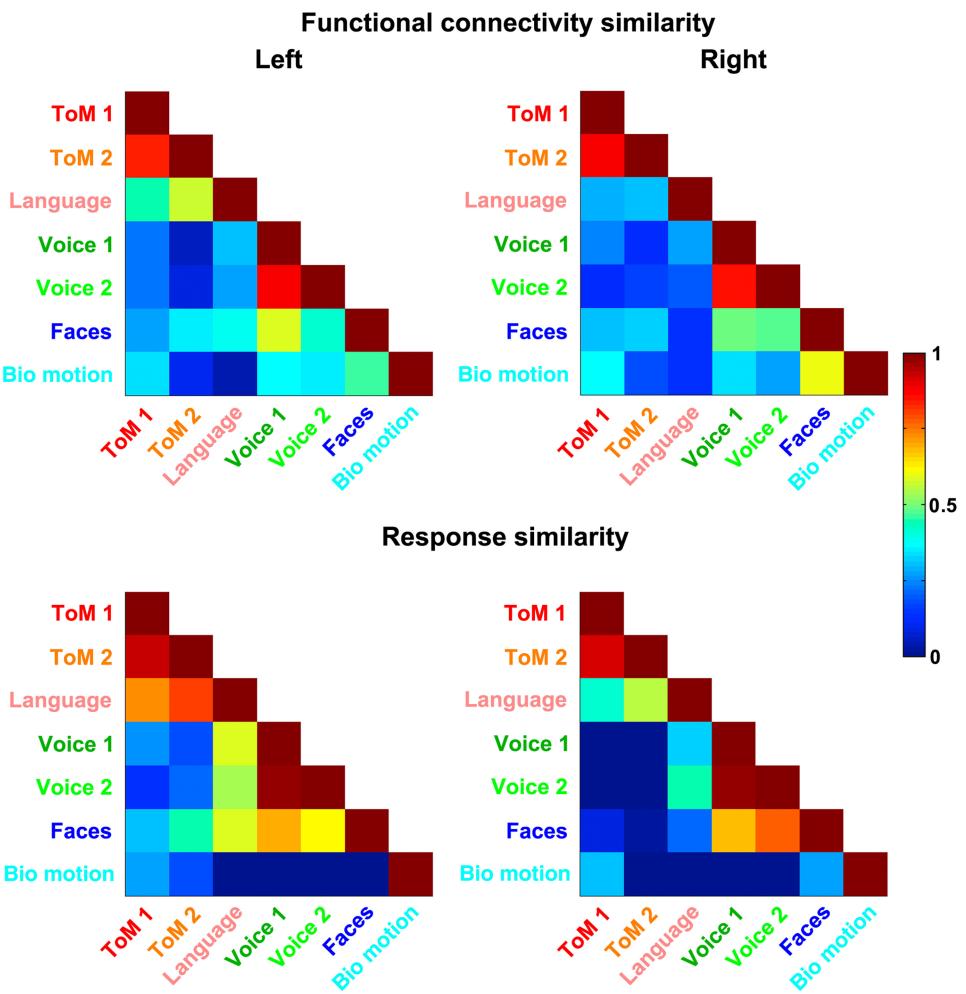
Matrices of functional connectivity and response similarity are shown in Figure 4 (additionally, whole-brain functional connectivity maps for each seed region are shown in Supplementary Fig. 4). Generally, positive correlations were observed between functional connectivity patterns. Excluding correlations between regions defined by similar contrasts (ToM 1/ToM 2, Voice 1/Voice 2), these correlations ranged from 0.05 to 0.58 (LH), and 0.11 to 0.60 (RH). The broad range in these correlation values indicates that some pairs of STS subregions share similar functional connectivity patterns, while others diverge.

Does this variability in functional connectivity similarity relate to variability in the response similarity of pairs of regions? A linear mixed model showed a significant relationship both in the left hemisphere ( $z = 2.52, P < 0.05$ ) and the right hemisphere ( $z = 3.29, P < 0.001$ ), after accounting for effects of spatial proximity. These findings indicate that there are multiple functional connectivity patterns and response profiles associated with STS subregions, and that pairwise similarity along these 2 measures is related: regions with more similar response profiles also have more similar patterns of functional connectivity.

