

Inferring false beliefs from the actions of oneself and others: an fMRI study

J. Grèzes,^{a,*} C.D. Frith,^a and R.E. Passingham^{a,b}

^aWellcome Department of Imaging Neuroscience, University College London, London WC1N 3BG, UK

^bDepartment of Experimental Psychology, University of Oxford, Oxford, UK

Received 22 July 2003; revised 6 October 2003; accepted 6 October 2003

The ability to make judgments about mental states is critical to social interactions. Simulation theory suggests that the observer covertly mimics the activity of the observed person, leading to shared states of mind between the observer and the person observed. We tested this hypothesis by investigating the neural networks activated while subjects watched videos of themselves and of others lifting a box, and judged the beliefs of the actors about the weight of the box. A parietal premotor circuit was recruited during action perception, and the activity started earlier when making judgments about one's own actions as opposed to those of others. This earlier activity in action-related structures can be explained by simulation theory on the basis that when one observes one's own actions, there is a closer match between the simulated and perceived action than there is when one observes the actions of others. When the observers judged the actions to reflect a false belief, there was activation in the superior temporal sulcus, orbitofrontal, paracingulate cortex and cerebellum. We suggest that this reflects a mismatch between the perceived action and the predicted action's outcomes derived from simulation.

© 2003 Elsevier Inc. All rights reserved.

Keywords: False belief; Self; Other; Action observation; Judgement; Intention; Expectancy; Simulation theory; fMRI

Introduction

The ability to make judgments about mental states is critical to social interactions. We are able to perceive and predict other people's intentions from their nonverbal behavior. It is suggested that we understand other people's behaviors by simulation (Decety and Grèzes, 1999; Gallese and Goldman, 1998; Jeannerod, 1994; Wolpert et al., 2003), particularly in cases where the information available is not easily encoded into language (Adolphs, 1999). The assumption is that when we observe others, we mentally simulate the actions that we see.

There is behavioral evidence that the action representation system contributes to the perception of actions (Knoblich and Prinz, 2001; Liberman and Mattingly, 1985; Orliaguet et al., 1997; Runeson and Frykholm, 1983; Shiffrar and Freyd, 1993; Viviani and Stucchi, 1992), and this idea has found recent support from neurophysiological data, both from cell recording in monkeys and neuroimaging studies in humans (Decety and Grèzes, 1999; Gallese et al., 1996; Rizzolatti et al., 1996). In the macaque monkey brain, a class of neurons called "mirror" neurons was discovered both in the parietal and in the premotor cortex that discharge not only when the monkey performed an action but also when observing the experimenter or another monkey performing the same action (di Pellegrino et al., 1992; Fogassi et al., 1998; Gallese et al., 1996). Similarly, in humans, the parietal and the premotor cortex have been found to be activated during action observation (Buccino et al., 2001; Decety et al., 1997; Grafton et al., 1996; Grèzes and Decety, 2001; Rizzolatti et al., 1996). Those results suggested that action observation automatically triggers action simulation, a mechanism that could be at the basis of action understanding (Adolphs, 2003; Decety and Grèzes, 1999; Gallese, 2003; Jeannerod, 1994; Rizzolatti et al., 2001).

However, humans are not only able to recognize actions from observation, but can also predict and infer underlying causes, intentions and beliefs from the behavior of others. To do so, the information that is directly available consists mostly of the movements of the agent in space and time in the physical or social environment (Barresi and Moore, 1996). The present study explores the neural basis of everyday human competence to make judgments about mental states through the observation of the nonverbal behavior of other people, and aims to clarify to what extent this ability relies on simulation. We scanned subjects, using event-related fMRI, while they viewed actions being performed on video. We consider two issues.

First, it has been shown that observers can make judgments from kinematic patterns (point-lights displays) about the expectation of subjects when picking up boxes of different weights. Specifically, they can distinguish whether the actors have been deceived or not about the weight of the box (Runeson and Frykholm, 1983). In the present study, we therefore presented videos of actors picking up boxes. When preparing these videos, on some trials the actors were correctly informed of the weight of the box and therefore had a correct expectation, whereas on other

* Corresponding author. Wellcome Department of Imaging Neuroscience, University College London, 12 Queen Square, London WC1N 3BG, UK. Fax: +44-207-813-1420.

E-mail address: jgrezes@fil.ion.ucl.ac.uk (J. Grèzes).

Available online on ScienceDirect (www.sciencedirect.com.)

trials the actors had been misinformed and therefore had an incorrect expectation. We required the subjects to judge whether the actors had a true or false expectation. By comparing these trial types, we aimed to identify the neural circuits that are involved in the attribution of mental states from the observation of behavior.

