Neuropsychologia 42 (2004) 1435–1446

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

A region of right posterior superior temporal sulcus responds to observed intentional actions

R. Saxe[a,*], D.-K. Xiao[c], G. Kovacs[c], D. I. Perrett[c], N. Kanwisher[a,b]

[a] *BCS, MIT, 3 Cambridge Center, 77 Massachussetts Avenue NE20-464 Cambridge, MA 02139 USA*
[b] *MIBR, MIT, Cambridge, MA USA*
[c] *Department of Psychology, St. Andrews University, Scotland, UK*

Received 24 December 2003; received in revised form 19 April 2004; accepted 21 April 2004

Abstract

Human adults and infants identify the actions of another agent based not only on its intrinsic perceptual features, but critically on the contingent relationship between its motion path and the environmental context [Trends Cogn. Sci. 7 (1995) 287; Cognition 72 (2003) 237]. Functional neuroimaging studies of the perception of agents and intentional actions, on the other hand, have mostly focussed on the perception of intrinsic cues to agency, like a face or articulated body motion (e.g. [J. Neurosci. 17 (1997) 4302; Neuroimage 8 (1998) 221; Trends Cogn. Sci. 4 (2000) 267; Nat. Neurosci. 3 (2000) 80; Neuroimage 13 (2001) 775; Proc. Natl. Acad. Sci. U.S.A. 98 (2001) 11656; Neuron 35 (2002) 1167; Neuron 34 (2002) 149, Neuroscience 15 (2003) 991; J. Neurosci. 23 (2003) 6819; Philos. Trans. R Soc. Lond. B. Biol. Sci. 358 (2003) 435]. Here we describe a region of the right posterior superior temporal sulcus that is sensitive not to articulated body motion per se, but to the relationship between the observed motion and the structure of the surrounding environment. From this and other aspects of the region's response, we hypothesize that this region is involved in the representation of observed intentional actions.

Keywords: Posterior superior temporal sulcus; Neuroimaging; Biological motion

1. Introduction

Both adults and infants can perceive complex social interactions, including emotions, goals, and personality traits, in the motions of simple geometrical shapes (Heider & Simmel, 1944). However, motion cues alone are poor predictors of such perceived agency. In particular, so-called "violations of Newtonian mechanics", such as a spontaneous start, stop, or change in acceleration, receive intentional explanations from fewer than half of observers (Gelman, Durgin, & Kaufman, 1995, see also Tremoulet & Feldman, 2000). Rather, subjects ascribe agenthood if the motion is contingent on the environment (e.g. apparently avoiding an obstacle) or on the motion of another object (e.g. two objects change acceleration [almost] simultaneously). Similarly, 9-month-old infants treat the inverted-U motion path of a circle as an intentional action (with the goal "to get to the other side") only if there is a block underneath the U; that is, if the U-shaped path can be interpreted as the circle "jumping over" an obstacle (Csibra et al., 1997).

Neuroimaging studies have tried to characterise the neural substrate for representing human action. Many of these studies have followed a different tradition in psychophysics and developmental psychology of investigating the perception of "biological motion" – that is, the characteristic articulated motion of chordate animal bodies (e.g. Vaina, Solomon, Chowdhury, Sinha, & Belliveau, 2001; Grossman & Blake, 2002; Beauchamp, Lee, Haxby, & Martin, 2002; Pelphrey, Mitchell, McKeown, Goldstein, Allison, & McCarthy, 2003) or body and face parts (e.g. Hoffman & Haxby, 2000; Hooker et al. 2003; Kilts et al. 2003; Pelphrey et al., 2003). Biological motion can also be perceived from the relative motion of just a few dots ("point-light walkers", Johansson, 1973); if the dots are spatially or temporally rearranged, the percept is destroyed. These neuroimaging studies suggest that one brain region, the posterior superior temporal sulcus, is particularly involved in the representation of biological motion.

Two sets of recent neuroimaging data suggest that the role of the posterior superior temporal sulcus (pSTS) may extend beyond a response to biological motion, to more abstract representations of intentional action. First, Castelli, Happe, Frith, & Frith, (2000) and Schultz, Grelotti, Klin, Kleinman, Van der Gaag, Marois, & Skudlarski, (2003)

---

[*] Corresponding author. Tel.: +1 617 425 3127; fax: +1 617 258 8654.
E-mail address: saxe@mit.edu (R. Saxe).

0028-3932/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2004.04.015

reported that a region of the pSTS showed a significantly higher response to animations of moving geometric shapes that depicted complex social interactions than to animations depicting inanimate motion. Second, using movies of human actors engaged in structured goal-directed actions (e.g. cleaning the kitchen), Zacks, Braver, Sheridan, Donaldson, Snyder, Ollinger, Buckner, & Raichle, (2001) found that activity in the pSTS was enhanced when the agent switched from one action to another, suggesting that this region encodes the goal-structure of actions. Both of these results are consistent with a role for a region of pSTS cortex in representing intentional action, and not just biological motion.

However, in all of these studies, a difference in the observed intentional action corresponded to a substantial difference in the visible motion of the agent. The motion paths of the geometric shapes in the Heider and Simmel—style animations were considerably more complex, and non-linear, in the social interaction stimuli than in the control condition. Similarly, the boundaries between intentional actions in the movies used by Zacks, Braver, Sheridan, Donaldson, Snyder, Ollinger, Buckner, & Raichle, (2001) correspond to changes in the biological motion of the actor. When Koski, Wohlschlager, Bekkering, Woods, Dubeau, Mazziotta, & Iacoboni, (2002) presented exactly the same biological (finger) motion either with a target dot or with no target dot, the pSTS response was not higher in the goal-directed motion condition. Finally, none of these studies tested whether the response of the pSTS, like the percept of intentionality, depends particularly on the contingency between the agent's motion and the environment.

To disentangle the role of the pSTS in representing the intentional action of an agent from its response to biological motion per se, we designed novel stimuli in which the visible biological motion was exactly matched, but the observed action varied. Subjects watched short movies in which a target person traversed a room, passing behind a central occluding object (e.g. bookcase). In the critical comparison, we varied how long the person remained invisible (occluded): either the person re-emerged immediately (Short Occlusion), with apparently no chance in velocity, or else the person remained hidden for a few seconds before re-emerging (Long Occlusion). The visible biological motion was exactly matched (Fig. 1). However, in the Long Occlusion movies, the action was different, in a way that was strongly contingent on the local environmental context—the walking person stopped just when he/she was behind the bookcase. (We consider two other interpretations of the difference between Long and Short Occlusion, both related to the perception of intentional action, in Section 4).