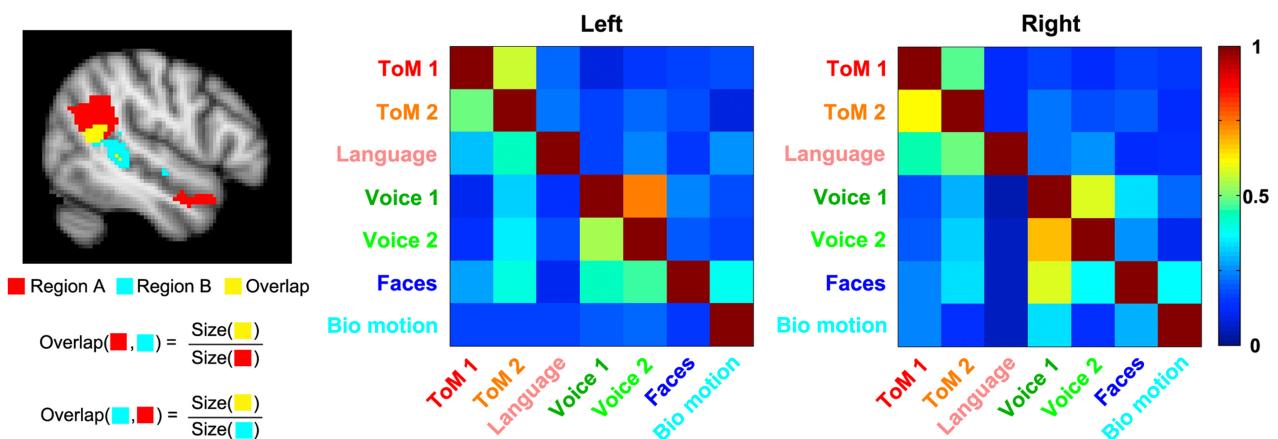
### Overlap Analysis

Having investigated the response profiles of maximally responsive focal regions in individual subjects, we next asked whether the full STS response to each contrast is spatially distinct or overlapping across contrasts. To answer this question, we computed the proportion of the STS activation to one contrast that overlapped with activations to each other contrast.

Results from the overlap analysis are shown in Figure 5. As expected, the strongest overlap values (47–75%) were found for the ToM 1 and ToM 2 contrasts, as well as the Voice 1 and Voice 2 contrasts, intended to elicit activity in similar regions. Overlap values for other pairs of contrasts ranged from 4 to 59%. Particularly strong overlap was observed between face and voice responses (19–59%, mean = 36%), face and biological motion responses (15–39%, mean = 30%), and ToM and language responses (11–50%, mean = 29%). Relatively high values were also observed for overlap between ToM and face responses (16–39%, mean = 24%), and ToM and voice responses (9–35%, mean = 20%). This indicates in addition to focal regions with selective response



**Figure 4.** Matrices of functional connectivity similarity (correlations between whole-brain resting-state functional connectivity maps of seed ROIs defined by each contrast) and response similarity (correlations between vectors of task responses from each seed ROI). ROIs are defined to consist of a focal region of maximal activation to a given contrast.



**Figure 5.** Overlap matrices for regions of activation defined by each task contrast. Each cell in a given overlap matrix is equal to the size of the overlapping region for the tasks on the corresponding row and column, divided by the size of the region of activation for the task on that row, as shown in the graphic on the left-hand side.

profiles, the STS contains parts of cortex that respond significantly to social information from multiple domains.

We next assessed how overlap values vary as a function of amount of smoothing and the statistical threshold used to define

regions. *Supplementary Figures 5* and *6* show overlap matrices at smoothing kernels of 5-, 3-, and 0-mm FWHM, and thresholds of  $q < 0.05$ ,  $q < 0.01$ , and  $q < 0.005$ . As expected, using stricter thresholds and less spatial smoothing lead to numerically smaller

overlap values, with spatial smoothing appearing to have a greater influence in the range of parameters tested. For example, face/voice overlap had a mean value of 42% at  $q < 0.05$  and 5-mm FWHM, and a mean value of 28% at  $q < 0.005$  and no smoothing. Nevertheless, a similar pattern of overlap values across pairs of contrasts was observed across thresholds and smoothing kernels. In particular, relatively strong overlap between responses to language and ToM, faces and voices, and faces and biological motion were consistently observed. In contrast, overlap between responses to ToM and faces, as well as ToM and voices, decreased substantially as less smoothing was used. This may indicate that this overlap was introduced by spatial blurring of distinct regions; alternatively, it is possible that the increased signal-to-noise ratio afforded by smoothing is necessary to detect this overlap.

Lastly, we investigated response profiles of overlapping regions, by defining overlapping ROIs in one half of the dataset, and extracting responses from left-out, unsmoothed data (Fig. 6). We focused on pairs of contrasts for which substantial overlap was consistently observed.

The language and ToM region responded substantially above baseline for all language conditions, with a near-zero response to all other conditions, and was additionally modulated by the presence of mental state content. Significant effects of the language contrast were observed in both hemispheres (RH:  $t_{(10)} = 6.31$ ,  $P < 10^{-4}$ ; LH:  $t_{(11)} = 4.56$ ,  $P < 10^{-3}$ ). Effects of the ToM 2 contrast were also significant bilaterally (RH:  $t_{(10)} = 4.59$ ,  $P < 10^{-3}$ ; LH:  $t_{(11)} = 2.74$ ,  $P < 0.01$ ), while effects of ToM 1 were significant in the right hemisphere ( $t_{(10)} = 3.08$ ,  $P < 0.01$ ) and marginal in the left ( $t_{(11)} = 1.86$ ,  $P < 0.05$ ).

