



Perceiving social interactions in the posterior superior temporal sulcus

Leyla Isik^{a,b,1,2}, Kami Koldewyn^{c,1}, David Beeler^{a,b}, and Nancy Kanwisher^{a,b,2}

^aDepartment of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; ^bMcGovern Institute of Brain Research, Massachusetts Institute of Technology, Cambridge, MA 02139; and ^cSchool of Psychology, Bangor University, Bangor LL57 2DG, United Kingdom

Contributed by Nancy Kanwisher, September 14, 2017 (sent for review August 16, 2017; reviewed by Alex Martin and Kevin Pelphrey)

Primates are highly attuned not just to social characteristics of individual agents, but also to social interactions between multiple agents. Here we report a neural correlate of the representation of social interactions in the human brain. Specifically, we observe a strong univariate response in the posterior superior temporal sulcus (pSTS) to stimuli depicting social interactions between two agents, compared with (i) pairs of agents not interacting with each other, (ii) physical interactions between inanimate objects, and (iii) individual animate agents pursuing goals and interacting with inanimate objects. We further show that this region contains information about the nature of the social interaction—specifically, whether one agent is helping or hindering the other. This sensitivity to social interactions is strongest in a specific subregion of the pSTS but extends to a lesser extent into nearby regions previously implicated in theory of mind and dynamic face perception. This sensitivity to the presence and nature of social interactions is not easily explainable in terms of low-level visual features, attention, or the animacy, actions, or goals of individual agents. This region may underlie our ability to understand the structure of our social world and navigate within it.

social perception | social interaction | superior temporal sulcus | fMRI | social brain

Humans perceive their world in rich social detail. We see not just agents and objects, but also agents interacting with each other. The ability to perceive and understand social interactions arises early in development (1) and is shared with other primates (2–4). Although considerable evidence has implicated particular brain regions in perceiving the characteristics of individual agents—including their age, sex, emotions, actions, thoughts, and direction of attention—whether specific regions in the human brain are systematically engaged in the perception of third-party social interactions is unknown. Here we provide just such evidence of sensitivity to the presence and nature of social interactions in the posterior superior temporal sulcus (pSTS).

To test for the existence of a brain region preferentially engaged in perceiving social interactions, we identified five neural signatures that would be expected of such a region. First, the region in question should respond more to stimuli depicting multiple agents interacting with each other than to stimuli depicting multiple agents acting independently. Second, this response should occur even for minimalist stimuli stripped of the many confounding features that covary with social interactions in naturalistic stimuli. Third, the response to social interactions should not be restricted to a single set of stimulus contrasts, but rather should generalize across stimulus formats and tasks. Fourth, the presence of a social interaction should be unconfounded from the presence of an agent's animacy or goals. Fifth, the region in question should not merely respond more strongly to the presence of social interactions, but should also contain information about the nature of those interactions. To test for these five signatures, we scanned subjects while they viewed two different stimulus sets that reduce social interactions to their minimal features: two agents acting with temporal and semantic contingency.

Previous studies have reported neural activations during viewing of social interactions (e.g., refs. 5 and 6; see also ref. 7 for a review of related studies), but have not provided evidence for the selectivity of this response. Two other studies (8, 9) found activations in numerous brain regions when people viewed social interactions between two humans vs. two humans engaged in independent activities, both depicted with point-light stimuli. However, in both studies, the task was to detect social interactions, so the interaction condition was confounded with target detection, making the results difficult to interpret. Numerous other functional magnetic resonance imaging (fMRI) studies have shown activation in and around the pSTS when subjects view social interactions depicted in shape animations (10–16), based on the classic stimuli of Heider and Simmel (17). However, those studies have generally interpreted the resulting activations in terms of the perception of animacy or goal-directed actions, or simply in terms of their general “social” nature, without considering the possibility that they are specifically engaged in the perception of third-party social interactions.

Of note, a recent study in macaques found regions of the frontal and parietal cortex that responded exclusively to movies of monkey social interactions and not to movies of monkeys conducting independent actions or of interactions between inanimate objects (4). If humans have a similarly selective cortical response to social interactions, where might it be found in the brain? One region that seems a likely prospect for such a response is the pSTS, which has been previously shown to respond during the perception of a wide variety of socially significant stimuli, including biological motion (18), dynamic faces (19), direction of gaze (20), emotional expressions (21), goal-directed actions (22), and communicative intent (23).

Significance

Humans spend a large percentage of their time perceiving the appearance, actions, and intentions of others, and extensive previous research has identified multiple brain regions engaged in these functions. However, social life depends on the ability to understand not just individuals, but also groups and their interactions. Here we show that a specific region of the posterior superior temporal sulcus responds strongly and selectively when viewing social interactions between two other agents. This region also contains information about whether the interaction is positive (helping) or negative (hindering), and may underlie our ability to perceive, understand, and navigate within our social world.

Author contributions: L.I., K.K., and N.K. designed research; L.I., D.B., and N.K. performed research; L.I. analyzed data; and L.I. and N.K. wrote the paper.

Reviewers: A.M., National Institute of Mental Health; and K.P., George Washington University.

The authors declare no conflict of interest.

Published under the PNAS license.

¹L.I. and K.K. contributed equally to this work.

²To whom correspondence may be addressed. Email: lisik@mit.edu or ngk@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1714471114/-DCSupplemental.

To test for a cortical region sensitive to the presence of social interactions, fitting some or all of the five criteria listed above, we first contrasted responses to point-light displays of two individuals who were interacting with each other vs. acting independently. We replicated previous unpublished findings from our laboratory, with most subjects showing a preferential response in the pSTS to social interactions in this contrast (24). We then asked whether this response generalizes to social interactions depicted using very different stimuli, shape animations. To unconfound responses to social interactions in the shape stimuli from responses to merely animate agents, or the goal-directed actions of those agents, we included a contrasting condition in which a lone animate agent pursues individual goals (22, 25). Finally, we tested whether the region showing a preferential response to social interactions contains information about the nature of that social interaction (helping vs. hindering). We found sensitivity to the presence and nature of social interactions in a region of the pSTS that cannot be explained by sensitivity to physical interactions or to the animacy or goals of individual agents.

Results

Experiment 1.

A region in the pSTS is sensitive to the presence of social interactions. To identify brain regions sensitive to the presence of social interactions, we scanned 14 participants while they viewed video clips of point-light walker dyads engaged either in a social interaction or in two independent actions (Fig. 1A). We used three of four runs from each subject to perform a whole-brain random-effects group analysis. This group analysis revealed a region in the right pSTS (MNI coordinates of voxel with peak significance: [54, -43, 18]) that responded significantly more strongly to social interactions than to independent actions (Fig. 2A). Apart from a weaker spread of this activation more anteriorly down the right STS and weaker activity in superior medial parietal regions bilaterally (Fig. S1), no other cortical region reached significance in this contrast.