Second, it has been demonstrated that subjects can recognize kinematic displays of self-generated drawings and handwriting strokes as being their own, and are better at predicting the endpoint of an action that is their own than an action that is performed by others (Knoblich and Prinz, 2001; Knoblich et al., 2002). Based on the assumption that both action production and action perception are driven by the representation of the outcomes of the action (Hommel et al., 2001; Prinz, 1997), it has been proposed that when one perceives one's own actions, there is a closer match between the predicted and actual outcomes of the action than there is when we perceive the action of others (Knoblich and Prinz, 2001). In the present study, we therefore compared the perception of one's own actions and those of others. Unlike others studies, in the present experiment, subjects were not required to explicitly take a first-person or third-person perspective (Ruby and Decety, 2001) or to make judgment about whether an action was performed by themselves or by others (Farrer and Frith, 2002). The task was to infer mental states through the observation of actions. If simulation subserves the ability to attribute mental states through action observation, we predicted differences within the motor system for the comparison between the perception of one's own actions and those of others.

Methods

Subjects

Six subjects (four men and two women; range: 25–39 years) with no neurological or psychiatric history participated in the imaging study. All gave informed consent according to procedures approved by the Joint Ethics Committee of the National Hospital for Neurology and Neurosurgery (UCLH NHS Trust) and Institute of Neurology (UCL).

Stimuli

Eight actors (four men and four women) were videotaped with over 50 trials of lifting and carrying a box of varying weights (1, 6, 12 or 18 kg). Each recorded act began with the actor off-screen to the left, and the box placed in the middle of the scene and about 1 m from the right edge of the table. The actor entered on the left, lifted the box by its side handles, walked toward the table, placed the box on the table and went out of the room. The entire act usually lasted 6–9 s. During most of the filming session, the experimenter gave the correct exact weight of the box so that the actor had a correct expectation of the weight. However, on some occasions (10% of the trials), the experimenter misled the actor, and gave him the wrong weight, so that the actor had a false expectation about the weight of the box. Eight actors were used to assess variability in observer performance due to individual difference in actor styles of lifting and carrying. We edited video recordings of 3.5 s concentrating on the lifting phase. Using the After-effect software, we blurred the face of the actors on each

movie to hide facial and emotional expression so that only body information was provided.

As the subjects of the fMRI experiment were the actors, the video recording session and the fMRI study were 8 months apart to rule out memory effects. For technical reasons, two actors among the eight could not participate in the fMRI study. The stimuli used during the fMRI experiment consisted of color video recordings of the actors performing the action of lifting and carrying a box of different weights. During the scanning session, the videos were viewed by projection onto a mirror mounted onto the head coil in the scanner. The subjects viewed videos (3.5 s each) of themselves or of two other actors. At the end of each video, a screen was presented for 1.5 s on which were displayed the words 'Correct' and 'False' (see Fig. 1). The subjects were required to decide (forced choice) whether the actor has a correct or a false expectation of the weight by making key press with their right fingers. The location of the two words (on the right and left part of the screen) was randomized. We delayed the responses on purpose so as to ensure that the subjects responded at the same time at the end of the observation period. Furthermore, by varying the location of the response cues, we deliberately prevented the subjects from preparing their responses during the observation period. As a result, our design was not appropriate for detecting any reaction time advantages in making judgements about oneself compared with others. The study was carried out in a single continuous scanning session. One hundred and twenty video films (40 self, 40 actor 1, 40 actor 2) intermixed with 80 null events were randomly presented with an intertrial interval of 5 s. Fifty percent of the video films involved false expectation.

Behavioral analysis

Trials with reaction times of more than 1.5 s (maximum time to answer) were excluded from the analysis. ANOVAs were performed on mean accuracy, with within participant factors of actors (self or other) and actions (correct or false expectation). Only effects surviving $P < 0.05$, in the absence of higher-order interactions, are reported. In addition, the mean judgements per subject were converted into d' sensitivity measures (Macmillan and Creelman, 1991). This is a measure of the ability to make correct detections that are independent of any response bias. Two different d' values were computed; one for the ability to correctly detect the expectation for one's own actions and one for the ability to correctly detect expectation for the actions of others.

fMRI imaging

Imaging was performed using a 2-T scanner (Siemens Vision, Erlangen, Germany). The functional images sensitive to blood oxygenation level-dependent (BOLD) contrasts were acquired by T_2^* -weighted echo planar (EPI) images (TR = 2.74 s, TE = 40 ms, 385 sequential whole brain volume acquisitions, $64 \times 64 \times 48$ at 3 mm isotropic resolution). High-resolution structural T1-weighted MPRAGE images (TR = 9.5 s, TE = 4 ms, T1 = 600 ms, voxel size $1 \times 1 \times 1.5$ mm, 108 axial slices) were also acquired on all subjects.