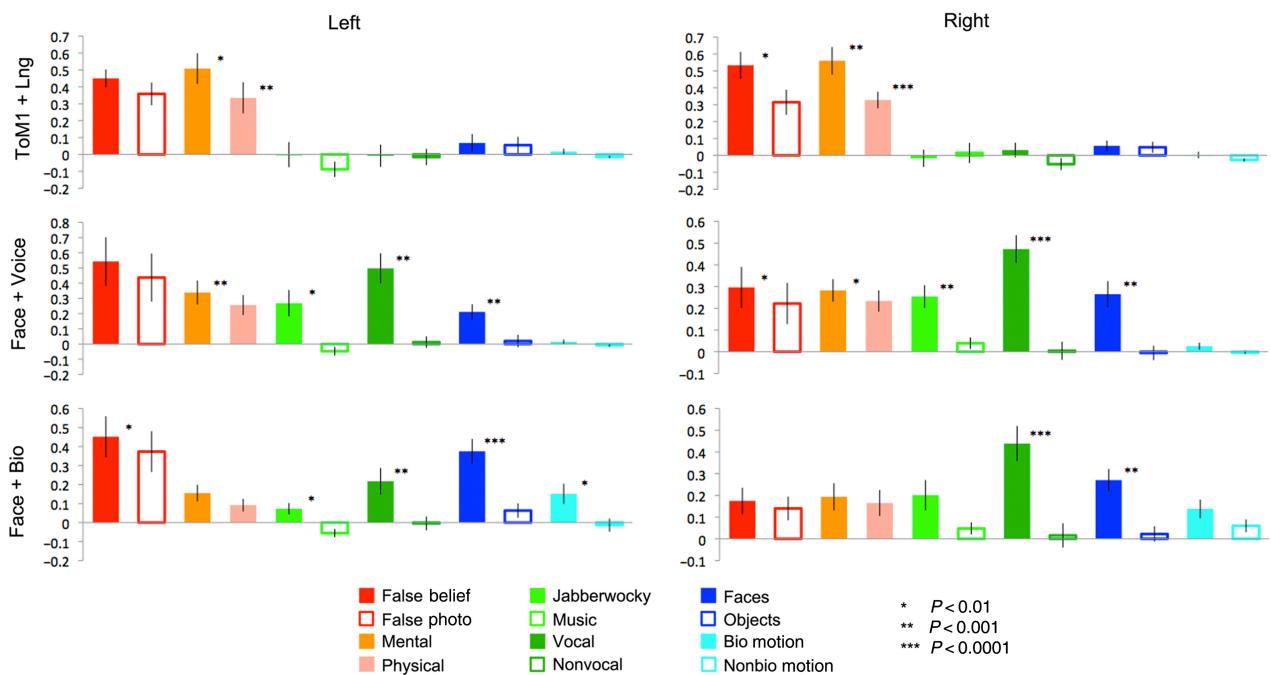
The face and voice region had a roughly similar response profile to the pSTS region defined by a face contrast, with a moderately sized response to both dynamic faces and vocal sounds. The effect of the Voice 1 contrast was significant bilaterally (RH:  $t_{(12)} = 9.00$ ,  $P < 10^{-6}$ ; LH:  $t_{(11)} = 5.38$ ,  $P < 10^{-3}$ ), as was the effect of the Voice 2 contrast (RH:  $t_{(12)} = 4.50$ ,  $P < 10^{-3}$ ; LH:  $t_{(11)} = 2.93$ ,  $P < 0.01$ ). There was also a significant effect of faces over objects bilaterally (RH:

$t_{(12)} = 5.12$ ,  $P < 10^{-3}$ ; LH:  $t_{(11)} = 4.68$ ,  $P < 10^{-3}$ ). This region also had a weak but reliable response to the ToM contrasts: ToM 2 bilaterally (RH:  $t_{(12)} = 3.50$ ,  $P < 0.01$ ; LH:  $t_{(11)} = 4.45$ ,  $P < 10^{-3}$ ) and ToM 1 in the right hemisphere ( $t_{(12)} = 3.75$ ,  $P < 0.01$ ), and marginally in the left hemisphere ( $t_{(11)} = 2.05$ ,  $P < 0.05$ ). This response appeared to be driven by overlap with the mid-STS ToM response.

