

From social-signal detection to higher social cognition: an fMRI approach

Motoaki Sugiura,^{1,2} Yukihiro Yomogida,^{3,4} Yoko Mano,⁵ Yuko Sassa,¹ Toshimune Kambara,^{3,4,6} Atsushi Sekiguchi,^{1,7} and Ryuta Kawashima¹

¹Institute of Development, Aging and Cancer, ²International Research Institute of Disaster Science, Tohoku University, Sendai 980-8575, ³Japan Society for the Promotion of Science, Tokyo 102-0083, ⁴Brain Science Institute, Tamagawa University, Tokyo 194-8610, Japan, ⁵Department of Psychology, Northwestern University, Evanston, IL 60208-2710, USA, ⁶Faculty of Environment and Information Studies, Keio University, Fujisawa 252-0882, and ⁷Tohoku Medical Megabank Organization, Tohoku University, Sendai 980-8573, Japan

Implicit or automatic detection of social signals, which discriminate animate, intentional objects in the environment, is essential for higher social cognition and its development. Using functional magnetic resonance imaging, we identified the neural substrate of detecting simple visual social signals and examined its functional link with the mechanism of inferring another's mental state. Healthy participants were presented with the eye-gaze shift (EG) and self-propelling motion (SP) under both implicit and explicit task conditions. They also performed a social role-playing game in which mental inference (MI) was implicitly prompted during the presentation of faces (implicit MI). Implicit detection of EG and SP activated the posterior middle temporal gyrus (pMTG) bilaterally, whereas the right posterior superior temporal sulcus was activated during the explicit conditions. We revealed that the individual variation in neural response in the right pMTG during implicit eye-gaze detection explains the individual tendency to recruit the regions implicated in mental-state inference (medial prefrontal cortex, temporal pole and striatum) during the implicit MI task. Our results suggest that the implicit detection of visual social signals involves the pMTG and underlies the development of higher social cognition.

Keywords: animacy; intention; fMRI; mentalizing; theory of mind

INTRODUCTION

However sophisticated human cognitive abilities for social survival may be, they must originate from the detection of rather simple perceptual signals that discriminate animate, intentional objects from the rest of the environment. Psychologists have been looking for the essential features of such social signals from a developmental viewpoint. One stream of research identified eye gaze as one of the simplest examples: infants preferentially view faces, particularly eyes (Maurer and Salapatek, 1976), and have the ability to discriminate between directed and averted eye gaze (Vecera and Johnson, 1995). Another stream focused on the nature of object motion. The ability to discriminate the movement of objects that cannot be explained by an external energy source, namely self-propelling motion, was first identified in preschool children (Dasser *et al.*, 1989) and then in infants (Rochat *et al.*, 1997). The cognitive modules that process these two exemplars of social signals are hypothesized to be the bases for the development of the theory of mind (ToM), the ability to make inferences of another's mental state (Premack, 1990; Baron-Cohen, 1994). This developmental hypothesis is supported by the observation of an impaired ability to detect these types of social signals in children with autism (Senju *et al.*, 2004; Rutherford *et al.*, 2006), which is a developmental disorder characterized by impaired social cognitive abilities, including the ToM.

In this functional magnetic resonance imaging (fMRI) study using healthy human subjects, we aimed to identify the neural link between the detection of simple visual social signals and higher social-cognitive faculties. Here, we dissociate implicit (i.e. automatic, not necessarily unconscious) and explicit (i.e. required by a 'task') detection. We were

particularly interested in implicit detection because social-signal detection is not explicitly required in infants undergoing social-cognitive development or in adults in daily life.

We first identified cortical regions responsive to social signals in implicit and explicit detection tasks. We used the two distinct exemplars of social signals, eye-gaze shift (EG) and self-propelling motion (SP) of a geometric object. Based on findings in single-unit recording studies of monkeys and functional imaging studies of humans on the perception of social signals (Allison *et al.*, 2000), the cortical region of the initial stage for processing visual social signals was previously considered to be the cortex surrounding the posterior part of the superior temporal sulcus (pSTS). The pSTS has accordingly been hypothesized to be the primary provider of relevant information to higher social-cognitive systems, such as the ToM (Frith and Frith, 2003) and the mirror neuron system (MNS) (Nelissen *et al.*, 2011), implicated in action understanding, imitation and communication (Rizzolatti and Craighero, 2004; Iacoboni, 2005). Recently, however, the findings of several functional imaging studies appear to suggest that the human pSTS is located at an advanced, rather than 'initial' stage, of social-signal processing; activation in this region is modulated by the congruency of presented gaze-shift or action with the prediction (Pelphrey *et al.*, 2004a,b) and task demand (Jastorff and Orban, 2009). Instead, recent neuroimaging evidence suggests that a region in the lateral occipito-temporal cortex is more compatible with implicit detection. This region responds to human action presented using point-light displays (Grossman *et al.*, 2000), and activation is not modulated with the prediction or task demand (Pelphrey *et al.*, 2004b; Jastorff and Orban, 2009). A similar region is activated during viewing of eye gaze (Hoffman and Haxby, 2000) or the interaction of two objects (Castelli *et al.*, 2000) under no task demand. It was recently suggested that functionally homologous regions for not only the human pSTS but also these lateral occipito-temporal cortices are located in the superior temporal sulcus (STS) of monkeys (Tsao *et al.*, 2003; Jastorff *et al.*, 2012). We therefore expected that social-signal sensitive regions

Received 13 November 2012; Revised 8 April 2013; Accepted 18 July 2013

Advance Access publication 24 July 2013

We thank Ai Fukushima for her support in preparing the visual stimuli.

This study was supported by MEXT KAKENHI Grant Number 25560347 and JSPS KAKENHI Grant Number 23119702.