The random-effects group analysis showed that a preferential response to social interactions is significant and anatomically consistent across subjects. However, group analyses are not ideal for characterizing the functional response of the region, because most functional regions do not align perfectly across subjects, and so the responses of specific regions are usually blurred with those of their cortical neighbors (26). To more precisely characterize the functional response of the region, we defined functional regions of interest (fROIs) in each subject individually. To do this, we selected the top 10% of interaction-selective voxels (i.e., the voxels with the lowest P values in the contrast of interacting vs. independent point-light conditions) for each subject within the region identified by the group analysis, using the same three runs of data as in the earlier group analysis. We term this individually defined fROI the “social interaction functional ROI” (SI-fROI). As in other previous group-constrained subject-specific analyses (27, 28), this method is an algorithmic way to select individual subject fROIs without subjective judgment calls while allowing for individual variation between subjects’ fROI locations, yet still broadly constraining them to the region defined by the group analysis.

We quantified the response to social interaction in each subject’s fROI using the held out run from the point-light experiment, and found a significantly greater response to social interactions over independent actions ($P = 9.5 \times 10^{-5}$, paired t test) (Fig. 3A). Twelve of the 14 subjects exhibited the presence of this SI-fROI, as defined with a threshold of $P < 0.005$. All 14 subjects showed a greater response to interacting vs. independent videos in held out data in the top 10% of voxels, indicating the presence of this sensitivity in all subjects. In addition, 6 of the 14 subjects showed a significantly greater ($P < 0.005$) response to socially interacting vs. independent point-light displays in the left hemisphere near the pSTS. Because this re-

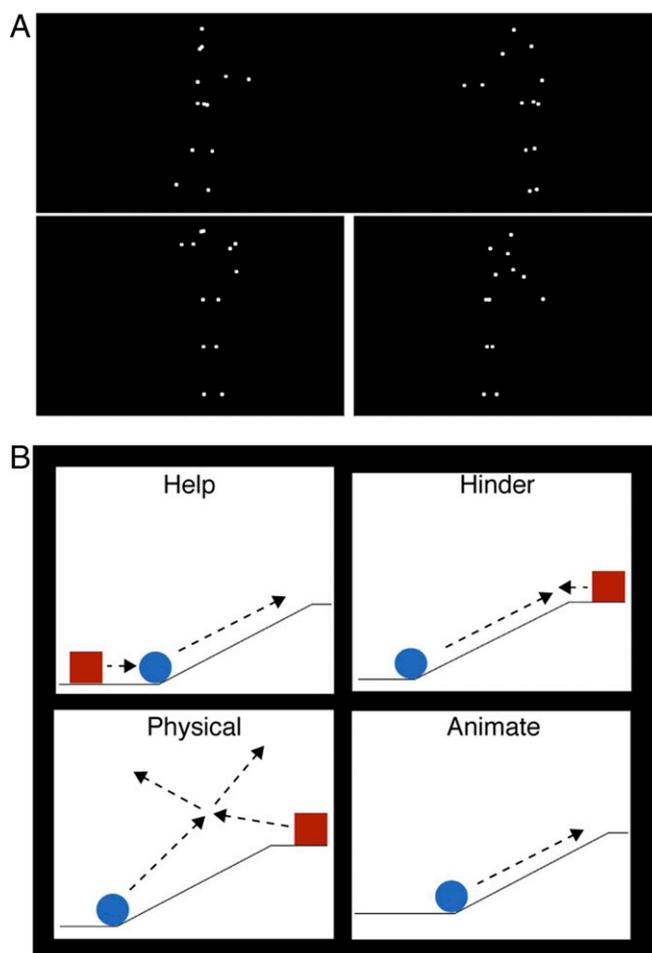


Fig. 1. Experimental stimuli. (A) In experiment 1, subjects viewed videos of two point-light figures either engaged in a social interaction (Top) or conducting two independent actions with a white line drawn between the two actors to increase the impression that they were acting independently (Bottom). (B) In experiment 2, subjects viewed videos of two animate shapes engaged in either a helping or a hindering interaction (Top). The first shape (in this example, blue) had a goal (e.g., climb a hill), and the second shape either helped (Top, Left) or hindered (Top, Right) the first shape. These two interaction conditions were then contrasted with two other conditions: a physical interaction condition, in which the two shapes moved in an inanimate fashion, like billiard balls (Bottom, Left), and an animate condition containing a single goal-oriented, animate shape (Bottom, Right). Movies S1–S4 provide examples of the four types of shape videos.

gion was not found consistently across subjects and did not reach significance in the group analysis, we did not analyze it further. **Relationship to nearby ROIs.** We next asked how the SI-fROI compares in location and response profile to established nearby regions engaged in other social tasks, namely the right temporal parietal junction (TPJ) (29, 30) and the right pSTS “face” region [referred to as such because that is the contrast by which it is defined, even though this region is now known to respond similarly to voices (31–33)]. To define these fROIs, we ran standard face (34) and theory of mind (35) localizers and used a similar group-constrained subject-specific method as described above. We again selected each individual subject’s top 10% of voxels for the relevant contrast within a group map defined based on a large number of subjects in previous studies (28, 36). We also identified motion-sensitive middle temporal (MT) region using the top 10% of voxels that responded more to moving shapes than to static task instructions in experiment 2 (see below) within the Freesurfer anatomic MT parcel (Methods).

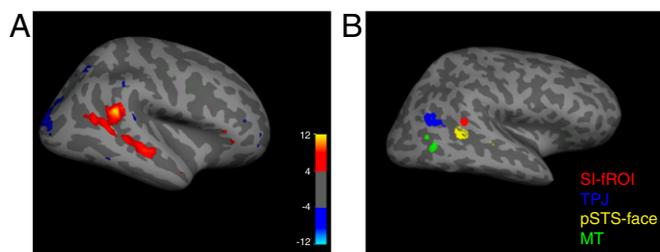


Fig. 2. Selectivity to social interactions in the pSTS. (A) Group random-effects map for the interaction vs. independent point-light walker contrast in experiment 1 showing a peak of activity in the right pSTS, with weaker activity along the STS. The color bar indicates the negative log of the P value for the interaction > independent contrast in that voxel. (B) Locations of the individually defined fROIs for one subject, including the SI-fROI in red, the TPJ in blue, the pSTS face region in yellow, and the MT in green. Individual-subject fROIs, defined with a group-constrained subject-specific analysis (Methods), show a consistent spatial organization across subjects, with the SI-fROI falling anterior to the TPJ and superior to the pSTS face region.

