International Journal of Psychophysiology 83 (2012) 357–364

Contents lists available at SciVerse ScienceDirect

International Journal of Psychophysiology

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

False belief and verb non-factivity: A common neural basis? ☆

Him Cheung [a,*], Lan Chen [b], Ching-Yee Szeto [a], Gangyi Feng [b], Guangming Lu [c], Zhiqiang Zhang [c], Zude Zhu [b], Suiping Wang [b,**]

[a] Department of Psychology, Chinese University of Hong Kong, Shatin, N.T., Hong Kong
[b] Department of Psychology, South China Normal University, Guangzhou, China
[c] Department of Medical Imaging, Jinling Hospital, Nangjing University School of Medicine, Nanjing, China

ARTICLE INFO

Article history:
Received 15 June 2011
Received in revised form 15 September 2011
Accepted 10 December 2011
Available online 26 December 2011

Keywords:
False belief
Theory of mind
Mentalizing
Non-factive verb
Complement

ABSTRACT

Using fMRI, the present study compares the brain activation underlying false belief thinking induced by pictorial, nonverbal material to that instigated by strong non-factive verbs in a sample of adult Chinese speakers. These verbs obligatorily negate their complements which describe the mind content of the sentence agent, and thus may activate part of the false belief network. Some previous studies have shown a behavioral correlation between verb non-factivity/false complementation and conventional false belief but corresponding neural evidence is lacking. Our results showed that the non-factive grammar and false belief commonly implicated the right temporo-parietal junction (TPJ), which had been shown by past studies to play a role in general mentalizing. Regions that were unique to nonverbal false belief were the left TPJ and right middle frontal gyrus (MFG), whereas the unique regions for the non-factive grammar were the left inferior frontal gyrus (IFG) and right superior temporal gyrus (STG). Hence, conventional nonverbal false belief and verb non-factivity have both shared and unique neural representations.

1. Introduction

In day to day living we frequently consider others' and our own mental states to make sense of behavior so as to smooth social interaction. Such mentalizing is based on a socio-cognitive capacity commonly known as theory of mind (ToM). The concept of ToM is broad; it involves the attribution of intentions, thoughts, and beliefs, as well as recognizing the fact that different individuals, or the same individual at different times, can perceive, understand, and interpret reality in very different ways (Premack and Woodruff, 1978). In fact, such a capacity to recognize multiple and sometimes conflicting mental representations of reality has been regarded as the core of ToM; it requires representing others' thoughts about reality as mental representations only, not reality itself (Suddendorf, 1999).

Because of the importance of ToM in social functioning, recent years have witnessed a growing interest in identifying the unique neural circuits that underlie it. The present research aims to pinpoint the brain networks that are uniquely activated in thinking about others' false beliefs, i.e., beliefs that are at variance with known reality, using functional magnetic resonance imaging (fMRI). Furthermore, we compare the active network for false beliefs conveyed through nonverbal pictorial material to that activated by strong non-factive verbs that negate their complements. For example, in “John falsely thinks Mary is ill” the verb phrase “falsely thinks” dictates that “Mary is ill” should be false. Some previous studies have linked verb non-factivity and false complement understanding to false belief at a behavioral level (Cheung, 2006; Cheung et al., 2009; de Villiers and Pyers, 2002; Hale and Tager-Flusberg, 2003). We now ask the question: Do strong non-factives and false beliefs expressed nonverbally activate similar neural circuits, since both entail false mental representations that are in conflict with reality?

ToM studies reported in the brain imaging literature have generally agreed that the medial prefrontal cortex (mPFC) is implicated in general mentalizing as well as more specific perspective taking (Amodio and Frith, 2006). The participants' mentalizing tasks in these studies included thinking about the functions of unfamiliar objects assuming the perspective of another person (Goel et al., 1995), thinking and talking about mental states (Calarge et al., 2003), viewing nonverbal comic strips in which access to a character's intention was necessary (Brunet et al., 2000), and simply thinking about others (Mitchell et al., 2005; Mitchell et al., 2006). In addition to the mPFC, previous studies have also agreed that the temporo-parietal junction (TPJ) is responsible for ToM (e.g., Young and Saxe, 2008). Saxe and Powell (2006) demonstrated right TPJ activation in their participants' reading of another person's thoughts but not bodily sensations nor physical characteristics. Note

☆ Author note: This research is supported by a General Research Fund granted to the first author by the Research Grants Council, Hong Kong Government (project number 441809). We thank the action editor and reviewers for their very constructive comments on an earlier version of the paper.

* Correspondence to: H. Cheung, Dept of Psychology, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong SAR, China. Tel.: +852 2609 6457; fax: +852 2603 5019.
** Corresponding author. Fax: +86 20 8521 6507.
E-mail addresses: hcheung@psy.cuhk.edu.hk (H. Cheung), suiping@scnu.edu.cn (S. Wang).

0167-8760/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpsycho.2011.12.002

that in this study the posterior cingulate cortex (PCC) was also shown to be involved in processing others' thoughts.

Of special relevance to the current study is the more specific capacity of false belief understanding, which is traditionally regarded as the hallmark of a mature ToM. This is because the capacity involves the advanced ability to separate a less from a more accurate representation of reality, and to allow their co-existence in different minds although they apparently conflict with each other. Hence, mental representations are regarded as representations only, which need to be separated or "decoupled" from reality. Similar to more general ToM skills, false belief have been shown to correlate with activities in the mPFC and TPJ. In an early review, Frith and Frith (2003) argued that the mPFC constituted the neural basis for the decoupling mechanism in false belief or deception, necessary for separating false thoughts from reality. This claim was later supported by Lissek et al. (2008), who found that the mPFC was active when the participants considered deceptive as opposed to cooperative intention. Such activation was attributed to the mismatch between an agent's intention and another person's expectation in the context of deception. Similar prefrontal activities were also recorded by Kobayashi et al. (2006) who required their participants to answer questions about second-order false belief stories (i.e., X thinks Y thinks that…..).

On the other hand, Saxe and Kanwisher (2003) highlighted the TPJ as the unique area responsible for understanding the content of mental states, because the region responded only to false mental representations, not false but non-social representations (i.e., false photographs) nor physical, non-mental characteristics of people (see also Mitchell, 2008; Saxe and Wexler, 2005; Scholz et al., 2009). This conclusion is in agreement with the findings later reported by Sommer et al. (2007), Gobbini et al. (2007), and Bedny et al. (2009), who identified the right TPJ as the main area responsible for considering mental states that did not match reality in a false belief context. Saxe et al. (2006) further showed that the TPJ responded selectively to false belief attribution but not to inhibitory control demand which was a component common to many false belief tasks. Young et al. (2010a, 2010b) found TPJ activation that was unique to processing false belief as opposed to physical stories, regardless of how attention-catching these stories were, and hence rejected attentional demand as a factor for heightened TPJ activation.

