NeuroImage 54 (2011) 1406–1414

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Anatomical and temporal architecture of theory of mind: A MEG insight into the early stages

Damien Vistoli [*,1], Eric Brunet-Gouet [1], Emilie Baup-Bobin,
Marie-Christine Hardy-Bayle, Christine Passerieux

EA 4047, Université de Versailles Saint Quentin, Service de Psychiatrie Adulte, Centre Hospitalier de Versailles, Fondation FondaMental, 177 route de Versailles, 78150 Le Chesnay, France

ARTICLE INFO

Article history:
Received 11 June 2010
Revised 27 August 2010
Accepted 8 September 2010
Available online 17 September 2010

Keywords:
Attribution of intentions
Magneto-encephalography
Theory of mind
Mirror neuron system

ABSTRACT

Representation of others' actions and mental states leads to the activation of several brain networks: the mentalizing and the "mirror neuron" systems as well as a "low level" social perception component. However, respective activations of the regions belonging to these networks remain unknown with respect to chronometrical data when static drawing stimuli are presented. To determine anatomical and temporal characteristics of theory of mind processes, magnetic signals were measured in 21 subjects during a validated nonverbal attribution of intentions task. Minimum norm estimation provides chronometric and localization data showing that regions known to be involved in the mentalizing, "mirror neuron" and social perception networks have simultaneous activations between 100 and 700 ms post-stimulus, a period which may be thought as corresponding to early stages of social processes. Among some regions, different profiles as well as modulations regarding experimental conditions suggest functional distinctions between these structures, pleading for a cooperative nature of these networks. While the left temporo-parietal area and superior temporal sulcus seem more specialized in social cues coding, we demonstrate that their right homologues, as well as the right inferior parietal cortex, are preferentially recruited during attribution of intentions stimuli compared to scenarios based on physical causality from 200 to 600 ms.

Introduction

Social cognition encompasses the processes contributing to our ability to understand and interact with others. One of the most important and complex strategies used by people to represent and predict others' behavior consists in raising hypotheses on their putative mental states, a capability called "theory of mind" (ToM) or "mentalizing". It is widely acknowledged that ToM relies on a brain network whose extension has increased with new findings but which is still distinct from other forms of cognitive processes such as executive functions. A large proportion of studies based on mental state attribution paradigms (i.e. based on comparison of conditions with mental states inferences with non-mentalist conditions) reports activations in temporo-parietal junction (TPJ) and the anterior part of the paracingulate cortex (aPCC, also called medial prefrontal cortex), with a notable contribution of the right hemisphere (for reviews, see Brunet-Gouet and Decety, 2006; Frith and Frith, 2006; Van Overwalle, 2009). This network is referred to as the mentalizing system.

Another set of regions is implicated in "recognizing the goal of a perceived action by matching it to a representation in our memory of our own actions" as reviewed by Van Overwalle and Baetens (2009). This network may be called the "mirror neuron" system (MNS) by reference to findings of specific neurons that responded to both the observation of simple hand movements and the initiation of a similar movement in macaque monkeys (Rizzolatti et al., 1996). The human MNS encompasses the inferior parietal sulcus (IPS), inferior parietal lobule (IPL) and inferior frontal cortex (IFC) (Rizzolatti and Craighero, 2004). The MNS has been shown to respond to human agents in action (Montgomery et al., 2007) and to the intention of an observed action (Iacoboni et al., 2005). In the present study, such an anatomical distinction of functional networks will be used and referred to as the "MNS regions".

Finally, the posterior part of the superior temporal sulcus (pSTS) is also reported to be activated in biological motion perception (see Allison et al., 2000 for a review; Materna et al., 2008) and, especially in the right hemisphere, in intentional action understanding (Pelphrey et al., 2004; Pelphrey and Morris, 2006; Saxe et al., 2004). Interestingly, Proverbio and collaborators found significant activations in the pSTS, as well as in the MNS, even when photographs of human characters were used as stimuli (2009). From these different results, it may be hypothesized that the activation in the MNS regions reflects a low-level processing allowing a mapping of others' behavior on our own action repertoire whereas the mentalizing system involves a high-level inferential mechanism.

However, the contribution of these different systems and their interactions during social cognition remain unclear. Van Overwalle

---

[*] Corresponding author. Fax: +33 1 39 63 90 80.
E-mail address: damien.vistoli@gmail.com (D. Vistoli).

[1] These authors have equally contributed to this work.

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.09.015

and Baetens (2009) note the lack of "evidence about precisely how these two systems [mentalizing and MNS] may cooperate and inform each other". Methodologically speaking, the vast majority of neuroimaging studies has explored brain correlates of ToM with either positron emission tomography (PET) or functional MRI (fMRI). These techniques are powerful to localize brain responses and to bring into light the influence of conditions on the amplitude of activations using experimental designs based on cognitive subtraction. However, temporal resolution is limited to seconds and does not allow finding at which stage each brain system intervenes which is a crucial information to the current debate on the functional relations between these systems.

To overcome these limitations, in the present study we propose the use of magnetoencephalography (MEG) which benefits from an excellent .8 ms resolution while maintaining a centimetric precision. It is necessary to note that in ToM domain, no data on the characteristics of MEG signals and the efficient way to adapt paradigms are available yet. In a step by step approach, we chose to use a task previously validated (Brunet et al., 2000) and replicated in neuroimaging conditions (Ciaramidaro et al., 2007; Völlm et al., 2006; Walter et al., 2004) and pathology assessment (Benedetti et al., 2009; Brunet et al., 2003; Walter et al., 2009). These earlier studies showed that this task based on comic-strips, requiring attribution of intentions to human figures, elicited robust hemodynamic and metabolic responses in several regions commonly reported in ToM literature such as the right aPCC, pSTS, temporal pole (TP), orbito-frontal cortex (OFC), bilateral TPJ and right IFC. Furthermore, the procedure used allowed varying the nature of the logical inferences by contrasting sequences with intentions (AI), sequences with human characters involved in physical events (PCCH), and sequences only depicting objects (PCOB). Here, MEG signals were recorded while subjects were presented with a similar material. Minimum norm algorithm was used in order to provide localization information (Baillet et al., 2001).

Having laid theses methodological bases which improve the level at which we can examine brain activity, several questions and hypotheses were raised and implemented experimentally:

#1: Attribution of intentions (AI) to others, as a core component of ToM, is based on inferences about mental states leading to measurable activations in the mentalizing network, including the TPJ and the aPCC, preferentially in the right hemisphere. AI will elicit stronger activations in these regions compared to stimuli depicting physical causality with human characters (PCCH).

#2: Does the observation of static and symbolic series of drawings involving attribution of intentions to characters in action elicit activations within all or parts of the MNS regions? As participating to a low-level non-inferential mechanism to represent and understand others, MNS regions, if activated, should precede that of the mentalizing system as defined in #1.

