False belief vs. false photographs: a test of theory of mind or working memory?

Alicia Callejas[1]*, Gordon L. Shulman[1] and Maurizio Corbetta[1,2,3]

[1] Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA
[2] Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA
[3] Department of Anatomy and Neurobiology, Washington University School of Medicine, St. Louis, MO, USA

Theory of mind (ToM), the ability to reason about other people's thoughts and beliefs, has been traditionally studied in behavioral and neuroimaging experiments by comparing performance in "false belief" and "false photograph" (control) stories. However, some evidence suggests that these stories are not matched in difficulty, complicating the interpretation of results. Here, we more fully evaluated the relative difficulty of comprehending these stories and drawing inferences from them. Subjects read false belief and false photograph stories followed by comprehension questions that probed true ("reality" questions) or false beliefs ("representation" questions) appropriate to the stories. Stories and comprehension questions were read and answered, respectively, more slowly in the false photograph than false belief conditions, indicating their greater difficulty. Interestingly, accuracy on representation questions for false photograph stories was significantly lower than for all other conditions and correlated positively with participants' working memory span scores. These results suggest that drawing representational inferences from false photo stories is particularly difficult and places heavy demands on working memory. Extensive naturalistic practice with ToM reasoning may enable a more flexible and efficient mental representation of false belief stories, resulting in lower memory load requirements. An important implication of these results is that the differential modulation of right temporal–parietal junction (RTPJ) during ToM and "false photo" control conditions may reflect the documented negative correlation of RTPJ activity with working memory load rather than a specialized involvement in ToM processes.

Keywords: theory of mind, working memory, social cognition, false belief

INTRODUCTION

Cognitive neuroscientists are intensively studying the cognitive processes and neural systems underlying theory of mind (ToM; Saxe and Kanwisher, 2003), the ability to understand that other people have separate mental representations (i.e., beliefs or desires) that guide their behavior. Most recent studies (e.g., Saxe and Kanwisher, 2003; Mitchell, 2008) have used a "false belief" (FB) task (Wimmer and Perner, 1983) to measure people's ability to reason about other people's beliefs and a "false photograph" (FP) task (Zaitchik, 1990) as a control story that putatively relies on the same reasoning processes as the FB task without involving references to another person's mental representations[1]. Another set of studies has compared FB stories to stories in which the physical cause

of an event has to be deciphered (Fletcher et al., 1995; Gallagher et al., 2000). All of these tasks have been inherited from the field of developmental psychology, where they have been used to study normal and abnormal (i.e., autism) development.

Neuroimaging studies that have used the FB and FP material published by Saxe and Kanwisher (2003) have consistently reported that the right temporo-parietal-junction (RTPJ), an area between the posterior superior temporal sulcus and the inferior parietal lobule, is activated by FB reasoning (Aichhorn et al., 2009; Saxe et al., 2009). The RTPJ, however, is also activated by changes in working memory load (Todd et al., 2005), multisensory conflict (Balslev et al., 2005), attentional functions (Corbetta et al., 2008), sense of agency (Farrer and Frith, 2002), and out-of-body experiences (Blanke and Arzy, 2005; see Decety and Lamm, 2007 for a meta-analysis).

This heterogeneous list emphasizes the importance of designing ToM paradigms that control for auxiliary, non-ToM processes that activate the RTPJ. ToM paradigms involve linguistic analysis, the maintenance, retrieval, and manipulation of information within working memory, reasoning, inhibition of competing interpretations, and attention. Comprehension of FB and FP stories may differentially involve processes such as working memory that are separate from ToM but modulate RTPJ activity. Therefore, a

---

[1] Example of a FB story and comprehension question: "Jenny put her chocolate away in the cupboard. Then she went outside. Alan moved the chocolate from the cupboard into the fridge. Half an hour later, Jenny came back inside." Representation question: "Jenny expects to find her chocolate in the (cupboard/fridge)" Reality question: "When Jenny returns, she finds her chocolate in the fridge (true/false)." Example of a FP story and comprehension question: "A photograph was taken of an apple hanging on a tree branch. The film took half an hour to develop. In the meantime, a strong wind blew the apple to the ground." Representation question: "The developed photograph shows the apple on the (ground/branch)" Reality question: "Actually, the apple remained on the tree branch (true/false)."

deeper understanding of the cognitive processes recruited by the FB and FP tasks is necessary before conclusions regarding brain mechanisms can be drawn.

Theory of mind studies have reported that reaction times are faster to comprehension questions concerning FB than FP stories (Saxe and Kanwisher, 2003; Saxe and Powell, 2006), suggesting that FB stories are easier to comprehend. This result has been used to rule out arguments that brain activations for FB stories could be due to greater difficulty or time-on-task. Here, we confirmed differences in difficulty using a broader set of measures and evaluated several factors that might underlie this result. We checked the cognitive equivalence of FB and FP texts by measuring the time taken to read the story as well as to read and respond to the comprehension question. We also obtained working memory span (WMS) scores to check for possible correlations between WMS and text comprehension. Finally, we conducted a linguistic analysis of FB and FP texts to test their linguistic equivalence.