Correspondence should be addressed to Motoaki Sugiura, Department of Functional Brain Imaging, Institute of Development, Aging and Cancer, Tohoku University, Seiryomachi 4-1, Aoba-ku, Sendai 980-8575, Japan. E-mail: motoaki@idac.tohoku.ac.jp

would be identified in the lateral occipito-temporal cortex for the implicit detection and for the pSTS during explicit detection.

We then examined whether the functioning of such social-signal-sensitive regions is linked to a higher social-cognitive faculty. We exploited intersubject variation in neural response to demonstrate this link based on the following reasoning: if this developmental hypothesis holds, individual sensitivity in the (upstream) implicit detection of social signals explains the tendency to implicitly adopt a (downstream) higher social-cognitive process. As a task that allows such an individual difference to be fully expressed, we devised a naturalistic social role-playing game in which mental inference (MI) was implicitly required. We expected that the pattern of individual variation in neural response during implicit social-signal detection would be exhibited in certain cortical regions involved in the mental-state inference during the task. Early neuroimaging studies typically pointed to the medial prefrontal cortex (MPFC), temporal pole (TP) and pSTS, including the adjacent temporoparietal junction (TPJ), as common neural bases of this process (Frith and Frith, 2003). As growing acceptance of the complexity or multiplicity of the ToM concept, recent studies have adopted a variety of cognitive tasks and suggested additional regions, such as the posterior cingulate cortex (PCC), including the adjacent precuneus, amygdala and striatum (Carrington and Bailey, 2009; Abu-Akel and Shamay-Tsoory, 2011).

MATERIALS AND METHODS

Participants

A total of 45 healthy right-handed (Oldfield, 1971) volunteers (aged 19–28 years, 29 males) participated. Written informed consent was obtained from each subject. No subject had a history of neurological or psychiatric illness. The experimental protocol was approved by the ethics committee of the Tohoku University School of Medicine.

Experimental overview

Each subject participated in five fMRI sessions. Implicit and explicit detection tasks were administered in separate sessions for two types of social signals, EG and SP. An implicit MI session was administered to examine the link from the social-signal-sensitive regions to a higher social-cognitive process. Each subject participated in these sessions in the following order: implicit SP, implicit EG, implicit MI, explicit SP and explicit EG (lasted 498 s, 420 s, 1671 s, 498 s and 420 s, respectively). The session order was fixed (i) to make the detection of social signals incidental in the implicit task sessions (i.e. the implicit sessions preceded the explicit sessions) and (ii) to avoid within-session intersubject variability caused by counter-balancing of the session order (i.e. time/order effect). We did not perform direct between-session comparisons due to this fixed session order.

Subjects comfortably laid on the bed of the MRI scanner with their heads fixed in the head coil using elastic blocks. The visual stimulus was back-projected onto a semi-lucent screen behind the head coil and was viewed via a mirror. The size of the visual stimuli in all sessions subtended a visual angle of less than 5°.

Social-signal-detection sessions

All stimuli were created by the authors and prescreened (see [Supplementary Methods](#) for details).

In the EG sessions (Figure 1A), the sequence of two line drawing pictures (1 s presentation for each) for EG (E; 35 trials) and daily object (O; 35 trials) was presented in a random order with jittering of the sequence onset from 3 s to 12 s. Almost half of the E or O trials showed rightward rotations (around the vertical axis), and the other half leftward. We were not concerned about the difference in visual complexity between E and O pictures because we planned to examine the common

activation of the EG and SP sessions, and the activation due to the complexity should have been higher for the O as compared with the E stimuli.

In the SP sessions (Figure 1B), viewing of self-propelling motion (S) was contrasted against that of externally moved motion (X). Each film clip lasted 9 s, during which changes in the direction/speed of the triangle's motion occurred 14 times. The movement of the triangle was the alternation of a constant linear procession and a direction change near the frame border. While maintaining the equality of the average speed and the number of direction changes between the two conditions, two critical motion characteristics (Tremoulet and Feldman, 2000) differentiated them: (i) the principal axis of the triangle was aligned to the direction of movement in S but not in X, and (ii) the direction change following the Newtonian law (i.e. always hitting the frame at a reasonable angle of reflection) was present in X but not in S. There were 15 S, 15 X, and in addition, 10 vague (V) film clips in which the characteristics of the motion were in between that of the S and X film clips. The V film clips were included to prevent the subjects from becoming aware of the two distinct categories of motion characteristics during the implicit SP session.

During the implicit sessions (i.e. both EG and SP), the subject was required to detect the color change of part of a picture and to press a button with the right index finger as quickly as possible. Therefore, the social signals were incidental with respect to the task demand. During the EG session (Figure 1C), a conspicuous part of the second picture was colored in cyan in five E and five O trials. During the SP session (Figure 1D), the color of the triangle was temporarily (0.5 s) changed to yellow at an unpredictable time in each clip. Participants were not informed of the difference in the nature of the motion before the session. After the session, subjects were first asked to report any impressions about the motion, and they were then asked whether they were aware of the difference in motion characteristics after being informed of the two types.

During the explicit EG and SP sessions, the subject was explicitly required to detect social signals. During the EG session, the subject was asked to judge the direction of eye gaze (E) or of an object (O) changed, and to press one of two buttons with the right index or middle finger when the rotation was rightward or leftward, respectively. During the SP session, the subjects were required to press one of two buttons with their right index or middle finger when the motion appeared to be self-propelling (S) or externally moved (X), respectively.

Implicit MI session

Each subject played the role of a detective in a role-playing game that involved 24 episodes of trivial crime cases (e.g. a destroyed wall, broken cup, etc.). Each episode (Figure 1E) started with the Introduction period (7 s), during which a case was explained. In the subsequent Investigation period (15 s), three sets of items were presented sequentially; each set (5 s) included a picture of a suspect (a bust), a statement from the suspect and a picture of a relevant tool or object (e.g. a hammer, a marker pen). It was then followed by the Choice period (10 s), in which subjects were required to choose the most likely suspect. Periods for the disclosure of the criminal and feedback on the subject's choice (Feedback; 12 s) and an eye-fixated rest period (Rest; 12 s) followed (see [Supplementary Methods](#) and Figure 1 for details).