#3: Mentalization from observation of visual sequences requires fast detection and extraction of social cues related to the presence of human characters. Activity should be found in bilateral pSTS with a cumulative effect of characters and intentions, especially in the right hemisphere. From other authors (Hirai et al., 2003; Jokisch et al., 2005), we expect that the pSTS activation occurs before 300 ms post-stimulus.

Material and methods

Subjects

Twenty-one healthy volunteers (14 males), aged from 21 to 40 years old (mean age: 28; sd: 5.74), right-handed, native French speakers with normal or corrected to normal vision were included. All of them gave their informed consent in agreement with the French Ethical Committee (Comité de Protection des Personnes).

Task

The material consisted of 74 comic-strips each made of four black and white pictures. The images were extracted from previous works (Brunet et al., 2000), cropped and simplified (some graphic details were removed) to maintain subject's attention on the center of the images in order to reduce eye movements. The paradigm was modified to fit with a sequential presentation of the stimuli that allowed extracting time courses locked to pictures presentation. Each comic strip (or trial) consisted of four pictures appearing in a short succession on a white screen placed 1 m in front of the participants (visual angle = 10°). Subjects were required to answer on a two-button pad (right hand) whether the fourth picture was a logical ending of the sequence or not. Each of the first three pictures were presented for 2500 ms. The fourth picture was surrounded by a red border to facilitate its identification and was shown until the subject provided an answer or for a maximum of 4000 ms. Pictures were separated by a blank screen of random duration (between 300 and 500 ms). A fixation-cross appeared at the center of the screen during 1000 ms between trials to focus subject's gaze and minimize eye-artifacts. A schematic representation of the design is given in Fig. 1.

Three experimental conditions were proposed in order to vary the type of causality (i.e. intentional vs. physical) and the presence of characters (i.e. presence of human figures vs. presence of objects only): 1) attribution of intention (AI), 2) physical causality with characters (PCCH) and, 3) physical causality with objects (PCOB, i.e. without characters). In the AI condition (34 trials), participants saw short comic strips depicting one (or two) character whose behavior was driven by a specific intention. In this condition, participants had to infer the agent's intention to understand the scenario. In order to make sure that participants adopt an intentional stance in this condition, they were instructed to determine whether the last picture of each scenario was congruent or not regarding characters' goals. Fig. 2 gives an example of an AI comic-strip. In PCCH and PCOB (20 comic-strips each), only physical and mechanical properties of human bodies or objects were involved in the causality and no mental state attribution was promoted. To avoid biases due to the different number of stimuli in AI and physical causality conditions, 20 AI comic-strips were randomly selected among the 34 and were used for comparisons between AI and the two other conditions.

Before beginning MEG recordings, participants performed a short training task of 10 trials. Then, they were administered in separate sessions either the AI or the physical causality comic-strips. The physical causality session contained an equal proportion of PCCH and PCOB comic-strips, randomly presented. The whole material was presented twice. The order of comic-strips within a session as well as the order of the session type (intentional or physical causality) and the congruity of the fourth picture was counterbalanced across subjects. During the task, we recorded the subjects' responses using E-prime software (www.pstnet.com) to check their proportion of correct responses.

MEG recording

Magnetic event-related signals were recorded with a CTF Omega 151 apparatus (CTF Inc., Vancouver) at the MEG Center, Hôpital de la Pitié-Salpétrière (Paris, France). This system allows simultaneous recording of 151 MEG channels covering the whole head. The signals were continuously recorded and were digitized on-line at a 1250 Hz sampling rate with a 200 Hz low-pass filter. 3D head referential was defined using three magnetic coils attached to three anatomical landmarks, i.e. nasion and preauricular points. Locations of these coils were checked before each recording session to measure the subject's

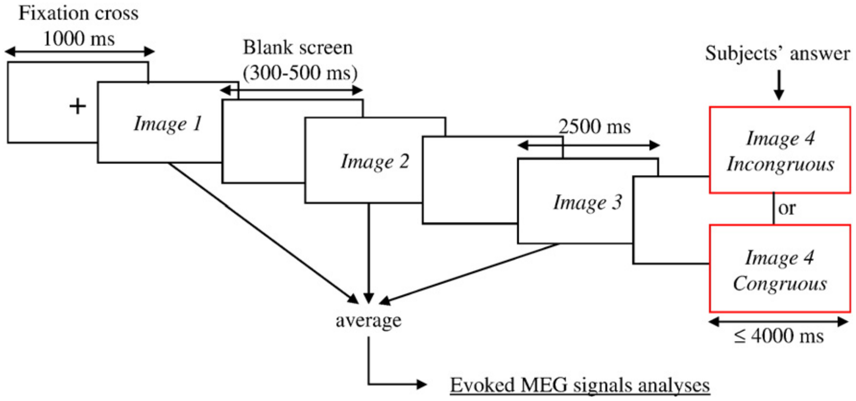

Fig. 1. Design of a comic-strip presentation. Each comic-strip contains 4 images presented successively. Each image is displayed for 2500 ms and separated from the next image by a blank screen of variable duration (from 300 to 500 ms). Image 4 is either congruous or incongruous (counterbalanced) with respect to the three preceding images. Subjects have to indicate on a two-button pad whether the 4th image is congruous or not. Image 4 disappears with the subject's answer and is presented for a maximum of 4000 ms. Each comic-strip is separated from the other by a fixation cross presentation during 1000 ms. The first three images are merged for subsequent evoked MEG signal analysis.

head position relative to the magnetometer. Heart activity was recorded using two electrodes placed around the collarbone and the lower abdomen and electrical potentials generated by eye movements were recorded with four electrodes placed around the right eye.

MEG signals processing

MEG artifacts induced by heart activity was corrected on each channel using linear regression. The raw MEG signals were filtered off-line with a .4–35 Hz band-pass filter. Each trial was reviewed by the experimenter (DV) and those contaminated by eye-movement artifacts were removed. Eight subjects out of 29, who had more than 50% of signal removed this way, were excluded from the study. To avoid contaminations by decision and motor-preparation, processes which could occur during the fourth picture presentation, we processed evoked magnetic signals of the first three pictures (signals of the fourth pictures were not analyzed here). Time origin was set to pictures' onset and the baseline corresponded to the preceding 200 ms.

Localization of the neural generators was performed using a linear minimum norm algorithm implemented in BrainStorm (http://neuroimage.usc.edu/brainstorm). White and grey matter segmentation was performed on the MNI standard brain (Montreal Neurological Institute), providing the white/grey matter interface. Cortical surface was decimated to 15,000 sources evenly distributed on this surface, i.e., at each node of the cortical tessellation. Sources' orientation was constrained perpendicularly to the white/grey matter interface. This cortical surface served as the solution space for the estimated current generators, constraining the MEG solutions to the white/grey matter interface. Source estimation was performed using a spherical homogenous conductivity head model. Source localization analysis was transformed into each subject's head coordinate system using the three coils' positions. Minimum norm approach was used as described in Baillet et al. (2001), to estimate the respective contributions of sources to the recorded evoked magnetic signals (inverse problem) at each time step during the −200 to +1000 ms time window.