In addition to the mPFC and TPJ, the posterior cingulate cortex (PCC) was also highlighted as a false belief region by Fletcher et al. (1995) and Gobbini et al. (2007), whereas the anterior cingulate cortex (ACC) was suggested by others (Kobayashi et al., 2006; Sommer et al., 2007). Other relevant areas included the anterior paracingulate cortex (APC) (Gobbini et al., 2007), precuneus (PC), and anterior temporal sulci (aSTS) (Bedny et al., 2009).

To summarize, the mPFC, TPJ, ACC, APC, PCC, and PC have been identified as the main regions implicated in a variety of ToM tasks, with the mPFC and TPJ being the most frequently linked to general mentalizing and false belief reasoning. Would similar regions be activated by language that conveys mental states and false beliefs? There are two levels to this question. First, would similar results emerge if we contrast nonverbal with verbal tasks? The answer appears to be positive, as studies using verbal material have yielded similar results as those using nonverbal material (e.g., Calarge et al., 2003; Kobayashi et al., 2006; Saxe and Powell, 2006). There are a few studies directly contrasting verbal with nonverbal tasks, identifying significant overlapping regions of activation such as the mPFC and TPJ (Gallagher et al., 2000; Kobayashi et al., 2007). In a recent review, Carrington and Bailey (2009) have argued with good evidence that verbal and nonverbal ToM tasks do not give rise to systematic differences in regions of activation between studies.

The second sense of the question has to do with a class of linguistic items specializing in expressing mental states, such as the mental terms "think" and "want" and their complements. For example, in "John thinks Mary is ill", the mental term "thinks" opens up a (John's) mental world the content of which is described in the complement "Mary is ill". Hence "Mary is ill" needs to be decoupled from reality because it is only a description of John's thought. Previous studies have shown a behavioral correlation between young children's use of mental terms and ToM performance (Brown et al., 1996; Furrow et al., 1992; Moore et al., 1990). de Villiers and Pyers (2002) further argue that children's distinction between false complements and story reality constitutes a foundation for their false belief thinking development. Hale and Tager-Flusberg's (2003) results support this causal interpretation.

Another line of research has focused on the semantic nature of the mental verb preceding the complement. Some verbs are described as "factive" because they presuppose the veracity of the complement, such as "know". On the other hand, complements following non-factive verbs, such as "think" and "guess", can either be true or false. Lee et al. (1999), and Tardif et al. (2004) reported that including non-factive mental verbs in standard false belief questions enhanced false belief performance. Cheung et al. (2009) showed a unique correlation between children's understanding of false belief and the strong Cantonese–Chinese non-factive verb/ji5-wai4/[1] ("falsely think"), which negates the following complement.

The current study was the first to compare via brain imaging the neural correlates of false belief thinking to those underlying strong non-factive verbs which negate their complements. The lexicalized semantics of these verbs dictates that what follows (i.e., its complement content) is in conflict with reality, which parallels a false belief expressed nonverbally. In the current study we compared standard nonverbal false belief conveyed through pictures to strong non-factives, using non-mentalistic picture and verbal stories (i.e., the fillers) as the respective controls. Correspondingly, pictorial true belief was also compared to factives employing the same filler items as controls.

2. Material and methods

2.1. Participants

Twenty-three right-handed healthy adults, aged between 22 and 26 years (mean = 23.5 years; *sd* = 1.1 years), were paid a small sum of money to participate in this experiment. All were native Chinese (Mandarin) speakers having normal or corrected-to-normal vision with no history of psychiatric or neurological disorders. Three participants were excluded from data analysis because of low response accuracies and excessive motion. Written informed consent was obtained from each participant following a protocol approved by the ethics committee of the local hospital system, which was consistent with the American Psychological Association guidelines.

2.2. Material

Material consisted of experimental stimuli and fillers. Experimental stimuli were picture sequences and short Chinese sentences depicting and narrating either true or false belief stories. Following each story a question was asked about the chief character's belief. Four experimental conditions therefore resulted: picture-true-belief (PT), picture-false-belief (PF), sentence-true-belief (ST), and sentence-false-belief (SF). There were 20 trials in each condition. In addition, 40 picture and 40 sentence filler trials were also included, in which the participant was to answer non-belief questions about the story content. The filler stories were about physical events rather than beliefs.

In each trial four slides formed a complete sequence representing a story, followed by one question slide and finally one response slide. In the picture condition the four story slides were a sequence of pictures whereas in the sentence condition they were constituents of

[1] Cantonese transcriptions are in Jyutping, the romanization system adopted by the Linguistic Society of Hong Kong in 1993. Numbers indicate lexical tones, of which there are six.