The face and biological motion region had a moderate response to faces, a relatively weak response to biological motion, and a response to vocal sounds of variable effect size. This region responded significantly to faces over objects bilaterally (RH:  $t_{(15)} = 10.15$ ,  $P < 10^{-7}$ ; LH:  $t_{(14)} = 6.64$ ,  $P < 10^{-5}$ ). There was also a significant effect of biological motion in the left hemisphere ( $t_{(15)} = 3.33$ ,  $P < 0.01$ ), but not in the right. The lack of an effect in the right hemisphere, however, appeared to be driven by a single outlier with a strongly negative effect of biological motion; there was a significant effect after removing this participant ( $t_{(13)} = 3.82$ ,  $P < 0.01$ ). Additionally, there was a significant effect of the Voice 1 contrast bilaterally (RH:  $t_{(13)} = 5.01$ ,  $P < 10^{-3}$ ; LH:  $t_{(12)} = 4.61$ ,  $P < 10^{-3}$ ), and the Voice 2 contrast in the left hemisphere ( $t_{(15)} = 3.15$ ,  $P < 0.01$ ) and marginally in the right hemisphere ( $t_{(14)} = 1.83$ ,  $P < 0.05$ ). Lastly, there was a weak but significant effect of the ToM 1 contrast in the left hemisphere ( $t_{(15)} = 2.84$ ,  $P < 0.01$ ). These results indicate that, while there is a pSTS region responsive to both dynamic faces and biological motion, it only responds weakly to biological motion, with a much stronger response observed in the more selective biological motion area. Furthermore, this region appears to also have a substantial response to vocal sounds, of magnitude similar to or stronger than the response to biological motion.

## Discussion

We investigated STS responses to a number of social cognitive and linguistic contrasts, and found that patterns of response along the length of the STS differed substantially across each contrast. Furthermore, we found largely selective subregions of



**Figure 6.** Responses (in percent signal change) of overlapping regions responsive to multiple contrasts, across all conditions. Responses were measured in data independent of those used to define the ROIs.

the STS for ToM, biological motion, voice perception, and linguistic processing, as well as a region that specifically responds to dynamic faces and voices. Contrary to claims that there is little systematic spatial organization to the STS response to different tasks and inputs (Hein and Knight 2008), these findings argue for a rich spatial structure within the STS. In addition to these selective areas, regions responsive to multiple contrasts were observed, most clearly for responses to language and ToM, faces and voices, and faces and biological motion. These results indicate that the STS contains both domain-specific regions that selectively process a specific type of social information, as well as multifunctional regions involved in processing information from multiple domains.

Our analysis of resting-state functional connectivity of STS subregions supports the argument for systematic spatial organization in the STS. Our results point to distinct patterns of functional connectivity within the STS, suggesting the presence of fine-grained distinctions within the 2–3 patterns observed in prior studies (Power et al. 2011; Shih et al. 2011; Yeo et al. 2011). Furthermore, we show that these connectivity differences are linked to differences in response profiles, consistent with the broad claim that areas of common functional connectivity also share common function (Smith et al. 2009). Contrary to the claim that STS subregions are recruited for different functions based on their spontaneous coactivation with other brain regions (Hein and Knight 2008), these results paint a picture in which STS subregions have stable, distinct response profiles and correspondingly distinct patterns of coactivation with the rest of the brain.

The STS regions found to respond to each contrast in the current study are broadly consistent with regions reported in prior studies. Prior studies using ToM tasks have most consistently reported the posterior-most TPJ region (Fletcher et al. 1995; Gallagher et al. 2000; Saxe and Kanwisher 2003; Saxe and Powell 2006; Ciaramidaro et al. 2007; Saxe et al. 2009; Dodell-Feder et al. 2011; Bruneau et al. 2012; Gweon et al. 2012), but some have also reported middle and anterior STS regions like those observed in the present study (Dodell-Feder et al. 2011; Bruneau et al. 2012; Gweon et al. 2012). At least 2 prior studies have investigated responses to both ToM and biological motion, and both found that the response to biological motion was anterior to the TPJ region elicited by ToM tasks, as observed in the present results (Gobbini et al. 2007; Saxe et al. 2009). Consistent with prior arguments, the right TPJ region observed in the current study was strongly selective for mental state reasoning, among the tasks used here.

Studies of face and biological motion perception have found responses in a similar region of pSTS (Allison et al. 2000; Grossman et al. 2000; Pelphrey, Mitchell, et al. 2003; Pitcher et al. 2011; Engell and McCarthy 2013). In the present study, we observe overlapping responses to faces and biological motion in the pSTS, consistent with a recent study on responses to biological motion and faces in a large set of participants (Engell and McCarthy 2013). However, we find that the pSTS region responding to faces is slightly but reliably anterior to the region responding to biological motion, and that it is possible to define a maximally biological motion-sensitive region of pSTS that has no response to faces over objects. Thus, pSTS responses to faces and biological motion, while overlapping, also differ reliably. The finding of a consistent difference in position of face and biological motion responses diverges from the results of Engell and McCarthy (2013); this difference could result from the use of dynamic face stimuli in our study. While studies using static faces have typically observed a single face-responsive region of posterior STS, the

current results and prior evidence (Pitcher et al. 2011) indicate that dynamic faces engage several regions along the length of the STS, and may engage a posterior STS region with slightly different spatial properties.