Selection of regions of interest

Averaged *z*-maps in the three conditions were used to define a set of regions of interests (ROIs) allowing a dramatic reduction of statistical comparisons. To distinguish evoked signals from noise, time-series of each source were normalized into *z*-scores with regard to the mean and standard deviation estimated from the baseline epoch. This analysis was performed for each subject and for each condition, and then averaged into *z*-maps across subjects, for each condition. We ensured the independence between ROIs selection and ROIs comparisons between conditions. To avoid any bias, ROIs were defined on the common activity across all conditions, i.e., the averaged data of AI, PCCH and PCOB. Each ROI had a spherical shape constrained by the cortical surface and consisted of 20 reconstructed sources, centered on a local maximum of activation, that had averaged *z*-scores superior to 4 for more than 10 ms.

Statistical analyses

Time-courses of each ROI were then extracted from raw data (non-normalized) for each condition and subject, leading to one time-course per ROI, per condition and per subject. We performed

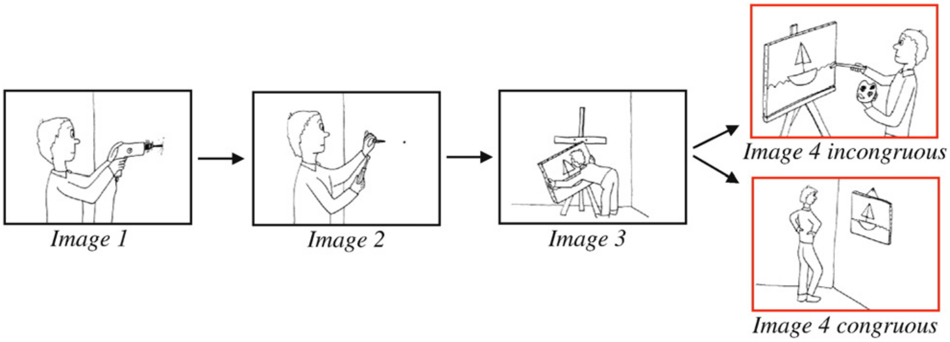

Fig. 2. Example of an AI (attribution of intention) comic-strip. During the first three images, subjects have to construct a representation of the character's intention in order to give a correct response during the 4th image. In this example, subjects have to infer the character's intention: to hang something (the picture) on the wall.

two types of analyses. First, we focused on the relative activation amplitude of each region according to the experimental conditions. As mentioned previously, the number of stimuli was not equal in AI versus PCCH and PCOB conditions, so the number of trials in AI was reduced to 20 so that conditions were equilibrated. Averaged magnetic activations were computed with a temporal resolution reduced to 50 ms epochs (62 samples per epoch), between 200 and 600 ms post-stimulus. Two comparisons were performed with a paired permutation test not assuming a specific distribution of data: AI versus PCCH (intention effect) and PCCH versus PCOB (character effect). Alpha risk was set to .001 to avoid type-I errors due to multiple comparisons. We also report results at a less stringent threshold (.05) as they may also be relevant and make sense.

Second, we focused on the chronometric properties of activities in the AI condition (raw data). Two characteristics of evoked brain responses were extracted for each couple of subject and ROI: 1) the latency of the peak of activation, 2) the latency at which magnetic signal reached 50% of its maximum amplitude (we will refer to this latency as FWHM latency). To reduce noise effect on these measures, latencies were computed from Gaussian functions that were fitted on the subjects*ROI time-courses (i.e. unconstrained derivative-free norm minimization with four variables on functions $y\ (a, \text{mu}, \text{sigma}, y0) = a * \text{Normal Function}\ (\text{mu}, \text{sigma}) + y0$). The algorithm was set to reject automatically bad fits (i.e. negative going bell curves, or estimated latency outside the 50 to 1000 ms time-window, or absence of convergence). Effects of ROIs (those reported in Table 1) on the latencies were tested using one-way ANOVAs. In addition, hypotheses-driven ANOVAs were performed to compare latencies of regions belonging to ToM, MNS networks as well as pSTS within each hemisphere. Two factors were entered in the model: 1) hemisphere: right vs. left, 2) group of ROIs: TPJ vs. aPCC vs. IPL-IPS vs. pSTS.

Results

Participants completed the task with a mean accuracy of 91.5%. The accuracy for the AI condition reached 90.2%, while the PCCH and PCOB conditions reached 90.4% and 94.1% respectively. A one-way ANOVA showed an absence of condition effect ($F(2;54) = 2.03$, $p = .14$). Mean reaction time for the average of congruous and incongruous trials was 1640 ms in AI, 1523 ms in PCCH and 1464 ms in PCOB condition. Here again, a one-way ANOVA revealed no effect of experimental conditions (AI vs. PCCH vs. PCOB) on mean reaction times ($F(2;60) = .96$, $p = .39$). These results demonstrate an equivalent level of complexity between experimental conditions and rules out any biases in terms of cognitive effort in the following interpretations of comparisons of interest.

Localizing the "AI network" (ROIs definition)

Minimum norm estimation was performed to determine the neural network recruited during attribution of intentions to others. The ROI criteria were met in a set of 20 regions: the bilateral occipital lobes, TP, pSTS extended to the supramarginal gyrus in the left hemisphere, IPL, IPS, TPJ, OFC and the paracingulate cortex in its anterior (aPCC) and middle part (mPCC). Moreover, some regions showed unilateral activity such as the right fusiform gyrus, and the right middle temporal gyrus (their left homologues did not reach the z threshold). Some ROIs localizations are listed in Table 1B. Fig. 3 represents some of the regions activated in the AI condition (an animated illustration is available at http://www.predicto.fr/meg/ai_21subjects.avi). IPS and IPL exhibited very similar time-courses, suggesting that they possibly belong to a common parietal cluster.

Effect of condition on ROIs' time courses

To test the effect of condition, the mean amplitude of each ROI between AI and PCCH conditions and between PCCH and PCOB conditions were compared at different time-intervals. Fig. 4 illustrates the cortical activations of the contrast between AI and PCCH and between PCCH and PCOB at different time steps. Permutation test showed that the right IPL as well as the right pSTS and TPJ were significantly more activated during AI than PCCH between approximately 200 and 600 ms (see Table 1). In addition, the right IPS showed enhanced activation in AI compared to PCCH in the 240–540 ms time-window. Moreover, we found that the left TPJ, left IPL and left pSTS were strongly recruited when stimuli contained characters (PCCH) compared to objects only (PCOB) during the 350–450 ms time-

Table 1
AI network localizations, chronometry and modulation by condition. (A) ROIs names. (B) Talairach coordinates (Talairach and Tournoux, 1988). (C) Mean and standard deviation (sd) of FWHM latencies in milliseconds. (D) Mean and standard deviation (sd) of peak latencies in ms. (E) Mean peak amplitudes in $pA.m^{-1}$. (F) Effect of condition on mean amplitude. (G) Time intervals of significant differences ($p<.05$).