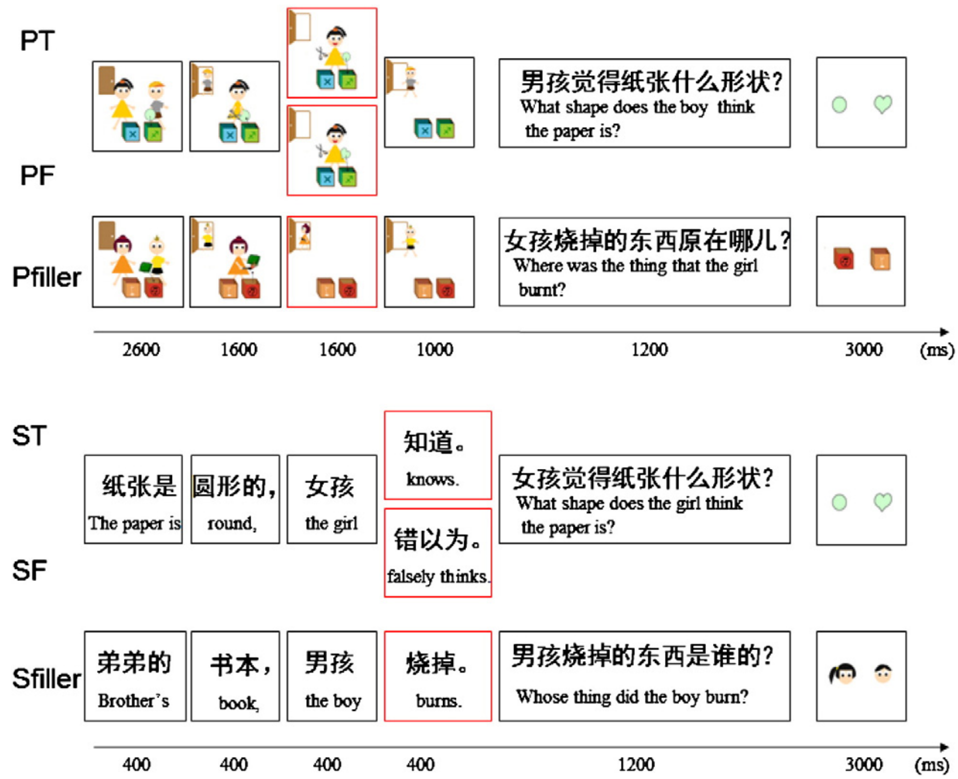

Fig. 1. Illustration of the stimuli and procedure. There were four experimental conditions in which beliefs were induced: picture-true-belief (PT), picture-false-belief (PF), sentence-true-belief (ST), sentence-false-belief (SF), and two filler conditions in which beliefs were *not* induced: picture filler (Pfiller), sentence filler (Sfiller). The participants were asked to view the four story pictures or phrases before a question slide was presented. Then they would choose one out of two options depicted in the final response slide to answer the question. A 400-ms blank slide was presented after each slide. The critical manipulation of false versus true belief was done via the third and fourth slide (i.e., the critical slides, highlighted in red) for the pictures and sentences, respectively.

one complete Chinese sentence (i.e., a story was told through only one sentence).

The experimental stories were modeled after the classic unseen displacement task (Baron-Cohen et al., 1985) with modification. All the pictures were in the style of computer-drawn cartoon. Each false belief story was about a character, Y, intentionally changing the color, shape, or location of a target object without another character, X, knowing the change. At a later time point X comes back and thus holds a false belief about the color, shape, or location of the target. In each false belief trial, slide 1 depicted X, standing next to Y, putting a target object of a certain color and shape into one of two available boxes. Slide 2 showed X leaving the room and Y taking out a pair of scissors or some paint intending to change the shape or color of the target, or taking the target out from the original box intending to shift it to the other box. In slide 3 (i.e., the critical slide) Y was shown to modify the color/shape of the target, or to shift it to another location. Slide 4 showed X coming back into the room. Slides 1, 2, and 4 were shared by the true and false belief stories. Slide 3 in a true belief trial depicted Y *not* changing the color/shape/location of the target, and thus X's knowledge about it remains accurate. Examples are shown in Fig. 1.

For the sentence trials, each story slide contained a constituent phrase of a complete Chinese sentence, and the critical mental verb which determined the veracity of its complement was always presented in the last (fourth) story slide (i.e., the critical slide). These verbs were the factives "realize" and "know" used in the true belief condition, and the strong non-factives "falsely think" and "wrongly think" used in the false belief condition.[2] In contrast, there were no mental verbs in the filler sentences, which described physical events only (see Fig. 1).

After the four story slides, a question and lastly a response slide were presented. These two slides were identical across the picture and sentence conditions. The question slide contained the interrogative sentence "What shape/color/location does X think the [target] is?". Finally, the response slide was presented containing two options from which the participant was asked to choose the best answer by pressing one of two pre-designated keys.

2.3. Procedure

A practice session was conducted outside of the MRI scanner to familiarize the participants with the procedure. Practice ceased at 80% of performance accuracy in every experimental condition when the participant was introduced to the fMRI session, which lasted approximately an hour.

Using an event-related design, four runs each including 40 trials were presented during fMRI data collection. Each run was preceded and followed by a 30-s passive viewing of a fixation cross to enhance the estimation of baseline signals. The reason for including this passive viewing period was to gather further resting baseline data in addition to the baseline data from the non-critical slides in each condition, to contrast with those acquired when the participant was viewing the critical slide in the picture (i.e., slide 3) and sentence condition (i.e., slide 4), as did in previous studies (Wang et al., 2008; Zhu et al., 2009). Stimuli were presented in a pseudo-random sequence within each run. In the picture condition, the four story slides in each trial were presented for 2.6, 1.6, 1.6 and 1 s, respectively, each followed by a 400-ms blank slide. In the sentence condition each story slide was presented for 400 ms followed by a 400-ms blank

[2] These Chinese non-factive verbs do not have translation equivalents in English; "falsely think" and "wrongly think" are the closest translations.

slide. Then after a 400-ms fixation the question and response slide were presented for 1.2 s and 3 s respectively. The interval between any two successive trial onsets varied randomly between 10 and 18 s (10 s, 12 s, 14 s for sentence trials; 14 s, 16 s, 18 s for picture trials). Trials from each condition were equally distributed across the runs.

The slide durations in both the picture and sentence conditions were determined in accordance with the results from a pilot behavioral experiment using the self-paced paradigm. In this pilot experiment the participants were asked to do exactly the same picture and sentence-false-belief tasks as in the present study, except that slide presentation was paced by the participants themselves. To go on to the next slide the participant would always need to press a certain key, and the presentation or viewing duration of each slide was recorded for each participant. Thirty participants were recruited in the pilot test, who would not attend the present fMRI experiment. In the pilot test no fMRI responses were recorded. An average presentation duration and standard deviation for each slide were obtained across the thirty participants and subsequently used to determine the slide presentation times in the fMRI experiment.

Stimuli were presented on a computer screen using E-Prime (Psychology Software Tools), which was visible via mirror-glasses by the participants lying supine inside the scanner. Responses were recorded using two buttons of a four-button fMRI compatible response pad. Participants were told not to move their heads (restrained in padding) inside the scanner and that they could close their eyes for a short rest between two successive runs.

2.4. fMRI data acquisition

Scanning was conducted on a 3 T Siemens scanner using a standard headcoil. Thirty axial slices covering the whole brain were acquired with a T2*-weighted gradient-echo echo-planar imaging (EPI) pulse sequence (TR = 2 s, TE = 30 ms, flip angle = 90°, FOV = 224 mm, matrix = 64 × 64, in-plane resolution of 3.5 × 3.5 mm, slice thickness = 4 mm, no gap). Coplanar anatomical images were acquired with a T1-weighted spin echo pulse sequence (matrix = 256 × 256, TR = 30 ms, TE = 3 ms) for co-registration. A high-resolution 3d volume was also acquired with spoiled gradient recalled echo (SPGR) sequence for spatial normalization.

