Frontal and Temporo-Parietal Lobe Contributions to Theory of Mind: Neuropsychological Evidence from a False-Belief Task with Reduced Language and Executive Demands

Ian A. Apperly, Dana Samson, Claudia Chiavarino, and Glyn W. Humphreys

Abstract

■ A model of the functional and anatomical basis of belief reasoning is essential for understanding the relationship between belief reasoning and other cognitive processes in both normal development and pathology. Studies of brain-damaged patients can give valuable insights into the nature of belief processing but pose unique methodological problems. The current study addresses these problems by using a nonlinguistic belief-reasoning task with substantially reduced executive demands. A case series of 12 brain-damaged patients is presented. The belief-reasoning errors of four patients with damage to the prefrontal cortex appeared to arise from these patients' executive function problems. The belief-reasoning errors of three patients with damage to the temporo-parietal junction could not easily be accounted for in this way, raising the possibility that this brain region has a necessary role in representing beliefs, rather than handling the executive demands of belief-reasoning tasks. We discuss the importance of gaining empirical evidence about the scope of ''theory of mind'' impairments, and the important role for neuropsychological studies in this project. ■

INTRODUCTION

It is increasingly recognized that reasoning about mental states, such as beliefs, desires, and knowledge (often referred to as ''theory of mind''), is central to a range of cognitive activities including our ability to communicate and to explain and predict behavior (e.g., Malle, Moses, & Baldwin, 2001; Baron-Cohen, Tager-Flusberg, & Cohen, 2000; Sperber, 2000a, 2000b). More than 20 years of research has developed techniques to explore these abilities in different species and in children at different ages (e.g., Baron-Cohen et al., 2000; Mitchell & Riggs, 2000; Astington, Harris, & Olson, 1988; Lewis & Mitchell, 1994). In comparison, relatively few studies have been conducted on adults. As a result, we lack a clear account of the cognitive and anatomical basis of adult mental state reasoning. Developing such an account is crucial if we are to understand the place of mental state reasoning in relation to other aspects of adult cognition. An adult model is also vital to understand what children are developing and to understand the breakdown of adult abilities in certain forms of mental illness and brain damage. In the current article, we describe a new method that addresses some of the unique requirements of a neuropsychological approach to the study of reasoning about beliefs. We then present the findings from a study of a case series of 12 brain-damaged patients that suggest a role not only for the prefrontal cortex but also for the left temporo-parietal junction (TPJ) in belief reasoning.

A task commonly used to examine mental state reasoning in children requires the inference that someone has a false belief (e.g., Wimmer & Perner, 1983; see also, e.g., Mitchell & Riggs, 2000). For example, the child might be told a story where Billy puts his chocolate in the cupboard, then goes outside to play. While he is away, his mother moves the chocolate to the fridge. The child is asked where Billy will first look for his chocolate when he returns. To answer correctly, the child must infer that Billy thinks that the chocolate is still in the cupboard. Many 4- and 5-year-olds answer correctly, while many 3-year-olds judge incorrectly that Billy will look in the fridge (i.e., they answer from their own knowledge and not the perspective of the other person). Neuroimaging studies have shown activation in the frontal lobes (e.g., Gallagher et al., 2000; Fletcher et al., 1995), as well as more posterior regions such as the TPJ (e.g., Saxe & Kanwisher, 2003; Gallagher et al., 2000), when neurologically intact adults perform such tasks. It has been hypothesized that the role of the frontal lobes could be in holding separate perspectives (e.g., Gallagher & Frith, 2003) or in resisting interference from one's own perspective (Ruby & Decety, 2003). As

University of Birmingham

for the TPJ, there is some debate about whether this region is involved in mental state reasoning per se (e.g., Saxe & Kanwisher, 2003) or just in lower level processing of socially relevant stimuli such as human movements (Allison, Puce, & McCarthy, 2000; Frith & Frith, 1999). However, in isolation, neuroimaging data do not show whether particular brain regions are necessary for belief reasoning, nor do they provide any direct evidence about the functions of these areas in solving belief-reasoning tasks. Studies of adults with neurological damage are, therefore, a potentially valuable source of complementary evidence on which to base a model of belief reasoning.