A/ROIs	B/Talairach coordinates			C/FWHM latencies in AI condition (sd)	D/Peak latencies in AI condition (sd)	E/Peak amplitude in AI condition	F/Condition effect on amplitude	G/Significant time-intervals
	x	y	z					
Temporal lobe								
R pSTS	44	−53	13	244 (80)	450 (105)	8.7	AI>PCCH	200–600
L pSTS	−63	−39	15	247 (74)	397 (116)	9.3	PCCH>PCOB	330–430
Frontal lobe								
R aPCC	7	36	17	205 (64)	396 (33)	3.1	PCCH>PCOB	290–340
L aPCC	−10	41	6	272 (142)	415 (141)	4	PCCH>PCOB	200–350
R OFC	20	23	−11	268 (228)	395 (250)	5.2	PCCH>PCOB	240–290
L OFC	−18	15	−17	277 (144)	436 (146)	5.1	PCCH>PCOB	300–350
R mPCC	8	−17	45	260 (62)	440 (89)	5.6	–	–
L mPCC	−7	−11	45	272 (84)	434 (117)	4.2	PCCH>PCOB	290–340
Parietal lobe								
R IPS	30	−40	36	264 (59)	401 (89)	6.6	AI>PCCH	240–540
L IPS	−29	−44	39	310 (100)	441 (91)	4.7	AI>PCCH	540–600
							PCCH>PCOB	390–440
R IPL	45	−50	45	256 (84)	404 (102)	11	AI>PCCH	200–600
L IPL	−36	−55	44	270 (69)	428 (85)	7.5	PCCH>PCOB	340–440
R TPJ	58	−55	25	218 (77)	430 (81)	10	AI>PCCH	240–600
L TPJ	−49	−55	30	246 (83)	391 (94)	11	PCCH>PCOB	390–440

Abbreviations: L: Left; R: Right; pSTS: posterior part of the superior temporal sulcus; aPCC: anterior paracingulate cortex; OFC: orbitofrontal cortex; mPCC: middle paracingulate cortex; IPS: intraparietal sulcus; IPL: inferior parietal lobule; TPJ: temporoparietal junction.

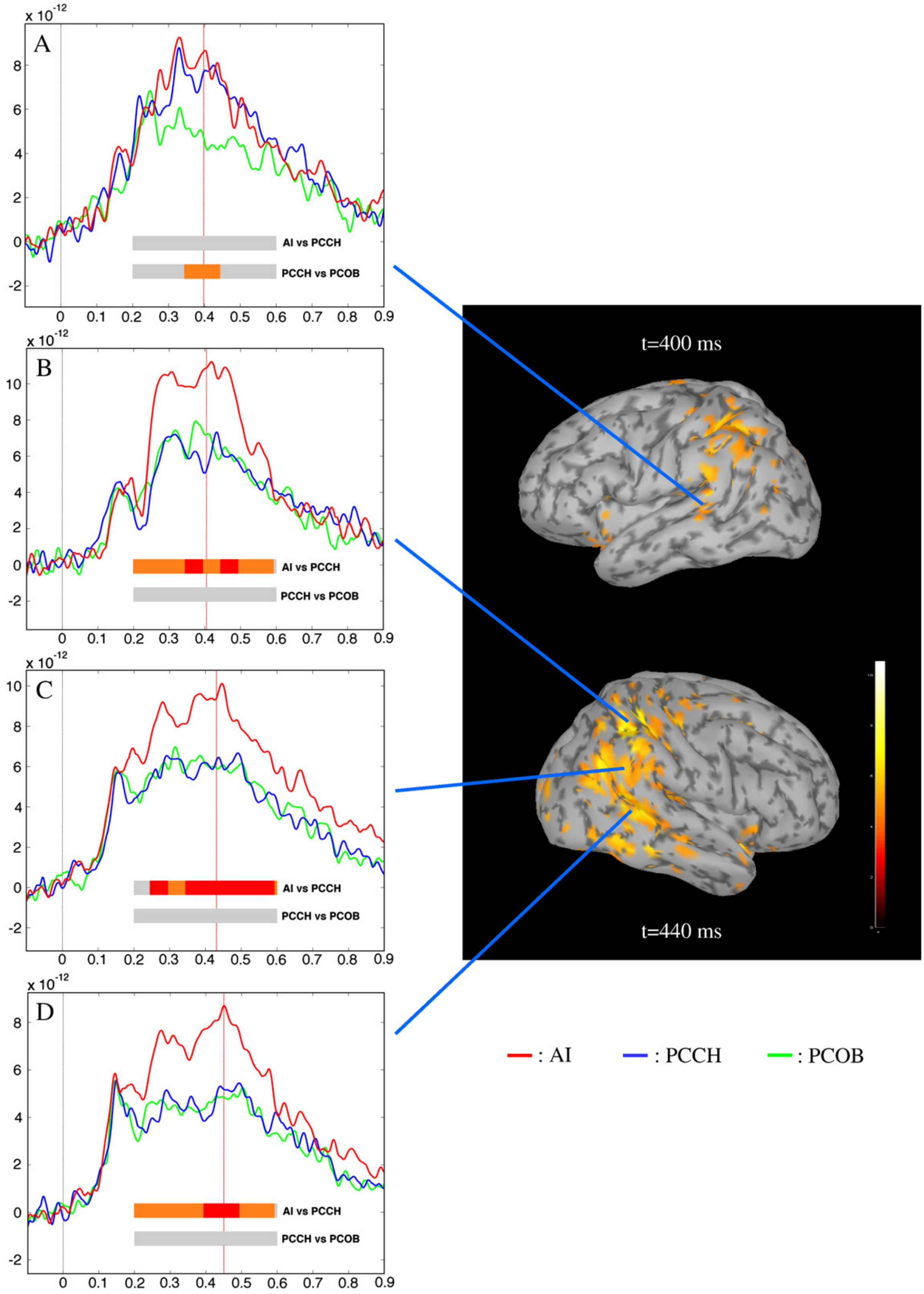

Fig. 3. Reconstruction of cortical activations and time-courses of selected ROIs. Right and left cortical surface with normalized magnetic activation at 440 ms (right hemisphere) and 400 ms post-stimulus (left hemisphere) in AI condition. Colour scale: arbitrary units (z values). A, B, C, D: Time-courses (raw data) of the left pSTS (A), right IPL/IPS cluster (B), right TPJ (C) and right pSTS (D) in the three conditions (AI in red, PCCH in blue and PCOB in green). Horizontal axis in seconds (−0.1 to 0.9 s). Time origin corresponds to pictures' presentation. Vertical axis in A.m^{-1}. Red lines represent peak latencies. Horizontal coloured bars represent time intervals in which comparisons between conditions are found significant: orange stands for $p<.05$ and red stands for $p<.001$ (paired permutation test). Upper bars: AI versus PCCH comparison; lower bars: PCCH versus PCOB comparison.