These fROIs showed a systematic spatial organization across subjects (Fig. 2B and Fig. S2), with the SI-fROI generally residing anterior to the TPJ and superior to the pSTS face region. The SI-fROI showed some overlap with these other fROIs in individual subjects. The overlap between the SI-fROI and the TPJ constituted on average 8% of the SI-fROI (the number of overlapping voxels divided by the number of voxels in the SI-fROI) and 2% of the TPJ. The overlap with the face STS region was on average 1% of the size of the SI-fROI and 1% of the size of the pSTS face region (Table S1). To examine the extent to which these regions represent distinct information, we removed these few overlapping voxels between the SI-fROI and the other fROIs in subsequent analyses.

The pSTS face region responded significantly more strongly to social interactions than to independent actions ($P = 0.014$), but this contrast was not significant in either the TPJ or the MT ($P = 0.21$ and 0.56 , respectively). Furthermore, a two-way ANOVA with fROI (SI-fROI vs. TPJ vs. pSTS face vs. MT) and the social interaction contrast (point-lights interacting vs. independent) as repeated-measures factors revealed a significant interaction [$F(3,13) = 10.16$, $\eta_p^2 = 0.43$, $P = 4.5 \times 10^{-5}$]. This two-way interaction reflected significantly greater sensitivity to social interactions in the SI-fROI than in each of the other fROIs [$F(1,13) = 17.02$, $\eta_p^2 = 0.57$, $P = 0.0012$ for TPJ; $F(1,13) = 16.14$, $\eta_p^2 = 0.55$, $P = 0.0015$ for pSTS face; $F(1,13) = 23.13$, $\eta_p^2 = 0.64$, $P = 0.00034$ for MT].

Sensitivity to other social dimensions. The foregoing analyses indicate that the SI-fROI is significantly more selective for social interactions compared with each of the three nearby regions (pSTS face, TPJ, and MT), and that the pSTS face region is the only other fROI showing a significant effect in this contrast. Does the SI-fROI differ from these nearby fROIs in other aspects of its response profile? On one hand, the SI-fROI shows a small but significant response to the theory of mind contrast (false belief > false photo; $P = 0.0042$) and face contrast (faces > objects; $P = 0.01$). On the other hand, the SI-fROI is significantly less sensitive to these contrasts compared with its cortical neighbors, as demonstrated by significant interactions of (i) SI-fROI vs. TPJ \times false belief vs. false photo [$F(1,13) = 22.93$, $\eta_p^2 = 0.64$, $P = 0.00035$] and (ii) SI-fROI vs. pSTS face \times faces vs. objects [$F(1,13) = 6.95$, $\eta_p^2 = 0.35$, $P = 0.021$]. Overall, these results indicate that while the social interaction region is both spatially close to and shares some functional information with the TPJ and pSTS face regions, its functional response profile differs significantly from that of each of these regions.

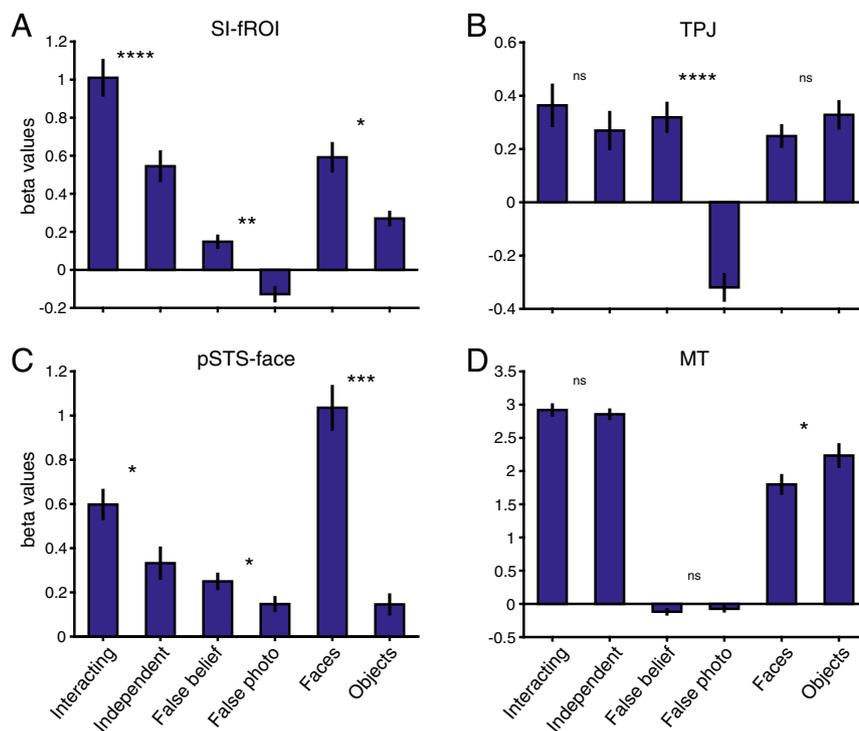


Fig. 3. fROI responses to experiment 2 shape stimuli. Shown are the average beta values (mean \pm SEM) across subjects in each individually defined fROI (A, SI-fROI; B, TPJ; C, pSTS face; D, MT) for the first 6 s of the help, hinder, physical interaction, and animate videos. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$; ns, not significant, $P > 0.05$.

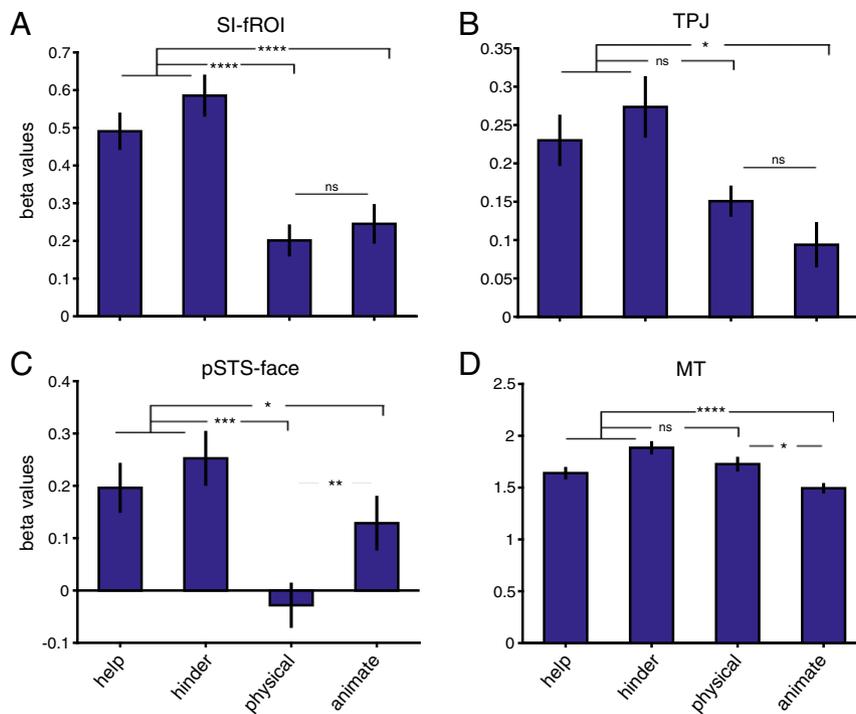


Fig. 4. Decoding helping vs. hindering conditions. Shown is the average classifier accuracy (mean \pm SEM) across subjects in each fROI for decoding help vs. hinder. A linear SVM classifier was trained on the beta values from nine pairs of videos in each individual subject's fROIs and tested on the tenth held-out pair. * $P \leq 0.05$; ** $P \leq 0.01$; *** $P \leq 0.001$; **** $P \leq 0.0001$; ns, not significant, $P > 0.05$.