To foreshadow our results, a region of right posterior STS produced a reliable increased BOLD response to the Long (versus Short) Occlusion of a walking person. In subsequent manipulations, we further established that this differential activity did not reflect other differences between Long and Short Occlusion movies, such as the extended presence of an occluded person or a break in a smooth motion. The same region was also more strongly engaged by simple geometric

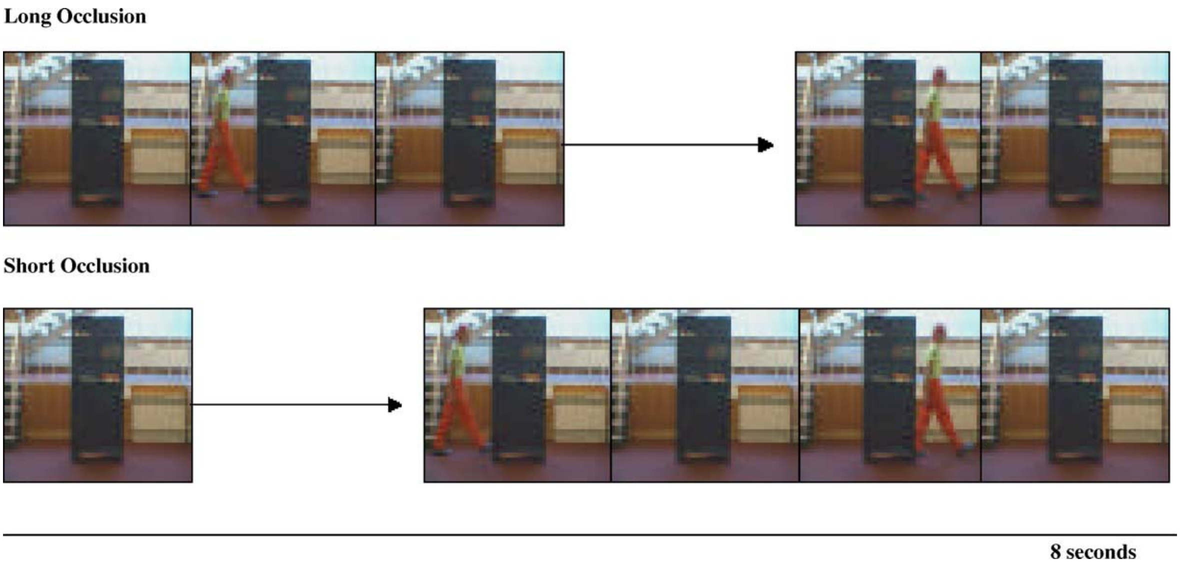

Fig. 1. Experiment 1: Schematic illustration of a movie from each of the two critical conditions, 'Short Occlusion', and 'Long Occlusion'. In each movie, a person walked across the scene, passing behind a large bookcase. Movies in the 'Short Occlusion' and 'Long Occlusion' conditions were constructed from identical movie fragments. The five components of each movie are illustrated: (1) an empty room before the actor emerged, (2) the actor walked to the bookcase, (3) an empty room while the actor was occluded, (4) the actor walked from the bookcase to offstage, and (5) the empty room again. 'Long Occlusion' and 'Short Occlusion' movies differed only in the relative duration of the empty scene in (1) and (3). In Short Occlusion movies, the room was empty only briefly while the actor passed behind the bookcase. The actor appeared not to stop at all behind the bookcase. In the Long Occlusion movies, the occlusion was extended in time (the actor appeared to stop), and the duration of the empty room at the start of the movie was brief. Each movie lasted 8 s; three movies from the same condition presented consecutively made up each block.

animations depicting goal-directed action than by control animations. We therefore hypothesised that this human brain region is involved in the perception of agents' intentional actions.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Sixteen healthy right-handed adults volunteered or participated in this study for payment. All subjects (Experiments 1–4) had normal or corrected to normal vision and gave informed consent to participate in the study.

2.1.2. Equipment

Subjects were scanned in the Siemens 1.5 T scanner at the MGH-NMR center in Charlestown, MA, using a head coil. Standard echoplanar imaging procedures were used ($\mathrm{TR} = 2\,\mathrm{s}$, $\mathrm{TE} = 30\,\mathrm{ms}$, flip angle 90°). Twenty 5 mm thick axial slices covered the whole brain, excluding the cerebellum.

2.1.3. Materials

Stimuli consisted of 8.0 s long movies, presented in colour using the Matlab PsychToolbox[57,58] in Quicktime format (QT. mex, 30 frames per second). Each movie began with an empty room, containing a large bookcase. Then, a human actor or target object crossed the room, from one side to the other (left to right in 50% of movies) leaving the room empty once again (Fig. 1).

Subjects saw 18 movies from each of three conditions (see Table 1). A fourth condition was not analysed here). No Occlusion movies began with 4 s of the empty room. Then an actor emerged, walked across the room for 3.5 s, passing in front of the bookcase, and then leaving the room empty for the final 0.5 s of the movie. The long period of empty room was placed at the beginning of the trial (rather than at the end) to ensure that subjects remained equally vigilant throughout the 8.0 s of movie in all three conditions. Short Occlusion movies used the same timing parameters as No Occlusion movies, except that the actor passed behind the bookcase, rendering him/her very briefly invisible. Finally, in the Long Occlusion movies, the room began empty for 1 s, and then the actor appeared and walked to behind the bookcase. Here the actor paused for 3 s, and then emerged, and walked off stage. These movies were created by digitally editing the Short Occlusion movies, so that the observed biological motion in these two conditions (Short and Long Occlusion) was identical.

Table 1
Summary of movie conditions from Experiment 1

Condition	Target object	Manner of motion	Occlusion
No Occlusion	Person	Walking	None
Short Occlusion	Person	Walking	Brief
Long Occlusion	Person	Walking	Long

Movies were presented in blocks of three movies from a single condition, producing blocks of 24.0 s. Each run lasted 464 s and contained four blocks of each condition, and five periods of fixation lasting 16.0 s. The order of conditions was counterbalanced within and across runs. Subjects viewed 2–5 runs of these stimuli. Each movie appeared between one and three times over the course of the experiment.

Subjects were instructed to press a button when the actor first appeared on the stage, and when the actor left the stage, in each movie. These responses were monitored during the scanning, to ensure that the subjects were awake and attending. However, due to technical limitations of QT. mex, the timing of the responses could not be recorded for off-line analysis.

2.1.4. FMRI data analysis

MRI data were analysed using SPM 99 (http://www.fil.ion.ucl.ag.uk/spm/spm99.html) and in-house software. Each subject's data were motion corrected and then normalized onto a common brain space (the MNI template). Date were then smoothed using a Gaussian filter (full width half maximum = 8 mm), and high-pass filtered during analysis. Every experiment used a blocked design, and was modelled using a boxcar regressor. Thus, the reported response to a given condition was the average response across a whole block of that condition (three consecutive movies).

We carried out a whole brain random-effect analysis ($P < 0.0001$ uncorrected), for two comparisons: No Occlusion versus Short Occlusion movies, and Short Occlusion versus Long Occlusion movies. Further analyses were conducted using a region of interest (ROI) approach. The region of interest was defined as all voxels that showed a significantly higher response during Long than Short Occlusion movies ($P < 0.0001$ uncorrected).