Data analysis

We used the SPM99 software (<http://www.fil.ion.ucl.ac.uk/spm>) for image processing and analysis. The first five volumes

were discarded to allow for T1 equilibration effects. The remaining 380 image volumes were realigned to the first volume by rigid body transformation, sinc interpolated over time to correct for phase advance during acquisition, normalized to the Montreal Neurological Institute (MNI) reference brain (voxel size = 2 mm³) and spatially smoothed by a Gaussian kernel of FWHM 8 mm (Ashburner and Friston, 1997). Statistical parametric maps (SPMs) of *t* statistics were calculated for condition-specific effects within a general linear model.

Effects of interest were defined as follows: the perception of self with correct or with false expectation, and the perception of others with correct or false expectation. The data for true and false expectation were analyzed in two ways. First, the trials were divided by stimulus presentation, that is, into trials in which the actors actually had a true expectation and trials in which the actor actually had a false expectation. Second, the trials were divided by the judgement made by the subjects, that is, the trials on which the subjects judged the actor to have a true expectation and those on which the subjects judged the actor to have a false expectation.

The BOLD response to the stimulus onset for each event type was modelled by a box-car waveform of 5 s (1.86 TR) convolved with the hemodynamic response function and its temporal derivative to account for any temporal shifts in the response of the stimuli (Friston et al., 1998). Also included for each session were six covariates to capture residual movement-related artifacts (the three rigid-body translations and the three rotations determined from initial registration), and a single covariate representing the mean (constant) over scans. The data were high-pass filtered with a frequency cut-off at 140 s.

We performed a random effects analysis. Images of parameter estimates for each contrast of interest were created for each subject (first-level analysis) and were then entered into a second-level analysis using ANOVAs with the two HRFs comprising a factor. Each analysis tested for a specific effect on the shape of the BOLD impulse response. Statistical parametric maps (SPMs) of the *F* statistic were constructed using a generalized Greenhouse–Geisser correction and of the *T* statistic. The statistical parametric maps were thresholded at $P < 0.001$ (uncorrected for multiple comparisons). These maps were overlaid on the MNI template and on the normalized structural images of each subject and labelled using the atlas of Duvernoy (1999).

Results

Behavioral results

Subjects were able to correctly detect the expectation of the actors from their nonverbal behaviors and were significantly above chance ($d' = 0.78$, *t* test $P = 0.025$). Subjects tended to be less accurate in detecting when they have been deceived ($d' = 0.63$) than when the others were deceived ($d' = 0.99$), but the difference in the d' measures for judgements about one's own actions and of others was not significant (paired *t* test, $P = 0.288$). The ANOVAs revealed a significant main effect of action type. The subjects judged more often the action as reflecting a true expectation (Mean correct = 79%) than a false expectation (Mean accuracy = 51%), [$F(1,5) = 25.04$, $P = 0.004$]. There was no significant difference for the judgements about one's own actions (Mean accuracy = 63%) and of others (Mean accuracy = 67%) [$F(1,5) = 0.786$, $P =$

0.416]. As explained in the methods, the design was not suitable for looking for differences in RT.

Neuroimaging results

We tested the hypothesis that within the motor system, there would be differences between the perception of one's own actions and the actions of others. When contrasting the perception of one's own actions with the actions of others, we found differences in the BOLD signal bilaterally in the dorsal premotor cortex, the left frontal operculum, the left intraparietal sulcus and the right cerebellum (Table 1, Figs. 2a and c). These findings were based on an *F* test for the canonical HRF and its temporal derivative.

This difference could be explained by a difference in the amplitude or in the latency of the BOLD signal in these regions. Therefore, the latency effect was further assessed by exclusively masking ($P = 0.001$) the *F* contrast for the temporal derivative with the corresponding *F* contrast on the canonical HRF. A significant difference in the temporal derivative (i.e., in the latency), in the absence of a significant difference in amplitude, was detected in all the regions previously identified above. Those regions showed earlier responses for the perception of one's own actions as compared to the perception of the actions of others (Fig. 2c). This result is consistent across subjects (Fig. 2b). It will be seen from these figures that though there was a difference in the latency of the BOLD signal between perception of oneself and others, the BOLD amplitude in these areas is similar for both conditions (Fig. 2c).

The present study aimed to identify the neural network subserving the detection of true and false expectation. As explained in the methods, the data for true and false expectation were analyzed in two ways. In the first analysis, the trials were divided by stimulus presentation, that is, into trials in which the actors actually had a true expectation and trials in which the actor actually had a false expectation. When the data were analyzed in this way, activations were found in the posterior cingulate gyrus ($xyz = 8, -38, 28$, *Z* score = 4.24) and in the right posterior part of the superior temporal sulcus ($xyz = 42, -64, 26$, *Z* score = 3.80) for false vs. real expectation.