Experiment 2. To investigate the nature of the social interaction information represented in the SI-fROI, we scanned the same subjects while they viewed 12-s videos of animations containing moving shapes from four different conditions: help, hinder, physical interactions, and animate (Fig. 1B and Movies S1–S4). The help and hinder videos consisted of two shapes engaged in a social interaction and could be divided into two segments. During the first ~ 6 s, one shape moved in a clearly goal-directed fashion. The other shape was either stationary or moved very little during this period, but in the context of the experiment, the percept of a social interaction was nonetheless clear during this period, with the second shape apparently “watching” the first shape. During the second 6 s of each video, the first shape was either helped or hindered in its goal by the second shape. We modeled each of these 6-s periods separately in a generalized linear model analysis. The first 6-s period of these videos was better controlled for motion and designed to provide a clean contrast between the interacting (help and hinder) and non-interacting (animate and physical interaction) conditions, while the second 6-s period was designed to most vividly depict helping and hindering. The two interaction conditions (help and hinder) were contrasted with the physical interaction videos, which consisted of two shapes moving in an inanimate fashion, like billiard balls colliding with each other and their background. Finally, to measure the extent to which activity in the SI-fROI is driven by animacy and goals of individuals in the absence of social interactions, the fourth set of animate videos consisted of a single goal-driven shape; five videos used the shape trajectory from shape 1 from a random half of the help videos, and five videos used the shape trajectory from the half of the hinder videos with unused help video trajectories. Two subjects (S1 and S2) saw a different version of the animate videos and were not included in the subsequent analysis, but were included in the final help vs. hinder analyses.

Social interaction sensitivity generalizes to shape stimuli. To test whether this new set of stimuli elicited social interaction responses in the pSTS, we compared the responses to the help and hinder

videos (social interaction) to the physical interaction videos in our four fROIs. For this contrast, we used the responses to the first 6 s of the shape videos, which were better controlled for low-level motion across the different conditions than the second 6 s. Although the second shape moved little or not at all during this initial period of each video, the percept of a social interaction was nonetheless

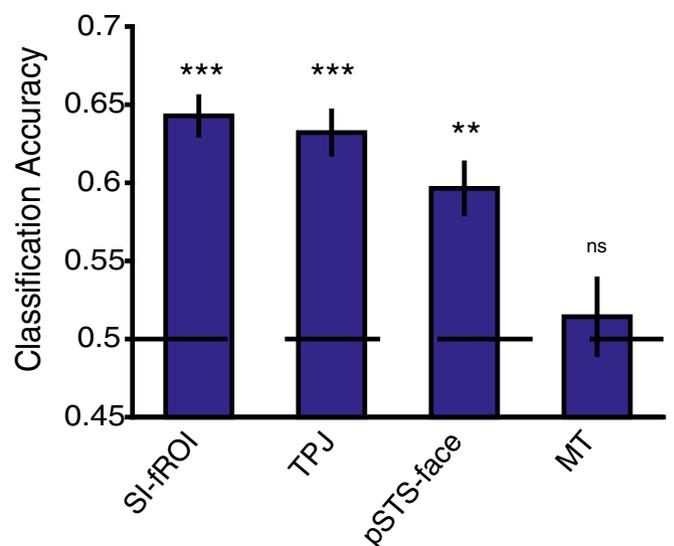


Fig. 5. fROI responses to experiment 1 stimuli. Shown are the average beta values (mean \pm SEM) across subjects in each individually defined fROI (A, SI-fROI; B, TPJ; C, pSTS face; D, MT) for each condition from the three fROI-defining contrasts: point light walkers interacting vs. independent (experiment 1), false belief vs. false photo stories (standard theory of mind localizer), and faces vs. objects. All beta values are calculated from a held-out localizer run that was not used to define the fROI. ** $P \leq 0.01$; *** $P \leq 0.001$; ns, not significant, $P > 0.05$.

clear during this period, an impression validated with ratings of naïve viewers on Amazon Mechanical Turk (Fig. S3C). The SI-fROI showed a significantly greater response to this first 6 s of the interaction videos (help and hinder) than physical interactions ($P = 1.3 \times 10^{-4}$, paired t test between the average of help and hinder and the physical interaction condition). The STS face region also showed a significantly greater response to socially interacting shapes ($P = 2.9 \times 10^{-4}$). The TPJ showed a trend toward a greater response to social interaction, but this did not reach significance ($P = 0.056$), and the MT did not show a significant difference ($P = 0.67$) (Fig. 4).

A two-way ANOVA, with ROI (SI-fROI vs. TPJ vs. pSTS face vs. MT) and social vs. physical interaction videos as repeated-measures factors, revealed a significant interaction [$F(3,11) = 8.03$, $\eta_p^2 = 0.42$, $P = 3.7 \times 10^{-4}$]. This two-way interaction reflects significantly greater sensitivity to social interactions over physical interactions in the SI-fROI than in the TPJ and MT [$F(1,11) = 27.5$, $\eta_p^2 = 0.71$, $P = 2.75 \times 10^{-4}$ for TPJ and $F(1,11) = 13.17$, $\eta_p^2 = 0.55$, $P = 0.004$ for pSTS face], but not the STS face region [$F(1,11) = 2.7$, $\eta_p^2 = 0.20$, $P = 0.13$]. These results generalize the sensitivity of the SI-fROI to social interactions found in experiment 1 to a new and very different stimulus set.