Within the ROI, the average percent signal change (PSC) relative to fixation baseline (PSC = 100 × raw BOLD magnitude for (condition − fixation)/raw BOLD magnitude for fixation) was calculated for each condition (averaging across all voxels in the ROI, all TRs in the block, and all blocks of the same condition). This calculation yielded a single grand average PSC value per ROI for each condition. These values were then entered in a repeated measures ANOVA. Because the data defining the ROIs were independent from the data used in the repeated measures statistics, Type I errors were drastically reduced.

3. Results and discussion

A single region, in right posterior superior temporal sulcus (pSTS), showed a significantly higher BOLD response to a simple walking motion when the target person stopped behind the bookcase (Long Occlusion > Short Occlusion

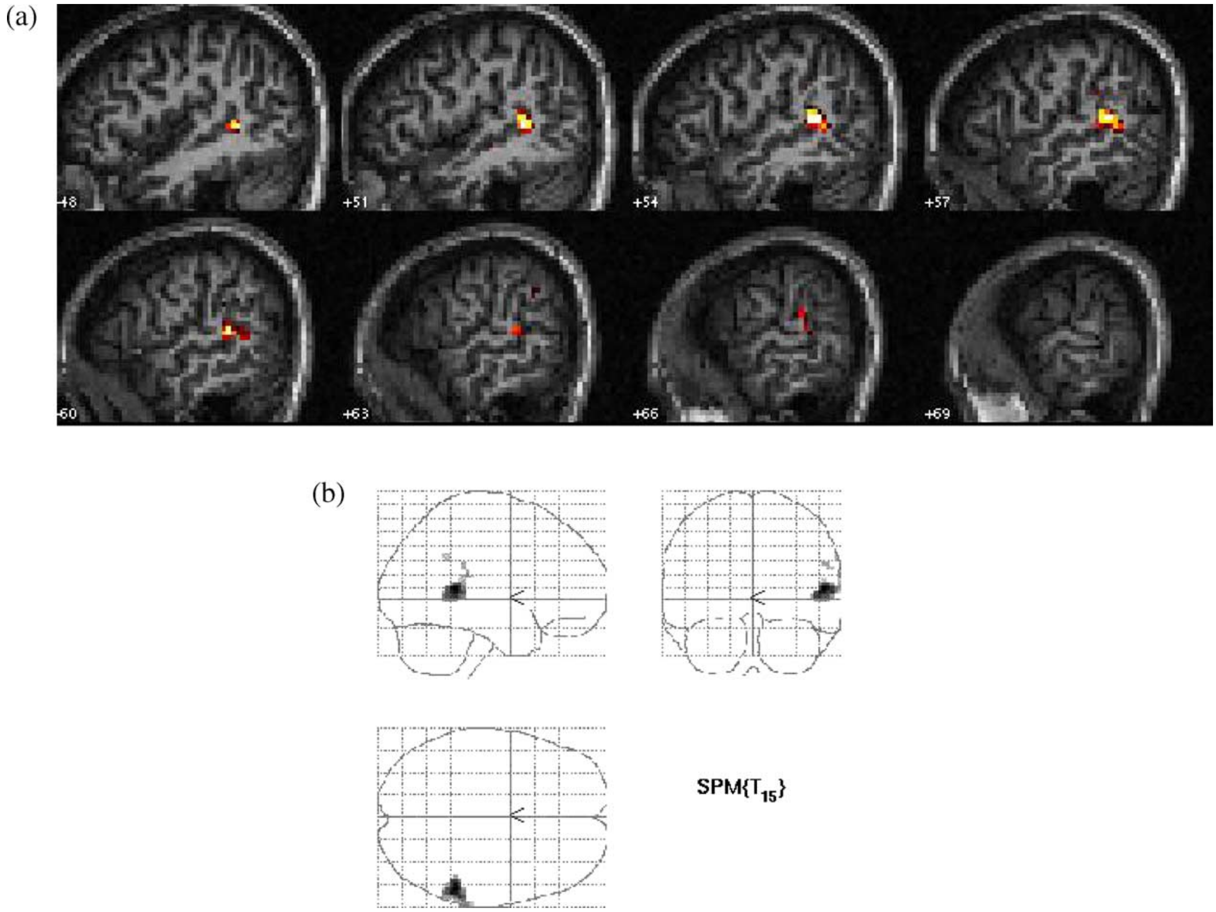

Fig. 2. Experiment 1: The pSTS-A. (a) Saggital slices show the group pSTS-A ROI overlaid on the anatomy of a single subject. The ROI was defined as all voxels that responded more during Long Occlusion than Short Occlusion movies, in Experiment 1 (Random effects analysis, $n = 16$, $P < 0.0001$). (b) Glass brain showing the same ROI. Slices covered the whole brain.

random effects $n = 16$, $P < 0.0001$ uncorrected, peak voxel: [54–42 9], $Z = 5.197$, Fig. 2; the differential activity in this region remained significant ($P < 0.05$) after a correction for multiple comparisons over the whole brain).[1] Voxels in this region were significantly more active during Long Occlusion compared to Short Occlusion movies in twelve out of sixteen individual subjects ($P < 0.0001$ uncorrected in each). There was no significant difference in the response of the pSTS region, or of any other region, to a walking person who was briefly occluded (Short Occlusion) versus never occluded (No Occlusion), suggesting that this pSTS region does not respond to the occlusion itself (cf. Baker, Keysers, Jellema, Wicker, & Perrett, 2001). For convenience of exposition, and to distinguish the current functional region of interest from the (much larger) entire STS, we will refer to this region as the pSTS-A (A for action).

The pSTS-A showed a robust response to the observed action—walking across the room—in all three conditions. An analysis of the BOLD time-course revealed no difference in the pSTS-A response to Long, Short and No Occlusion conditions during the portion of the display when the actor was visible, walking (Fig. 3a). The enhancement of the response in Long Occlusion movies occurred in the period when no person was visible on the screen (interaction $P < 0.001$, repeated measures ANOVA, $n = 16$, Fig. 3a). In the No Occlusion and Short Occlusion conditions, each movie began with a long pause before the target person entered the room. In these periods, with no person present, the response of the pSTS-A dropped (Fig. 3b: note the pronounced dips in the response to these two conditions). In the Long Occlusion movies, on the other hand, during the corresponding long period in each movie when a person was not visible in the room, the person was still present, standing behind the bookcase. In this case, the response of the pSTS-A remained high throughout the block (Fig. 3b: note the relatively flat profile of the response to this condition across three movies).

---

[1] Lesion studies have implicated parietal cortex, rather than superior temporal cortex, in perception of (e.g. Schenk & Zihl, 1997; Sirigu, Daprati, Pradat-Diehl, Franck, & Jeannerod, 1999) and reasoning about (Godbout & Doyon, 1995) intentional action. However, lesion studies have not yet focussed on patients' ability to perceive the structure of another person's intentional action. The current results predict that right superior temporal damage should impair subjects' ability to form, or at least to revise, representations of others' intentional actions.