To date, studies with patients have produced conflicting results about the lateralization of belief reasoning (e.g., Channon & Crawford, 2000; Happé, Brownell, & Winner, 1999; Stone, Baron-Cohen, & Knight, 1998; Winner, Brownell, Happé, Blum, & Pincus, 1998) and/or whether the frontal lobes (or executive functions) are necessary (e.g., Bird, Castelli, Malik, Frith, & Husain, 2004; Fine, Lumsden, & Blair, 2001). It seems possible that this pattern of findings is the result of difficulty in finding appropriate tasks for testing patients. The selection of appropriate tasks for neuropsychological testing frequently poses a dilemma: The tasks must be difficult enough to generate errors yet simple enough that errors are not merely due to more general processing demands of the task. Patients who present with a social impairment often do not make errors on first-order false-belief tasks (such as predicting where Billy will look for his chocolate). To overcome this problem, neuropsychological studies have commonly employed more complicated tasks designed to place a heavier load on the participants' mental state reasoning abilities. Such tasks may require the participant to evaluate what one person thinks another person is thinking (second-order belief reasoning, Perner & Wimmer, 1985) or to relate two different mental states, such as a belief and a resulting emotion, or a belief and an intention. However, these tasks often entail comprehension of linguistically complex narratives and questions. This makes it difficult to know whether errors reflect difficulty with belief reasoning per se or with meeting high incidental demands on language and executive functions. A number of studies support this concern, showing that patients may make errors on story-based test stimuli because of the incidental demands that they pose on memory (Stone et al., 1998), pragmatic processing (Surian & Siegal, 2001; Siegal, Carrington, & Radel, 1996), and executive function (Channon & Crawford, 2000). Hypotheses about the functional and anatomical basis of mental state reasoning are likely to remain difficult to test using batteries of diverse and complicated belief-reasoning tasks with high incidental processing demands.

A key aim of the research reported in the current article was to use a simple, first-order false-belief task with highly regular trials in which processing demands were either very much reduced or closely controlled. The task was adapted from a false-belief task devised by Call and Tomasello (1999) to make it suitable for work with adult participants. Participants watched a series of short videos where they knew that there is an object in one of two identical boxes but did not initially know which. Instead, a helpful female character in the video gave them a clue to the object's location by pointing to one of the boxes. In the majority of trials, this clue was accurate. On false-belief trials, a male character swapped the boxes while the woman was absent; thus, when she returned, she inadvertently indicated the wrong box. However, this was still a useful clue, provided participants took account of the woman's false belief. Working memory control trials followed the same sequence except that the woman indicated one of the boxes before leaving the room. In this case, inferring the correct location of the object did not require the attribution of a false belief, but as in false-belief trials, participants had to remember that the location of the object changes when the boxes are swapped. Inhibition control trials followed the same sequence as false-belief trials, but instead of swapping the boxes, the man performed a visible transfer of the object from one box to the other. Participants did not need to infer a belief to locate the object, but to respond correctly, they need to inhibit pointing to the wrong location indicated by the woman when she returned to the room. True-belief trials followed the same sequence as false-belief trials, but the boxes were not swapped while the woman was out of the room; thus, the correct response was to point to the same box as the woman. These trials served to check that correct answers on false-belief trials were not the result of the superficial strategy of pointing to the opposite box from the one indicated by the woman. On clue confirmation trials, the woman indicated one box before leaving the room and, in full view of the participant, the man took the object from this box and placed it in the other box. These trials provided evidence that the woman pointed accurately when she was well informed and served as filler trials.