MATERIALS AND METHODS
PARTICIPANTS
Fifteen participants (nine males; mean age = 25) performed the experiment in exchange for payment. Participants were native English speakers between the ages of 18 and 32, were naïve concerning the purpose of the experiment, and gave written consent following the guidelines set by the Human Studies Committee of Washington University.

MATERIALS AND PROCEDURE
Twenty-four FB and 24 FP stories that have been widely used in ToM studies[2] (Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Saxe et al., 2006; Mitchell, 2008; Scholz et al., 2009; etc.) were presented in one experimental block. Four more blocks were conducted with stories that have been used in other ToM studies. However, because these stories have not been used in many studies or did not have comprehension questions associated with them, they were not included in the present analysis. Block order was counterbalanced and story order was randomized. Each FB and FP story had two associated comprehension questions, a representation question (RP) and a reality question (RL). The RP question probed the character's mental state in the FB text or the state of the world portrayed by the photograph in the FP text, while the RL question probed the actual outcome of the story in the FB text or the current state of the world (as opposed to when the photograph was taken) in the FP text. An example of each type of story and comprehension question is given in footnote 1. Participants saw both questions in a randomized order. Therefore the experimental design was 2 (text type: FB vs. FP) × 2 (comprehension question: RP vs. RL).

---

[2]The original set of localizer stories used in the above referenced studies only contains 12 FB and 12 FP stories. A larger set (24 stories/condition) has been developed by Dr. Saxe and is now being used as a functional localizer in their new investigations. Since this new set contains most of the stories from the original set (8/12 FB and 10/12 FP) and has two comprehension questions per story, results are reported for this set. Results for the original set are reported in footnote 5. A complete list of all the stories and questions included in the original localizer as well as in the new localizer can be found in Dr. Saxe's laboratory website.

A moving window paradigm (Just et al., 1982) was used to present the text material. Words in the text were hidden behind letter place-holders. Each button-press revealed a new word and covered the previous word with place-holders. The reaction time for a button-press was the estimate of processing time for a word. This method of presentation allowed participants to use peripheral vision to obtain information about the spatial distribution of the text, as in normal reading, but prevented participants from rereading previous sections of the text. As a consequence, reading times were better controlled than under a free reading procedure. Texts were presented on a gray screen with a black font. After the text was read, a question appeared on a blank screen with two answers located on the bottom left and right. Participants selected an answer by pressing a key with the index finger located on the matching side. Once a response had been given, a second question was presented on the screen. Response time to the comprehension questions was measured from question onset.

A working memory span (WMS) test (Daneman and Carpenter, 1980) was run prior to the task in order to estimate WM capacity for each participant. In this test, sentences were presented on the screen, one at a time, and participants had to read each sentence out loud. Following the presentation of the last sentence, participants recalled the last word of each sentence. Initially, a five trial set was presented in which each trial consisted of two sentences. Therefore, two words had to be recalled on each trial. If a correct response was given in three of the five trials, task difficulty was increased by presenting three sentences per trial. Again, if a correct response was given in three of five trials, task difficulty was increased. For the highest difficulty, six sentences were presented on a trial. Words could be recalled in any order with the restriction that the word belonging to the last sentence could not be recalled first.

RESULTS
We first compared the structure of the FB and FP stories. FB and FP texts contained a similar number of grammatical sentences per story (FB = 2.6 and FP = 2.4, $t(14) = 1.17$, $p = 0.247$, $d = 0.34$), but a different number of clauses per story (that is, the number of grammatical units consisting of a subject and a predicate) (FB = 4.3 and FP = 3.5, $t(14) = 3.22$, $p = 0.002$, $d = 0.93$). This difference indicated that FB and FP stories had non-equivalent syntactic structures. Syntactic structure is known to influence comprehension processes (Friederici and Weissenborn, 2007), as manifested in end-of-sentence wrap-up effects (Balogh et al., 1998) that are thought to reflect mechanisms for integrating information or for checking the completeness of the sentence and its arguments. As a result, the time taken to read each word of a sentence in a story is influenced by the specific syntactic structure of the sentence in which it is embedded. Sentences in a text are individual entities when it comes to comprehension mechanisms. Given the lack of equivalence between the grammatical structure of the sentences in FB and FP stories and the fact that FB and FP texts had, on average, the same number of sentences per story, we chose sentence reading time rather than story reading time as the index of linguistic processing/comprehension

times[3]. An analysis of sentence reading time showed that FB stories were read faster than FP stories ($t(14) = -6.77$, $p = 0.001$, $d = -0.46$).