In each episode, two suspects appeared potentially guilty and one appeared very unlikely in terms of the tool/object that accompanied them. We could, therefore, manipulate the feedback (subject's success or failure in making an accusation) post hoc by telling the subject that of the two potential suspects, the one chosen (or not chosen) was the criminal. The success rate for each subject was manipulated to 50% to adjust for emotional reaction and motivation. We confirmed that none

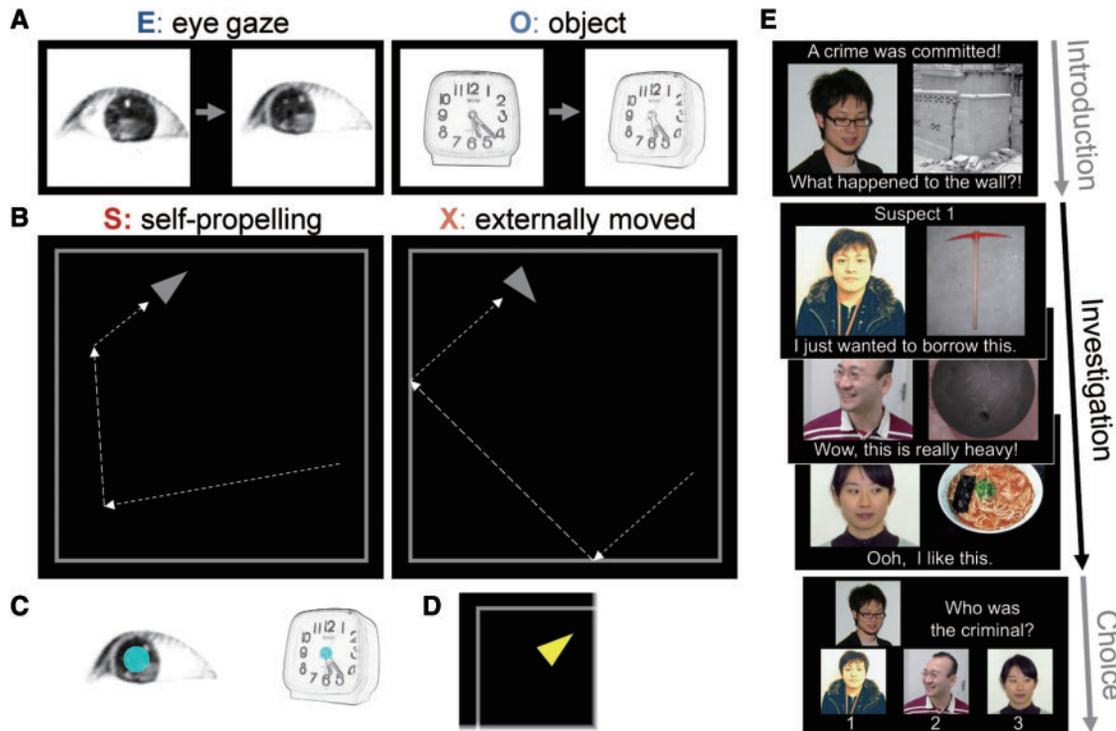


Fig. 1 Examples of stimuli. (A) For the EG sessions, a change in the direction of the eye gaze (E) or object (O) was presented using a sequence of two slightly different line drawings. (B) For the SP sessions, a short movie clip was presented showing a triangle moving around in a square frame that gave the impression of self-propelling (S) motion or of being externally moved (X). The principal axis of the triangle was aligned to the direction of movement in S, and the direction change followed the Newtonian law in X (Tremoulet and Feldman, 2000). The color of part of the pictures was occasionally changed in the implicit EG (C) and SP (D) sessions. (E) For the implicit MI session, following a client’s description of a case (Introduction period), three suspects were sequentially introduced with a set of pictures, including a face and a potentially relevant tool or object, as well as the suspect’s statement to the detective (Investigation period). Then the subject was asked to specify one suspect (Choice period). Although the decision could logically be made by judging whether the suspect’s tool could feasibly cause the problem in the case, subjects were expected to automatically infer the mental state of each suspect from the face and statement, to varying degrees.

of the subjects was aware of this manipulation in the interview after the session. The subject’s choice of either potential suspect was counted as a correct response, irrespective of the feedback given.

The target of the analysis was the individual variance in cortical activation during the Investigation period. Theoretically, the most efficient way to comply with the explicit task requirement during this period was to simply think about a tool/object that was most likely to cause the consequence. However, while reasoning whether each presented person committed the crime, each subject was expected to infer the mental states of the suspect to varying degrees. We assumed that the process is in part triggered by the implicit eye-gaze detection and that variation in the process is in part caused by the detection.

We prepared a control condition for this analysis using a concept different from conventional subtraction analysis. Because we would apply a multivariate exploratory analysis, we did not need to subtract out the cognitive processes of no interest (e.g. lower visual and linguistic processes). On the other hand, it was important for the control condition to include a minimal degree of intersubject variability in the activation of interest. The resting state was not a good control condition in this respect. Some brain regions are spontaneously active during the resting state (i.e. default system), and they largely overlap with those involved in social-cognitive processes (Schilbach *et al.*, 2008). Therefore, we included an attention-demanding reaction-time task as a control condition in which such activation was expected to be minimal. During the task (13 s), a digit (1, 2, or 3) was randomly presented every 1.2 s, and the subjects were required to press the corresponding button as quickly as possible. This control condition was included between the episodes as well as at the beginning and the end of each session.

fMRI data acquisition and preprocessing

For all sessions, 48 transaxial gradient-echo images (echo time = 30 ms, flip angle = 85°, slice thickness = 2.5 mm, slice gap = 0.5 mm, FOV = 192 mm, matrix = 64 × 64, voxel size = 3 × 3 × 3 mm) covering the whole cerebrum were acquired at a repetition time of 3 s, using an echo planar sequence and a Philips Achieva (3T) MR scanner.