In the second analysis, the trials were divided from the judgement made by the subjects, that is, the trials on which the subjects judged the actor to have a true expectation and those on which the subjects judged the actor to have a false expectation. When the data were analyzed in this way, in several regions, the amplitude of the BOLD signal was significantly higher when subjects rated a perceived action as reflecting false expectation as compared to correct expectation. There was activation bilaterally in the lateral orbitofrontal cortex, in the anterior part of the insula and in the superior temporal sulcus. Differences in amplitude were

Table 1
Brain regions showing latency effects for the perception of one's own action vs. the perception of the action of others

Brain regions	Talairach coordinates			<i>Z</i> score	Nb of voxels
	<i>x</i>	<i>y</i>	<i>z</i>		
L Dorsal premotor cortex	-30	-12	64	3.91	37
R Dorsal premotor cortex	32	-8	50	3.65	14
L Inferior parietal lobule	-42	-38	48	3.47	18
L Operculum	-46	2	0	3.83	23
R Cerebellum	22	-38	-40	3.49	6

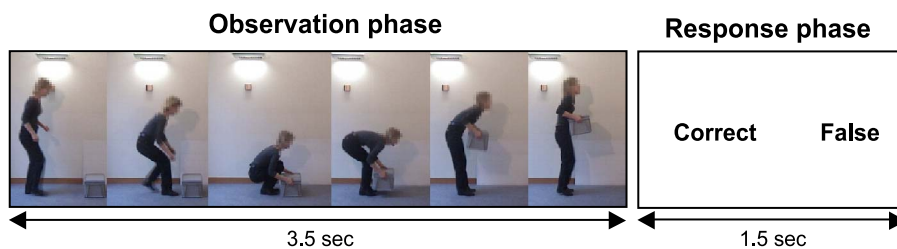


Fig. 1. The subjects watched videos of themselves and of others lifting a box and were instructed to judge whether the actor had a correct or a false expectation about the weight of the box. Each trial corresponds to the presentation of a color movie of 3.5 s following by a screen of 1.5 s on which were displayed the words 'Correct' and 'False'. The position on the screen of the two words was randomized across trials so as to avoid motor preparation during the observation phase. The subjects were required to decide (forced choice) whether the actor had a correct or a false expectation of the weight by making a key press with the right fingers.

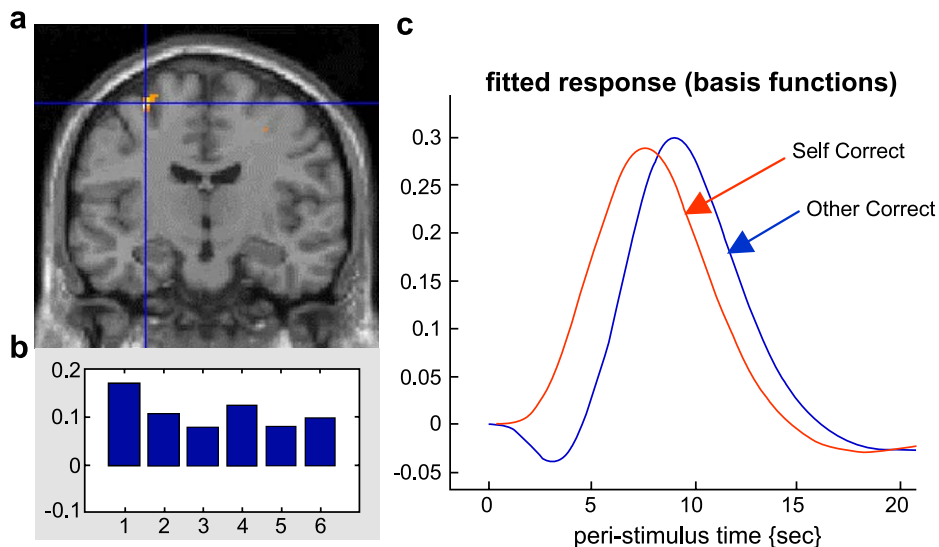


Fig. 2. (a) Left dorsal premotor cortex ($-30, -12, 64$). Group ($n = 6$) average activation of the left dorsal premotor cortex superimposed on a coronal section of the MNI brain. (b) Histograms representing, for the six subjects, the contrast parameter for the temporal derivative, for this voxel for the perception of one's own actions vs. the perception of the actions of others. Note that the latency effect is consistent across all subjects. (c) Fitted responses at the local maxima in the left dorsal premotor cortex showing latency difference between the perception of one's own action and the perception of the actions of others.