This new task solves a number of problems with existing belief-reasoning tasks. First, test trials can be administered entirely without language, enabling us to test a wider range of participants than is normally possible and eliminating the danger that the language-processing demands of the overall task could be responsible for participants' errors. Second, our task eliminates a key executive demand that is typically confounded with belief reasoning. In standard false-belief tasks, the participant knows the correct answer (that the chocolate is in the fridge in the earlier example). Thus, to evaluate a character's false belief, the participants must resist interference from their own knowledge. Eliminating this demand is particularly important for exploring the functional relation between belief reasoning and executive functions (or the anatomical relation between belief reasoning and the frontal lobes),

because it is well known that patients with frontal lesions and impaired executive function can have difficulty resisting interference (Stuss, Floden, Alexander, Levine, & Katz, 2001; Stuss & Benson, 1986). Third, separate trials control for specific working memory and inhibitory control demands that were not eliminated from the belief-reasoning trials. Errors on these control trials would indicate that any belief-reasoning errors could potentially be attributed to more general processing demands of the task. Moreover, we might expect participants showing such error patterns to show impairment on independent tests of working memory and/or executive function. Fourth, false-belief trials in our task are highly regular, unlike those of conventional story-based tasks. The heterogeneity of stimuli in conventional tasks is likely to pose varying executive and memory demands, meaning that the number of errors on such a stimulus set could reflect varying difficulty of solving each belief-reasoning problem, rather than difficulty in belief reasoning per se. By making the trials of our task highly regular, the number of errors should index the difficulty of inferring a belief in the face of identifiable and consistent processing demands. Fifth, control trials require the participant to comprehend exactly the same socially relevant cues as false-belief trials. Thus, a pattern of pure errors on false-belief trials could not be due to a difficulty with understanding socially relevant cues.

Preliminary data from this task have provided the first evidence that damage to the TPJ can lead to a relatively specific deficit in inferring someone else's belief (Samson, Apperly, Chiavarino, & Humphreys, 2004). In the current study, we presented the task to a larger range of patients. This is the first study of frontal patients using a method that deconfounds belief reasoning from the key executive demand of resisting interference from one's own knowledge of the correct answer.

Participants also completed a set of conventional story-based first-order false-belief tasks (see, e.g., Mitchell & Riggs, 2000; Stone et al., 1998). Our first reason for including story-based tasks was to examine whether our new false-belief tasks had indeed enabled us to test patients who would not have been able to pass the control questions of standard tasks. However, the story-based tasks also incorporated a counterfactual reasoning control condition that has not previously been used in neuropsychological studies. In the above example, the matched counterfactual question would be ''What if Billy's mother had not moved the chocolate, where would it be? In the fridge or in the cupboard?'' Studies of children suggest that false-belief reasoning and counterfactual reasoning tasks of comparable complexity correlate and are of similar difficulty (e.g., German & Nichols, 2003; Riggs, Peterson, Robinson, & Mitchell, 1998). Moreover, counterfactual reasoning may share important formal similarities with false-belief reasoning (e.g., Peterson & Riggs, 1999). For our current purposes, these findings suggest that a counterfactual question is a valid control for the inferential complexity of a false-belief question, while not requiring the participant to infer a belief.

RESULTS

Video-Based False-Belief Task

Because the task consisted of a binary-choice response, performance was evaluated against a 50% chance baseline. For an individual to score statistically above chance on a particular trial type, they needed to give 10 or more out of a possible 12 correct responses (10/12 correct has a one-tailed probability of .019 by binomial test).

There were three distinct patterns of performance across the patients in the false belief, working memory control, and inhibition control trials (see Figure 1). As reported earlier, a group of three patients (P.F., R.H., and D.B.) did not perform significantly above chance on false-belief trials but made no errors on control trials, suggesting a relatively specific impairment with belief reasoning (Samson et al., 2004). A second group of four patients also did not perform significantly above chance on false-belief trials and, in addition, made errors on working memory control trials. Two of these patients (D.S. and F.K.) did not perform above chance on the working memory control trials, suggesting that their errors on false-belief trials could reflect a more general difficulty with the incidental working memory demands of the task. Two other patients (G.A. and P.W.) performed just above chance on the working memory control trials, with 10/12 correct answers. On this evidence, we are cautious about whether these two patients have a pure belief-reasoning impairment or whether they are indeed more impaired on the working memory control measure than the patients who only show belief-reasoning errors. The third group of five patients (W.B.A., J.B., C.N., M.H., and P.H.) performed above chance on false-belief trials, with no patient giving fewer than 11/12 correct answers. Importantly, all of these patients were also above chance for true-belief filler trials, indicating that success on false-belief trials reflected genuine belief reasoning and not a superficial strategy of pointing to the opposite box from that indicated by the character in the video. These patients also performed well on control trials, with only W.B.A. not performing above chance on the working memory control trial.