---

[3]In a 30-word-2-sentence story with 10 words in the first sentence and 20 words in the second one, story reading time would be calculated as the addition of the reading time of each of the 30 words. Alternatively, the sentence reading time would be calculated by adding up the reading time of the first sentence's 10 words, separately adding up the reading time of the 20 words in the second sentence and then obtaining the mean of both summations. Therefore, the reading time of a single word is weighted differently depending on the sentence it belongs to and thus provides more information as to the comprehension processes happening for that sentence.
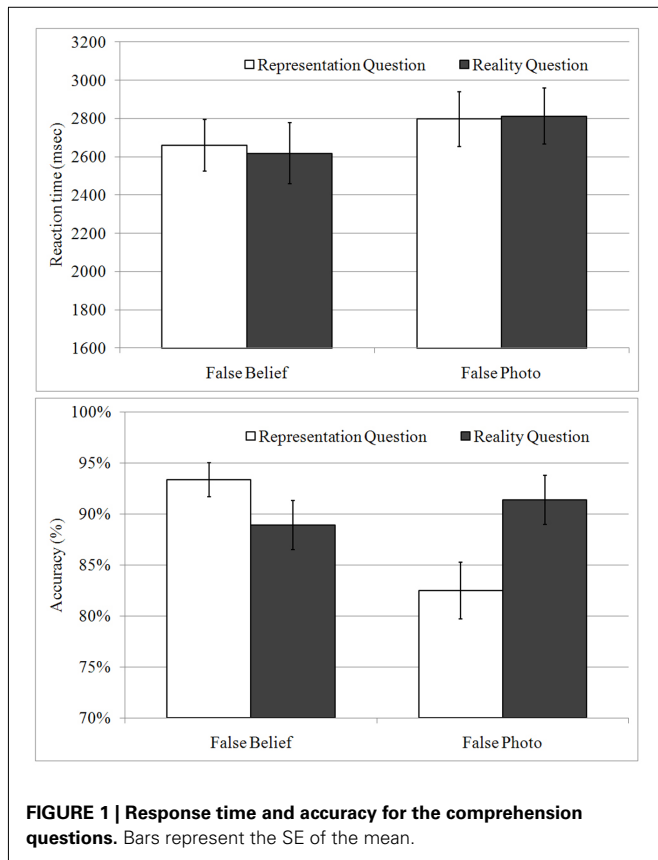

FIGURE 1 | Response time and accuracy for the comprehension questions. Bars represent the SE of the mean.

To analyze the comprehension questions, separate ANOVAs [2 (Story Type: FB vs. FP) × 2 (Comprehension Question: RL vs. RP)] on response times and accuracy were performed. Results for these analyses are shown in **Figure 1**. Response times to correctly answered questions were faster in the FB than FP condition ($F(1,14) = 8.20$, $p = 0.013$, $\eta_p^2 = 0.37$). No main effect of Comprehension Question or interaction was found (both $Fs < 1$). The accuracy analysis showed that FB questions were answered more accurately than FP questions ($F(1,14) = 9.55$, $p = 0.008$, $\eta_p^2 = 0.41$). Moreover, there was a significant interaction of Story Type by Comprehension Question ($F(1,14) = 13.24$, $p = 0.003$, $\eta_p^2 = 0.49$). *Post hoc* LSD tests showed that Representation questions concerning an FP story were more difficult than questions in the other three conditions (all $ps < 0.030$).

We conducted several analyses to check that differences in response times to questions associated with different story types were not caused by linguistic features of the questions. As shown in **Table 1**, FB and FP questions contained the same number of sentences and words. Five linguistic indexes showed significantly larger values (i.e., greater difficulty) for FB than FP questions, while only two showed the reverse effect. There is no clear evidence in the literature that the latter two indices are more predictive of difficulty than the five that showed the opposite effect. Overall, these analyses suggest that any effect of the linguistic difficulty of the question on response time would likely have gone in the opposite direction to the observed effect, which indicated longer response times to FP than FB questions. Therefore the observed differences between story types very likely reflected processes involved in manipulating and reconstructing the story to access the information probed by the question.

We also performed a linguistic analysis to check for differences between FB and FP stories. **Table 1** shows that a clear pattern was not found. Some indexes indicated that FB stories were more difficult than FP stories, while a similar number pointed in the opposite direction. Therefore, differences in linguistic difficulty between FP and FB texts do not account for the observed difference in the sentence comprehension measure.

Alternatively, the main factor driving the results could be the degree to which the material presented in the story had to be manipulated in order to answer the comprehension question.

Table 1 | Linguistic indexes for both stories and comprehension questions.