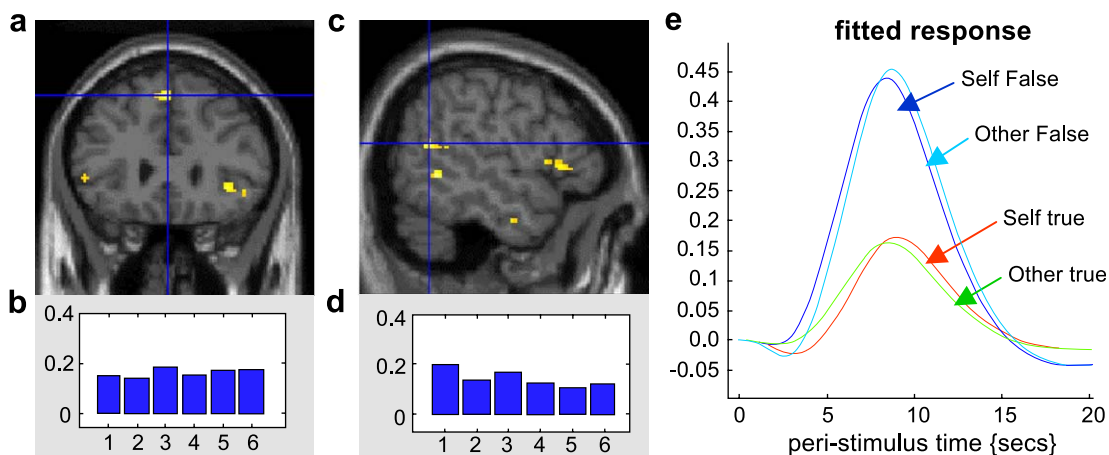


Fig. 3. Regions showing an amplitude difference for the main effect for the perception of false vs. correct expectation. (a) Group ($n = 6$) average activation of the region of orbitofrontal and neighboring insula cortex and dorsomedial frontal cortex superimposed on a coronal section of the MNI brain. (b) Histograms representing, for the six subjects, the contrast parameter over the hemodynamic response function at the local maxima in the right orbitofrontal cortex between the perception of false vs. correct expectation. (c) Group ($n = 6$) average activation of the right superior temporal sulcus and right ventral frontal cortex superimposed on a sagittal section of the MNI brain. (d) Histograms representing, for the six subjects, the contrast parameter for the hemodynamic response function at the local maxima in the right superior temporal sulcus for the perception of false vs. correct expectation. (e) Fitted responses at the local maxima in the right superior temporal sulcus showing an amplitude difference between the perception of false vs. correct expectation both for the perception of one's own action and other's actions.

Table 2
Brain regions showing amplitude effects when subjects judged the actions to reflect false expectation vs. correct expectation

Brain regions	Talairach coordinates			Z score	Nb of voxels
	x	y	z		
R Inferior frontal gyrus, Pars opercularis/orbitalis	50	22	10	3.51	32
L Inferior frontal gyrus, pars orbitalis	−56	34	0	3.44	100
R Anterior Insula/lateral orbital gyrus	40	30	−4	3.67	112
L Anterior Insula/lateral orbital gyrus	−32	24	−10	4.01	88
Dorso-medial frontal cortex	−2	26	52	3.49	49
R Head of caudate	16	10	10	3.59	18
L Superior temporal sulcus, Ascending posterior segment	−48	−46	14	3.92	149
R Superior temporal sulcus/ Middle temporal gyrus	60	−44	2	3.52	114
L Superior temporal sulcus, Posterior horizontal segment	−54	−48	0	3.26	13
L Cerebellum	−22	−76	−42	3.57	24

also detected in the dorsomedial frontal or paracingulate cortex, in the right head of the caudate and in the left cerebellum (Table 2, Figs. 3a, b, c, d and e).

We also looked at trials on which the actor actually had a false expectation and divided them into trials on which the observer made a correct judgement and trials in which the observer made an incorrect judgement. We again found activations bilaterally in the superior temporal sulcus (left hemisphere (LH), $Z = 4.52$, right hemisphere (RH), $Z = 3.96$), the lateral orbitofrontal cortex (LH, $Z = 4.70$, RH, $Z = 4.27$), the left cerebellum ($Z = 4.07$) and dorsomedial frontal cortex ($Z = 3.73$). The coordinates are the same as Table 2.

Discussion

This study explores the neural basis of everyday human competence in predicting and explaining human behavior, including the capacity to understand the mental states of others through the observation of their nonverbal behavior. One possible explanation of this competence is given by simulation theory according to which human beings are able to attribute mental states to others by using the resource of their own behavior control system. The theory is that to adopt the perspective of others, we mentally simulate their behaviors. This simulation process allows us to make predictions concerning the action. In this way, the observer can identify the mental state of others (Gallese and Goldman, 1998; Gordon, 1995).