The task was also presented to three neurologically intact control subjects (aged 55–64) and none of these controls made errors on any trial.

Story-Based False-Belief Task

The story-based task also consisted of a binary-choice response and, thus, performance was evaluated against a 50% chance baseline, with the above-chance perform-

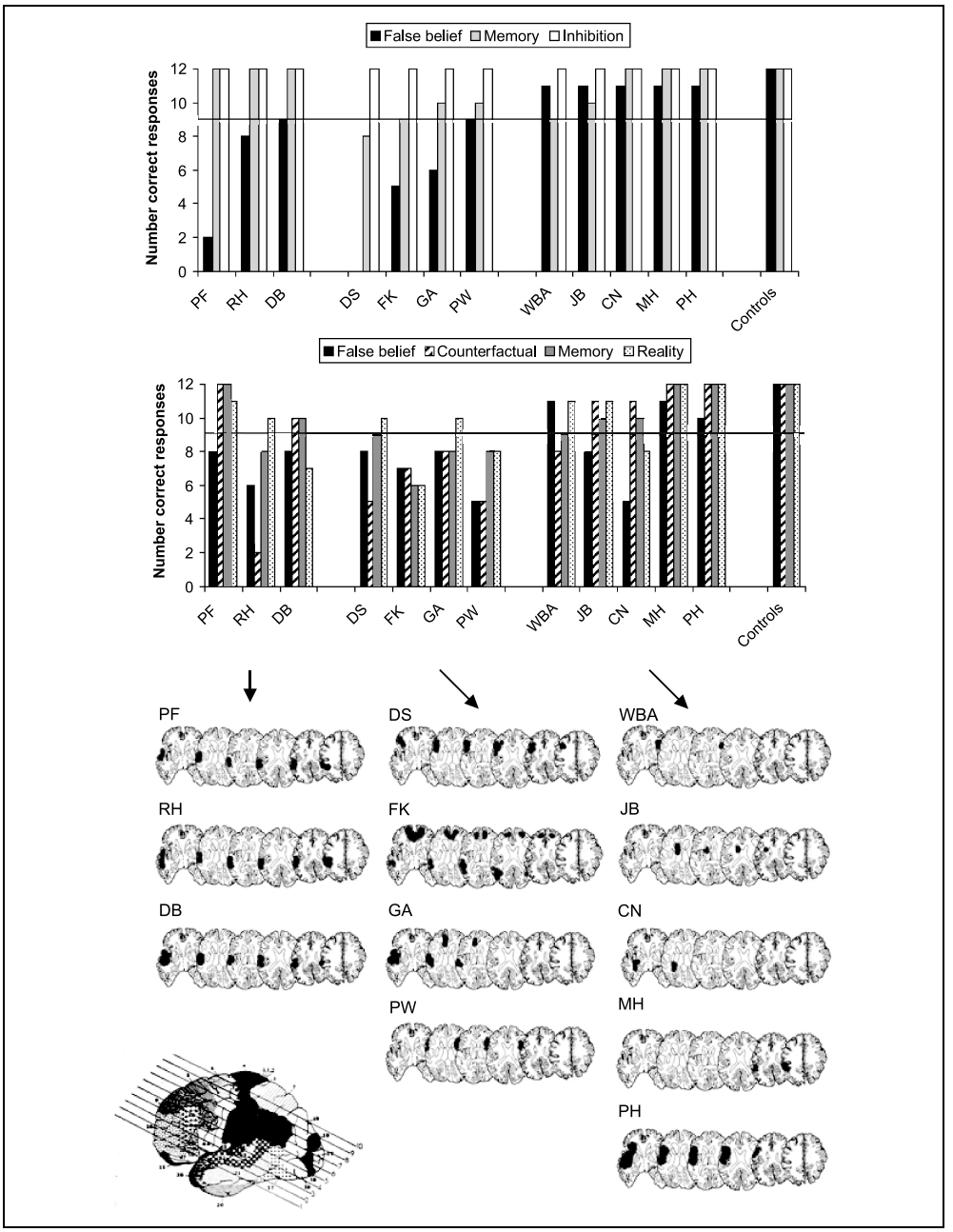

Figure 1. Behavioral performance for the video- (upper graph) and story-based (lower graph) false-belief tasks as well as brain lesion reconstruction for the 12 patients. The number of correct responses is reported in black for the false-belief trials, in gray for the memory control trials, in white for the inhibition control trials, in striped black for the counterfactual control trials, and in dotted white for the reality control trials. The horizontal lines shows the level above which the score is significantly higher than chance. Lesions have been drawn onto standard slices from Gado, Hanaway, and Frank (1979). Only slices 3–8 are reported. From left to right and based on the performance in the video-based task: the three patients with a pure deficit in false-belief reasoning, the four patients who make errors on both the false belief as well as the control trials, and the five patients who show no impairment for the false-belief trials.

ance corresponding to 10 or more out of a possible 12 correct responses (see Figure 1). Of the 12 patients, 7 (R.H., D.B., D.S., F.K., P.W., C.N., and G.A.) were not above chance on the false-belief questions nor on one or more control questions; thus, their false-belief errors could have been due to failure to understand the task or the questions or to remember crucial information. Thus, although our story-based task was substantially simpler than many that have been used in previous neuropsychological studies (we used a first-order rather than a second-order belief-reasoning task), it still posed insurmountable processing demands for many patients. It is noteworthy that five of these patients did perform above chance on both control trials in the video-based task (only D.S. and F.K. did not) with three (C.N., R.H., and D.B.) making no errors.

Two patients (P.F. and J.B.) were above chance on control and counterfactual trials, but not above chance on false-belief trials. This could be seen as evidence for domain specificity in these patients' belief-reasoning deficit. However, this finding must be treated with caution, since one of these patients (J.B.) performed significantly above chance on false-belief trials of the video-based task. Notably, P.F. was impaired on both false-belief tasks while performing above chance on all control conditions.