With the present event-related design, picture and sentence data were collected from the same participant within a session. This design therefore eliminated the possibility that picture-sentence differences could be attributed to the different scanner noise levels across the sessions. In addition, the trials were jittered (with varied SOAs between 10 and 18 s) and thus the duration of a given trial could overlap with that of the previous and also the following trial. Hence, the number of TRs was not straightforward for each event and event type. Although presentation times differed across picture and sentence stimuli, the signal-noise ratio (SNR) of a given material type depended strictly on the number of trials in which the experiment conditions did not differ. Hence any signal changes observed in the analysis should be a result of experimental manipulation and could not be attributed to SNR differences.

2.5. fMRI data analysis

MRI data were analyzed using SPM8 (http://www.fil.ion.ucl.ac.uk/spm) and Matlab R2008b (The MathWorks Inc, Natick, MA, USA). Standard pre-processing was performed (realignment, slice timing, co-registration, segmentation, normalization to MNI space, and smoothing with an 8-mm full-width at half-maximum (FWHM) Gaussian kernel) (Friston et al., 1995). Voxel size was interpolated during pre-processing to isotropic $2 \times 2 \times 2$ mm^3.

For the four story slides in each trial, we were most interested in the manipulative slide which was the only slide differing between the true and false belief condition (slide 3 for the pictures and slide 4 for the sentences, or the critical slides). These critical slides were modeled as regressors of interest, such that the slide onsets completely coincided with the regressor onsets, and the slide and regressor durations were identical (i.e., 1600 ms and 400 ms for the picture and sentence condition, respectively). Trials with incorrect responses were excluded from the main analyses. The incorrect responses, non-critical story slides, and question slides were modeled as regressors of no interest. Note that the interval between any two successive trial onsets varied randomly between 10 and 18 s (10 s, 12 s, 14 s for sentence trials; 14 s, 16 s, 18 s for picture trials), and such variation served the purpose of a jittered baseline before the stimuli of interest.

An event-related design was modeled using the canonical hemodynamic response function and low-frequency drifts were removed using a high-pass filter (128 s), and the six motion correction parameters from pre-processing (x, y, z translations, roll, pitch, and yaw) as covariates to remove residual motion artifacts.

We examined the following contrasts for each participant:

(1) The condition versus rest contrast involved comparing the resting baseline data (30-second passive viewing + all the non-critical slides) to all the critical slide data (slide 3 in PT, PF, and Pfiller + slide 4 in ST, SF, and Sfiller). The purpose of this comparison was to ascertain that heightened brain activities could be induced by the critical manipulation under the present paradigm, compared to baseline activities.
(2) To examine the similarities and differences between picture-false-belief and strong verb non-factivity using physical, non-mentalistic stories as baselines, we compared the difference between PF and Pfiller (PF-Pfiller) against that between SF and Sfiller (SF-Sfiller). The corresponding comparison involving true belief was also done, i.e., PT-Pfiller against ST-Sfiller. This was to ascertain that belief processing, true or false, was indeed induced in the various belief conditions, compared to the fillers.[3]

A random-effect model was used for the group analysis to reveal these anticipated differences. The activation threshold was $p = 0.001$, and was $p = 0.05$ when corrected with false discovery rate (FDR) for multiple comparisons.

3. Results

3.1. Behavioral results

Response accuracies and response times (RTs) were recorded. Trials with omissions, incorrect responses, and RTs more than ± 3 *sds* away from the respective condition means were excluded from the subsequent analyses. This procedure resulted in 7.8% of the data being discarded. A 2 (Material: picture vs. sentence) × 3 (Belief: true belief vs. false belief vs. fillers) repeated measures ANOVA (with Green-house–Geirsser correction) was used to analyze the accuracy and RT data.

Mean accuracies for PT, PF, Pfiller, ST, SF, and, Sfiller were 95% ($sd = 5\%$), 95% ($sd = 6\%$), 89% ($sd = 6\%$), 92% ($sd = 5\%$), 93% ($sd = 6\%$), and 96% ($sd = 4\%$), respectively. Neither of the main effects was significant ($ps > .05$). The interaction was significant, $F(1, 19) = 32.3$, $p < .001$.

[3] We did not compare PF-Pfiller directly with SF-Sfiller (nor PT-Pfiller with ST-Sfiller) at the stage of deriving the regions of activation because such comparison usually applies to conditions that share a lot of commonalities. With such comparison following a subtractive logic, the commonalities are "canceled out" and what is left would indicate the uniqueness associated with one of the two conditions. The current procedure, however, involved verbal and nonverbal material differing in so many aspects in terms of processing, and thus a direct subtractive comparison between them might produce many differences that could be hard to interpret. For this reason, neither a direct comparison nor the exclusive masking procedure was performed between the pictures and sentences at the stage of deriving the active regions. Direct comparisons were only done within the modalities.

Pairwise comparisons showed that while the participants were less accurate with the Pfiller than the PF and PT trials, $ps<.001$, they were more accurate with the Sfiller than the SF and ST trials, $ps<.05$. Mean RTs for PT, PF, Pfiller, ST, SF, and Sfiller were 829 ms ($sd=149$ ms), 873 ms ($sd=159$ ms), 970 ms ($sd=108$ ms), 920 ms ($sd=100$ ms), 1017 ms ($sd=138$ ms), and 886 ms ($sd=97$ ms), respectively. The main effects and interaction were all significant, $Fs>6.2$, $ps<.05$. Pairwise comparisons showed that the mean RT for Pfiller was longer than those for PF and PT, and the mean RT for SF was longer than those for ST and Sfiller, all $ps<.01$.

3.2. fMRI results

3.2.1. Condition versus rest

Fig. 2 shows the significant brain activations for each condition relative to the resting baseline. The three picture conditions elicited very similar patterns of activities in a set of brain regions, including the bilateral inferior, medial, and superior frontal cortex, the anterior and posterior cingulate, the presupplementary motor area, the bilateral posterior temporal cortex, bilateral temporo-parietal junction, the posterior cortical areas in the parietal and occipital lobes, and also some subcortical areas such as the thalamus and caudate. The three sentence conditions elicited similar regions, except the bilateral posterior temporal cortex.

3.2.2. False belief versus fillers

3.2.2.1. PF-Pfiller. For the PF-PFiller contrast, heightened activations (regions colored blue and white in Fig. 3(a); see Table 1 for a list of these regions) were recorded in the right middle frontal gyrus (right MFG, Brodmann area BA 9), left superior frontal gyrus/supplementary motor area (BA 6) in the frontal cortex, and left and right inferior parietal lobule (left/right TPJ, BA 40). Extensive activation was also revealed in the left lingual gyrus, left cingulate gyrus, left thalamus, and left lentiform nucleus.