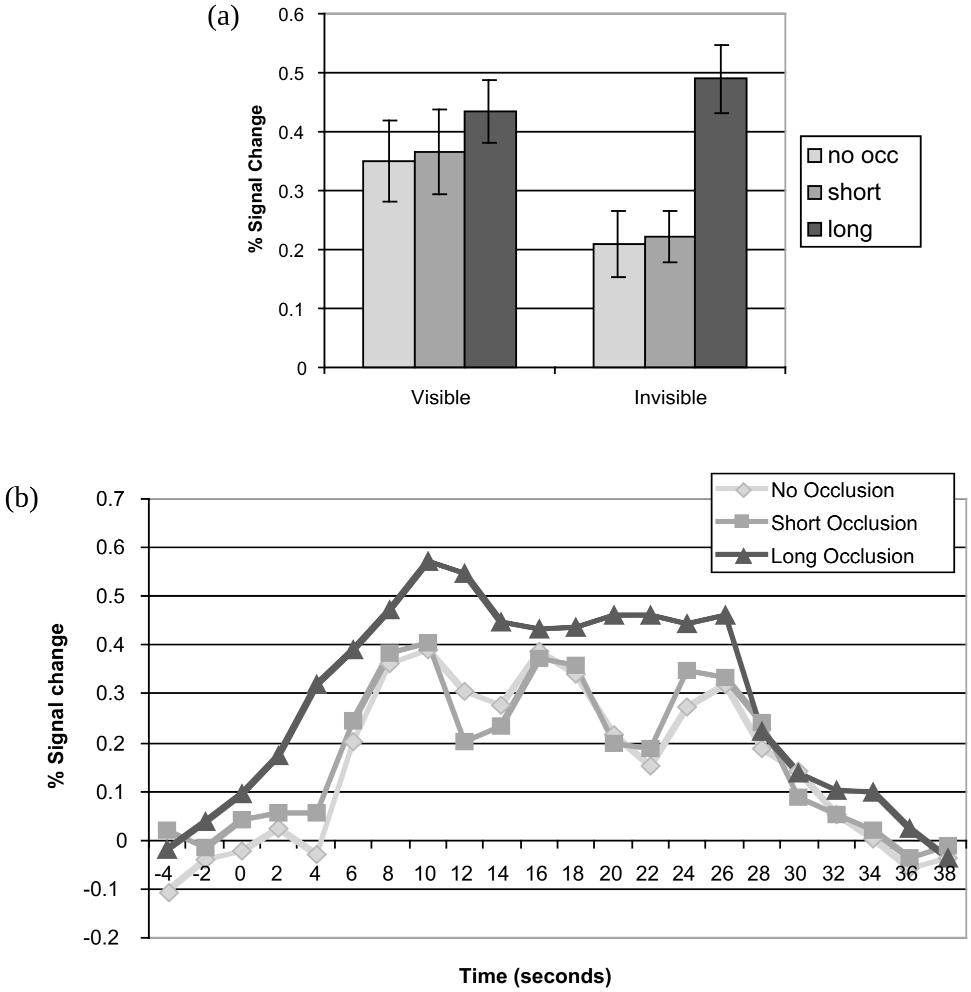

Fig. 3. Experiment 1: Visible versus Invisible. Response of the pSTS-A (group ROI) for phases of each trial when the actor was visible, or invisible but anticipated ($n = 16$). (a) The averaged response of the pSTS-A to the visible biological motion (walking, before being occluded) in the No Occlusion, Short Occlusion and Long Occlusion movies did not differ. When the person was invisible, the pSTS-A responded significantly more to the occluded person in the Long Occlusion movies than to the anticipated but not yet present person in the No Occlusion and Short Occlusion movies (both paired-samples t-test $P < 0.001$; interaction of Visibility by Occlusion Duration $P < 0.002$, repeated measures ANOVA). Note that the visual displays were identical across conditions, for both visible and invisible phases. Bars indicate standard error. (b) Time series of the BOLD response in the pSTS-A. Each block consisted of three consecutive movies. In both Short Occlusion and No Occlusion conditions, the response is noticeably lower to the empty room at the beginning of each movie, producing three troughs and three pronounced peaks in the response to the whole block. The response did not decrease when the subject saw an empty room for an equal duration, if there was a person occluded in the room (i.e. the Long Occlusion condition).

The posterior STS response to intentional action did appear to be right lateralised. Whole brain analyses revealed no effects in the left STS, even at a lower threshold of $P < 0.001$. To confirm this laterality, we created an ROI in left ST, the symmetric position in MNI coordinates ([−54 –42 9]) to that of the pSTS-A, with a radius of 6 mm. There was no hint of an effect of Long versus Short Occlusion in this left STS ROI.

4. Experiment 2

What does the increased response in the pSTS-A during Long Occlusion movies reflect? Since the movies in the Long Occlusion and Short Occlusion conditions were visually matched, the differential response of the pSTS-A to these two conditions cannot be explained by low-level physical differences in the stimuli, nor by a categorical response to biological motion or moving people. We hypothesised that pSTS-A was involved in the representation of agency (or intentional action). The pSTS-A may treat “stopping behind the bookcase” as an extra action in the Long Occlusion movies. But there remained at least two alternative features of the Long Occlusion movies that could be responsible for the pSTS-A response that we observed: (1) the extended presence of an occluded person or (2) the break in a smooth motion trajectory. We addressed these possibilities in two subsequent experiments.

Experiment 2 allowed us to replicate the results of Experiment 1 in an independent group of subjects, and to further characterise the function of the pSTS-A. In Experiment 2, we compared movies in which the intentional appearance of the action was disrupted by having the target person glide passively across the scene, with movies to a person walking. If the high response to the Long Occlusion movies in Experiment 1 reflect a response to an intentional action, we would predict no effect of occlusion duration for the gliding people. On the other hand, if the pSTS-A responded to a break in velocity or the presence of a person per se, then we predicted that the effect of occlusion for gliding people would be equivalent to that for walking people. In fact, we had initially believed that the pSTS-A responded to the presence of another person (a more abstract representation than the visual appearance of a person), and designed Experiment 2 to test this hypothesis.

4.1. Methods

4.1.1. Participants

Fourteen new healthy right-handed adults participated in this study.

4.1.2. Equipment

As in Experiment 1.

4.1.3. Materials

Stimulus parameters were similar to those used in Experiment 1. Subjects saw six movies in each of eight conditions (Table 2): *No Occlusion* movies were constructed as in the No Occlusion condition above. Two sets of No Occlusion movies were included to decrease the overall proportion of movies containing an occlusion event. *Walk Long Occlusion* and *Walk Short Occlusion* were designed to match the Long Occlusion and Short Occlusion conditions of Experiment 1, respectively. However, these movies used slightly lower velocities and different actors than original movies, to test the generality of brain response beyond a particular set of stimuli.

In the *Glide Short Occlusion* movies, a still cut-out of the actor in profile, facing the direction of motion, moved smoothly across the room with the same velocity and timing as the walking actor in the *Walk Short Occlusion* movies, and undergoing apparent occlusion (gradual occlusion and accretion) at the bookcase edges. *Glide Long Occlusion* movies were created to match the timing parameters of the *Walk Long Occlusion* condition, with the same gliding motion. Finally *Object Short Occlusion* and *Object Long Occlusion* movies were created, as in the Glide movies, but replacing the cut-out of a person with a non-human object of similar aspect ratio to the human targets.