The following preprocessing procedures as well as subsequent statistical analysis were performed using Statistical Parametric Mapping (SPM5) software (Wellcome Department of Imaging Neuroscience, London, UK) and MATLAB: the adjustment of acquisition timing across slices, correction for head motion, spatial normalization using the EPI-MNI template and smoothing, using a Gaussian kernel with a full-width at half-maximum of 8 mm.

Data selection

Subjects for the data analyses were selected based on the following criteria: a detection or judgment accuracy of more than 80% in the implicit EG, implicit SP and explicit EG sessions; judgment accuracy of more than 60% in the explicit SP session; and head-motion of less than 3.5 mm for any of the sessions analyzed. A lower judgment-accuracy threshold was adopted for the explicit SP session because in this session only, judgment was based on subjective feeling rather than on explicit perceptual cues.

Estimation of activation (first-level analysis)

For the social-signal-detection sessions, expected signal changes were modeled for the E, O and target (color change) trials for the implicit EG session, E and O trials for the explicit EG session and S, X and V

trials for both the implicit and explicit SP sessions. For the implicit MI session, the six periods (Introduction, Investigation, Choice, Feedback, Rest and Control) were modeled. We did not exclude trials with an incorrect response, because we were primarily interested in the relationship between the stimulus and neural response, rather than the processes reflecting behavioral response. A model of the expected signal change was constructed using the hemodynamic response function provided by SPM5. A high-pass filter with a cutoff period of 128 s (default setting of SPM5) for the social-signal-detection sessions and that with a 138-s cutoff period for the implicit MI session (twice the task cycle; i.e. episode and control task) were used to eliminate the artifactual low-frequency trend.

Identification of social-signal-responsive regions

Four social-signal-detection sessions were included in the second-level analysis adopting one-sample *t*-tests. Thirty-one subjects (aged 19–23 years, 18 males) satisfied the data selection criteria for the four sessions. To identify regions responsive to social signals for the implicit tasks, the contrast E – O in the implicit EG and the contrast in S – X in the implicit SP were summed (i.e. [E – O] + [S – X]) and tested at the statistical threshold of $P < 0.001$. To ensure responses to both types of stimuli, the test was restricted (i.e. masked) to the areas where both contrasts (i.e. E – O and S – X) were significant at $P < 0.05$ (uncorrected). When we had no a priori hypothesis, a correction ($P < 0.05$) for multiple comparisons due to the voxel-wise search was performed using cluster size (mask did not affect the correction). Social-signal-sensitive regions for the explicit tasks were similarly identified using the same contrasts in the explicit EG and SP sessions. To examine the profile of each activation peak, the contrasts E – O and S – X in the sessions not included in the voxel-by-voxel analysis were also tested post hoc ($P < 0.05$, uncorrected).

Here, we could adopt conjunction analysis (Nichols *et al.*, 2005) to identify overlap of activated areas between two contrasts. We, however, considered this option too conservative (Friston *et al.*, 2005) for our neural response of interest, for which we did not expect high statistical sensitivity due to the implicit task requirement and large individual variability. See [Supplementary Methods](#) for further details.

Extraction of individual variation patterns

To extract the individual variation patterns of interest from the identified social-signal-responsive regions, we applied a principal component analysis (PCA) to the activation estimates in the two EG sessions (i.e. implicit and explicit) and the implicit MI session. Although theoretically the data from the implicit EG session alone were sufficient to represent the individual variation pattern, the data would also include several variance components of no interest. By applying the PCA to the implicit and explicit EG and the implicit MI sessions, we expected the variance component of interest to be isolated; the component should be identified by its existence in the sessions in which eye-gaze detection was implicit (implicit EG and MI) but not in the explicit EG session. Variation of no interest, such as session-common (e.g. subject-specific physiological factors), O-stimulus-dependent and session-unique (e.g., arousal) variations should be eliminated as other components.

Thirty-five subjects (aged 19–28 years, 20 males) satisfied the data selection criteria for the three sessions; 28 of the subjects overlapped with those included in the identification of social-signal-responsive regions. Three identified social-signal-responsive regions (see Results) were defined as regions of interest (ROIs). The contrasted parameter estimates E – O for both EG sessions and Investigation – Control for the implicit MI session at each peak voxel were used. Therefore, three ROIs and three sessions constituted the multivariate response (i.e. nine variables) and the 35 subjects constituted the

observations, resulting in a matrix of 35×9 . The PCA was performed using a singular value decomposition of the mean-corrected matrix in MATLAB. Each obtained principal component (PC; eigenvariate) represented a variation component (see [Supplementary Methods](#) for details). We first selected a small number of significant PCs based on their contributions in scree plot (Cattell, 1966). We then identified the PC in which the loading (eigenvector) was large on the implicit EG and MI sessions but small on the explicit EG session.

Functional-link analysis

Then, as a second-level analysis, using SPM5, we searched for cortical regions that were remote from ROIs and exhibited individual variation patterns identified in each PC during each session. We performed the voxel-by-voxel regression analysis using PC score (eigenscore), which represented the individual variation. Specifically, our PC score of interest, which was expected to represent individual variation in neural responsiveness to eye gaze in the implicit EG session, would also be expressed in remote ToM-related regions in the implicit MI session. Scores were entered as a regressor of the multiple regression model; all the retained PCs were included to regress out the individual variation of no interest. The cortical networks showing positive or negative effects of our PC of interest were identified by testing for significant regressions using *t*-tests. The statistical threshold was $P < 0.001$ in height, corrected to $P < 0.05$ for multiple comparisons using cluster size.