The most striking case of overlap observed in the current study was that between responses to dynamic faces and vocal sounds. This result manifested as a strong voice response in a region defined to be maximally responsive to faces, and substantial overlap between the set of voxels responding significantly to each contrast. This is consistent with prior work finding a region of posterior STS that responds to faces and voices (Wright et al. 2003; Kreifels et al. 2009; Watson et al. 2014), as well as individual neurons in macaque STS that respond to faces and voices (Barracough et al. 2005; Ghazanfar et al. 2008; Perodin et al. 2014). The strikingly high voice response of a region defined by a face contrast was nevertheless unexpected, and indicates that this region should not be characterized as a “face region,” but rather a fundamentally audiovisual area. This case of overlap seems most plausibly interpreted as suggesting a common underlying process elicited by the 2 broad contrasts used in this study. One possibility is that this region is involved in human identification using both facial and vocal cues. However, this hypothesis cannot easily explain the strong preference for dynamic over static faces in this region (Pitcher et al. 2011). Another possibility is that this region is involved in audiovisual processing of speech and/or other human vocalizations. Alternatively, this region might be more generally involved in processing communicative stimuli of different modalities. Further research will be necessary to tease apart these possibilities.

In addition to a region of overlapping response to faces and voices, a further-anterior region responded highly selectively to vocal stimuli. This region was centered on the upper bank of the middle STS, consistent with prior reports (Belin et al. 2000).

Language responses have been observed along the entire length of the left STS (Binder et al. 1997; Fedorenko et al. 2012), consistent with our results. In our data, language responses were strongest bilaterally in a far-anterior region of STS, a pattern that prior studies have not noted. Unlike most prior studies, the sentences used in our language contrast involved no human characters whatsoever (nor any living things at all), to dissociate language from social reasoning. This difference might account for the slight divergence between our results and those of prior studies.

Another case of particularly strong overlap occurred between responses to language and ToM. The left TPJ region defined by the ToM contrast was modulated by linguistic content, and the bilateral anterior language regions were both modulated by mental state content. We also observed substantial overlap between language and ToM responses across the STS: roughly half of voxels with a language response also had an effect of ToM content. While not observed previously, this observation is consistent with prior reports that regions elicited by ToM and semantic contrasts both bear a rough, large-scale resemblance to default mode areas (Buckner and Carroll 2007; Binder 2012). While relationships between language and ToM in development have been extensively documented (De Villiers 2007), lesion evidence indicates that these functions can be selectively impaired in adults (Apperly et al. 2004, 2006). Nevertheless, the finding of strong overlap between language and ToM responses is intriguing and should be explored further. One potential account of the effect of ToM in language regions is that the presence of agents is an organizing principle of semantic representations in these regions.

Based on fMRI overlap results, we have argued that, at the spatial resolution of the present study, STS responses to certain types of social information overlap, in some cases substantially. However, a number of caveats must be made regarding the interpretation of fMRI overlap data. For a number of reasons, it is impossible to directly infer the presence of overlap in underlying neural responses from overlap in fMRI responses, which measure a hemodynamic signal at relatively low resolution. For one, fMRI measures signal from blood vessels, and stronger signal is obtained from larger vessels that pool blood from larger regions of cortex (Polimeni et al. 2010). Next, fMRI is a relatively low-resolution measure (typically ~3 mm), pooling responses over hundreds of thousands of individual neurons, and further structure likely exists within the resolution of a single voxel. Lastly, fMRI data are both intrinsically spatially smoothed by the use of k-space sampling for data acquisition, and typically smoothed further in preprocessing, to increase signal-to-noise ratio. In the present study, we mitigate this concern by measuring overlap at different smoothing kernels, and assessing response profiles of overlapping regions using spatially unsmoothed data.

The above points establish that fMRI can miss spatial structure at a fine scale. Thus, in principle, functionally specific STS subregions within the overlapping regions observed here could exist at finer spatial scales. Nevertheless, the present results argue for substantial overlap at the typical resolution of fMRI studies, which is not merely induced by spatial smoothing during preprocessing, and is robust to differences in the statistical threshold used to define regions.

We have argued that the STS contains subregions with distinct profiles of fMRI responses, resulting from distinct underlying neuronal selectivity profiles. Could the differences observed here instead relate to other influences on contrast-to-noise ratio, such as differences in signal reliability across regions, or differences in the efficacy of different stimulus sets in driving neural responses? It is unlikely that signal reliability contributes substantially to regional differences, given that these regions are roughly matched on tSNR (and despite slightly lower tSNR in voice regions, these areas had among the strongest category-selective responses observed). The paradigms we used were designed to be similar in basic ways (e.g., mostly blocked designs, with ~5 min of stimulation per condition), but could nevertheless differ in their ability to drive strong responses. For instance, they might differ in variability along stimulus dimensions encoded in a given region, which is impossible to judge without knowing these dimensions. We consider it unlikely that this accounts for the substantial differences across regions observed here, given that 1) each task drove strong responses in at least some STS subregion; and 2) for tasks that evoked the most spatially extensive responses (ToM and voices), we found similar responses across 2 tasks, suggesting that these were somewhat robust to stimulus details. Nevertheless, it is important to note that differences in the efficacy of different tasks could have influenced the response magnitudes reported here.