3.2.2.2. SF-Sfiller. For the SF-SFiller contrast, activations (regions colored red and white in Fig. 3(a); see Table 1 for a list of these regions) were recorded in the right MFG (BA 9), left superior frontal gyrus/supplementary motor area (BA 6), left inferior frontal gyrus (left IFG, BA 47), and right superior temporal gyrus (STG, BA 21). In addition, activations were also detected in the right inferior parietal lobule (right TPJ, BA 40).

The active regions shared by the above two contrasts were the right TPJ, which had been shown to be involved in thinking about others' thoughts (Kanwisher, 2010), and the left supplementary motor area (see Fig. 3(a), the regions colored white). Some overlapping was also noticed in the right MFG but the two peaks were distinct. The peak for PF-Pfiller closely matched that found in previous false belief studies (36, 42, 34, compared to 38, 40, 31 in Kobayashi et al., 2007), whereas the peak for SF-Sfiller was more lateral and posterior (48, 16, and 40).

The left TPJ was unique to the PF-Pfiller comparison. This region has been shown to be responsible for false belief reasoning in some previous studies (Aichhorn et al., 2009; Apperly et al., 2004). In contrast, the SF-Sfiller comparison revealed unique activations in regions known to be involved in linguistic processing: the left IFG (e.g., Hagoort, 2003, 2005; Hagoort and van Berkum, 2007; Jung-Beeman, 2005) and right STG (e.g., Humphries et al., 2005; Humphries et al., 2006). The different activations between the pictures and sentences are not attributable to differential activations in the respective baseline conditions.

3.2.3. True belief versus fillers

3.2.3.1. PT-Pfiller. For PT-Pfiller, the results revealed activations (regions colored blue and white in Fig. 3(b); see Table 1 for a list of these regions) in the right superior and middle frontal gyrus (BA 8, 11), right TPJ (BA 40), left inferior parietal lobule (BA 40), left occipital gyrus (BA17), and left and right insula (BA 13).

3.2.3.2. ST-Sfiller. For ST-SFiller, activations (regions colored red and white in Fig. 3(b); see Table 1 for a list of these regions) were recorded in the right TPJ (BA 40), right STG (BA 21), right precuneus (BA 7), and right insula (BA 13).

The active region shared by the above two contrasts was the right TPJ (BA 40) (see Fig. 2(b), the regions colored white). Therefore, the right TPJ was activated by both false and true beliefs conveyed through pictorial and sentence materials, compared to the processing of physical, non-mentalistic events represented by the fillers. Hence, the right TPJ appears to be involved in general mentalizing.

4. Discussion

4.1. Overview

The present study compares the neural basis of picture-induced (nonverbal) false belief thinking to that of strong non-factive verb processing, which according to some previous theorization and results is closely associated with false belief because both critically involve separating a false mental representation from reality. The results have important implications for the nature of false belief reasoning and its relation with language.

Comparing PF-Pfiller against SF-Sfiller revealed both common and unique active regions, suggesting that the neural network for standard nonverbal false belief is not entirely the same as that for strong non-factivity. Consistent with many previous findings on conventional false belief thinking, our results showed that picture-induced false belief was uniquely associated with the left TPJ (e.g., Gallagher et al., 2000; Gobbini et al., 2007; Grezes et al., 2004; Kobayashi et al., 2007; Mitchell, 2008; Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Sommer et al., 2007; Young and Saxe, 2008) and right MFG (Kobayashi et al., 2007). On the other hand, processing of strong non-factive verbs was associated with the left IFG and right STG, which are regarded as language-related areas (e.g., Bookheimer, 2002; Ferstl et al., 2008; Friederici et al., 2003; Hagoort et al., 2009). The only active region shared by both analyses was the right TPJ.

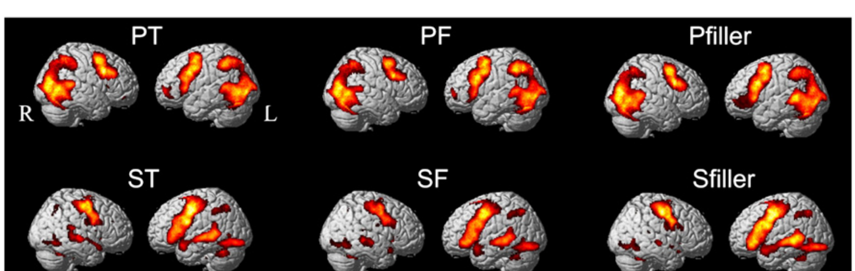

Fig. 2. Brain regions showing significant activities in the different conditions, relative to the resting baseline (L = left, R = right).

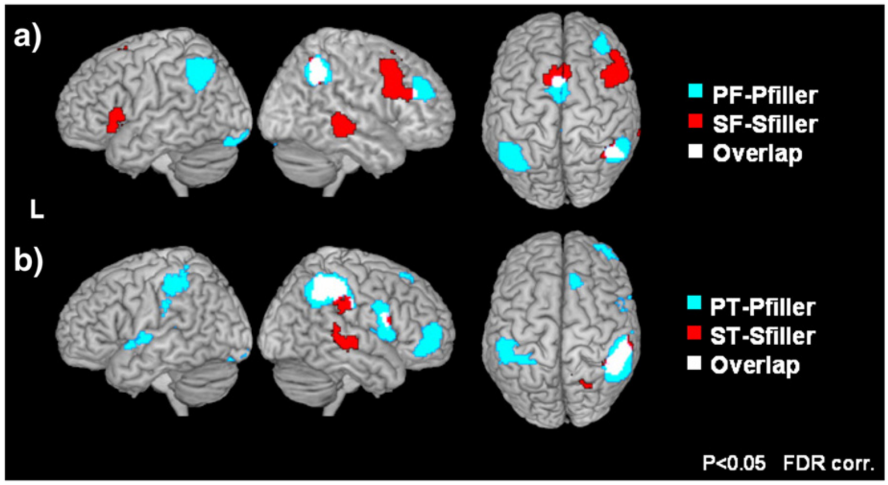

Fig. 3. (a) Activations revealed by the PF-Pfiller and SF-Sfiller contrasts. (b) Activations revealed by the PT-Pfiller and ST-Sfiller contrasts (L = left).

4.2. The false belief network with picture material

4.2.1. Right TPJ

Our results showed that the right TPJ was activated in both the PF-Pfiller and PT-Pfiller contrasts, supporting the conclusion that it is the selective cortical region for handling the beliefs of another person (Kanwisher, 2010). This region is engaged in thinking about other's thoughts independent of whether they are true or false (Aichhorn et al., 2009; Jenkins and Mitchell, 2009; Scholz et al., 2008), justified or unjustified (Young et al., 2010a, 2010b), positive or negative (Young and Saxe, 2008, 2009), and expected or unexpected (Young et al., 2010a, 2010b).