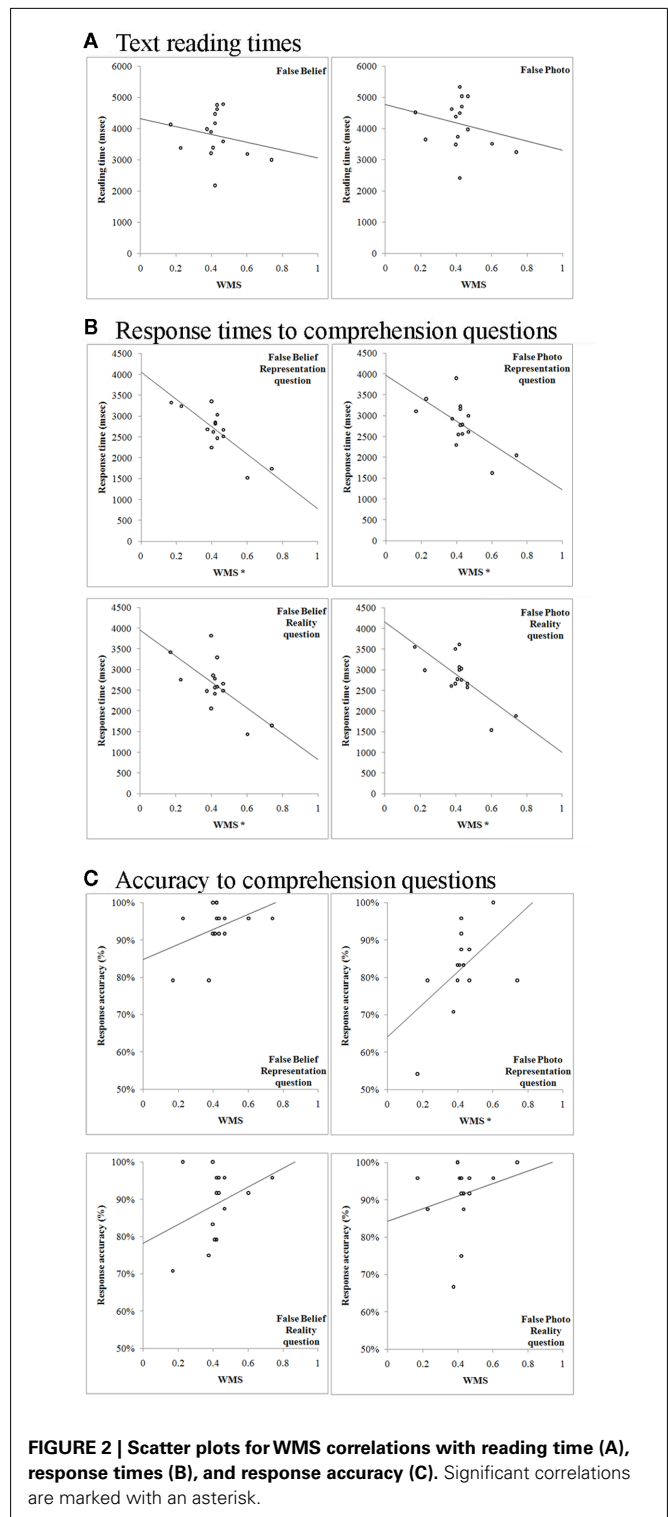|	Stories					Comprehension questions				
	FB	**FP**	***t*-test**	***p***	***d***	**FB**	**FP**	***t*-test**	***p***	***d***
Sentences per text	2.58	2.38	1.17	0.25	0.34	1.00	1.00	–	–	
Words per sentence	12.81	13.08	−0.37	0.72	−0.11	10.52	10.04	0.91	0.37	0.19
Clauses per sentence	1.76	1.52	2.08	0.04*	0.60	1.77	1.08	6.16	0.00*	1.37
Complex clauses per sentence	1.40	0.98	2.08	0.04*	0.60	1.38	0.17	6.42	0.00*	1.39
Passive clauses per sentence	0.08	0.45	−3.43	0.00*	−1.01	0.06	0.04	0.46	0.65	0.09
Relative clauses per sentence	0.18	0.03	2.72	0.01*	0.87	0.46	0.06	4.90	0.00*	1.06
Subjects per clause	0.60	0.82	−3.33	0.00*	−0.97	1.52	1.06	5.31	0.00*	1.16
Pronouns per clause	0.18	0.10	1.55	0.13	0.45	0.09	0.00	3.43	0.00*	0.99
Distance subject–verb per clause	0.20	0.40	−1.07	0.29	−0.32	0.13	0.74	−2.66	0.01*	−0.59
Adjectives per clause	0.59	1.00	−2.63	0.01*	−0.79	0.38	0.75	−3.15	0.00*	−0.67

Significant differences are marked with an asterisk.

Responding to some questions may require more manipulation of the story information in order to extract the facts related to the question. Since the story was not displayed on the screen at the time the comprehension question was presented, the accuracy of subjects' responses was determined by their ability to manipulate the story information in working memory and reconstruct the state of reality probed by the question. Consequently, the working memory demands for different types of stories may have depended on the degree to which the story information had to be manipulated.