Another general limitation of this study is that while we would like to identify regions of the STS that are involved in specific cognitive and perceptual processes, we instead use the proxy of identifying regions of the STS that respond in a given task contrast, as in any fMRI study. There is presumably no simple one-to-one mapping between processes or representations and broad pairwise contrasts. Given that we have little knowledge of the specific processes underlying social perception and cognition, it is difficult to determine the appropriate mapping. For instance, our face stimuli presumably elicit processes related to the perceptual processing of faces (which itself is a high-level

description that likely comprises multiple computations), but they also contain specific types of biological motion (eye and mouth motion), which might be processed via separate mechanisms. These stimuli could also trigger the analysis of intentions or emotional states of the characters in the clips. Likewise, our ToM contrasts are intended to target mental state reasoning, but the ToM condition in both contrasts is more focused on human characters, and thus some of the responses we observe may relate to general conceptual processing of humans or imagery of human characters (although note that prior studies have found TPJ responses using tighter contrasts (Saxe and Kanwisher 2003; Saxe and Powell 2006)). Generally, it is important to emphasize that the activation maps we describe presumably comprise responses evoked by a number of distinct processes, which future research will hopefully tease apart.

In sum, the present study converges with prior research to indicate that the STS is a key hub of social and linguistic processing. Our method of testing each subject on multiple contrasts enables us to rule out prior claims that the STS represents a largely homogeneous and multifunctional region (Hein and Knight 2008), instead revealing rich spatial structure and functional heterogeneity throughout the STS. Specifically, the STS appears to contain both subregions that are highly selective for processing specific types of social stimuli, as well as regions that respond to social information from multiple domains. These findings paint a picture in which the extraordinary human capacity for social cognition relies in part on a broad region that computes multiple dimensions of social information over a complex, structured functional landscape. This work opens up myriad questions for future research, including which specific computations are performed within each subregion of the STS, how the functional landscape of the STS differs in disorders that selectively disrupt or selectively preserve social cognition (e.g., autism and Williams' syndrome respectively), and whether the spatial overlap observed here is functionally relevant, or whether it reflects distinct but spatially interleaved neural populations.

## Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>

## Funding

This study was supported by grants from the Packard Foundation (to R.S.), and NSF (graduate research fellowship to B.D., CCF-1231216 to N.K. and R.S.). Funding to pay the Open Access publication charges for this article was provided by The Ellison Medical Foundation.

## Notes

We thank Hilary Richardson and Nicholas Dufour for help with data acquisition. Conflict of Interest: None declared.

## References

- Allison T, Puce A, McCarthy G. 2000. Social perception from visual cues: role of the STS region. *Trends Cogn Sci.* 4:267–278.
- Ambady N, Rosenthal R. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: a meta-analysis. *Psychol Bull.* 111:256.
- Apperly IA, Samson D, Carroll N, Hussain S, Humphreys G. 2006. Intact first-and second-order false belief reasoning in a