Nevertheless, the right TPJ has also been shown by some past studies to be implicated in more general cognitive functions in addition to belief attribution. For instance, heightened activation in the right TPJ has been reported during the detection of low-frequency (thus unfamiliar and unexpected) targets (Bledowski et al., 2004), and invalidly cued targets as compared to validly cued ones (Vossel et al., 2009). These effects are supposed to stem from the extra attention and inhibition demand associated with the stimuli. Therefore, while we would argue that the right TPJ is commonly implicated in nonverbal false belief and strong non-factive verb processing, we do not attempt to eliminate the possibility that other cognitive processes, especially attention and inhibition, are also attributable to the region. In fact, it is rather likely that these general cognitive functions are involved in false belief thinking, verbal or nonverbal. Hence, we argue that the right TPJ is the most significant region shared by nonverbal false belief and strong non-factivity; the role it plays in other cognitive functions lies beyond the scope of the current study.

4.2.2. Left TPJ

Our results showed that the left TPJ was activated in PF-Pfiller but not in PT-Pfiller. This region has been associated with deception and cooperation (Lissek et al., 2008). A previous lesion study has shown that the left TPJ mediates mental-state processing and sustains higher-level social reasoning, such as false belief reasoning (Apperly et al., 2004). However, the left TPJ was also activated in non-mental tasks involving misinformation (e.g., the false sign task) (Aichhorn et al., 2009; Perner et al., 2006) and tasks requiring the recognition of perspective differences (Aichhorn et al., 2009). Although many

Table 1
Detailed information about the active regions shown in Fig. 3.

Regions	Picture			Sentence		
	Coordinate	Voxels	Z	Coordinate	Voxels	Z
False belief condition vs. filler condition						
L Superior Frontal Gyrus (6)	[−4 8 64]	384	5	[−8 14 62]	510	4.41
R Middle Frontal Gyrus(9)	[36 42 34]	298	4.29	[48 16 40]	994	5.05
R Inferior Parietal Lobule(40)	[56–54 44]	475	4.25	[54–56 36]	242	3.39
L Inferior Parietal Lobule(40)	[−46 −52 42]	866	4.78			
L Lingual Gyrus(18)	[−18 −86 −18]	513	4.6			
L Cingulate Gyrus (23)	[2–32 26]	85	4.49			
L Thalamus	[0–8 16]	116	4.16			
L Lentiform Nucleus	[−18 4–2]	131	4.04			
L Thalamus	[−6 −30 12]	36	3.69			
L Inferior Frontal Gyrus(47)				[−54 20 2]	333	5.18
R Superior Temporal Gyrus(21)				[52–28 −6]	716	5.71
True belief condition vs. filler condition						
R Inferior Parietal Lobule(40)	[56–40 44]	1620	5.88	[54–40 52]	1004	5.19
R Insula(13)	[48 4 16]	955	5.11	[48 8 18]	168	4.42
R Superior Frontal Gyrus(8)	[10 30 54]	101	4.11			
R Middle Frontal Gyrus(11)	[46 46–12]	808	5.68			
L Inferior Parietal Lobule(40)	[−50 −36 56]	873	4.82			
L Inferior Occipital Gyrus(17)	[−10 −92 −12]	301	4.05			
L Insula(13)	[−40 2 12]	455	4.34			
R Superior Temporal Gyrus(21)				[54–24 −4]	635	5.12
R Precuneus(7)				[22–60 56]	131	3.69

Coordinates refer to the Montreal Neurological Institute (MNI) reference brain. The number in brackets shows Brodmann areas. $p<.05$, corrected with FDR, $df=19$.

false belief studies have reported bilateral TPJ activity (e.g., Gallagher et al., 2000; Kobayashi et al., 2007; Saxe and Kanwisher, 2003; Saxe and Powell, 2006), the exact role of the left TPJ in false belief reasoning is still unclear. In the current study, the left TPJ activation found with picture false belief may have to do with the presence of misinformation in the PF condition.

4.2.3. Right MFG

In the current study right MFG activation was found with the PF-Pfiller comparison. Such activation may be attributed to the inhibitory control that was needed to suppress the reality representation while selecting the false belief. This explanation is consistent with the findings of Kobayashi et al. (2007), also showing right MFG activation by comparing second-order false belief with non-belief. Understanding of second-order false belief (i.e., X thinks that Y thinks that……) requires double inhibitory operations and hence leads to enhanced right MFG activation.

4.3. Strong non-factivity

Both the strong non-factive and factive verbs in the sentence condition, compared to the non-mental filler verbs, activated the right TPJ, which was also associated with the corresponding picture material. Some previous fMRI studies have suggested that the TPJ/STS is associated with the resolution of semantic incongruity and integrating inconsistent information in linguistic processing (e.g., Ferstl et al., 2008). In addition, the right TPJ has also been shown to be selective to thinking about what other people are thinking (Kanwisher, 2010), independent of the different manipulations on belief reasoning (e.g., Aichhorn et al., 2009; Young et al., 2010a, 2010b).

In the current SF and ST conditions the sentences described the belief of a certain character, whereas in the Sfiller condition the filler sentences described physical events. The right TPJ activity in this study suggested that our belief manipulations in the SF and ST conditions were valid. Belief representations were elicited by the factive and non-factive sentences but not by the filler sentences. Young et al. (2010a, 2010b) also reported activity in the right TPJ by comparing sentences describing beliefs with those describing physical events or objects.

Our results showed that activation in the right TPJ was common among PF-Pfiller, SF-Sfiller, PT-Pfiller, and ST-Sfiller. The pattern is consistent with previous behavioral data pointing to an association between children's false belief understanding and mental verb/false complement comprehension (e.g.., Cheung et al., 2009; de Villiers and Pyers, 2002; Hale and Tager-Flusberg, 2003). The behavioral relationship between strong non-factives and false belief understanding in young children previously found may have its root at the neural level to a certain extent.