In order to check this possibility, we tested whether subjects' comprehension indices (reading time, response time, and response accuracy) correlated with their scores in the WMS task[4]. If the results we had found were due to the difference in the amount of information that needed to be stored and manipulated in order to correctly answer the question, we would expect to find a negative correlation between WMS and RTs and/or a positive correlation between WMS and accuracy. WMS did not correlate with reading time (all $ps > 0.400$, **Figure 2A**) but it did correlate with response time to the comprehension question for all four conditions: FB–RP $r = -0.82$, $p = 0.001$; FB–RL $r = -0.67$, $p = 0.006$; FP–RP $r = -0.65$, $p = 0.009$; FP–RL $r = -0.73$, $p = 0.002$ (**Figure 2B**). In all cases, higher WMS was associated with faster responses. Most importantly, WMS correlated with accuracy in one of the four experimental conditions (**Figure 2C**). Accuracy for representation questions about false photos (FP–RP), the most difficult question type[5], was higher for those individuals with higher WMS ($r = 0.53$, $p = 0.040$).

---

[4]The categorical WMS scores were not discriminative (score bracket 2–2.5 = 13.3% participants; 3–3.5 = 73.3%; and 4–4.5 = 13.3%) so we re-scored the tests using a linear method: responses were scored as percentages of the maximum possible score (i.e., if participants got a perfect WMS score with a perfect performance they would have recalled 88 words. We therefore counted the number of words they had remembered and used the ratio of remembered words over the total as the score).
[5]The parallel analysis of the original set of stories (Saxe and Kanwisher, 2003) shed very similar results. Story reading time: FB = 3793.52 vs. FP = 4147.66 ms, $t(14) = -6.36$, $p = 0.001$, $d = -0.45$. Comprehension question response time: FB = 2799.69 vs. FP = 3317.79 ms, $t(14) = -3.48$, $p = 0.004$, $d = -0.67$. Comprehension question accuracy: FB = 91.40% vs. FP = 85.06%, $t(14) = 2.22$, $p = 0.044$, $d = 0.41$. ANOVA for response time with 2 (Story Type: FB vs. FP) × 2 (Question Probe: RP vs. RL; only six stories populate each condition due to the originally incomplete factorial design): main effect of Story Type ($F(1,14) = 12.60$, $p = 0.003$, $\eta_p^2 = 0.47$) where FB stories were faster; main effect of Comprehension Question ($F(1,14) = 4.79$, $p = 0.046$, $\eta_p^2 = 0.26$) where Representation questions was faster. Non-significant interaction ($F(1,14) = 1.48$, $p = 0.243$, $\eta_p^2 = 0.10$) although the effect of Comprehension Question was numerically smaller for FP than FB (117.5 vs. 414.7 ms). ANOVA for accuracy, 2 (Story Type: FB vs. FP) × 2 (Question Probe: RP vs. RL): main effect of Story Type with responses to FB marginally more accurate ($F(1,14) = 4.06$, $p = 0.064$, $\eta_p^2 = 0.23$); no effect of Comprehension Question ($F(1,14) = 1.28$, $p = 0.277$, $\eta_p^2 = 0.08$); significant interaction ($F(1,14) = 15.10$, $p = 0.002$, $\eta_p^2 = 0.52$) with *post hoc* LDS tests showing that FP–RP questions were clearly more difficult than any other condition (all $ps < 0.050$). Correlations: non-significant for WMS and sentence reading time: FB sentences $r = -0.25$, $p = 0.373$ and FP sentences $r = -0.20$, $p = 0.464$; WMS and response time to questions: the higher the WMS, the faster the question was answered. True for FB stories with RP probe ($r = -0.68$, $p = 0.006$) or RL probe ($r = -0.70$, $p = 0.004$) and for FP stories with RP probe ($r = -0.56$, $p = 0.028$) or RL probe ($r = -0.50$, $p = 0.056$); WMS and response accuracy to questions: FP stories with RP probe tended to be responded more accurately by those with a higher WMS ($r = 0.48$, $p = 0.073$; all other $r < 0.22$ and $ps > 0.400$).

FIGURE 2 | Scatter plots for WMS correlations with reading time (A), response times (B), and response accuracy (C). Significant correlations are marked with an asterisk.