- patient with severely impaired grammar. *Soc Neurosci.* 1:334–348.
- Apperly IA, Samson D, Chiavarino C, Humphreys GW. 2004. Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. *J Cogn Neurosci.* 16:1773–1784.
- Barraclough NE, Xiao D, Baker CI, Oram MW, Perrett DI. 2005. Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *J Cogn Neurosci.* 17:377–391.
- Beauchamp MS, Lee KE, Argall BD, Martin A. 2004. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron.* 41:809–824.
- Behzadi Y, Restom K, Liau J, Liu TT. 2007. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage.* 37:90.
- Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. 2000. Voice-selective areas in human auditory cortex. *Nature.* 403:309–312.
- Binder JR. 2012. Task-induced deactivation and the “resting” state. *Neuroimage.* 62:1086–1091.
- Binder JR, Frost JA, Hammeke TA, Cox RW, Rao SM, Prieto T. 1997. Human brain language areas identified by functional magnetic resonance imaging. *J Neurosci.* 17:353–362.
- Bonda E, Petrides M, Ostry D, Evans A. 1996. Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *J Neurosci.* 16:3737–3744.
- Brass M, Schmitt RM, Spengler S, Gergely G. 2007. Investigating action understanding: inferential processes versus action simulation. *Curr Biol.* 17:2117–2121.
- Brett M, Johnsrude IS, Owen AM. 2002. The problem of functional localization in the human brain. *Nat Rev Neurosci.* 3:243–249.
- Bruneau EG, Dufour N, Saxe R. 2012. Social cognition in members of conflict groups: behavioural and neural responses in Arabs, Israelis and South Americans to each other’s misfortunes. *Philos Trans R Soc B Biol Sci.* 367:717–730.
- Buckner RL, Carroll DC. 2007. Self-projection and the brain. *Trends Cogn Sci.* 11:49–57.
- Calvert GA, Hansen PC, Iversen SD, Brammer MJ. 2001. Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect. *Neuroimage.* 14:427–438.
- Chai XJ, Castañón AN, Ongür D, Whitfield-Gabrieli S. 2012. Anticorrelations in resting state networks without global signal regression. *Neuroimage.* 59:1420–1428.
- Ciaramidaro A, Adenzato M, Enrici I, Erk S, Pia L, Bara BG, Walter H. 2007. The intentional network: how the brain reads varieties of intentions. *Neuropsychologia.* 45:3105–3113.
- Corbetta M, Shulman GL. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci.* 3:201–215.
- De Villiers J. 2007. The interface of language and theory of mind. *Lingua.* 117:1858–1878.
- Dodell-Feder D, Koster-Hale J, Bedny M, Saxe R. 2011. fMRI item analysis in a theory of mind task. *Neuroimage.* 55:705–712.
- Engell AD, McCarthy G. 2013. Probabilistic atlases for face and biological motion perception: an analysis of their reliability and overlap. *Neuroimage.* 74:140–151.
- Fedorenko E, Behr MK, Kanwisher N. 2011. Functional specificity for high-level linguistic processing in the human brain. *Proc Natl Acad Sci USA.* 108:16428–16433.
- Fedorenko E, Kanwisher N. 2009. Neuroimaging of language: why hasn’t a clearer picture emerged? *Lang Linguistics Compass.* 3:839–865.
- Fedorenko E, Nieto-Castanon A, Kanwisher N. 2012. Lexical and syntactic representations in the brain: an fMRI investigation with multi-voxel pattern analyses. *Neuropsychologia.* 50:499–513.
- Fletcher PC, Happe F, Frith U, Baker SC, Dolan RJ, Frackowiak RS, Frith CD. 1995. Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. *Cognition.* 57:109–128.
- Gallagher H, Happé F, Brunswick N, Fletcher P, Frith U, Frith C. 2000. Reading the mind in cartoons and stories: an fMRI study of “theory of mind” in verbal and nonverbal tasks. *Neuropsychologia.* 38:11–21.
- Ghazanfar AA, Chandrasekaran C, Logothetis NK. 2008. Interactions between the superior temporal sulcus and auditory cortex mediate dynamic face/voice integration in rhesus monkeys. *J Neurosci.* 28:4457–4469.
- Gobbini MI, Koralek AC, Bryan RE, Montgomery KJ, Haxby JV. 2007. Two takes on the social brain: a comparison of theory of mind tasks. *J Cogn Neurosci.* 19:1803–1814.
- Greve DN, Fischl B. 2009. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage.* 48:63.
- Grossman E, Donnelly M, Price R, Pickens D, Morgan V, Neighbor G, Blake R. 2000. Brain areas involved in perception of biological motion. *J Cogn Neurosci.* 12:711–720.
- Grossman ED, Blake R. 2002. Brain areas active during visual perception of biological motion. *Neuron.* 35:1167–1175.
- Gweon H, Dodell-Feder D, Bedny M, Saxe R. 2012. Theory of mind performance in children correlates with functional specialization of a brain region for thinking about thoughts. *Child Dev.* 83:1853–1868.
- Haxby JV, Hoffman EA, Gobbini MI. 2000. The distributed human neural system for face perception. *Trends Cogn Sci.* 4:223–233.
- Hein G, Knight RT. 2008. Superior temporal sulcus—it’s my area or is it? *J Cogn Neurosci.* 20:2125–2136.
- Kanwisher N. 2010. Functional specificity in the human brain: a window into the functional architecture of the mind. *Proc Natl Acad Sci USA.* 107:11163–11170.
- Kreifelts B, Ethofer T, Shiozawa T, Grodd W, Wildgruber D. 2009. Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice-and-face-sensitive regions in the superior temporal sulcus. *Neuropsychologia.* 47:3059–3066.
- Lindstrom MJ, Bates DM. 1990. Nonlinear mixed effects modeling for repeated measures data. *Biometrics.* 46:673–687.
- Marchini JL, Ripley BD. 2000. A new statistical approach to detecting significant activation in functional MRI. *Neuroimage.* 12:366–380.
- Morris JP, Pelphrey KA, McCarthy G. 2005. Regional brain activation evoked when approaching a virtual human on a virtual walk. *J Cogn Neurosci.* 17:1744–1752.
- Pelphrey KA, Mitchell TV, McKeown MJ, Goldstein J, Allison T, McCarthy G. 2003. Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *J Neurosci.* 23:6819–6825.
- Pelphrey KA, Morris JP, McCarthy G. 2004. Grasping the intention of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J Cogn Neurosci.* 16:1706–1716.
- Pelphrey KA, Morris JP, Michelich CR, Allison T, McCarthy G. 2005. Functional anatomy of biological motion perception in posterior temporal cortex: an fMRI study of eye, mouth and hand movements. *Cereb Cortex.* 15:1866–1876.
- Pelphrey KA, Singer JD, Allison T, McCarthy G. 2003. Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia.* 41:156–170.