Sensitivity to animacy and goals. We have argued that the SI-fROI is specifically sensitive to the interaction of two shape stimuli, but is this region also driven by the animacy or goals individual agents? To find out, we measured the response of the SI-fROI to the single animated shape videos, where the shapes were both animate and goal-driven, and contrasted this with the physical interaction condition, where shapes were neither animate nor goal-driven, over the first 6 s of each video. Twenty independent raters on Amazon Mechanical Turk rated this segment of the video as significantly more animate (mean rating, 3.14/4) and goal-directed (mean rating, 3.22/4) compared with the physical interaction condition (mean rating, 2.0/4 for animacy and 1.6/4 for goal-directed; $P = 5.4 \times 10^{-7}$ and 8.5×10^{-11} , respectively). The SI-fROI showed no difference in response to the animate vs. physical interaction conditions ($P = 0.44$, paired t test). Similarly, the TPJ did not show a higher response to the animate vs. the physical videos ($P = 0.15$), and MT showed a higher response to the physical video compared with the animate videos ($P = 0.018$). The STS face region did show a significantly greater response to animate conditions than to physical interaction conditions ($P = 0.002$). In the second 6 s of the shape movies, the SI-fROI showed a slightly higher response to the animate condition than to the physical condition (Fig. S4), perhaps indicating a response to success or failure in attaining goals. The lack of such an effect in the first 6 s, when the goals and animacy of the shapes in the animate videos were very clear (Fig. S3), indicates that the presence of animacy and individual goals on their own is not sufficient to activate this region.

A whole-brain random-effects group analysis revealed a region in the pSTS that responds significantly more to the animate than physical videos (Fig. S5). This activation is anterior to the highly significant region observed for the social interaction contrast (Fig. 2A). This group analysis, combined with the lack of response to the animate video conditions in the SI-fROI, suggests that separate regions in the pSTS process social interactions vs. animacy and goal-directed actions.

Representation of helping and hindering. The foregoing analyses reveal a clear univariate sensitivity to presence of social interactions in the SI-fROI in the pSTS. Does this region also contain information about the nature of that social interaction? To answer this question, we used multivariate pattern analysis (MVPA) to decode whether the video depicted a helping or a hindering interaction (1). We used the beta values for each voxel from the second 6 s of each of the 10 help and hinder movies (when the helping or hindering action occurs) as input features to a linear support vector machine (SVM) classifier. We trained this classifier on data from nine pairs of matched help and hinder videos (i.e.,

two videos that begin very similarly but end with a helping action or a hindering action), and tested it on data from a tenth held out pair of help/hinder videos. We repeated this analysis for each held out video pair. This analysis provides a strong test of generalization across our different stimulus pairs, because the objects on screen and location and movement patterns of the shapes are more similar within a matched help/hinder pair than they are across different help videos, or across different hinder videos.

We can robustly decode helping vs. hindering in the SI-fROI ($P = 1.2 \times 10^{-4}$) and TPJ ($P = 5.0 \times 10^{-4}$), as well as to a lesser extent in the STS-face region ($P = 0.0086$). Importantly, we cannot decode help vs. hinder in the MT ($P = 0.39$), and we also do not observe a univariate difference in the univariate MT response (Fig. 5 and Fig. S4).

Discussion

Here we report a region of the pSTS, detectable in most subjects individually, that responds approximately twice as strongly when viewing simple point-light videos of two people interacting compared with two people acting independently. This selective response to social interactions is unlikely to result from differences in attentional engagement or low-level differences in the stimuli, because it is not found in visual motion area MT, which is sensitive to both (37). Moreover, the response to social interactions cannot be reduced to a response to the animacy, actions, or goals of individual actors, because these attributes alone do not drive this region to a greater degree than inanimate shapes. Furthermore, the strong sensitivity to social interactions depicted in the point-light displays in experiment 1 generalizes to the very different depictions of social interactions in animated shapes in experiment 2. Finally, the same region contains information about the nature of the social interaction—specifically, whether it is positive (helping) or negative (hindering). All told, this region exhibits all five signatures that we predicted for a region selectively engaged in perceiving social interactions.

Our present findings are further strengthened by the fact that an independent study with different stimuli and subjects yielded highly similar results, including a preferential response to social interactions compared with independent actions that generalizes from point-light displays to shape animations, and the ability to decode cooperative vs. competitive social interactions from the same region (38).

The selective response to social interactions reported here does not appear to take the form of a discrete cortical region with sharp edges that is exclusively engaged in perceiving social interactions. Although the peak activation to social interactions is largely nonoverlapping with the TPJ, as suggested by previous work (15), it shows a significantly higher response to false beliefs than false photos (the standard theory of mind localizer contrast). This region also shows a significantly higher response to faces than to objects. On the other hand, responses overall are much lower in all of these conditions than for social interactions, and some of the response to theory of mind stories and faces may be due to the social interactions implied by these stimuli. The sensitivity to social interactions also spills over into the nearby TPJ and pSTS face region, albeit in weaker form. Thus, the sensitivity to social interactions reported here may be better considered not as a discrete module, but rather as a peak in the landscape of partially overlapping sensitivities to multiple dimensions of social information in the STS.

The sensitivity to social interactions reported here may further inform our understanding of previously reported cortical responses to social stimuli. The pSTS face region has remained an intriguing mystery ever since it was shown to respond threefold more strongly to dynamic faces than to static faces (34), and to respond equally to videos of faces and recordings of voices (31, 33, 39). Here we found that this region also shows some sensitivity to both the presence and nature of social interactions depicted with point-light displays and animated shapes, implicating this region in

the perception of third-party social interactions. It is possible that the strong response of this region to dynamic face stimuli may be related to the fact that most of the faces in our study were clearly interacting with a third person off-screen. This hypothesis predicts a greater response in this region to dynamic faces interacting with off-screen third-party agents than to dynamic faces engaged in individual noninteractive activities.

A number of previous studies that reported activations in the pSTS during viewing of animations of interacting shapes (10, 13, 14, 16, 40, 41) have interpreted these activations as reflecting inferences about the intentions or animacy of individual actors. Our data suggest that these activations cannot be driven primarily by either the animacy or goals of these shapes, but instead likely reflect the perception of social interactions. Importantly, however, many other studies have found sensitivity in nearby regions to the intentions of individuals that cannot be straightforwardly accounted for in terms of a social interaction (25, 42), and we also observed portions of the STS that responded to individual agents pursuing goals (Fig. S5). Thus, the currently available evidence suggests the existence of at least four areas of dissociable responses in this general region: the TPJ, specialized for inferring the thoughts of others; the pSTS-face region; another STS region responsive to the agency and/or goal-directed actions of individual actors; and the selective response to social interactions reported here (see also ref. 15).

Why might an analysis of social interactions between third-party agents be so important that a patch of cortex is allocated to this task? Clearly, humans care a great deal about social interactions, and recognizing the content and valence of others' interactions plays several important roles in our daily lives. First, social interactions reveal information about individuals; we determine whether a person is nice or not nice by how that person treats others. Social interactions also improve the recognition of individual agents and their actions (43). In addition, social interactions reveal the structure of our social world: who is a friend (or foe) of whom, who belongs to which social group, and who has power over whom. Understanding these social relationships is crucial for deciding how to behave in the social world, particularly for deciding whether and when to trade off our individual self-interests for the potential benefits of group cooperation (44). To inform such complex decisions, the perception of social interactions likely interacts with other relevant social dimensions, such as the gaze direction, emotions, thoughts, and goals of others, perhaps providing a clue as to why these functions reside nearby in cortical space.