DISCUSSION

The results presented here show that processing of FB and FP stories is not equivalent, raising concerns that these two conditions are not suitable for comparison in ToM paradigms. We found that FB stories were read faster than FP stories and replicated

previous findings of faster response times to comprehension questions following FB than FP stories. These response times were correlated across subjects with WMS scores, indicating the importance of working memory during the comprehension phase of the task. The most interesting result, however, was that comprehension questions following FP stories were particularly difficult when they required subjects to manipulate the information provided by the story in order to reconstruct the original state of the world, as in the case of representation questions (i.e., in the FP story presented in footnote 1, the apple hanging on the tree as shown in the photo as opposed to the apple lying on the ground after being blown off the tree by the wind). Only in this condition was comprehension accuracy correlated with WMS scores, consistent with a particularly heavy involvement of working memory. Finally, linguistic analyses showed that the increased difficulty and greater working memory requirements of representation questions following FP than FB stories were not due to the linguistic features of the questions, consistent with the hypothesis that they instead arose from the need to reorganize how information from FP stories was structured in memory in order to arrive at a correct response. It is possible that the moving window paradigm, in which words were presented one at a time, placed increased demands on working memory. Since these increased demands applied to both FB and FP stories, however, the use of this paradigm does not explain the observed difference between story types. At most, it may have made this difference easier to measure.

Since the information targeted by representation questions was presented at the beginning of the text in 87.5% of FP stories, the greater difficulty and larger working memory loads associated with representation questions following FP stories might have reflected a recency effect (Deese and Kaufman, 1957). However, 100% of FB stories also presented representation information prior to belief information without any cost in accuracy, reinforcing the view that the representation of FB stories was more flexible and less structured for a particular type of question than the representation of FP stories. A more flexible representation of FB information would also explain the faster reading times for this condition.

Developmental studies also suggest that FB and FP stories are not equivalent. While FB and FP tasks are comparable in difficulty in 3- to 5-year-olds (Zaitchik, 1990; Leekam and Perner, 1991; Leslie and Thaiss, 1992) no correlation between performance in the two tasks is found when children's age is partialed out. Therefore, the fact that they tend to be similar in difficulty in one age bracket seems more a coincidence than an index of a common underlying reasoning process (Perner and Leekam, 2008). Perner and Leekam (2008) also point out the lack of equivalence between FB and FP tasks since both beliefs and photos are representations of a reality, but the FB misrepresents its target (in the chocolate story, the place where the chocolates are currently stored) while the FP shows its target correctly (in the apple story, the apple hanging on the tree at the time the photo was taken).

IMPLICATIONS FOR NEUROIMAGING RESEARCH

The results reported here suggest two factors separate from ToM reasoning that might be responsible for reported activation differences between FB and FP stories. The first is related to language processing, since we found that FB stories were read faster than FP stories and the two types of stories were not matched in their syntactic structure. It is widely agreed that the right hemisphere plays an important role in language processing (Jung-Beeman, 2005), specifically during high level tasks involving comprehension of metaphors (Mashal et al., 2007) and jokes (Coulson and Wu, 2005), drawing inferences (Mason and Just, 2004), generating sentence endings (Kircher et al., 2001), and detecting inconsistencies in stories (Meyer et al., 2000). Interestingly, manipulations of the syntactic structure of sentences produce activity differences in right posterior STS (Friederici et al., 2009). Therefore, the RTPJ activations reported in ToM studies could be related to uncontrolled differences in linguistic features of the stories. This possibility is supported by the fact that results originally attributed to differences in mental processes have been later shown to be due to linguistic features of the experimental material. Happé et al. (1999) reported that right hemisphere damage (RHD) patients showed difficulty with ToM stories but not with control stories. Tompkins et al. (2008) retained Happé et al.'s original ToM set but created a new set of control stories that were better matched in linguistic difficulty. They found that RHD patients were not selectively impaired in ToM, although they were significantly worse overall than age-matched controls. Similarly, neuroimaging studies that have used material different than the stories of Saxe and colleagues (e.g., Saxe and Kanwisher, 2003) have tended to report less clear results (Fletcher et al., 1995; Happé et al., 1999; Gallagher et al., 2000; Vogeley et al., 2001).

A second factor that is separate from ToM processing and distinguishes FP and FB conditions is working memory load. Response times to comprehension questions were slower following FP than FB stories and comprehension accuracy was significantly worse for FP stories followed by representation questions than for the other conditions. WMS scores correlated with the time to answer comprehension questions in all four conditions and with the accuracy in answering representation questions following FP stories, indicating that working memory load is greater under FP conditions. Importantly, the magnitude of RTPJ deactivation co-varies with working memory load (Todd et al., 2005), consistent with other results showing that deactivations become larger as task difficulty increases (McKiernan et al., 2003). The dependence of RTPJ responses on working memory load was observed in a non-ToM paradigm, consistent with the view that RTPJ activity is modulated by a more basic process than ToM that is nonetheless likely operative during ToM paradigms. Therefore, RTPJ activity may be more positive in FB than FP conditions because it involves a lower working memory load.