Here, a significant effect (regression) in or close to the ROIs for the PCA should be cautiously interpreted because the source of the regressors was variation of activation in these regions, and therefore, the test could be circular. However, such a finding is still informative, as it tells us which ROI contributed to each PC.

RESULTS

Behavioral data

The judgment accuracy and reaction times are summarized separately for the identification of social-signal-responsive regions and the functional-link analysis in Tables 1 and 2, respectively, considering differences in the included subjects. Between-condition comparisons of these behavioral data suggest that they were unlikely to affect the fMRI findings (see [Supplementary Results](#)).

In the interview after the implicit SP session, only three subjects reported that they were aware of the existence of different types of motion characteristics, even after being debriefed about the existence of such differences. During the explicit SP session, the subjects were largely successful in discriminating S and X film clips in terms of intentionality (Table 1).

Table 1 Behavioral data for identification of social-signal-responsive regions

	EG		SP		
	E	O	S	X	V
Implicit					
Reaction time (ms)	424 ± 65	437 ± 53	367 ± 58	361 ± 73	372 ± 47
Explicit					
Judgment (%)	97 ± 4	87 ± 14	79 ± 18	11 ± 12	60 ± 31
Reaction time (ms)	533 ± 236	689 ± 183	4853 ± 1238	4893 ± 1414	4847 ± 1029

Average reaction time in color-change detection in each trial category is shown for the implicit EG and SP sessions. For the explicit sessions, judgment accuracy and average reaction time in the change-direction judgment are given for the EG session, as are the ratio of detecting intentionality (e.g. self-propelling rather than being externally moved) and the average reaction time of the judgment for the SP session. V denotes the trials where vague film clips were presented. Values are mean ± s.d.

Table 2 Behavioral data for functional-link analysis

	Implicit EG		Explicit EG		Implicit MI	
	E	O	E	O	Detective task	Reaction-time task
Accuracy (%)			97 ± 4	88 ± 14	96 ± 3	97 ± 2
Reaction time (ms)	412 ± 54	430 ± 53	514 ± 226	672 ± 180	1978 ± 922	431 ± 63

For the implicit EG session, average reaction time in color-change detection for E and O trials are given. Accuracy as well as average reaction times are shown for other sessions. Accuracy refers to accuracy in the direction-change judgment for the explicit EG session, on choosing one of two possible suspects rather than the unlikely one for the detective role-playing game, and on pressing the button corresponding to the presented number for the reaction-time task in the implicit MI session. Values are mean ± s.d.

Table 3 Social-signal-responsive regions

Structure	Coordinates			Peak <i>T</i>	Cluster size		
	<i>x</i>	<i>y</i>	<i>z</i>		<i>k</i>	<i>P</i>	
Implicit Middle temporal gyrus (posterior)	L	-58	-62	2	4.45	39	*
	R	60	-62	6	4.42	20	*
Explicit Superior temporal sulcus (posterior)	R	54	-42	8	4.61	64	*

Regions commonly responsive to the two types of social signals are separately reported for implicit and explicit sessions. For each region, the coordinates (*x y z*) of the activation peak in the MNI space, peak *T*-value, size of the activated cluster in a number (*k*) of voxels ($2 \times 2 \times 2 \text{ mm}^3$), and *P*-value are shown. Significant activation was thresholded at $P < 0.001$ and corrected to $P < 0.05$ for multiple comparisons using cluster size except for regions with an a priori hypothesis (*). L: left; R: right.

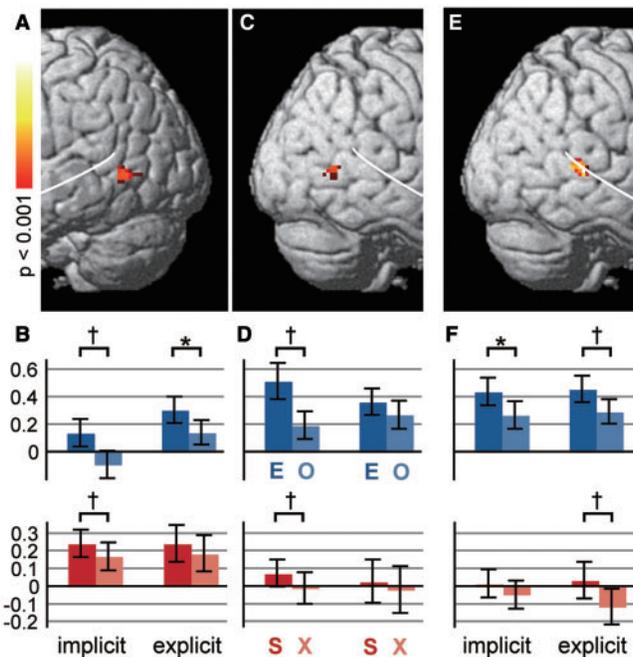


Fig. 2 Social-signal-responsive regions. The contrast [E - O] + [S - X] in the implicit tasks generated significant activation in the left (A and B) and right (C and D) pMTG and the same contrast generated significant activation in the explicit tasks in the right pSTS (E and F). The statistical threshold was $P < 0.001$ uncorrected. Activation is rendered on the surface of the standard anatomical image of SPMS5 (A, C and E). The superior temporal sulcus is indicated with a white line. Activation profiles (B, D and F) show the parameter estimate (partial regression coefficient or β value) in the EG (top) and SP (bottom) sessions at each peak voxel. Error bars indicate standard errors of mean (SEM). * $P < 0.05$ in the post hoc *t*-test for the contrasts not included in the voxel-by-voxel analysis. † $P < 0.05$ assured by the mask contrasts applied in the voxel-by-voxel analysis.

Social-signal-responsive regions

Regions commonly responsive to the two types of social signals were separately identified for the implicit and explicit sessions (Figure 2 and Table 3). The response during implicit tasks was bilaterally found in the posterior part of the middle temporal gyrus (pMTG; Figure 2A–D). In the left pMTG, the response was also significant during the explicit EG session (Figure 2B). Response during explicit tasks was identified in the right pSTS (Figure 2E and F), where activation was also significant in the implicit EG session (Figure 2F).