Perception of the actions of oneself and others

The contrast between perceiving the actions of oneself and perceiving the actions of others revealed activations in action-related structures, namely the dorsal premotor cortex, the intraparietal sulcus and in the cerebellum (Table 1, Fig. 2c). Several functional imaging studies have shown that the same regions are

involved not only in the observation of action but also in various aspects of motor behavior such as movement execution, motor planning and motor imagery (for reviews, see Decety and Grèzes, 1999; Grèzes and Decety, 2001; Jeannerod, 2001). In macaque monkeys, there are cells in both the premotor and parietal cortex that fire not only when the monkey observes the actions of another but also when the monkey carries out similar actions; these are the ‘mirror neurones’ (Gallese et al., 1996; Rizzolatti et al., 2001). Thus, the activity both in humans and monkeys is consistent with the hypothesis that the observer simulates the actions of others. In the present study, the activations in the premotor and parietal cortex as well as in the cerebellum could therefore reflect the automatic triggering of an off-line process of simulation as the result of the perception of actions.

Our second finding was that within these action-related structures, there was a clear advantage in the latency of the BOLD signal for the perception of one’s own actions. This difference in BOLD latency could either reflect a difference in the onset of the underlying neural activity or from a difference in the duration of activity between two conditions with the same onset (Henson et al., 2002). Fig. 2c suggests that in the premotor cortex, there was a difference in onset. We can explain the difference in latency (onset) within the common coding framework between perception and action (Knoblich and Prinz, 2001). This model assumes that when one observes an action, the perceptual inputs activate the motor codes that would produce the same outcomes. Given that the system used for simulating the perceived action and the one used for producing it is the same when one observes one’s own actions, there is a closer match between perceptual inputs and the motor codes than when one perceives the action of others, and therefore the activation in the motor system appears sooner when one observes one’s own actions. The internal–forward dynamical model (Wolpert et al., 2001) also claims that actions are coded in terms of their outcomes, and a similar assumption concerning closer matching can also be derived from this model.

Inferring false vs. correct expectation

It is suggested that by mental simulation, subjects are able to make predictions concerning the perceived action and compare the predicted and observed outcomes (Wolpert et al., 2003). This allows the observer to draw an inference concerning the mental states of the actor. In the present experiment, the subjects were instructed to decide whether the action kinematics of the actors while lifting a box reflected a correct or a false expectation about the weight of the box. When contrasting the trials that were rated as false expectation with those rated as correct expectation, activations were detected bilaterally in the superior temporal sulcus, dorsomedial frontal cortex or paracingulate cortex, anterior insula, lateral orbitofrontal cortex and cerebellum.

One explanation of these results is that they reflect the perception of kinematic changes. Indeed, the only clues that the subjects could rely on were changes in the kinematic pattern of the actions. It has been previously shown with point-light displays that this pattern conveys information not only about the weight of the box but also about the lifter’s expectation (Runeson and Frykholm, 1983). When the actual weight of a box differs from the expected one, postural readjustments occur, and the amount of readjustments specifies the difference between the expected and actual weight. Several neuroimaging studies have shown that the

superior temporal sulcus is involved in the perception of biological motion, such as hand, mouth and eye movements (for a review, see Allison et al., 2000). When we divided the trials by the actual expectations (true or false) of the actors, which separates the data according to trials on which there were or were not postural readjustments, the only area in which activity was detected was the superior temporal sulcus. This activation may reflect kinematic differences.

However, the activations found in the present study cannot be fully explained as simply reflecting detection of kinematic changes. When we directly compared trials on which the expectation was false and the subjects made the correct judgement and trials on which the expectation was false and the subjects made an incorrect judgement, activations were revealed not only in the superior temporal sulcus, but also in the dorsomedial frontal cortex or paracingulate cortex, orbital frontal cortex and neighboring anterior insula and cerebellum. These additional activations cannot be explained by the kinematics; the crucial factor is the inference made by the subjects.

We found activations in the superior temporal sulcus and the dorsomedial frontal cortex or paracingulate cortex. Activations have been reported in these areas when subjects make judgements about the mental states of others (for review, see Frith and Frith, 1999; Gallagher and Frith, 2003). For example, the STS has been reported to be activated when subjects make explicit judgements about trustworthiness, even though the stimuli did not involve biological motion (Winston et al., 2002). Furthermore, these activations were only found when the judgements were social and not when they were about physical characteristics. Activation has also been reported in the paracingulate cortex when subjects make judgements about the mental states from either verbal material (stories) or visual material (cartoons) (Gallagher et al., 2000). It is clear from this paper, as well as from a meta-analysis (Frith and Frith, 1999), that peaks can be found both anteriorly and posteriorly. They tend to cluster around the paracingulate sulcus. The peak in the present experiment lay along the back of this sulcus.