The absolute polarity of responses in RTPJ is somewhat inconsistent across studies. While most papers find a relative activation for FB and a relative deactivation for FP (Saxe and Kanwisher, 2003; Saxe et al., 2006, 2009; Aichhorn et al., 2009; Scholz et al., 2009), some studies find two activations (Perner et al., 2006; Saxe and Powell, 2006; Kobayashi et al., 2007) or two deactivations (Mitchell, 2008). The reasons for this inconsistency are unclear, but may reflect the fact that studies typically use a block-like design that does not differentiate task phases or components such as text processing, memory maintenance, or the manipulation of information in order to respond to the comprehension question.

Different processes within a single trial have been shown to activate and deactivate the RTPJ, with the polarity of the observed response at any time point as the sum of the responses to the component processes (Shulman et al., 2003). Therefore, it may not be surprising that the overall polarity of the RTPJ response in ToM paradigms differs across studies depending on the duration and timing of the complex mixture of processes involved in FB and FP conditions. Irrespective of the observed polarity of the overall response, however, conditions involving higher working memory loads, such as the FP condition, will produce less positive RTPJ responses.

Theory of mind studies suggest that RTPJ is a key component of the network for belief reasoning. According to this proposal, FB reasoning is carried out by a specialized module (i.e., RTPJ) that is very effective in reasoning about mental states while the brain module specialized for other types of reasoning is not as effective. We suggest instead that the demands on basic processes needed to successfully perform both conditions are not equivalent and as a result, the areas modulated by these processes are differentially activated. Mind reading is a highly valued ability in any social environment and the amount of practice that we carry out from early childhood likely far exceeds that for other types of causal reasoning. We suggest that this extensive practice allows for a more flexible mental representation of FB than FP stories in terms of the frame of reference and for a more straightforward manipulation of information when a reality check is necessary, and results in the use of fewer resources during processing. Therefore, RTPJ activity may not be related to mental state reasoning or any type of reasoning *per se*, but to a more basic process that is shared to different extents by FB and FP stories, such as the use of working memory.

Additional evidence for this proposal can be found in studies of children. If extensive practice makes us more proficient when reasoning about mental states, we would expect children to show no differences in RTPJ activation, since they are not yet experts

and reasoning about FB requires as much resources as reasoning about FP. As they become more proficient with the former type of reasoning, differences in RTPJ activity should be more evident. Saxe et al. (2009) observed this pattern of results in children aged 6–11. While RTPJ was recruited equally for FB and control stories in younger children, it was only recruited (i.e., RTPJ "activation") for mental reasoning in older children. Although the results indicated "activations" for each child in both conditions, the baseline for this comparison was another condition that had shown a large deactivation in the group analysis. Therefore, it is quite possible that if the task data were referenced to a resting baseline, most of the younger children would actually show a deactivation for both tasks (indexing a greater difficulty in both of them), while older kids would only show a deactivation for the control stories, which continued to be difficult, but not for the FB stories.

Taken together, the current results show that comparisons of FB and FP tasks may not cleanly isolate ToM processes. The behavioral differences between the two types of stories indicate that the inferential processes necessary to comprehend the text and the difficulty level of these processes are not comparable. While we do not claim that WM differences are the sole determinant of the lack of equivalence between FB and FP conditions, our results do show that these conditions differ in their cognitive processing requirements. In conclusion, more attention should be given to the lack of equivalence between contrasting conditions in ToM paradigms before drawing general conclusions about the functional role of brain areas such as the RTPJ.

ACKNOWLEDGMENTS

REFERENCES

Aichhorn, M., Perner, J., Weiss, B., Kronbichler, M., Staffen, W., and Ladurner, G. (2009). Temporo-parietal junction activity in theory-of-mind tasks: falseness, beliefs, or attention. *J. Cogn. Neurosci.* 21, 1179–1192.

Balogh, J., Zurif, E., Prather, P., Swinney, D., and Finkel, L. (1998). Gap-filling and end-of-sentence effects in real-time language processing: implications for modeling sentence comprehension in aphasia. *Brain Lang.* 61, 196–182.

Balslev, D., Nielsen, F. A., Paulson, O. B., and Law, I. (2005). Right temporoparietal cortex activation during visuo-propirioceptive conflict. *Cereb. Cortex* 15, 166–169.

Blanke, O., and Arzy, S. (2005). The out-of-body experience: disturbed self-processing at the temporo-parietal junction. *Neuroscientist* 11, 16–24.

Corbetta, M., Patel, G., and Shulman, G. L. (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306–324.

Coulson, S., and Wu, Y. C. (2005). Right hemisphere activation of joke-related information: an event-related brain potential study. *J. Cogn. Neurosci.* 17, 494–506.

Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* 19, 450–466.

Decety, J., and Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how lo.level computational processes contribute tometa-cognition. *Neuroscientist* 13, 580–593.

Deese, J., and Kaufman, R. A. (1957). Serial effects in recall of unorganized and sequentially organized verbal material. *J. Exp. Psychol.* 54, 180–187.