Humans are not the only animals with a strong interest in third-party social interactions, and recent work has identified regions in the macaque cortex that respond exclusively during viewing of such interactions (4). However, the regions selectively responsive to social interactions in macaques are situated in the frontal and parietal lobes, not in the temporal lobe, and thus are unlikely to be strictly homologous to the region described here. Macaques do show a sensitivity to social interactions in the lateral temporal lobe, but that region responds similarly to interactions between inanimate objects, in sharp contrast to the region reported here in humans. Nonetheless, it is notable that the perception of social interactions is apparently important enough in both humans and macaques that a region of cortex is allocated largely or exclusively to this function, even if the two regions are not strict homologs.

This initial report leaves open many questions for future research. First, as is usual with fMRI alone, we do not yet have evidence that this region is causally engaged in the perception of social interactions. In particular, while MVPA presents a powerful tool for reading out neural patterns, such as those distinguishing helping vs. hindering, the fact that scientists can read out a certain kind of information from a given region does not necessarily mean that the rest of the brain is reading out that information from that region (45). Future studies might investigate this question with transcranial magnetic stimulation or studies of patients with brain damage. Second, it is unknown when or how the selective response

to social interactions develops, and whether its development requires experience in viewing social interactions. Behaviorally, human infants are highly attuned to social interactions and can distinguish between helping and hindering by 6 mo of age (1), perhaps suggesting that this region may be present by that age. Third, the structural connectivity of this region and its interactions with the rest of the brain are unknown. Beyond the obvious hypothesis that this region is likely connected to other parts of the social cognition network, it may also be connected with brain regions implicated in intuitive physics (46), since the distinction between helping and hindering fundamentally hinges on understanding the physics of the situation.

Finally, and most importantly, we have barely begun the to characterize the function of this region and the scope of stimuli to which it responds. Will it respond to a large group of people interacting with each other, as at a party, a football game, or a lecture hall? It is also unknown what exactly this region represents about social interactions (the mutual perceptual access of two agents; social dominance relations between two people; the temporal contingency of actions; all of the above?), and whether these representations are calculated directly from bottom-up cues or from top-down information about goals and social judgments (47). More fundamentally, is this region a unimodal visual region or will it respond to other types of stimuli, such as an audio description or verbal recording of an interaction? This region's functional dissociation from the low-level visual motion MT region and proximity to other regions integrating multimodal social information in the STS lead to the intriguing possibility that it responds to abstract, multimodal representations of social interactions. Although considerable further work is needed to precisely characterize the representations and computations conducted in this region, the initial data reported here suggest that this work is likely to prove fruitful.

Methods

Participants. Fourteen subjects (age 20–32 y, 10 females) participated in this study. All subjects had normal or corrected-to-normal vision and provided informed written consent before the experiment. MIT's Committee on the Use of Humans as Experimental Subjects approved the experimental protocol.

Paradigm. Each subject performed four experiments over the course of one to three scan sessions. The first experiment consisted of point-light dyads that were either engaged in a social interaction (stimuli from ref. 48) or performing two independent actions (stimuli from ref. 49). Individual videos ranged in length from 3 to 8 s, and three videos were presented in each 16-s block. Each run consisted of eight blocks of each condition and two 16-s fixation blocks presented at the middle and end of each run, for a total time of 160 s per run. Stimulus conditions were presented in a palindromic order. This experiment was split over the course of two runs and was repeated twice, for a total of four runs. The subjects passively viewed these videos.

In the second experiment, each subject viewed 12-s videos of one or two simple shapes moving in one of four conditions: help, hinder, animate, or physics. In the first three conditions, the shapes were portrayed as animate and one shape in each video had a goal (e.g., the blue square wants to climb a hill) (Fig. 1). In the help and hinder conditions, a second shape was present, which either helped (e.g., pushed the first shape up the hill) or hindered (e.g., blocked the first shape from the top of the hill) the first shape in achieving its goal. In the third, animate condition, the first shape's motion was kept the same as in the help or hinder videos, but the second shape was removed, leaving only one shape on the screen either achieving or failing at its goal (one half of the videos were shape 1 from the help videos and the other half were shape 1 from the hinder videos). In all three conditions, the first shape's goal was kept constant across each set of three videos (help, hinder, and animate), and each 12-s video consisted of two parts: the first 6 s, during which one shape establishes a goal, and the second 6 s, during which that shape is either helped or hindered, or does or does not achieve its goal alone (in the animate condition). In a fourth, physics condition, the shapes were depicted as inanimate billiard balls moving around the same scene as shown in the first three videos and having physical collisions with each other and with the background. The videos with two shapes contained a red and blue shape (color counterbalanced between shape 1 and shape 2 in the help/hinder videos), and the animate videos contained one blue shape. After viewing each video, the subject was given 4 s to answer the question, "How much do you like the blue shape" on a scale of 1–4 (the

response order was flipped halfway between each run to avoid motor confounds). Each subject viewed 10 different sets of matched videos for each of the four conditions, for a total of 40 different videos presented over two runs. Each run lasted 320 s [20 videos \times (12-s video + 4-s response period)]. Each subject saw each video a total of four times, over eight total runs.

Finally, we performed two localizer experiments to identify nearby regions in the pSTS also known to process socially relevant stimuli: the STS face region (which responds equally to voices) and the theory of mind selective region in the TPJ. To localize the face region, subjects viewed 3-s videos of moving faces or moving objects as described in ref. 34. Stimuli were presented in 18-s blocks of six videos that subjects passively viewed. For six subjects, additional blocks of bodies, scenes, and scrambled scenes were presented, but their responses were not analyzed for this study. Stimuli were presented in two runs, each containing four blocks per condition presented in palindromic order. Each run also contained two 18-s fixation blocks at the start, middle, and end, for a total run time of 180 s.

Each subject also performed a theory of mind task, as described in ref. 35 and available at saxelab.mit.edu/superloc.php. The subject read brief stories describing beliefs (theory of mind condition) or physical descriptions (control condition) and answered a true/false question about each story. Stories were presented for 10 s, followed by a 4-s question period, with a 12-s fixation period at the beginning of end of each run, for a total run time of 272 s. Stories were presented in two counterbalanced palindromic runs.