Each movie appeared once in each run; each subject saw two runs of these stimuli. The blocked design used the same parameters as in Experiment 1. Subjects were instructed to press a button when the target actor/object first appeared on the stage, and when the target reappeared from occlusion (a slight variation from the task in Experiment 1). This task was designed to demand similar spatial attention to the empty room, at the beginning of the Short occlusion movies, and in the middle of the Long occlusion movies, since in each case the subject would be monitoring in order to make a response.

Table 2
Summary of movie conditions from Experiment 2

Condition	Target object	Manner of motion	Occlusion
No Occlusion (2×)	Person	Walking	None
Walk Short Occlusion	Person	Walking	Brief
Walk Long Occlusion	Person	Walking	Long
Glide Short Occlusion	Person	Gliding	Brief
Glide Long Occlusion	Person	Gliding	Long
Object Short Occlusion	Inanimate Object	Gliding	Brief
Object Long Occlusion	Inanimate Object	Gliding	Long

4.1.4. FMRI data analysis

Was conducted as in Experiment 1. We carried out a whole brain random-effect analysis ($P < 0.0001$ uncorrected) for Walk Long Occlusion versus Walk Short Occlusion. For consistency, all other analyses used the same independent group ROI used in Experiment 1.

5. Results and discussion

We replicated the basic finding of the pSTS-A in a second group of 14 subjects with new stimuli. The BOLD response in the pSTS-A was higher to Walk Long Occlusion (average percent signal change from fixation (PSC) =0.33), than to Walk Short Occlusion movies (PSC: 0.22, $P < 0.005$,

Fig. 4. Experiment 2: Walking versus Gliding. Percent signal change from fixation of the pSTS-A (independent group ROI from Experiment 1), to five new conditions in Experiment 2 ($n = 14$). Dark bars indicate a long occlusion; light grey bars indicate an uninterrupted walking path (short or no occlusion). Two different motions were observed: Walking and Gliding people (static cut-out human bodies moving passively across the room). The pSTS-A response was significantly higher for the Long occlusion only for Walking people ($P < 0.005$, interaction between motion (Walk versus Glide) and occlusion duration (Long versus Short) $P < 0.05$). Bars indicate standard error.

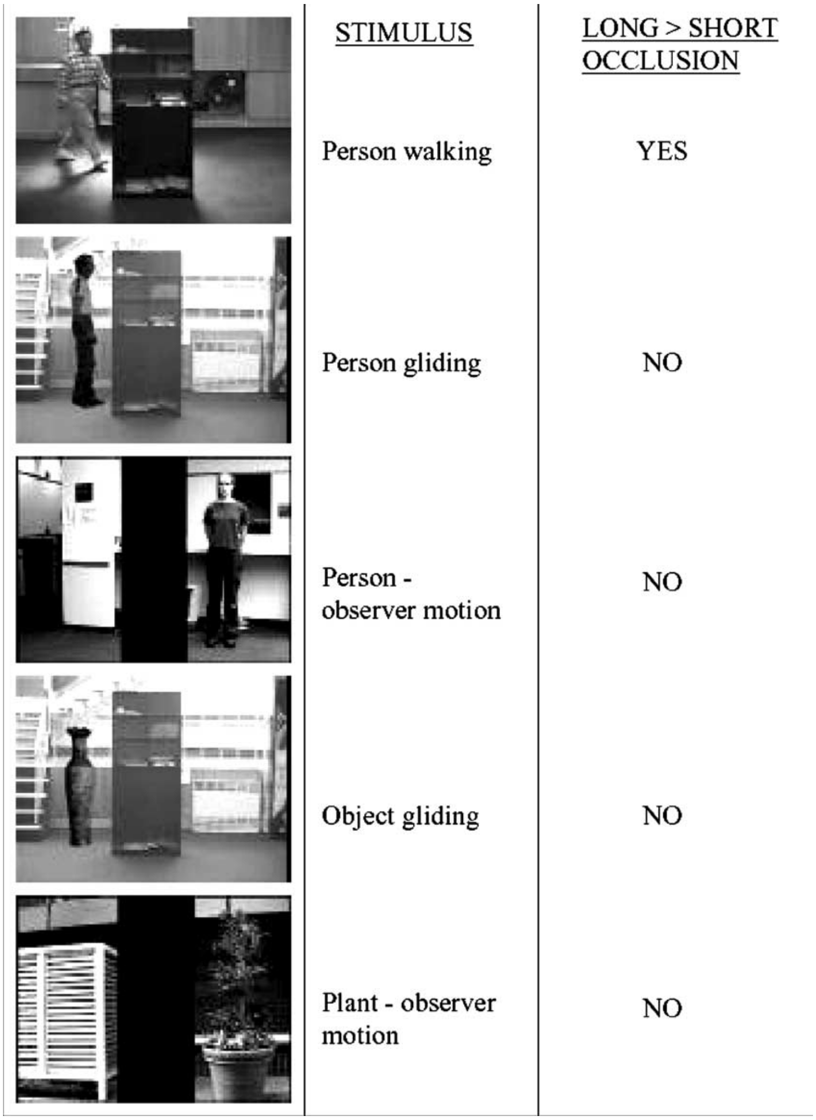

Fig. 5. Experiments 1–3: Summary of conditions. The pSTS-A responded significantly more to the Long than Short occlusion movies for walking people in all three groups of subjects. The pSTS-A did not respond more to the Long Occlusion of gliding people or objects (Experiment 2) or the observer motion of people or plants (Experiment 3).

paired samples *t*-test, Fig. 4). The results of the first experiment were also independently confirmed by an independent whole brain analysis of the activity in the second group of subjects (Random Effects $n = 14$, $P < 0.0001$ uncorrected, local maximum: [51 –42 18], $Z = 6.01$).

Long Occlusion movies with both Walking and Gliding people contained the same break in a smooth motion path, and an occluded person of equal duration. In spite of these similarities, the pSTS-A did not distinguish between Glide Long (PSC: 0.25) and Glide Short Occlusion movies (0.26; $P > 0.5$, paired-samples *t*-test, Figs. 4 and 5). The interaction in pSTS-A response between occlusion duration (Long versus Short) and motion type (Walk versus Glide) in a repeated measures ANOVA was significant ($P < 0.05$). The absence of any occlusion effect for the gliding people suggests that the occlusion effect observed for walking people did not reflect a response to any break in a smooth motion path, or the extended duration that a person was present in the scene. The pSTS-A did respond significantly more to the gliding people than to the gliding objects (repeated measures ANOVA, main effect of Person > Object, $P < 0.05$), but there was not a significant effect of Long versus Short Occlusion for the gliding objects (Fig. 6). Still, the gliding stimuli are very unnatural-looking, so we also included another more naturalistic control condition in Experiment 3.