Farrer, C., and Frith, C. D. (2002). Experiencing oneself vs. another person as being the cause of an action: the neural correlates of experience of agency. *Neuroimage* 15, 596–603.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., and 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., Makuuchi, M., and Bahlmann, J. (2009). The role of the posterior superior temporal cortex in sentence comprehension. *Neuroreport* 20, 563–568.

Friederici, A. D., and Weissenborn, J. (2007). Mapping sentence form onto meaning: the syntax-semantic interface. *Brain Res.* 1146, 50–58.

Gallagher, H. L., Happé, F., Brunswick, N., Fletcher, P. C., Frith, U., and 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.

Happé, F., Brownell, H., and Winner, E. (1999). Acquired "theory of mind" impairments following stroke. *Cognition* 70, 211–240.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends Cogn. Sci.* 9, 512–518.

Just, M. A., Carpenter, P. A., and Wooley, J. D. (1982). Paradigms and processes in reading comprehension. *J. Exp. Psychol. Gen.* 111, 228–238.

Kircher, T. T., Brammer, M., Tous Andreu, N., Williams, S. C., and McGuire, P. K. (2001). Engagement of right temporal cortex during processing of linguistic context. *Neuropsychologia* 39, 798–809.

Kobayashi, C., Glover, G. H., and Temple, E. (2007). Children's and adults' neural bases of verbal and nonverbal "theory of mind." *Neuropsychologia* 45, 1522–1532.

Leekam, S., and Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition* 40, 203–218.

Leslie, A. M., and Thaiss, L. (1992). Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition* 43, 225–251.

Mashal, N., Faust, M., Hendler, T., and Jung-Beeman, M. (2007). An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain Lang.* 100, 115–126.

Mason, R. A., and Just, M. A. (2004). How the brain processes causal inferences in text. A theoretical account of generation and integration component processes utilizing both cerebral hemispheres. *Psychol. Sci.* 15, 1–7.

McKiernan, K. A., Kaufman, J. N., Kucera-Thompson, J., and Binder, J. R. (2003). A parametric manipulation of factors affecting task-induced deactivation in functional neuroimaging. *J. Cogn. Neurosci.* 15, 694–408.

Meyer, M., Friederici, A. D., and von Cramon, D. Y. (2000). Neurocognition of auditory sentence comprehension: event related fMRI reveals sensitivity to syntactic violations and task demands. *Brain Res. Cogn. Brain Res.* 9, 19–33.

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

Perner, J., Aichhorn, M., Kronbichler, M., Staffen, W., and Ladurner, G. (2006). Thinking of mental and other representations: the roles of left and right temporo-parietal junction. *Soc. Neurosci.* 1, 245–258.

Perner, J., and Leekam, S. (2008). The curious incident of the photo that was accused of being false: issues of domain specificity in development, autism, and brain imaging. *Q. J. Exp. Psychol.* 61, 76–89.

Saxe, R., and 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., and Powell, L. J. (2006). It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol. Sci.* 17, 692–699.

Saxe, R., Schulz, L. E., and Jiang, Y. V. (2006). Reading minds versus following rules: dissociating theory of mind and executive control in the brain. *Soc. Neurosci.* 1, 284–298.

Saxe, R., Whitfield-Gabrieli, S., Scholz, J., and Pelphrey, K. A. (2009). Brain regions for perceiving and reasoning about other people in school-aged children. *Child Dev.* 80, 1197–1209.

Scholz, J., Triantafyllou, C., Whitfield-Gabrieli, S., Brown, E. N., and Saxe, R. (2009). Distinct regions of right temporo-parietal junction are selective for theory of mind and exogenous attention. *PLoS ONE* 4, e4869. doi:10.1371/journal.pone.0004869

Shulman, G. L., McAvoy, M. P., Cowan, M. C., Astafiev, S. V., Tansy, A. P., d'Avossa, G., and Corbetta, M. (2003). Quantitative analysis of attention and detection signals during visual search. *J. Neurophysiol.* 90, 3384–3397.

Todd, J. J., Fougnie, D., and Marois, R. (2005). Visual short-term memory load suppresses temporo-parietal junction activity and induces inattentional blindness. *Psychol. Sci.* 16, 965–972.

Tompkins, C. A., Scharp, V. L., Fassbinder, W., Meigh, K. M., and Armstrong, E. M. (2008). A different story on "theory of mind" deficit in adults with right hemisphere brain damage. *Aphasiology* 22, 42–61.

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happé, F., Falkai, P., Maier, W., Shah, N. J., Fink, G. R., and Zilles, K. (2001). Mind reading: neural mechanisms of theory of mind and self-perspective. *Neuroimage* 14, 170–181.

Wimmer, H. L., and Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 103–128.

Zaitchik, D. (1990). When representations conflict with reality: the preschooler's problem with false beliefs and "false" photographs. *Cognition* 35, 41–68.