Thus, the activity in these regions may reflect judgements about mental states. However, the question arises as to why the activity in these and other regions was greater for judgements that the expectation was false than for judgements that the expectation was true. The present study shows that our subjects judged more often the action as reflecting a true belief than a false belief. It is known that people have the tendency to attribute to others true beliefs (Bartsch and Wellman, 1995) and that a person's perception is shaped and guided by preexisting knowledge about the world leading therefore to expectations and predictions about other behaviors. A recent fMRI study, on the influence of context on perceived gaze shift, demonstrated that when an observer's prediction was violated, the activity in the STS was more prolonged compared with the situation in which the observer's predictions were met (Pelphrey et al., 2003). Thus, the activation was longer when there was a mismatch between the predictions of the subjects and the perceived action. The activations found in the orbital frontal cortex and neighboring anterior insula may also reflect the violation of expectations. Activity in these areas as well as in the temporo-parietal junction has been reported when subjects perceived a mismatch between what they expect and what actually happened (Downar et al., 2001, 2002). On the basis of these observations, we conclude when subjects judge that the expectation was false, this is because they detected a mismatch between the predicted outcomes of the action

and the perceived ones. Finally, it is suggested that the activity within those brain areas may either reflect an alerting system (Downar et al., 2002) or the change in ongoing cognitive activity (Corbetta and Shulman, 2002). When the observer's predictions are violated, we propose that the observer must update his representation of the mental state of the actor.

Conclusions

The aim of this experiment was to specifically examine the neural mechanisms that are involved in making judgements about mental states from nonverbal behavior, and to determine to what extent simulation subserves this ability. In the present experiment, the subjects made judgements about the expectations of the actor, where the video either showed their own actions or those of others. We found activations in the parietal and premotor cortex which we relate to the simulation of the observed action. However, the significant finding was that there was a clear difference for these areas in the onset of the BOLD signal for the perception of own actions as compared to the perception of the actions of others. We take this latency difference to reflect the fact that when one observes one's own actions, there is a closer match between the simulated and perceived action than there is when one observes the actions of others.

We were also able to compare the judgements made by the observers when they judged that the actor had a correct or incorrect expectation. In both cases, there was activity in the parietal and premotor cortex. When the observers judged the action to reflect false expectation, there was additional activity in the superior temporal sulcus, dorsomedial frontal cortex, orbito-frontal cortex and cerebellum. We have suggested that this activity reflected the violation of the observer's prediction that there was a mismatch between the observer's prediction about the outcomes of the action and the perceived outcomes. These predictions are based on the simulation by the observer of the actions that are observed. When such predictions are violated, the observer must update the representation of the mental state of the actor.

Acknowledgments

We are grateful for the assistance of the radiographers. This work was supported by a EU Fifth Framework Program (Contract No. QLG3-CT-2002-00746) and by the Wellcome Trust.

References

- Adolphs, R., 1999. Social cognition and the human brain. *Trends Cogn. Sci.* 3, 469–479.
- Adolphs, R., 2003. Cognitive neuroscience of human social behaviour. *Nat. Rev., Neurosci.* 4, 165–178.
- Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of the STS region. *Trends Cogn. Sci.* 4, 267–278.
- Ashburner, J., Friston, K., 1997. Multimodal image coregistration and partitioning—A unified framework. *NeuroImage* 6, 209–217.
- Barresi, J., Moore, C., 1996. Intentional relations and social understanding. *BBS* 19, 107–154.
- Bartsch, K., Wellman, H., 1995. *Children Talk About Mind*. Oxford Univ. Press, New York.
- Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., Freund, H.J., 2001. Action obser-