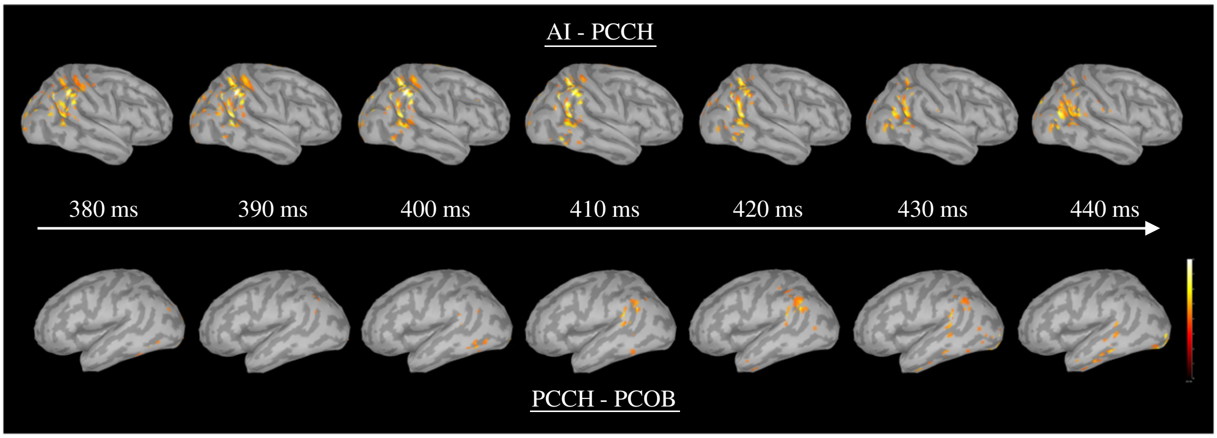

Fig. 4. Reconstructed cortical activations for AI vs. PCCH and PCCH vs. PCOB contrasts at different time-steps. Up: Right cortical surface with raw magnetic activation in regions showing more response in AI compared to PCCH condition. Down: Left cortical surface with raw magnetic activation in regions showing more response in PCCH than in PCOB condition. Activations were shown every 10 ms from 380 ms to 440 ms post-stimulus onset. Colour scale from 0 (dark red) to 8 pA.m^{-1} (yellow/white).

interval approximately (let's note however that the significance of the differences observed in the left pSTS and TPJ was uncorrected, i.e. at a .05 threshold, during two 50 ms epochs). The right aPCC as well as the left OFC and mPCC showed the same effect from 300 to 350 ms. Finally, the left aPCC exhibited this effect between 200 and 350 ms; the right OFC, from 240 to 290 ms. Fig. 3 displays the time-courses of some ROIs in which significant differences between AI and PCCH or between PCCH and PCOB were found.

ROIs' chronometry in AI condition

MEG data allowed the determination of chronometry in each region for the AI condition. Measures of latencies (FWHM and peak) for each ROI were reported in Table 1C and D. We noticed that activity started in the occipital lobe and spread in posterior regions including right fusiform area and TPJ. However, the pattern of activations seemed to be more complex than a posterior–anterior chronology. For instance, we observed that several posterior areas such as the left pSTS and bilateral IPL were co-activated with anterior regions (see Table 1C).

These general observations on chronometric variables appear relevant to our understanding of early ToM mechanisms. However, no significant influence of ROI on peak or FWHM latencies was found (respectively, $F(13;197) = .46$, $p = .94$ and $F(13;197) = .84$, $p = .61$). Taking into consideration hypothesis #2, two-ways ANOVAs were performed with hemisphere and group of regions as factors. No effect of ROIs groups ($F(3,143) = .19$; $p = .90$) and hemisphere ($F(1,143) = .36$; $p = .54$) on peak latencies was found. However, while no effect of group of ROIs was found ($F(3,143) = 2.20$; $p = .09$), FWHM latencies exhibited differences given the hemisphere ($F(1,143) = 4.42$; $p < .05$). Post-hoc LSD contrasts demonstrated that the right hemisphere had earlier onset of activation latencies than the left ($p < .05$).

Discussion

Inferring mental states such as desires and intentions to other people is a core ability required for appropriate human communication and fluent navigation into the social world. Several studies have elucidated the anatomical bases of this capacity with either hemodynamic, metabolic measures or neuropsychological investigations. In the present study, we add to this knowledge base by providing insights into the integration between the temporal dynamics and the spatial localization of ToM. MEG was used to measure the cortical magnetic response evoked by attribution of intentions to human characters represented in sequences of pictures. Methodologically speaking, results demonstrate the relevance of MEG to address new issues in social cognition. Remarkably, brain structures involved in ToM as seen in Table 1A are in accordance with those found with hemodynamic or metabolic measures (Van Overwalle and Baetens, 2009). For example, the comparison of Talairach coordinates (Talairach and Tournoux, 1988) of the right posterior regions (pSTS, TPJ and IPL/IPS) found in our study (see Table 1A) with those reported in the meta-analysis of Van Overwalle and Baetens (2009), demonstrated a difference approximately equal (IPL/IPS) or less (pSTS and TPJ) than 1 cm.

The main results show that, in agreement with hypothesis #1, the mentalizing system was mobilized during the task. Main significant activations were found between 100 and 700 ms, a period which may be thought as corresponding to early stages of social processes. As we will discuss later, data analyses described in this article did not allow addressing the issue of later cognitive components (i.e. occurring after 700 ms) and leave open hypotheses on brain activities related to more complex and less stimulus-locked aspects of ToM. Another important finding, in line with hypothesis #2, was the presence of activations in the parietal cortex during the same time-window, as indexed by peak latencies, which suggests that the complementary processes of inference making and shared-representations (MNS) are elicited during the task. Moreover, schematic "sequential" model positing the activation in the MNS regions as preceding those of the high-level mentalizing system did not receive support from these results as no significant latencies difference between these two systems was observed. Finally, we found earlier right hemisphere recruitment compared to the left hemisphere, reflecting its predominant role in attribution of intentions.