6. Experiment 3

The advantage of the gliding motion used in Experiment 2 was that the stimulus was visually similar to the original movies showing walking people. The disadvantage of

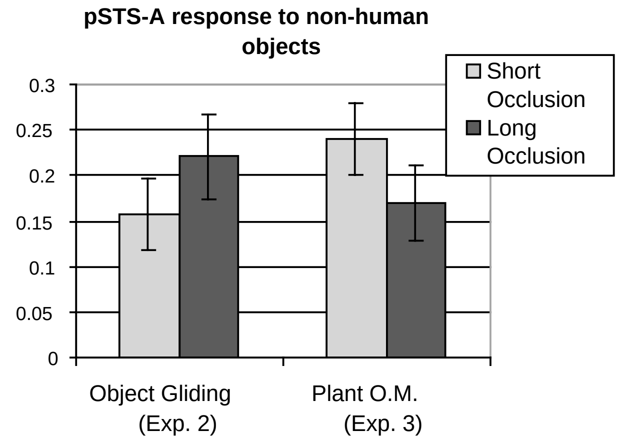

Fig. 6. Non-human objects (Experiments 2 and 3). The response of the pSTS-A to Long Occlusion was not significantly higher than the response to Short Occlusion in either of the pairs of conditions using non-human objects. In the Object Glide conditions, Long Occlusion elicited a non-significantly higher response than the Short Occlusion ($p = 0.1$), while in the Observer Motion conditions, the direction of the trend was reversed. Plant O.M.: Plant Observer Motion. Bars indicate standard error.

the gliding motion was that it looked unnatural. Therefore, we tested the occlusion effect for both people and inanimate objects in a third set of stimuli. In Experiment 3, we measured the response of the pSTS-A to natural passive occlusion events produced by apparent observer motion (i.e. the camera panned across the scene with a central occluding panel attached, like the view from a train window with a center divider, as the train drives out of the station). The person remained completely stationary within his/her environment throughout. In Observer Motion Short Occlusion movies, the camera was still at the beginning of the movie, and then panned continuously across the person. In the Observer Motion Long Occlusion movies, the camera panned until the person was completely occluded, and then paused. Again, if the pSTS-A response reflected a break in visible motion or the extended presence of a person, then we predicted that the Long Occlusion effect for the Observer Motion condition should be equivalent to that observed for the walking people in Experiments 1 and 2. If, on the other hand, the pSTS-A responded to an intentional action of “stopping behind the bookcase”, then we predicted no increase in the response for Long versus Short Occlusion in the Observer Motion movies.

Also, in Experiment 3, we sought to discriminate the pSTS-A from brain regions that respond simply to the visual appearance of a person, such as the nearby extrastriate body area (EBA, Downing, Jiang, Shuman, & Kanwisher, 2001).

6.1. Methods

6.1.1. Participants

Twelve new healthy right-handed adults participated in this study.

Table 3
Summary of movie conditions from Experiment 3

Condition	Target object	Manner of motion	Occlusion
New Walk Short Occlusion	Person	Walking	Brief
Observer Motion Short Occlusion	Person	Observer Motion	Brief
Observer Motion Long Occlusion	Person	Observer Motion	Long
Plant Short Occlusion	Plant	Observer Motion	Brief
Plant Long Occlusion	Plant	Observer Motion	Long

6.1.2. Equipment

As in Experiment 1.

6.1.3. Materials

Stimulus parameters were similar to those used in Experiment 1. Subjects saw 8 movies in each of five conditions (Table 3). In these movies, the scenes themselves did not include an occluding object. Instead, throughout all of the stimuli a black panel (1/4 the width of the movie) blocked the subject’s view of the centre of the scene, like a frame running down the centre of a train window. In *Observer Motion Short Occlusion* and *Observer Motion Long* Occlusion movies, the target was a person standing still, facing the camera. At the start of the film, the camera pointed off to one side (50% left, 50% right), showing an empty scene. Then, the camera panned past the target person so that the person moved smoothly across the screen but remained visibly stationary with respect to his/her environment. The motion of the camera paused at the centre of the scene, occluding the target, for 0.5 s (Short Occlusion) or 3 s (Long Occlusion). *Plant Short Occlusion* and *Plant Long Occlusion* were exactly matched to Observer Motion Short and Long movies, except that the person was replaced by a potted plant with approximately the same aspect ratio. Plants were chosen because they are familiar and definitely inanimate. We also included matched *New Walk Short Occlusion* movies, in order to establish the magnitude of response to intentional action in these subjects.

Movies were presented in blocks of three movies from a single condition, producing blocks of 24.0 s. The run lasted 368 s and contained two blocks of each condition, interleaved with five periods of fixation lasting 16.0 s. The order of conditions was counterbalanced within and across runs. Subjects were instructed as in group 2, except that a brief asterisk at the start of each movie cued the subject to the side of appearance of the target, to further reduce confounds with spatial attention.

Then, in two separate runs subjects in Experiment 3 saw the *Long Occlusion* and *Short Occlusion* movies from group 1. These movies were presented in alternating blocks of three movies from a single condition, interleaved with a fixation of 16.0 s. Each run lasted 228 s. Eleven subjects saw two runs of these stimuli; for one subject, the session had to be terminated before these runs.

6.1.4. FMRI data analysis

All ROI analyses used the same independent group ROI as in Experiments 1 and 2. We also carried out a whole brain random-effect analysis ($P < 0.0001$ uncorrected) for Person Observer Motion versus Plant Observer Motion.

7. Results and discussion

The increased response of the pSTS-A to the Long Occlusion of walking people was replicated in the new subjects in Experiment 3 (Long Occlusion PSC: 0.27; Short Occlusion PSC: 0.22, $P < 0.04$, paired-samples t-test, $n = 11$). As with the gliding people, the pSTS ROI did not respond more to long than short occlusion events caused by camera motion (Observer Motion Long Occlusion PSC: 0.26, $P > 0.5$, Observer Motion Short Occlusion PSC: 0.25 paired-samples t-test, $n = 12$, see also Fig. 5). The pSTS-A also did not respond more to the Long Occlusion than to the Short Occlusion of a Plant, in the Observer Motion condition (Fig. 6).

There was no main effect of Person > Plant in the Observer Motion condition, when the target (person or plant) was stationary with respect to its environment. A different response profile would be expected in the EBA (Downing, Jiang, Shuman, & Kanwisher, 2001), which responds to the visual appearance of a human body, whether moving or stationary, and indeed a different cluster of voxels in the vicinity of the EBA in right extrastriate cortex (peak: [57 –60 15]) was significantly more active (Random Effects analysis, $n = 12$, $P < 0.0001$ uncorrected) in the Observer Motion Short Occlusion condition for a stationary person (PSC: 0.25) than a stationary plant (PSC: 0.12). These conditions did not differ in the pSTS-A (repeated measures ANOVA interaction of Target [Person > Plant] by Region [putative EBA versus pSTS-A] $P < 0.006$). Thus, the EBA appears to make a categorical discrimination between people and inanimate objects for both moving and stationary stimuli, while the pSTS region does so only for moving stimuli.