- vation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur. J. Neurosci.* 13, 400–404.
- Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev., Neurosci.* 3, 201–215.
- Decety, D., Grèzes, J., 1999. Neural mechanisms subserving the perception of human actions. *Trends Cogn. Sci.* 3, 172–178.
- Decety, J., Grèzes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., Grassi, F., Fazio, F., 1997. Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain* 120, 1763–1777.
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 1992. Understanding motor events: a neurophysiological study. *Exp. Brain Res.* 91, 176–180.
- Downar, J., Crawley, A.P., Mikulis, D.J., Davis, K.D., 2001. The effect of task relevance on the cortical response to changes in visual and auditory stimuli: an event-related fMRI study. *NeuroImage* 14, 1256–1267.
- Downar, J., Crawley, A.P., Mikulis, D.J., Davis, K.D., 2002. A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *J. Neurophysiol.* 87, 615–620.
- Duvernoy, H.M., 1999. *The Human Brain. Surface, Blood Supply and Three-dimensional Sectional Anatomy.* Springer, New York.
- Farrer, C., Frith, C.D., 2002. Experiencing oneself vs. another person as being the cause of an action: the neural correlates of the experience of agency. *NeuroImage* 15, 596–603.
- Fogassi, L., Gallese, V., Fadiga, L., Rizzolatti, G., 1998. Neurones responding to the sight of goal-directed hand/arm actions in the parietal area PF7b of the macaque monkey. *Soc. Neurosci. Abs.* 24, 257.5.
- Friston, K.J., Fletcher, P., Josephs, O., Holmes, A., Rugg, M.D., Turner, R., 1998. Event-related fMRI: characterizing differential responses. *NeuroImage* 7, 30–40.
- Frith, C.D., Frith, U., 1999. Interacting minds—A biological basis. *Science* 286, 1692–1695.
- Gallagher, H.L., Frith, C.D., 2003. Functional imaging of 'theory of mind'. *Trends Cogn. Sci.* 7, 77–83.
- 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, 11–21.
- Gallese, V., 2003. The Manifold nature of interpersonal relations: the quest for a common mechanism. *Phil. Trans. R. Soc. London, Ser. B* 358, 517–528.
- Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. *Trends Cogn. Sci.* 2, 493–501.
- Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. *Brain* 119, 593–609.
- Gordon, R.M., 1995. Theories of Theories of Mind. In: Carruthers, P., Smith, P.K. (Eds.), Cambridge Univ. Press, UK, pp. 11–21.
- Grafton, S.T., Arbib, M.A., Fadiga, L., Rizzolatti, G., 1996. Localization of grasp representations in humans by positron emission tomography. 2. Observation compared with imagination. *Exp. Brain Res.* 112, 103–111.
- Grèzes, J., Decety, J., 2001. Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp.* 12, 1–19.
- Henson, R.N., Price, C.J., Rugg, M.D., Turner, R., Friston, K.J., 2002. Detecting latency differences in event-related BOLD responses: application to words versus nonwords and initial versus repeated face presentations. *NeuroImage* 15, 83–97.
- Hommel, B., Musseler, J., Aschersleben, G., Prinz, W., 2001. The Theory of Event Coding TEC: a framework for perception and action planning. *Behav. Brain Sci.* 24, 849–878.
- Jeannerod, M., 1994. The representing brain: neural correlates of motor intention and imagery. *BBS* 17, 187–245.
- Jeannerod, M., 2001. Neural simulation of action: a unifying mechanism for motor cognition. *NeuroImage* 14, S103–S109.
- Knoblich, G., Prinz, W., 2001. Recognition of self-generated actions from kinematic displays of drawing. *J. Exp. Psychol. Hum. Percept. Perform.* 27, 456–465.
- Knoblich, G., Seigerschmidt, E., Flach, R., Prinz, W., 2002. Authorship effects in the prediction of handwriting strokes: evidence for action simulation during action perception. *Q. J. Exp. Psychol., A* 55, 1027–1046.
- Lieberman, A.M., Mattingly, I.G., 1985. The motor theory of speech perception revised. *Cognition* 21, 1–36.
- Macmillan, N.A., Creelman, C.D., 1991. *Detection Theory: A User's Guide.* Cambridge Univ. Press, Cambridge.
- Orliaguet, J.P., Kandel, S., Böe, L.J., 1997. Visual perception of motor anticipation in cursive handwriting: influence of spatial and movement information on the prediction of forthcoming letters. *Perception* 26, 905–912.
- Pelphrey, K.A., Singerman, J.D., Allison, T., McCarthy, G., 2003. Brain activation evoked by perception of gaze shifts: the influence of context. *Neuropsychologia* 41, 156–170.
- Prinz, W., 1997. Perception and action planning. *Eur. J. Cogn. Psychol.* 9, 129–154.
- Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L., 1996. Localization of grasp representations in humans by PET: 1. Observation versus execution. *Exp. Brain Res.* 3, 131–141.
- Rizzolatti, G., Fogassi, L., Gallese, V., 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat. Rev., Neurosci.* 2, 661–670.
- Ruby, P., Decety, J., 2001. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci.* 4, 546–550.
- Runeson, S., Frykholm, G., 1983. Kinematic specification of dynamics as an informational basis for person- and-action perception: expectation, gender recognition, and deceptive intention. *J. Exp. Psychol. General* 112, 585–615.
- Shiffrar, M., Freyd, J.J., 1993. Apparent motion of the human body. *Psychol. Sci.* 1, 257–264.
- Viviani, P., Stucchi, N., 1992. Biological movements look uniform: evidence for motor-perceptual interactions. *J. Exp. Psychol. H. Percept. Perf.* 18, 603–623.
- Winston, J.S., Strange, B.A., O'Doherty, J., Dolan, R.J., 2002. Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nat. Neurosci.* 5, 277–283.
- Wolpert, D.M., Ghahramani, Z., Flanagan, J.R., 2001. Perspectives and problems in motor learning. *Trends Cogn. Sci.* 5, 487–494.
- Wolpert, D.M., Doya, K., Kawato, M., 2003. A unifying computational framework for motor control and social interaction. *Philos. Trans. R. Soc. Lond., B Biol. Sci.* 358, 593–602.