Interestingly, compared to previous works, we used an explicit instruction in the attribution of intention condition. Although this issue is largely understudied in the ToM domain, the strong similarities between the brain regions we found and those reported in previous ToM studies suggest that no major anatomical difference is induced by the use of an implicit versus explicit instruction. This observation is in line with reports from Iacoboni et al. (2005) who compared the two types of instructions in a task requiring intentional inferences from the observation of video-clips of simple actions. These authors reported no significant differences in the mirror neuron system but found nonspecific increased activation in the left mesial and cingulate prefrontal cortex in the explicit condition. According to these authors, the absence of difference within the mirror neuron system itself is a reflect of automated processing, poorly modulated by top-down influences. Similarly, in a recent review of ToM neuroimaging studies, Carrington and Bailey (2009) reported that, to date, no experimental evidence has been demonstrated about the influence of the instructions on the brain regions involved in ToM tasks. From the results presented here, we suggest that high temporal resolution, as provided by MEG, is an asset to address these questions as top-down influences may affect significantly brain responses at different epochs. Developing measures of brain activities occurring after 700 ms post-stimulus while varying the instructions would be relevant in future studies.

Attribution of intentions to others involves contribution of the mentalizing system

In agreement with hypothesis #1, magnetic activations were found in bilateral TPJ and aPCC in the 100–700 ms time-window. Previous fMRI and PET studies have included these regions in the mentalizing system which is involved in ToM high-level reflective processes (Carrington and Bailey, 2009; Frith and Frith, 2006; Van Overwalle and Baetens, 2009). Interestingly, we found that these different structures exhibited distinct responses to conditions, bringing into light the intrinsic complexity of the mentalizing system.

The major role of the right TPJ in the attribution of intentions was supported by our finding of greater magnetic activation in attribution of intention (AI) comic-strips than in stimuli depicting physical causality with characters (PCCH). Furthermore, statistical analysis showed that the intention effect may start around 240 ms post-stimulus. This is in accordance with the results of Van der Cruyssen (2009) who reported, using electroencephalography (EEG), early right temporo-parietal area participation in goal inferences. These findings may be interpreted within the scope of several theoretical proposals. Some authors argue that the right TPJ plays a specific high-level function in constructing a coherent model of others' mind independent of one's own mental states (Saxe and Kanwisher, 2003; Saxe and Wexler, 2005). An alternative view states that this region is involved in attention reorientation toward salient stimuli (Corbetta and Shulman, 2002; Decety and Lamm, 2007; Mitchell, 2008), a low-level process that should have been triggered in our study. Further researches on the influence of attentional parameters (i.e. use of distracters or double tasks) on TPJ time-courses will be necessary to disentangle these hypotheses.

Interestingly, when we compared the two TPJ areas, we observed that during the 390–440 ms time-window, right and left TPJ showed different modulation with regards to experimental conditions. While the right TPJ exhibited enhanced activation in AI compared to PCCH, its left homologue showed more activation in PCCH than PCOB. This suggests a functional dissociation between the two TPJ regions during attributing mental states from observation of a character's action: the left TPJ seems to respond to the presence of a character whereas the right TPJ participates in intentions coding.

Concerning the aPCC, although we found an early recruitment as Van der Cruyssen et al. (2009) did, this involved only a significant PCCH vs. PCOB comparison without a significant intention effect. However, the absence of such an effect was unexpected given that previous PET findings with an attribution of intention task were in support of a stronger participation of the aPCC in intention processing by comparison with understanding physical causality (Brunet et al., 2000). Moreover, it is well documented that this midline structure is implicated in ToM processes (Amodio and Frith, 2006; Gallagher and Frith, 2003; Van Overwalle and Baetens, 2009). It is worth noting that, compared to our previous investigation (Brunet et al., 2000), the characteristics of the paradigm were modified at least on two important dimensions: The pictures were presented sequentially rather than simultaneously and the subjects' task was to make a final congruity judgment instead of a choice between alternative endings. It cannot be ruled out that these changes weakened the cognitive load on strategic processes such as verification and generation of hypotheses, which could hypothetically engage the aPCC when social representations are processed. An alternative explanation would be that these strategic processes occur in a later time-window, suggesting that the early aPCC participation found in this study did not reflect the same cognitive process as the time-averaged activations recorded in previous fMRI and PET investigations. This interpretation is strengthened by the results of Liu and collaborators who measured EEG signals during a false-belief task which requires the involvement of high-level inferential processes (2004). Interestingly, they found an enhanced EEG component around 800 ms post-stimulus when participants thought about the mental states of a character, suggesting that ToM strategic processes would not occur earlier.

Attribution of intentions also generates early activations within regions belonging to the MNS

The paradigm used in this study is based on the observation of drawings depicting human characters performing intentional actions. Because this is a debated issue, we raised the question of the activation of MNS regions in such a non-ecological experimental condition. Anatomically, other authors have proposed a set of regions including the IPL, the IPS, and the IFC as the substratum of these representations (for reviews, see Agnew et al., 2007; Rizzolatti and Craighero, 2004; Van Overwalle and Baetens, 2009). In addition, Proverbio et al. (2009) reported that the MNS was implicated even when static representations were used as stimuli. Interestingly the present study shows that an early magnetic activation in bilateral IPL an IPS is found in the first 700 ms post-stimulus. Our results give some additional support for the functional dissociation among regions thought to constitute the MNS as frontal and parietal regions were differently recruited in the AI condition (IFC regions did not reach the z threshold).

Based on previous fMRI investigations, we expected that activation in frontal and parietal sites of this network would be sensitive to both the presence of human characters (Montgomery et al., 2007) and the intentionality of the observed action (Iacoboni et al., 2005). We did not find IFC activation in this study. It is unlikely that this absence is explained by the use of explicit instructions as Iacoboni et al. (2005) reported similar IFC recruitment using both implicit and explicit instructions. Although this structure is commonly found to be activated in action understanding paradigms (see Binkofski and Buccino, 2006 for a review), our result is in accordance with that of Chong et al. (2008) who found no evidence of mirror properties (i.e., action selectivity and action execution/observation co-activation) in the human IFC. Our result reinforces the interpretation of a functional dissociation between inferior frontal and parietal cortices.

Contrary to IFC, right IPS and IPL showed a stronger activation, in attribution of intention condition (AI) compared to stimuli depicting physical causality with character (PCCH). This effect was observed from 200 to 600 ms in right IPL and between 240 and 540 ms in right IPS, agreeing with recent fMRI/EEG results by Ortigue et al. (2009). There is no increased amplitude in PCCH relative to PCOB, suggesting that the role of the parietal areas is not related to the presence of human characters *per se*. Instead, our result is in support of a major role of the parietal cortex in early ToM processes. However, we cannot rule out the possibility that the parietal cortex detects low-level goals instead of high-level intentions in the AI comic-strips. To disentangle these alternatives hypotheses, future researches should control the amount of implied motion depicted in the AI and PCCH stimuli. Interestingly, fMRI results reported by Shmuelof and Zohary (2006), corroborated by Hamilton and Grafton (2008), suggested that the right IPS is implicated in the prediction of the outcome of a motor action. Our result may confirm the theoretical view that attribution of intentions from action observation could in part rely on such a mechanism (Blakemore and Decety, 2001; Shmuelof and Zohary, 2007) and that participants have used it to perform the task.