8. General discussion

We identified a region of human right pSTS that appears to be involved in the representation of intentional action, referred to here as the pSTS-A (group peak from the current study: [54 –42 9]). In three groups of subjects, the response of the pSTS-A to the same visible motion increased significantly, if a person observed walking across a room appeared to stop for a few seconds behind an occluding object (the Long Occlusion condition). The differential response to the Long Occlusion condition was eliminated when the perception of intentional action was disrupted (Fig. 5). We also found that the pSTS-A response was higher for a moving human body than for a moving object, but did not distinguish between bodies and objects when the target was stationary with respect to its own environment. This response profile is consistent with a role for the pSTS-A in the representation of the observed actions of other agents.

To check our hypothesis, in a separate group of subjects (Saxe, 2004) we used a direct comparison of goal-directed versus non-goal-directed motion, using animations of two-dimensional geometric shapes. In the “goal directed” motion condition, one shape appeared to be an actor moving other inanimate objects to achieve a simple goal (e.g. putting a ball in a box). In the “rotation” condition, a display of the same shapes underwent fast rigid rotation, changing direction of rotation unpredictably. Consistent with our prediction, the pSTS-A responded robustly to goal-directed motion (PSC: 0.88) compared with rotation (PSC: 0.25, repeated measures t-test $P < 0.002$, $n = 8$, 3T). Converging results from the tightly matched conditions in Experiments 1–3, and the robust though imprecise effect in the animations (Saxe, 2004) led us to conclude that the pSTS-A is indeed involved in the representation of intentional action.

Next we consider alternative accounts of the pSTS-A.

8.1. Biological motion

Brain regions in the posterior superior temporal sulcus are most commonly implicated in the perception of biological motion (Vaina et al., 2001; Grossman & Blake, 2001, 2002; Beauchamp, Lee, Haxby, & Martin, 2002, 2003; Pelphrey et al., 2003). Did the pSTS-A response reported here simply reflect the perception of the characteristic articulated motion of human and chordate animal bodies?

We found no evidence that the pSTS-A responded preferentially to articulated body motion. First, the Long and Short Occlusion movies, the contrast used to defined the pSTS-A above, differed only in the duration of an invisible cessation of motion: the visible biological motion was identical in these two conditions. Second, the pSTS-A responded more to the same rigid gliding motion (with no articulation) if the target object was a person, than if it was an object. Third, animations of rigid geometric (non-human) objects elicited a high response in the pSTS-A, only when these motions constituted goal-directed actions. Previous reports that the pSTS is involved in the perception of articulated biological motion may be reconciled with our findings in one of two ways: (1) the current results apply to a particular region of interest, the pSTS-A, and are consistent with the claim that a different (perhaps more anterior) region of the pSTS has a particular role in the representation of articulated biological motion; but (2) many of the previous contrasts used to study biological motion perception also included an intentional action in the experimental condition (e.g. point-light or full-body walking, kicking and jumping) and no intentional action in the control condition (e.g. scrambled point-lights, tool motion, Vaina et al., 2001; Beauchamp et al., 2002; Grossman & Blake, 2002). Therefore, some of the differential response to biological motion observed in the pSTS in previous studies may actually reflect the perception of intentional action, and not just of articulated body motion.

8.2. *Unexpected target detection*

The second alternative hypothesis that must be addressed is that the pSTS-A is involved in detecting unexpected targets. According to this hypothesis, in the Long Occlusion movies the pSTS-A responded simply to the unexpected appearance (or non-reappearance) of the target person from behind the bookcase. Prior reports indicate that the response of a nearby region in the right temporo-parietal junction increases following detection of simple geometric targets, especially when spatially or temporally unpredicted (e.g. the 'invalid' trials in a standard Posner spatial cueing paradigm, Posner & Walker, 1984; Corbetta, Kincade, & Shulman, 2000). The anatomical vicinity of regions involved in social cognition and target detection may reflect a true functional relationship (Leslie, 2000), and will be investigated in future work.

However, it is already clear that a simple response to an unexpected target cannot account for the full profile of response in the pSTS-A. First, in four sets of control conditions there was no increased response in the pSTS-A when a target object made an unexpected stop and reappearance. This was true even in the relatively natural and familiar Person Observer Motion control condition. The unexpected stop was significant only when it was perceived as part of an intentional action. Second, the pSTS-A showed a higher response to a person walking across a room than to a stationary person with a camera pan (observer motion), neither of which involved an unexpected event. Third, the pSTS-A responded more to gliding people than to gliding objects, when the trajectories of the targets were identical. These latter two patterns are consistent with a role in analysis of human action, but are not predicted by a response to an unexpected target per se.

8.3. *Theory of mind*

A final alternative hypothesis is that the pSTS-A was engaged when subjects reasoned about the mental states that caused the observed action; that is, when the subject considered what the actor wanted and believed that made him/her stop behind the bookcase. Previous studies have found that a region in the nearby temporo-parietal junction is more active when subjects read stories that involved a character's mental states, than stories about physical transformations (e.g. Fletcher, Happe, Frith, Baker, Dolan, Frackowiak, & Frith, 1995; Gallagher, Happe, Brunswick, Fletcher, Frith, & Frith, 2000; Saxe & Kanwisher, 2003). This is an unlikely explanation of the pSTS-A, though. Saxe, Perrett, & Kanwisher (in preparation) found that the pSTS-A was functionally and anatomically distinct from the TPJ region, and did not respond at all when subjects read stories that elicit mental-state attribution.

In all, the pSTS-A appears to be involved in constructing abstract visual descriptions of another agent's intentional actions. The Long Occlusion movies elicit an enhanced and extended response in the pSTS-A compared with the Short Occlusion movies that cannot be explained in terms of biological motion perception, or unexpected target detection, or the known presence of a person, per se. Nevertheless, at least three distinct interpretations of this result are possible, each consistent with a role for the pSTS-A in the representation of agents' actions.

First, in the Long Occlusion movies, the walk-stop-walk action was aligned with the structure of the environment. Results from both adult psychophysics (e.g. Gelman et al., 1995; Durgin & Kaufman, 1995; Tremoulet & Feldman, 2000) and developmental psychology (Gergely & Csibra, 2003) suggest that the contingent relationship between the pattern of motion and the environmental context is a particularly important perceptual cue for perceiving an agent or an intentional action. For instance, a dot that travels across a screen in a V-shaped path, changing direction and speed only once, is more likely to be perceived as an agent if the change of direction appears contingent on the structure of the environment—e.g. as if the dot changed direction to avoid an obstacle. Similarly, in the Long Occlusion movie, a single change in the motion path (stopping) is correlated with the structure of the environment (the bookcase). The presence of this extra cue to intentional action may account for the enhanced response of the pSTS-A during the Long Occlusion movies.