Functional link between social-signal detection and mental-state inference

Three PCs were retained based on their contributions (Figure 3A). Upon assessment of the loadings (Figure 3B), the third principal component (PC3) appeared to be the component of interest because it was commonly and specifically loaded on the implicit EG and MI sessions. In contrast, the first and second principal components (PC1 and PC2, respectively) appeared to represent between-session and session-specific variations, respectively, and they were therefore regarded as of no interest.

In the voxel-by-voxel search of the PC3 effect, a significant positive effect (i.e. positive slope of the partial regression line) was observed for the implicit EG and MI sessions only (Figure 3C–H, Table 4). For the

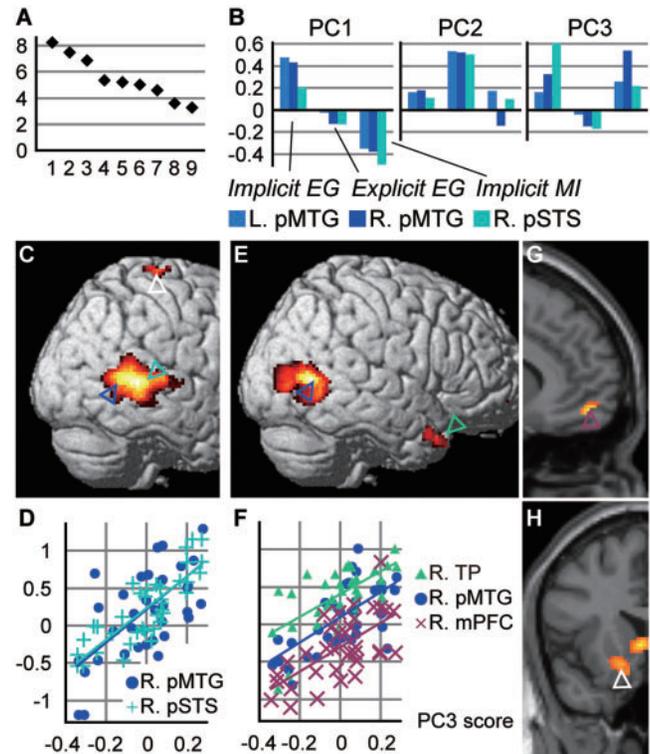


Fig. 3 Results of PCA and analysis of functional link. (A) Contribution (eigenvalue) of each PC (scree plot). The first three PCs lie away from the line composed of the rest of the PCs, which show a gradual decay. (B) Loadings of the retained PCs. A large absolute value indicates that the variance component was expressed in the ROI of that session. (C and D) Regions where the effect of PC3 was significant in the E – O contrast in the implicit EG session. Open arrowheads in blue, cyan and white indicate the activation peaks of the right pMTG, pSTS and precentral sulcus. Activation profiles (D) of the right pMTG and pSTS show the plot of the parameter estimate against the PC3 score for each subject. (E–H) Regions where the effect of PC3 was significant on activation during the Investigation period of the implicit MI session. Open arrowheads in blue, green, purple and white indicate activation peaks of the right pMTG, TP (E), mPFC (G) and left ventral striatum (H). Activation profiles are given for the right pMTG, TP and mPFC (F). Activated areas in the mPFC and ventral striatum are superimposed onto the sagittal and coronal section (at the peak voxel) of the standard anatomical image of SPMS5. Other details are the same as for Figure 2.

Table 4 Significant effect of PC3 on activation

Structure		Coordinates			Peak	Cluster size	
		x	y	z	T	k	P
Implicit EG							
Superior temporal sulcus (posterior) ^a	R	58	-40	4	9.23	1578	<0.001
Middle temporal gyrus (posterior) ^a	R	62	-54	8	5.69		
Precentral sulcus/SFG	R	16	-28	74	5.12	197	0.007
	L	-14	-34	78	4.47		
Explicit EG							
No significant activation							
Implicit MI							
Middle temporal gyrus (posterior) ^a	R	60	-62	6	7.09	920	<0.001
TP	R	54	12	-30	4.70	108	*
Superior frontal gyrus (ventral-medial)	R	10	54	-12	5.03	49	*
Caudate head	L	-2	16	0	5.12	303	0.005
Ventral striatum	L	-14	16	-12	4.43		

Regions showing a significant effect of PC3 are separately reported for implicit EG, explicit EG and implicit MI sessions.

^aSignificance of the effect was circular because the region was one of the sources of the regressors. Other details of the presentation were the same as for Table 3.

implicit EG session, the effect was identified in the right lateral temporal cluster involving the pMTG and pSTS and the cluster at the vertex covering the bilateral precentral sulci (Figure 3C and D). For the implicit MI session, the effect was identified in the right pMTG, TP (Figure 3E and F) and mPFC (Figure 3F and G), as well as in the subcortical cluster including the left caudate and ventral striatum (Figure 3H). The effects of PC1 and PC2 were also examined for reference purposes (Supplementary Figure 2 and Table 1).

DISCUSSION

We identified the responses to social signals in the bilateral pMTG for implicit detection and in the right pSTS for explicit detection. The observed difference in the involvement of these regions between the implicit and explicit tasks is consistent with previous observations of a point-light walker (Jastorff and Orban, 2009). From the activation data of these social-signal-responsive regions, we extracted a shared pattern of variation in response to eye gaze (i.e. PC3) across the two implicit sessions (i.e. EG and MI) for the subsequent functional-link analysis. In the implicit MI session, the extracted variation pattern was significant in the right pMTG as well as the mPFC, TP and striatum, which have often been implicated in ToM (Frith and Frith, 2003; Carrington and Bailey, 2009; Abu-Akel and Shamay-Tsoory, 2011). This suggests that subjects who engaged the right pMTG during perception of eye-gaze change under the implicit task requirement also engaged ToM-related regions under the implicit ToM requirement.