Importantly, this study is a direct example of a ToM task recruiting some regions thought to belong to the MNS in conjunction with the mentalizing system. This result reinforces the idea that these two systems are complementary rather than mutually exclusive (de Lange et al., 2008; Thioux et al., 2008). An important question is the relative onset latencies of the MNS and the mentalizing networks. Because the mentalizing system is classically categorized as subserving high-level processes (Van Overwalle and Baetens, 2009), we expected that the MNS activations would precede those of the mentalizing regions. We did not confirm this hypothesis. Indeed, we found a global overlap of activations in the two systems with peak latencies lying in the 390–450 ms time-window, and FWHM measures between 200 and 310 ms. This result suggests that the classic low versus high-level categories should be reconsidered in the light of chronometric data. Temporal overlaps between these levels may reflect a functional interaction between social perception which leads to activations of hypotheses on other's intentions and integrative/inferential processes which perform selection among these hypotheses. Nevertheless, one has to keep in mind that our analyses focused on the 200–600 ms time interval and that the mentalizing system may be solicited later, as discussed previously. Moreover, it's worth noting that several systems can be found co-activated but do not functionally interact. One promising approach in future researches would be to measure from MEG signals the functional connectivity between these systems using coherence or phase synchrony analyses.

The right posterior temporal cortex plays a major role in attribution of intentions

In line with the third hypothesis, we found bilateral pSTS activations during the 100–700 ms time-window. More precisely, FWHM latency measures showed a significant activation onset around 245 ms post-stimulus onset, with peak latencies around 400–450 ms. We found a later peak of activation than several studies (Conty et al., 2007; Hirai et al., 2003; Jokisch et al., 2005). However, these previous works focused on low-level aspects of social cognition (biological motion and eye-gaze detection) rather than mental states attribution. This methodological difference could explain the apparent discrepancies with their results.

We expected that the pSTS, especially on the right, would be responsive to both the presence of characters and the intentionality of the actions. This assumption was not verified as right pSTS activation was not significantly increased in PCCH compared to PCOB. However, this region was recruited more strongly in AI than in PCCH, during the 200–600 time interval. This result suggests that the right pSTS codes more the intentional aspect than the detection of social cues *per se*, reinforcing the findings of several neuroimaging results (Pelphrey et al., 2004; Pelphrey and Morris, 2006; Saxe et al., 2004). In agreement with this interpretation, Castelli et al. (2000) reported right pSTS involvement in subjects that observed geometrical forms which move in an intentional way.

Contrary to its right homologue, the left pSTS showed enhanced activation in PCCH compared to PCOB between 330 and 430 ms, but no difference was found compared to the AI condition. This pattern suggests that, contrary to its right homologue, the left pSTS is more implicated in the detection of characters than in the processing of intentions. This is in agreement with the classical interpretation about the pSTS function (see Allison et al., 2000 for a review; Materna et al., 2008). Unexpectedly, we found a similar pattern of response in bilateral anterior and left middle paracingulate regions (aPCC and mPCC) as well as in bilateral orbitofrontal (OFC) and left parietal cortices. To our knowledge, these regions have not been reported to be sensitive to the presence of human characters in previous studies. To date, it remains difficult to interpret such a finding and further studies, using similar technology, should investigate this issue.

Interestingly, by comparing the two pSTS regions, we observed that there was no statistical difference between their FWHM and peak latencies. Moreover, as mentioned previously, both showed different modulations regarding the experimental condition during at least the 330–430 ms time-interval. These characteristics lead us to the proposition that, as the two TPJ regions, these two structures underlie distinct but simultaneous processes during the attribution of intentions from action observation in the 100–700 ms time-period.

Conclusion and perspectives

This study used MEG associated with minimum norm algorithms to explore ToM processes in order to localize the activated brain network and to get temporal information. Although this technique suffers from the absence of methodological guidelines because of its innovative aspect, we show that it is promising to address important questions in social neurosciences. Indeed, localization results reported here demonstrate remarkable convergence with anatomical data obtained with either fMRI or PET. When compared to EEG studies that did not allow precise functional cartography, our findings add important information on the respective contribution of each region within the activated network. Importantly, while MNS, mentalizing and social perception regions exhibited different profiles of activation, we show that simultaneous involvements appear within parts of these systems that have been thought as separated and even mutually exclusive in the literature. Probably, advances in MEG signal processing, high-density array EEG and coupling fMRI with EEG will help reconsidering the current debate on these systems. As attribution of intentions to others was shown impaired in mental disorders such as schizophrenia, we suggest that the paradigm validated with MEG may prove successfully at showing patients' abnormalities in early processes of social cognition.

Funding

This work was supported by a grant of Neuropôle (DV), a grant of Fonds d'Etudes et de Recherche du Corps Médical (EBP), and an Institut National de la Santé Et de la Recherche Médicale interface contract (EBG). EA4047, CHV is member of Fondation FondaMental.

Acknowledgments

We are thankful to Bernard Renault, Antoine Ducorps, Denis Schwartz, Lydia Yahia-Cherif, Jean-Didier Lemarechal, Nathalie Georges (MEG Center, Hopital La Pitié Salpêtrière), Amélia Lemoalle, Audrey Angelard (Versailles), Philip Jackson (CIRRIS and CRULRG, Université Laval, Québec) and Pierre-Emmanuel Michon (CIRRIS, Université Laval, Québec), for their help at each stage of this experimental work. The EA 4047 (Université Versailles Saint-Quentin) — Service de Psychiatrie Adulte (Centre Hospitalier de Versailles) is a member of FondaMental (mental health foundation for research and care). Conflict of interest: none declared.

References

Agnew, Z.K., Bhakoo, K.K., Puri, B.K., 2007. The human mirror system: a motor resonance theory of mind-reading. Brain Res. Rev. 54, 286–293.

Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of the STS region. Trends Cogn. Sci. 4, 267–278.

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. Nat. Rev. Neurosci. 7, 268–277.

Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Mapping human brain functions using intrinsic electromagnetic signals. IEEE Signal Process Mag. 18, 14–30.

Benedetti, F., Bernasconi, A., Bosia, M., Cavallaro, R., Dallaspezia, S., Falini, A., Poletti, S., Radaelli, D., Riccaboni, R., Scotti, G., Smeraldi, E., 2009. Functional and structural brain correlates of theory of mind and empathy deficits in schizophrenia. Schizophr. Res. 114, 154–160.

Binkofski, F., Buccino, G., 2006. The role of ventral premotor cortex in action execution and action understanding. J. Physiol. Paris 99, 396–405.

Blakemore, S.J., Decety, J., 2001. From the perception of action to the understanding of intention. Nat. Rev. Neurosci. 2, 561–567.