A second possibility is that the pSTS-A is involved in representing the goal-structure of observed actions (Zacks, Braver, Sheridan, Donaldson, Snyder, Ollinger, Buckner, & Raichle, 2001). When subjects watched a movie of a sequence of actions, Zacks et al. found that the response of a region in the pSTS increased at the junctions between action components; i.e. when the target had completed one sub-goal. The Long Occlusion movies include a junction between two action components (i.e. walking to the bookcase, and then walking out of the room), while the Short Occlusion movies include only a single goal (i.e. walking across the room). The more complex goal structure of the Long Occlusion movies would also explain the increased response of the pSTS-A in this condition. Note though that this response to action structure is selective: it does not apply to motions of inanimate objects, or even to passively gliding people.

Finally, the pSTS-A may have been recruited for revision of the original action representation, when the action did not unfold as expected. At the beginning of the Long Occlusion movies, the target person did not slow down as he/she approached the bookcase, or look towards the bookcase. These behaviours are consistent with walking straight across the room, and are incongruent with a stop behind the bookcase. When the person did in fact stop, subjects may therefore have had to revise their action prediction/expectation. Consistent with this hypothesis for the pSTS-A, Pelphrey et al. (2003) reported that when subjects saw a person look left immediately after the presentation of a grating on the person's right, the response of a region of right pSTS was enhanced and extended, relative to the response of the same region to more predictable gaze shifts (towards the grating, or after a considerable delay). Also, Decety & Chaminade (2003) found a higher response in a nearby region when an actor recounted a negative personal experience using positive emotional facial and vocal expression compared with a congruent negative facial expression. Finally, Grezes, Frith, & Passingham (2004) found that increased activation in superior temporal sulcus when subjects watched an actor lifting a box that was heavier than he or she expected. The authors suggest that the actor may have made sudden postural adjustments to accommodate the unexpected weight, and that the increased STS response may reflect a corresponding adjustment in the subjects' action representation. In each of these three studies, and in our own, reformulation of expectations following an unexpected or incongruent action may have extended activity in a brain region involved in representing intentional actions, the pSTS-A.

In the context of previous findings, the current results suggest that sub-regions of the pSTS may play distinct roles in the detection of agency and the representation of intentional actions. At least one of these sub-regions, called here the pSTS-A, appears to be involved in the representation of agent's intentional actions and not just their biological motions, perhaps including sensitivity to the contingent relation between the action and the local environmental context.

Acknowledgements

This work was funded by grant NIHM 66696. Our thanks especially to David Badre, Susan Carey and Yuhong Jiang for comments and conversation, and to Amal Dorai, Nick Knouf, Christine Wang, Megan Daugherty, and Steve Lee for help with data collection and analysis.

References

Baker, C. I., Keysers, C., Jellema, T., Wicker, B., & Perrett, D. I. (2001). Neuronal representation of disappearing and hidden objects in temporal cortex of the macaque. *Experimental Brain Research*, *140*(3), 375–381.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*(1), 149–159.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2003). FMRI responses to video and point-light displays of moving humans and manipulable objects. *Journal of Cognitive Neuroscience*, *15*(7), 991–1001.

Castelli, F., Happe, F., Frith, U., & Frith, C. (2000). Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns. *Neuroimage*, *12*(3), 314–325.

Corbetta, M., Kincade, J. M., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, *3*(3), 292–297.

Decety, J., & Chaminade, T. (2003). Neural correlates of feeling sympathy. *Neuropsychologia*, *41*, 127–138.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*(5539), 2470–2473.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57*(2), 109–128.

Gallagher, H. L., Happe, F., Brunswick, N., Fletcher, P. C., Frith, U., & Frith, C. D. (2000). Reading the mind in cartoons and stories: An fMRI study of 'theory of mind' in verbal and nonverbal tasks. *Neuropsychologia*, *38*(1), 11–21.

Gelman, R., Durgin, F., Kaufman, L. (1995). Distinguishing between animates and inanimates: Not by motion alone. In Sperber, D., Premack, D. et al. (Eds.), *Causal cognition: A multidisciplinary debate*. Symposia of the Fyssen Foundation. New York, NY, US: Clarendon Press/Oxford University Press.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naive theory of rational action. *Trends Cognitive Science*, *7*(7), 287–292.

Godbout, L., & Doyon, J. (1995). Mental representation of knowledge following frontal-lobe or postrolandic lesions. *Neuropsychologia*, *33*(12), 1671–1696.

Grezes, J., Frith, C. D., & Passingham, R. E. (2004). Inferring false beliefs from the actions of oneself and others: An fMRI study. *Neuroimage*, *21*, 744–750.

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research*, *41*(10/11), 1475–1482.

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, *35*(6), 1167–1175.

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *Journal of American Psychology*, *57*, 243–259.

Hoffman, E. A., & Haxby, J. V. (2000). Distinct representations of eye gaze and identity in the distributed human neural system for face perception. *Nature Neuroscience*, *3*(1), 80–84.

Koski, L., Wohlschlager, A., Bekkering, H., Woods, R. P., Dubeau, M. C., Mazziotta, J. C., & Iacoboni, M. (2002). Modulation of motor and premotor activity during imitation of target-directed actions. *Cerebral Cortex*, *12*, 847–855.

Leslie, A. (2000). 'Theory of Mind' as a mechanism of selective attention. In: M. Gazzaniga (Ed.), *The new cognitive neurosciences*. Cambridge, MA: MIT Press, pp. 1235–1247.

Pelphrey, K. A., Mitchell, T. V., McKeown, M. J., Goldstein, J., Allison, T., & McCarthy, G. (2003). Brain activity evoked by the perception of human walking: Controlling for meaningful coherent motion. *Journal of Neuroscience*, *23*(17), 6819–6825.

Posner, M. I., & Walker, J. A. et al. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience*, *4*(7), 1863–1874.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in theory of mind. *Neuroimage*, *9*(4), 1835–1842.

Saxe, R. (2004). What fMRI can tell us about Theory of Mind. PhD thesis, MIT.

Saxe, R., Perrett, D. I., & Kanwisher, N. (in preparation). Distinct representations of bodies, actions and thoughts in posterior superior temporal sulcus.

Schenk, T., & Zihl, J. (1997). Visual motion perception after brain damage: II. Deficits in form-from-motion perception. *Neuropsychologia*, *35*(9), 1299–1310.

Schultz, R. T., Grelotti, D. J., Klin, A., Kleinman, J., Van der Gaag, C., Marois, R., & Skudlarski, P. (2003). The role of the fusiform face area in social cognition: Implications for the pathobiology of autism. *Philosophical Transactions of Royal Society of London, Series B: Biological Sciences*, *358*(1430), 415–427.

Sirigu, A., Daprati, E., Pradat-Diehl, P., Franck, N., & Jeannerod, M. (1999). Perception of self-generated movement following left parietal lesion. *Brain*, *122*, 1867–1874.

Tremoulet, P. D., & Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, *29*(8), 943–951.

Vaina, L. M., Solomon, J., Chowdhury, S., Sinha, P., & Belliveau, J. W. (2001). Functional neuroanatomy of biological motion perception in humans. *Proceedings of the National Academy of Sciences United States of America*, *98*(20), 11656–11661.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*(6), 651–655.