Data Acquisition. Data were collected at the Athinoula A. Martinos Imaging Center at MIT on a Siemens 3-T MAGNETOM Tim Trio Scanner with a 32-channel head coil. A high-resolution T1-weighted anatomic image (multiecho MPRAGE) was collected at each scan [repetition time (TR), 2,530 ms; echo time (TE), 1.64 ms, 3.44 ms, 5.24 ms, and 7.014 ms (combined with an RMS combination); echo spacing, 9.3 ms; bandwidth, 649 Hz/pixel; timing interval (TI), 1,400 ms; flip angle, 7; field of view (FOV) 220 \times 220 mm; matrix size, 220 \times 220 mm; slice thickness, 1 mm; 176 near-axial slices; acceleration factor of 3; 32 reference lines]. Functional data were collected using a T2*-weighted echo planar imaging EPI pulse sequence sensitive to blood oxygen level-dependent (BOLD) contrast (TR, 2,000 ms; TE, 30 ms; echo spacing, 0.5 ms; bandwidth, 2,298 Hz/pixel; flip angle, 90; FOV, 192 \times 192 mm; matrix, 64 \times 64 mm; slice thickness, 3 mm isotropic; slice gap, 0.3 mm; 32 near-axial slices).

Data Preprocessing and Modeling. Data preprocessing and generalized linear modeling were performed using the Freesurfer Software Suite (freesurfer.net). All other analyses were conducted in MATLAB (MathWorks). Preprocessing consisted of motion-correcting each functional run, aligning it to each subject's anatomic volume, and then resampling to each subject's high-density surface as computed by Freesurfer. After alignment, data were smoothed using a 5-mm FWHM Gaussian kernel. For group-level analyses, data were coregistered to standard anatomic coordinates using the Freesurfer FSAverage template. All individual analyses were performed in each subject's native surface. Generalized linear models included one regressor per stimulus condition, as well as nuisance regressors for linear drift removal and motion correction (x , y , z) per run.

Group Analysis. To test whether a systematic region across subjects responded more strongly to social interactions than to independent actions in experiment 1, we performed a surface-based random-effects group analysis across all subjects (holding out a single run) using Freesurfer. We first transformed the contrast difference maps for each subject to a common space (the Freesurfer fsaverage template surface). The random-effects group analysis yielded an activation peak in the right pSTS. In subsequent analyses, we used the contiguous significant voxels ($P < 10^{-4}$) around this peak as a group map to spatially constrain an individual subject's fROIs (described next).

ROI Definition. To examine the region in the pSTS showing a selective response to social interactions in each subject, and to compare it with nearby fROIs that have been previously implicated in the processing of dynamic and/or social stimuli, we defined four ROIs for each subject: right pSTS interaction region, pSTS face region, TPJ, and MT. Since the random-effects group

analysis found significant selective responses to social interactions only in the right hemisphere, we restricted our ROI analysis to the right hemisphere.

To define ROIs in individual subjects, we used a group-constrained subject-specific approach (27, 28), in which a functional or (in the case of MT) anatomic parcel was used to constrain an individual subject's fROIs. For each subject, we defined each fROI as the top 10% most significant voxels for the relevant contrast (holding out one run of data) within the relevant parcel.

The group-based parcel used to spatially constrain the selection of the SI-fROI was defined as the set of all voxels significant at the $P < 10^{-4}$ level (uncorrected) in the random-effects group contrast of social interaction $>$ independent actions, which were contiguous and included the peak voxel. In the same three of four runs used to define the group map, we selected the top 10% of voxels in each subject as that subject's SI-fROI.

To define the pSTS face region, we use a parcel from ref. 28 and identified the top 10% most significant voxels showing a greater response to faces than to objects in one run of each subject's face localizer. To define the TPJ, we used the group map from ref. 36 and selected the top 10% significant voxels representing false belief $>$ false physical task from one run of the theory of mind task. For the theory of mind task, we jointly modeled each story and question as a single event. Finally, to define MT, we used the Freesurfer anatomic MT parcel and the top 10% significant voxels from all shape videos $>$ task periods in experiment 2.

To examine each region's response to the social interaction, face, and theory of mind contrasts, we measured the magnitude of response to each condition in each ROI in the held-out run. Because MT was defined in a hypothesis-neutral manner (using all conditions), we did not hold out any data when defining the fROI.

Overlap Analysis. To assess to the extent to which our three pSTS ROIs overlap with one another, we calculated the overlap between each pair of fROIs with respect to each of the two ROIs: $\text{size}(A,B)/\text{size}(A)$ and $\text{size}(A,B)/\text{size}(B)$, representing the proportion of one region that is overlapping with the other region (33). To examine the extent to which these regions represent distinct information, we excluded overlapping voxels from subsequent analyses.

MVPA. To test whether pattern information in our ROIs could distinguish helping from hindering, we used the beta values from each ROI for the second 6 s (the period when the helping/hindering occurs) of each of the 10 help and 10 hinder videos. We trained a linear SVM (implemented with MATLAB) to perform the binary classification between help and hinder videos. We trained the classifier on 9 out of 10 help/hinder video pairs (leaving out one matched pair of help/hinder videos). We assessed the accuracy of the classifier by testing it on held-out help/hinder pairs of videos. We cycled through 10 held-out repetitions and averaged the accuracy for each subject and ROI across these 10 runs. For each ROI, we tested whether the average classification accuracy for all subjects was significantly better than chance by comparing it with a chance classification of 0.5 (one-tailed t test).

Amazon Mechanical Turk Ratings. To assess the saliency of the animacy, goals, and social interactions in the first 6 s of each shape video, we collected ratings from 20 independent raters for each video on Amazon Mechanical Turk. The raters first watched a video of a screen capture from one continuous run of the shape experiment (experiment 2) to familiarize them with the videos and the tasks performed by subjects in the scanner. They then viewed the first 6 s of each clip from the unseen run, and provided ratings (from 1 = least to 4 = most) for the animacy, goals, and social interactions of the shapes in each clip. We repeated this for both runs of the shape experiment to obtain 20 independent ratings for each video.

ACKNOWLEDGMENTS. We thank C. Robertson and A. Mynick for comments on the manuscript and L. Powell and R. Saxe for helpful discussions. This work was supported by a US/UK Office of Naval Research Multidisciplinary University Research Initiative Project (Understanding Scenes and Events Through Joint Parsing, Cognitive Reasoning, and Lifelong Learning); National Science Foundation Science and Technology Center for Brains, Minds, and Machines Grant CCF-1231216; and a National Institutes of Health Pioneer Award (to N.K.).

1. Hamlin JK, Wynn K, Bloom P (2007) Social evaluation by preverbal infants. *Nature* 450:557–559.
2. Cheney D, Seyfarth R, Smuts B (1986) Social relationships and social cognition in nonhuman primates. *Science* 234:1361–1366.
3. Bergman TJ, Beehner JC, Cheney DL, Seyfarth RM (2003) Hierarchical classification by rank and kinship in baboons. *Science* 302:1234–1236.
4. Sliwa J, Freiwald WA (2017) A dedicated network for social interaction processing in the primate brain. *Science* 356:745–749.