In addition to the right TPJ activity that was common to the sentences and pictures, there were activations unique to the two material types. The PF-Pfiller contrast revealed that other typical false belief regions were involved: the left TPJ (Aichhorn et al., 2009; Apperly et al., 2004) and right MFG (Kobayashi et al., 2007). On the contrary, the SF-Sfiller comparison revealed activations in the left IFG and right STG. Previous research has suggested that the left IFG and bilateral STG are selectively activated in tasks requiring the resolution of complex syntax (Christensen, 2008; Grodzinsky, 2000; Moro et al., 2001; Stromswold et al., 1996; Wang et al., 2008) and processing of semantic relationships (e.g., Friederici et al., 2003; Hagoort et al., 2009; Zhu et al., 2009). These regions are therefore regarded as critical for the integration and interpretation of linguistic information (Ferstl et al., 2008; Hagoort, 2003, 2005; Hagoort and van Berkum, 2007). The IFG and STG activation in our verbal contrasts could be attributed to the extra semantic integration incurred in processing the strong non-factive sentences, which required an additional negation operation turning their complements into false statements. Regarding this, strong non-factives are not unlike negative sentences, contrasting with affirmative sentences in terms of the truth value of the statement (Christensen, 2009). Hasegawa et al. (2002) found similar activation in the frontal cortex and left temporal gyrus comparing negative to affirmative sentences. Nieuwland and Kuperberg (2008) recorded the ERPs to negative versus affirmative sentences, and showed more negative waveforms at N400 (an ERP component for semantic variance) associated with the negative than affirmative sentences.

One factor that may have contributed to the different neural activations between the picture and sentence conditions is that reality was known to the participant in the former but not the latter condition. The strong non-factives in the sentence condition negated the story character's mind content, turning it into a false belief without revealing what reality actually was (i.e., false belief with reality unknown; Zhang et al., 2009). Hence the participant was to recognize the conflict between reality and a certain mind content, but did not need to inhibit the reality content which was empty. As mentioned earlier, the left TPJ could be involved in the recognition of misinformation, and the right MFG may be associated with the inhibitory demand in false belief reasoning. It could be that only partial decoupling was necessary in the verbal situation for there was no reality information to suppress, and thus the right MFG and left TPJ did not appear to be as active as processing false belief conveyed through pictures, which depicted the to-be-suppressed reality content.

We conclude from the current findings that the grammatical device of strong non-factivity, which implies false belief, implicates only one core region of false belief reasoning known to be responsive to standard nonverbal tasks, namely, the right TPJ. We therefore think that the previously found behavioral relationship between false complements/strong non-factives may have some neural basis. But this does not mean that strong non-factivity and false belief are neurologically equivalent, because the current results also point to regions that are unique to the two respective capacities. Therefore, conventional false belief and the strong non-factive grammar have both shared and unique neural representations.

References

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., Ladurner, G., 2009. Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. Journal of Cognitive Neuroscience 21, 1179–1192.

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. Nature Reviews. Neuroscience 7, 268–277.

Apperly, I.A., Samson, D., Chiavarino, C., Humphreys, G.W., 2004. Frontal and temporo-parietal lobe contributions to theory of mind: neuropsychological evidence from a false-belief task with reduced language and executive demands. Journal of Cognitive Neuroscience 16, 1773–1784.

Baron-Cohen, S., Leslie, A.M., Frith, U., 1985. Does the autistic child have a "theory of mind"? Cognition 21, 37–46.

Bedny, M., Pascual-Leone, A., Saxe, R.R., 2009. Growing up blind does not change the neural bases of theory of mind. Proceedings of the National Academy of Sciences of the United States of America 106, 11312–11317.

Bledowski, C., Prvulovic, D., Goebel, R., Zanella, F.E., Linden, D.E.J., 2004. Attentional systems in target and distractor processing: a combined ERP and fMRI study. NeuroImage 22, 530–540.

Bookheimer, S., 2002. Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. Annual Review of Neuroscience 25, 151–188.

Brown, J.R., Donelan-McCall, N., Dunn, J., 1996. Why talk about mental states? The significance of children's conversations with friends, siblings, and mothers. Child Development 67, 836–849.

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.

Calarge, C., Andreasen, N.C., O'Leary, D.S., 2003. Visualizing how one brain understands another: a PET study of theory of mind. The American Journal of Psychiatry 160, 1954–1964.

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

Cheung, H., 2006. False belief and language comprehension in Cantonese-speaking children. Journal of Experimental Child Psychology 95, 79–98.

Cheung, H., Chen, H.-C., Yeung, W., 2009. Relations between mental verb and false belief understanding in Cantonese-speaking children. Journal of Experimental Child Psychology 104, 141–155.

Christensen, K.R., 2008. Interfaces, syntactic movement, and neural activation: a new perspective on the implementation of language in the brain. Journal of Neurolinguistics 21, 73–103.