Downloaded from https://academic.oup.com/cercor/article-abstract/25/11/4596/2367595 by Universiteit Leiden / LUMC user on 08 March 2019

- Pelphrey KA, Viola RJ, McCarthy G. 2004. When strangers pass processing of mutual and averted social gaze in the superior temporal sulcus. *Psychol Sci.* 15:598–603.
- Perrodin C, Kayser C, Logothetis NK, Petkov CI. 2014. Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. *J Neurosci.* 34:2524–2537.
- Pitcher D, Dilks DD, Saxe RR, Triantafyllou C, Kanwisher N. 2011. Differential selectivity for dynamic versus static information in face-selective cortical regions. *Neuroimage.* 56:2356–2363.
- Polimeni JR, Fischl B, Greve DN, Wald LL. 2010. Laminar analysis of 7T BOLD using an imposed spatial activation pattern in human V1. *Neuroimage.* 52:1334–1346.
- Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL. 2011. Functional network organization of the human brain. *Neuron.* 72:665–678.
- Puce A, Allison T, Asgari M, Gore JC, McCarthy G. 1996. Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *J Neurosci.* 16:5205–5215.
- Saxe R, Brett M, Kanwisher N. 2006. Divide and conquer: a defense of functional localizers. *Neuroimage.* 30:1088–1096.
- Saxe R, Kanwisher N. 2003. People thinking about thinking people: the role of the temporo-parietal junction in “theory of mind”. *Neuroimage.* 19:1835–1842.
- Saxe R, Powell LJ. 2006. It’s the thought that counts: specific brain regions for one component of theory of mind. *Psychol Sci.* 17:692–699.
- Saxe RR, Whitfield-Gabrieli S, Scholz J, Pelphrey KA. 2009. Brain regions for perceiving and reasoning about other people in school-aged children. *Child Dev.* 80:1197–1209.
- Shih P, Keehn B, Oram JK, Leyden KM, Keown CL, Müller RA. 2011. Functional differentiation of posterior superior temporal sulcus in autism: a functional connectivity magnetic resonance imaging study. *Biol Psychiatry.* 70:270–277.
- Shultz S, Lee SM, Pelphrey K, McCarthy G. 2011. The posterior superior temporal sulcus is sensitive to the outcome of human and non-human goal-directed actions. *Soc Cogn Affect Neurosci.* 6:602–611.
- Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR. 2009. Correspondence of the brain’s functional architecture during activation and rest. *Proc Natl Acad Sci.* 106:13040–13045.
- Taylor KI, Moss HE, Stamatakis EA, Tyler LK. 2006. Binding cross-modal object features in perirhinal cortex. *Proc Natl Acad Sci USA.* 103:8239–8244.
- Triantafyllou C, Hoge RD, Wald LL. 2006. Effect of spatial smoothing on physiological noise in high-resolution fMRI. *Neuroimage.* 32:551–557.
- Triantafyllou C, Hoge RD, Krueger G, Wiggins CJ, Potthast A, Wiggins GC, Wald LL. 2005. Comparison of physiological noise at 1.5 T, 3 T and 7 T and optimization of fMRI acquisition parameters. *Neuroimage.* 26:243–250.
- Vander Wyk BC, Hudac CM, Carter EJ, Sobel DM, Pelphrey KA. 2009. Action understanding in the superior temporal sulcus region. *Psychol Sci.* 20:771–777.
- Vanrie J, Verfaillie K. 2004. Perception of biological motion: stimulus set of human point-light actions. *Behav Res Methods.* 36:625–629.
- Vigneau M, Beaucousin V, Herv P, Duffau H, Crivello F, Houd O, Mazoyer B, Tzourio-Mazoyer N. 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage.* 30:1414–1432.
- Watson R, Latinus M, Charest I, Crabbe F, Belin P. 2014. People-selectivity, audiovisual integration and heteromodality in the superior temporal sulcus. *Cortex.* 50:125–136.
- Woolrich MW, Behrens TEJ, Beckmann CF, Jenkinson M, Smith SM. 2004. Multilevel linear modelling for fMRI group analysis using Bayesian inference. *Neuroimage.* 21:1732–1747.
- Woolrich MW, Ripley BD, Brady M, Smith SM. 2001. Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage.* 14:1370–1386.
- Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G. 2003. Polysensory interactions along lateral temporal regions evoked by audiovisual speech. *Cereb Cortex.* 13:1034–1043.
- Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zöllei L, Polimeni JR, et al. 2011. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J Neurophysiol.* 106:1125–1165.

Downloaded from https://academic.oup.com/cercor/article-abstract/25/11/4596/2307585 by Universiteit Leiden / LUMC user on 08 March 2019