5. Quadflieg S, Gentile F, Rossion B (2015) The neural basis of perceiving person interactions. *Cortex* 70:5–20.
6. Petrini K, Piwek L, Crabbe F, Pollock FE, Garrod S (2014) Look at those two!: The precuneus role in unattended third-person perspective of social interactions. *Hum Brain Mapp* 35:5190–5203.
7. Quadflieg S, Koldewyn K (2017) The neuroscience of people watching: How the human brain makes sense of other people's encounters. *Ann N Y Acad Sci* 1396: 166–182.

8. Centelles L, Assaiante C, Nazarian B, Anton J-L, Schmitz C (2011) Recruitment of both the mirror and the mentalizing networks when observing social interactions depicted by point-lights: A neuroimaging study. *PLoS One* 6:e15749.
9. Sapey-Triomphe L-A, et al. (2017) Deciphering human motion to discriminate social interactions: A developmental neuroimaging study. *Soc Cogn Affect Neurosci* 12:340–351.
10. Castelli F, Happé F, Frith U, Frith C (2000) Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage* 12:314–325.
11. Castelli F, Frith C, Happé F, Frith U (2002) Autism, Asperger syndrome, and brain mechanisms for the attribution of mental states to animated shapes. *Brain* 125:1839–1849.
12. Martin A, Weisberg J (2003) Neural foundations for understanding social and mechanical concepts. *Cogn Neuropsychol* 20:575–587.
13. Schultz RT, et al. (2003) The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. *Philos Trans R Soc B Biol Sci* 358:415–427.
14. Schultz J, Friston KJ, O'Doherty J, Wolpert DM, Frith CD (2005) Activation in posterior superior temporal sulcus parallels parameter inducing the percept of animacy. *Neuron* 45:625–635.
15. Gobbi 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.
16. Santos NS, et al. (2010) Animated brain: A functional neuroimaging study on animacy experience. *Neuroimage* 53:291–302.
17. Heider F, Simmel M (1944) An experimental study of apparent behavior. *Am J Psychol* 57:243–259.
18. Grossman E, et al. (2000) Brain areas involved in perception of biological motion. *J Cogn Neurosci* 12:711–720.
19. Puce A, Allison T, Bentin S, Gore JC, McCarthy G (1998) Temporal cortex activation in humans viewing eye and mouth movements. *J Neurosci* 18:2188–2199.
20. Hoffman EA, Haxby JV (2000) Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nat Neurosci* 3:80–84.
21. Said CP, Moore CD, Engell AD, Todorov A, Haxby JV (2010) Distributed representations of dynamic facial expressions in the superior temporal sulcus. *J Vis* 10:11.
22. Pelphrey KA, Morris JP, McCarthy G (2004) Grasping the intentions of others: The perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *J Cogn Neurosci* 16:1706–1716.
23. Redcay E, Velnsky KR, Rowe ML (2016) Perceived communicative intent in gesture and language modulates the superior temporal sulcus. *Hum Brain Mapp* 37:3444–3461.
24. Koldewyn K, Weigelt S, Semmelmann K, Kanwisher N (2011) A region in the posterior superior temporal sulcus (pSTS) appears to be selectively engaged in the perception of social interactions. *J Vis* 11:630.
25. Wheatley T, Milleville SC, Martin A (2007) Understanding animate agents: Distinct roles for the social network and mirror system. *Psychol Sci* 18:469–474.
26. Saxe R, Brett M, Kanwisher N (2006) Divide and conquer: A defense of functional localizers. *Neuroimage* 30:1088–1096, discussion 1097–1099.
27. Fedorenko E, Hsieh P-J, Nieto-Castañón A, Whitfield-Gabrieli S, Kanwisher N (2010) New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *J Neurophysiol* 104:1177–1194.
28. Julian JB, Fedorenko E, Webster J, Kanwisher N (2012) An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage* 60:2357–2364.
29. 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.
30. 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.
31. Beauchamp MS, Argall BD, Bodurka J, Duyn JH, Martin A (2004) Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nat Neurosci* 7:1190–1192.
32. Kreifelts B, Ethofer T, Grodd W, Erb M, Wildgruber D (2007) Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *Neuroimage* 37:1445–1456.
33. Deen B, Koldewyn K, Kanwisher N, Saxe R (2015) Functional organization of social perception and cognition in the superior temporal sulcus. *Cereb Cortex* 25:4596–4609.
34. 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.
35. Dodell-Feder D, Koster-Hale J, Bedny M, Saxe R (2011) fMRI item analysis in a theory of mind task. *Neuroimage* 55:705–712.
36. Dufour N, et al. (2013) Similar brain activation during false belief tasks in a large sample of adults with and without autism. *PLoS One* 8:e75468.
37. O'Craven KM, Rosen BR, Kwong KK, Treisman A, Savoy RL (1997) Voluntary attention modulates fMRI activity in human MT-MST. *Neuron* 18:591–598.
38. Walbrin J, Downing PE, Koldewyn K (2017) The visual perception of interactive behaviour in the posterior superior temporal cortex. *J Vis* 17:990.
39. 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.
40. Gao T, Scholl BJ, McCarthy G (2012) Dissociating the detection of intentionality from animacy in the right posterior superior temporal sulcus. *J Neurosci* 32:14276–14280.
41. Blakemore S-J, et al. (2003) The detection of contingency and animacy from simple animations in the human brain. *Cereb Cortex* 13:837–844.
42. Wyk BC, Hudac CM, Carter EJ, Sobel DM, Pelphrey KA (2009) Action understanding in the superior temporal sulcus region. *Psychol Sci* 20:771–777.
43. Neri P, Luu JY, Levi DM (2006) Meaningful interactions can enhance visual discrimination of human agents. *Nat Neurosci* 9:1186–1192.
44. Rand DG, Nowak MA (2013) Human cooperation. *Trends Cogn Sci* 17:413–425.
45. de-Wit L, Alexander D, Ekroll V, Wagemans J (2016) Is neuroimaging measuring information in the brain? *Psychon Bull Rev* 23:1415–1428.
46. Fischer J, Mikhael JG, Tenenbaum JB, Kanwisher N (2016) Functional neuroanatomy of intuitive physical inference. *Proc Natl Acad Sci USA* 113:E5072–E5081.
47. Ullman T, Baker C, Macindoe O, Evans O (2009) Help or hinder: Bayesian models of social goal inference. Available at papers.nips.cc/paper/3747-help-or-hinder-bayesian-models-of-social-goal-inference. Accessed August 7, 2017.
48. Manera V, Becchio C, Schouten B, Bara BG, Verfaillie K (2011) Communicative interactions improve visual detection of biological motion. *PLoS One* 6:e14594.
49. Vanrie J, Verfaillie K (2004) Perception of biological motion: A stimulus set of human point-light actions. *Behav Res Methods Instrum Comput* 36:625–629.