Christensen, K.R., 2009. Negative and affirmative sentences increase activation in different areas in the brain. Journal of Neurolinguistics 22, 1–17.
de Villiers, J.G., Pyers, J.E., 2002. Complements to cognition: a longitudinal study of the relationship between complex syntax and false-belief-understanding. Cognitive Development 17, 1037–1060.
Ferstl, E.C., Neumann, J., Bogler, C., von Cramon, D.Y., 2008. The extended language network: a meta-analysis of neuroimaging studies on text comprehension. Human Brain Mapping 29, 581–593.
Fletcher, P.C., Happe, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S.J., Frith, C.D., 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition 57, 109–128.
Friederici, A.D., Ruschemeyer, S.-A., Hahne, A., Fiebach, C.J., 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. Cerebral Cortex 13, 170–177.
Friston, K.J., Holmes, A.P., Poline, J.-B., Grasby, P.J., Williams, S.C.R., Frackowiak, R.S.J., Turner, R., 1995. Analysis of fMRI time-series revisited. NeuroImage 2, 45–53.
Frith, U., Frith, C.D., 2003. Development and neurophysiology of mentalizing. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 358, 459–473.
Furrow, D., Moore, C., Davidge, J., Chiasson, L., 1992. Mental terms in mothers' and children's speech: similarities and relationship. Journal of Child Language 19, 617–631.
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.
Gobbini, M.I., Koralek, A.C., Bryan, R.E., Montgomery, K.J., Haxby, J.V., 2007. Two takes on the social brain: a comparison of theory of mind tasks. Journal of Cognitive Neuroscience 19, 1803–1814.
Goel, V., Grafman, J., Sadato, N., Hallett, M., 1995. Modeling other minds. Neuroreport 6, 1741–1746.
Grezes, J., Frith, C.D., Passingham, R.E., 2004. Inferring false beliefs from the actions of oneself and others: an fMRI study. NeuroImage 21, 744–750.
Grodzinsky, Y., 2000. The neurology of syntax: language use without Broca's area. The Behavioral and Brain Sciences 23, 1–21.
Hagoort, P., 2003. How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. NeuroImage 20, 18–29.
Hagoort, P., 2005. On Broca, brain, and binding: a new framework. Trends in Cognitive Sciences 9, 416–423.
Hagoort, P., van Berkum, J., 2007. Beyond the sentence given. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 362, 801–811.
Hagoort, P., Baggio, G., Willems, R.M., 2009. Semantic unification, In: Gazzanig, M.S. (Ed.), The New Cognitive Neurosciences, 4th ed. MIT Press, London, pp. 819–836.
Hale, C.M., Tager-Flusberg, H., 2003. The influence of language on theory of mind: a training study. Developmental Science 6, 346–359.
Hasegawa, M., Carpenter, P.A., Just, M.A., 2002. An fMRI study of bilingual sentence comprehension and workload. NeuroImage 15, 647–660.
Humphries, C., Love, T., Swinney, D., Hickok, G., 2005. Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. Human Brain Mapping 26, 128–138.
Humphries, C., Binder, J.R., Medler, D.A., Liebenthal, E., 2006. Syntactic and semantic modulation of neural activity during auditory sentence comprehension. Journal of Cognitive Neuroscience 18, 665–679.
Jenkins, A.C., Mitchell, J.P., 2009. Mentalizing under uncertainty: dissociated neural responses to ambiguous and unambiguous mental state inferences. Cerebral Cortex 20, 404–410.
Jung-Beeman, M., 2005. Bilateral brain processes for comprehending natural language. Trends in Cognitive Sciences 9, 512–518.
Kanwisher, N., 2010. Functional specificity in the human brain: a window into the functional architecture of the mind. PNAS 107 (25), 11163–11170.
Kobayashi, C., Glover, G.H., Temple, E., 2006. Cultural and linguistic influence on neural bases of 'theory of mind': an fMRI study with Japanese bilinguals. Brain and Language 98, 210–220.
Kobayashi, C., Glover, G.H., Temple, E., 2007. Children's and adults' neural bases of verbal and nonverbal 'theory of mind'. Neuropsychologia 45, 1522–1532.
Lee, K., Olson, D.R., Torrance, N., 1999. Chinese children's understanding of false beliefs: the role of language. Journal of Child Language 26, 1–21.
Lissek, S., Peters, S., Fuchs, N., Witthaus, H., Nicolas, V., Tegenthoff, M., Juckel, G., Brune, M., 2008. Cooperation and deception recruit different subsets of the theory-of-mind network. PloS One 3, e2023.
Mitchell, J.P., 2008. Activity in right temporo-parietal junction is not selective for theory-of-mind. Cerebral Cortex 18, 262–271.
Mitchell, J.P., Banaji, M.R., Macrae, C.N., 2005. The link between social cognition and self-referential thought in the medial prefrontal cortex. Journal of Cognitive Neuroscience 17, 1306–1315.
Mitchell, J.P., Macrae, C.N., Banaji, M.R., 2006. Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. Neuron 50, 655–663.
Moore, C., Pure, K., Furrow, D., 1990. Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. Child Development 61, 722–730.
Moro, A., Tettamanti, M., Perani, D., Donati, C., Cappa, S.F., Fazio, F., 2001. Syntax and the brain: disentangling grammar by selective anomalies. NeuroImage 13, 110–118.
Nieuwland, M.S., Kuperberg, G.R., 2008. When the truth is not too hard to handle: an event-related potential study on the pragmatics of negation. Psychological Science 19, 1213–1218.
Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., Ladurner, G., 2006. Thinking of mental and other representations: the roles of left and right temporo-parietal junction. Social Neuroscience 1, 245–258.
Premack, D., Woodruff, G., 1978. Does the chimpanzee have a theory of mind? The Behavioral and Brain Sciences 1, 512–526.
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., Powell, L.J., 2006. It's the thought that counts: specific brain regions for one component of theory of mind. Psychological Science 17, 692–699.
Saxe, R., Wexler, A., 2005. Making sense of another mind: the role of the right temporo-parietal junction. Neuropsychologia 43, 1391–1399.
Saxe, R., Schulz, L.E., Jiang, Y.V., 2006. Reading minds versus following rules: dissociating theory of mind and executive control in the brain. Social Neuroscience 1, 284–298.
Scholz, J., Gweon, H., Green, S.R., Pelphrey, K., Saxe, R., 2008. True or false: the RTPJ responds to task relevant beliefs. Poster presented at SAN 2008, Boston, MA.
Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E.N., Saxe, R., 2009. Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. PloS One 4, e4869.
Sommer, M., Dohnel, K., Sodian, B., Meinhardt, J., Thoermer, C., Hajak, G., 2007. Neural correlates of true and false belief reasoning. NeuroImage 35, 1378–1384.
Stromswold, K., Caplan, D., Alpert, N., Rauch, S., 1996. Localization of syntactic comprehension by positron emission tomography. Brain and Language 52, 452–473.
Suddendorf, T., 1999. The rise of the metamind. In: Corballis, M.C., Lea, S.E.G. (Eds.), The Descent of Mind. Oxford University Press, New York, pp. 218–260.
Tardif, T., Wellman, H.M., Cheung, K.M., 2004. False belief understanding in Cantonese-speaking children. Journal of Child Language 31, 779–800.
Vossel, S., Weidner, R., Thiel, C.M., Fink, G.R., 2009. What is "odd" in Posner's location-cueing paradigm? Neural responses to unexpected location and feature changes compared. Journal of Cognitive Neuroscience 21, 30–41.
Wang, S., Zhu, Z., Zhang, J.X., Wang, Z., Xiao, Z., Xiang, H., Chen, H.-C., 2008. Broca's area plays a role in syntactic processing during Chinese reading comprehension. Neuropsychologia 46, 1371–1378.
Young, L., Saxe, R., 2008. The neural basis of belief encoding and integration in moral judgment. NeuroImage 40, 1912–1920.
Young, L., Saxe, R., 2009. An fMRI investigation of spontaneous mental state inference for moral judgment. Journal of Cognitive Neuroscience 21, 1396–1405.
Young, L., Dodell-Feder, D., Saxe, R., 2010a. What gets the attention of the temporo-parietal junction? An fMRI investigation of attention and theory of mind. Neuropsychologia 48, 2658–2664.
Young, L., Nichols, S., Saxe, R., 2010b. Investigating the neural and cognitive basis of moral luck: it's not what you do but what you know. Review of Philosophy and Psychology 1 (3), 333–349.
Zhang, T., Sha, W., Zheng, X., Ouyang, H., Li, H., 2009. Inhibiting one's own knowledge in false belief reasoning: an ERP study. Neuroscience Letters 467, 194–198.
Zhu, Z., Zhang, J.X., Wang, S., Xiao, Z., Huang, J., Chen, H.-C., 2009. Involvement of left inferior frontal gyrus in sentence-level semantic integration. NeuroImage 47, 756–763.