Our finding on the bilateral pMTG demonstrated the role of these regions in the implicit detection of a simple social signal, consistent with previous findings (Grossman *et al.*, 2000; Hoffman and Haxby, 2000; Pelphrey *et al.*, 2004b; Jastorff and Orban, 2009). The pMTG region may serve as an landmark in the human lateral occipito-temporal cortices for making an inference about the evolutionary expansion from the cortices buried in the monkey STS (Tsao *et al.*, 2003; Jastorff *et al.*, 2012). The region seems to be located anterior to the motion-sensitive area MT/V5+ (Sunaert *et al.*, 1999) or bodily motion-sensitive area EBA (Downing *et al.*, 2001); in previous studies that used point-light walkers (Jastorff and Orban, 2009) and assessed sensitivity to action rationality (Jastorff *et al.*, 2011), activation identified close to the pMTG partially overlapped with the MT/V5+ or EBA, but most of the extension and activation peaks were located anteriorly outside these areas.

Considering that only a few of our subjects could retrospectively report the existence of self-propelling stimuli in the implicit SP session, the detection of social signals in the pMTG may be unconscious. The observed significant activation of the extrastriate visual cortex, despite unconscious processing, is in line with a previous study using patients with visual extinction (Driver *et al.*, 2001).

The effect of PC3 during the implicit EG session was identified in the right pSTS and bilateral precentral sulci, which are known to respond to EG and are implicated in spatial attention (Nummenmaa and Calder, 2009). This suggests that subjects who engage the right pMTG during perception of eye-gaze change under the implicit task requirement also engage the network for spatial attention under this requirement.

The effect of PC3 was thus identified in distinct brain systems between the implicit EG and MI sessions, whereas it was exhibited in the right pMTG in both sessions. The observed dissociation of identified systems between the implicit EG and MI sessions suggests a task-dependent switch of the functional link from the right pMTG. This may suggest the role of the pMTG in a common entrance node of social-signal processing. Although assuming that the primary provider of social information is the posterior temporal region and that it is linked with the higher social-cognitive systems is not new (Saxe, 2006), the current finding updates the functional anatomy of the entrance region and provides the first empirical evidence of a functional link. Although one may suspect that the right pMTG region where the PC3 effect was identified in the implicit EG session is anatomically distinct from that original social-signal-sensitive region, we considered the distance between the peak voxels too small (less than 10 mm) to assume the existence of different functional areas.

Given recent multiplicit or multi-componential concepts of the ToM (Frith and Singer, 2008; Carrington and Bailey, 2009; Abu-Akel and Shamay-Tsoory, 2011), it is interesting to speculate on which aspect or component of the ToM our finding is relevant. It has become a recent convention to divide the ToM into cognitive and affective domains, with the former pertaining to inferences about knowledge and belief and the latter to those about emotions. The process triggered by eye gaze during the implicit MI may be relevant to the affective domain because the mPFC, TP and ventral striatum comprise the system dedicated to the affective ToM (Abu-Akel and Shamay-Tsoory, 2011). Alternatively, the triggered process may be related to a mindset for social interaction. Similar regions in the mPFC and TP are activated during the perception of another's directed eye gaze and a voice calling one's own name (Kampe *et al.*, 2003). Activation of a similar mPFC region also occurs during joint attention, which entails detecting another person's gazing at the same object one is gazing at (Williams *et al.*, 2005; Schilbach *et al.*, 2010). The latter finding may be relevant to the concept of the shared attention mechanism (SAM) in the model of ToM development (Baron-Cohen, 1994); the SAM is assumed to detect joint attention and developmentally mediate the social-signal detector and ToM mechanism. Activation of the mPFC and TP may thus be related to the subject's awareness of being engaged in a real-time interactive relationship with the suspects.

The functional link demonstrated in this study is obviously not a causal link. We consider it comparable to a link between two different behaviors or abilities that co-emerge in a specific developmental stage, as is often demonstrated in developmental psychology. We show that individuals who sensitively detect eye gaze tend to recruit the ToM process in terms of neural activity. The causal relationship theoretically underlying this observation could be: (i) the former precedes the latter, (ii) the latter precedes the former or (iii) the two processes are preceded by a common (unknown) process. Relationship (i) corresponds to the hypothesis that social-signal processing modules underlie the

development of the ToM mechanism (Baron-Cohen, 1994). We are not aware of strong hypotheses for (ii) or (iii).

The functional link thus suggested is potentially supported by direct anatomical connections. The connections from the pMTG to TP and mPFC have been implicated by diffusion tensor imaging in humans (Ethofer *et al.*, 2011). In the monkey, projections have been identified from the TPO, which is a potential homolog of the pMTG (Jastorff and Orban, 2009), to the TP (Seltzer and Pandya, 1989b) and mPFC (Seltzer and Pandya, 1989a).

Although our findings basically support the ToM development model (Baron-Cohen, 1994), the discussion should incorporate a potential update of the model. The model assumed single-path development from social-signal detection to the ToM. Given recent multiplicity concepts (e.g. cognitive and affective) of the ToM, however, the model can be reasonably expected to include multiple paths, and the functional link that we identified seems to provide a neural basis for only one of them.

SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

REFERENCES

- Abu-Akel, A., Shamay-Tsoory, S. (2011). Neuroanatomical and neurochemical bases of theory of mind. *Neuropsychologia*, 49, 2971–84.
- Allison, T., Puce, A., McCarthy, G. (2000). Social perception from visual cues: role of the sts region. *Trends in Cognitive Sciences*, 4, 267–78.
- Baron-Cohen, S. (1994). How to build a baby that can read minds: cognitive mechanisms in mindreading. *Cahiers De Psychologie Cognitive/Current Psychology of Cognition*, 13, 513–52.
- Carrington, S.J., Bailey, A.J. (2009). Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Human Brain Mapping*, 30, 2313–35.
- 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, 314–25.
- Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 140–61.
- Dasser, V., Ulbaek, I., Premack, D. (1989). The perception of intention. *Science*, 243, 365–7.
- Downing, P.E., Jiang, Y.H., Shuman, M., Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science*, 293, 2470–3.
- Driver, J., Vuilleumier, P., Eimer, M., Rees, G. (2001). Functional magnetic resonance imaging and evoked potential correlates of conscious and unconscious vision in parietal extinction patients. *Neuroimage*, 14, S68–S75.
- Ethofer, T., Gschwind, M., Vuilleumier, P. (2011). Processing social aspects of human gaze: a combined fMRI-DTI study. *Neuroimage*, 55, 411–9.
- Friston, K.J., Penny, W.D., Glaser, D.E. (2005). Conjunction revisited. *Neuroimage*, 25, 661–7.
- Frith, U., Frith, C.D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 358, 459–73.
- Frith, C.D., Singer, T. (2008). The role of social cognition in decision making. *Philosophical Transactions of the Royal Society of London. Series B-Biological Sciences*, 363, 3875–86.
- Grossman, E., Donnelly, M., Price, R., et al. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, 12, 711–20.
- 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, 80–4.
- Iacoboni, M. (2005). Neural mechanisms of imitation. *Current Opinion in Neurobiology*, 15, 632–7.
- Jastorff, J., Clavagner, S., Gergely, G., Orban, G.A. (2011). Neural mechanisms of understanding rational actions: middle temporal gyrus activation by contextual violation. *Cerebral Cortex*, 21, 318–29.
- Jastorff, J., Orban, G.A. (2009). Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *Journal of Neuroscience*, 29, 7315–29.
- Jastorff, J., Popivanov, I.D., Vogels, R., Vanduffel, W., Orban, G.A. (2012). Integration of shape and motion cues in biological motion processing in the monkey sts. *Neuroimage*, 60, 911–21.
- Kampe, K.K., Frith, C.D., Frith, U. (2003). 'Hey john': signals conveying communicative intention toward the self activate brain regions associated with 'Mentalizing,' Regardless of modality. *Journal of Neuroscience*, 23, 5258–63.
- Maurer, D., Salapatek, P. (1976). Developmental changes in the scanning of faces by young infants. *Child Development*, 47, 523–7.
- Nelissen, K., Borra, E., Gerbella, M., et al. (2011). Action observation circuits in the macaque monkey cortex. *Journal of Neuroscience*, 31, 3743–56.
- Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, 25, 653–60.
- Nummenmaa, L., Calder, A.J. (2009). Neural mechanisms of social attention. *Trends in Cognitive Sciences*, 13, 135–43.
- Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9, 97–113.
- Pelphrey, K.A., Morris, J.P., McCarthy, G. (2004a). Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. *Journal of Cognitive Neuroscience*, 16, 1706–16.
- Pelphrey, K.A., Viola, R.J., McCarthy, G. (2004b). When strangers pass - processing of mutual and averted social gaze in the superior temporal sulcus. *Psychological Science*, 15, 598–603.
- Premack, D. (1990). The infant's theory of self-propelled objects. *Cognition*, 36, 1–16.
- Rizzolatti, G., Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, 27, 169–92.
- Rochat, P., Morgan, R., Carpenter, M. (1997). Young infants' sensitivity to movement information specifying social causality. *Cognitive Development*, 12, 441–65.
- Rutherford, M.D., Pennington, B.F., Rogers, S.J. (2006). The perception of animacy in young children with autism. *Journal of Autism and Developmental Disorders*, 36, 983–92.
- Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, 16, 235–9.
- Schilbach, L., Eickhoff, S.B., Rotarska-Jagiela, A., Fink, G.R., Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the 'Default system' of the brain. *Consciousness and Cognition*, 17, 457–67.
- Schilbach, L., Wilms, M., Eickhoff, S.B., et al. (2010). Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry. *Journal of Cognitive Neuroscience*, 22, 2702–15.
- Seltzer, B., Pandya, D.N. (1989a). Frontal lobe connections of the superior temporal sulcus in the rhesus monkey. *Journal of Comparative Neurology*, 281, 97–113.
- Seltzer, B., Pandya, D.N. (1989b). Intrinsic connections and architectonics of the superior temporal sulcus in the rhesus monkey. *Journal of Comparative Neurology*, 290, 451–71.
- Senju, A., Tojo, Y., Dairoku, H., Hasegawa, T. (2004). Reflexive orienting in response to eye gaze and an arrow in children with and without autism. *Journal of Child Psychology and Psychiatry*, 45, 445–58.
- Sugiura, M., Friston, K.J., Willmes, K., Shah, N.J., Zilles, K., Fink, G.R. (2007). Analysis of intersubject variability in activation: an application to the incidental episodic retrieval during recognition test. *Human Brain Mapping*, 28, 49–58.
- Sunaert, S., Van Hecke, P., Marchal, G., Orban, G.A. (1999). Motion-responsive regions of the human brain. *Experimental Brain Research*, 127, 355–70.
- Tremoulet, P.D., Feldman, J. (2000). Perception of animacy from the motion of a single object. *Perception*, 29, 943–51.
- Tsao, D.Y., Freiwald, W.A., Knutsen, T.A., Mandeville, J.B., Tootell, R.B. (2003). Faces and objects in macaque cerebral cortex. *Nature Neuroscience*, 6, 989–95.
- Vecera, S.P., Johnson, M.H. (1995). Gaze detection and the cortical processing of faces: evidence from infants and adults. *Visual Cognition*, 2, 59–87.
- Williams, J.H., Waiter, G.D., Perra, O., Perrett, D., Whiten, A. (2005). An fMRI study of joint attention experience. *Neuroimage*, 25, 133–40.