Brunet, E., Sarfati, Y., Hardy-Bayle, M.C., Decety, J., 2000. A PET investigation of the attribution of intentions with a nonverbal task. Neuroimage 11, 157–166.

Brunet, E., Sarfati, Y., Hardy-Bayle, M.C., Decety, J., 2003. Abnormalities of brain function during a nonverbal theory of mind task in schizophrenia. Neuropsychologia 41, 1574–1582.

Brunet-Gouet, E., Decety, J., 2006. Social brain dysfunctions in schizophrenia: a review of neuroimaging studies. Psychiatry Res. 148, 75–92.

Carrington, S.J., Bailey, A.J., 2009. Are there theory of mind regions in the brain? A review of the neuroimaging literature. Hum. Brain Mapp. 30, 2313–2335.

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, 314–325.

Chong, T.T., Cunnington, R., Williams, M.A., Kanwisher, N., Mattingley, J.B., 2008. fMRI adaptation reveals mirror neurons in human inferior parietal cortex. Curr. Biol. 18, 1576–1580.

Ciaramidaro, A., Adenzato, M., Enrici, I., Erk, S., Pia, L., Bara, B.G., Walter, H., 2007. The intentional network: how the brain reads varieties of intentions. Neuropsychologia 45, 3105–3113.

Conty, L., N'Diaye, K., Tijus, C., George, N., 2007. When eye creates the contact! ERP evidence for early dissociation between direct and averted gaze motion processing. Neuropsychologia 45, 3024–3037.

Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3, 201–215.

De Lange, F.P., Spronk, M., Willems, R.M., Toni, I., Bekkering, H., 2008. Complementary systems for understanding action intentions. Curr. Biol. 18, 454–457.

Decety, J., Lamm, C., 2007. The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. Neuroscientist 13, 580–593.

Frith, C.D., Frith, U., 2006. The neural basis of mentalizing. Neuron 50, 531–534.

Gallagher, H.L., Frith, C.D., 2003. Functional imaging of 'theory of mind'. Trends Cogn. Sci. 7, 77–83.

Hamilton, A.F., Grafton, S.T., 2008. Action outcomes are represented in human inferior frontoparietal cortex. Cereb. Cortex 18, 1160–1168.

Hirai, M., Fukushima, H., Hiraki, K., 2003. An event-related potentials study of biological motion perception in humans. Neurosci. Lett. 344, 41–44.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J.C., Rizzolatti, G., 2005. Grasping the intentions of others with one's own mirror neuron system. PLoS Biol. 3, e79.

Jokisch, D., Daum, I., Suchan, B., Troje, N.F., 2005. Structural encoding and recognition of biological motion: evidence from event-related potentials and source analysis. Behav. Brain Res. 157, 195–204.

Liu, D., Sabbagh, M.A., Gehring, W.J., Wellman, H.M., 2004. Decoupling beliefs from reality in the brain: an ERP study of theory of mind. NeuroReport 15, 991–995.

Materna, S., Dicke, P.W., Thier, P., 2008. The posterior superior temporal sulcus is involved in social communication not specific for the eyes. Neuropsychologia 46, 2759–2765.

Mitchell, J.P., 2008. Activity in right temporo-parietal junction is not selective for theory-of-mind. Cereb. Cortex 18, 262–271.

Montgomery, K.J., Isenberg, N., Haxby, J.V., 2007. Communicative hand gestures and object-directed hand movements activated the mirror neuron system. Soc. Cogn. Affect. Neurosci. 2, 114–122.

Ortigue, S., Thompson, J.C., Parasuraman, R., Grafton, S.T., 2009. Spatio-temporal dynamics of human intention understanding in temporo-parietal cortex: a combined EEG/fMRI repetition suppression paradigm. PLoS ONE 4, e6962.

Pelphrey, K.A., Morris, J.P., 2006. Brain mechanisms for interpreting the actions of others from biological-motion cues. Curr. Dir. Psychol. Sci. 15, 136–140.

Pelphrey, K.A., Morris, J.P., McCarthy, G., 2004. Grasping the intentions of others: the perceived intentionality of an action influences activity in the superior temporal sulcus during social perception. J. Cogn. Neurosci. 16, 1706–1716.

Proverbio, A.M., Riva, F., Zani, A., 2009. Observation of static pictures of dynamic actions enhances the activity of movement-related brain areas. PLoS ONE 4, e5389.

Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. Annu. Rev. Neurosci. 27, 169–192.

Rizzolatti, G., Fadiga, L., Gallese, V., Fogassi, L., 1996. Premotor cortex and the recognition of motor actions. Brain Res. Cogn. Brain Res. 3, 131–141.

Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". Neuroimage 19, 1835–1842.

Saxe, R., Wexler, A., 2005. Making sense of another mind: the role of the right temporo-parietal junction. Neuropsychologia 43, 1391–1399.

Saxe, R., Xiao, D.K., Kovacs, G., Perrett, D.I., Kanwisher, N., 2004. A region of right posterior superior temporal sulcus responds to observed intentional actions. Neuropsychologia 42, 1435–1446.

Shmuelof, L., Zohary, E., 2006. A mirror representation of others' actions in the human anterior parietal cortex. J. Neurosci. 26, 9736–9742.

Shmuelof, L., Zohary, E., 2007. Watching others' actions: mirror representations in the parietal cortex. Neuroscientist 13, 667–672.

Talairach, J., Tournoux, P., 1988. Co-Planar Stereotactic Atlas of the Human Brain. Thieme, New York.

Thioux, M., Gazzola, V., Keysers, C., 2008. Action understanding: how, what and why. Curr. Biol. 18, R431–R434.

Van der Cruyssen, L., Van Duynslaeger, M., Cortoos, A., Van Overwalle, F., 2009. ERP time course and brain areas of spontaneous and intentional goal inferences. Soc. Neurosci. 4, 165–184.

Van Overwalle, F., 2009. Social cognition and the brain: a meta-analysis. Hum. Brain Mapp. 30, 829–858.

Van Overwalle, F., Baetens, K., 2009. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. Neuroimage 48, 564–584.

Völlm, B.A., Taylor, A.N., Richardson, P., Corcoran, R., Stirling, J., McKie, S., Deakin, J.F., Elliott, R., 2006. Neuronal correlates of theory of mind and empathy: a functional magnetic resonance imaging study in a nonverbal task. Neuroimage 29, 90–98.

Walter, H., Adenzato, M., Ciaramidaro, A., Enrici, I., Pia, L., Bara, B.G., 2004. Understanding intentions in social interaction: the role of the anterior paracingulate cortex. J. Cogn. Neurosci. 16, 1854–1863.

Walter, H., Ciaramidaro, A., Adenzato, M., Vasic, N., Ardito, R.B., Erk, S., Bara, B.G., 2009. Dysfunction of the social brain in schizophrenia is modulated by intention type: an fMRI study. Soc. Cogn. Affect. Neurosci. 4, 166–176.