None of the three control participants made any errors.

Independent Measures of Language and Executive Function

Only one patient (C.N.) showed no impairment on any executive task (see Table 1). Performance was particularly poor in the group of four patients who, on the video-based task, were not above chance on false-belief trials and made errors on working memory control trials. One of these patients, F.K., showed particularly severe impairment with an impaired score on all measures. Other patients in this group accounted for the poorest individual performance on the Brixton (P.W.) and the second-poorest performance on one of the inhibition measures (G.A.) and on the working memory measures (D.S.). However, it was also the case that for every test there was always at least one patient in this group who was less impaired than a patient who was above chance on false-belief trials. Thus, no single executive impairment could specifically account for these patients' errors on the video-based task.

It was also the case that patients who showed pure false-belief-reasoning impairments on the video-based task did not show a distinctive profile of impairment on the independent tests of executive function. In addition, for every measure apart from W.M., there was at least one patient who performed at least as poorly yet scored above chance on all video-based false belief and control trials. In the case of working memory, two patients (R.H. and P.H.) could not be tested because their language impairments led to a verbal span (of two) that was too low for them to complete working memory tasks. Despite this, R.H. showed pure belief-reasoning errors whereas P.H. showed no significant impairment on belief reasoning or control trials.

Finally, there was no evidence of a specific link between a belief-reasoning deficit and either a semantic deficit (as assessed by the synonym task) or a grammatical deficit (as assessed by the sentence/picture matching task). In both groups of patients who made belief-reasoning errors, there was always someone that performed better on the language measures than patients who made no errors on the belief-reasoning trials. This is consistent with Varley, Siegal, and Want (2001), whose study of two severe aphasics suggested that severe impairment of language may not preclude the ability to reason about false beliefs.

Relation to Lesion Site

The patients' lesion reconstructions are shown in Figure 1. As reported by Samson et al. (2004), all three patients with a pure false-belief-reasoning deficit (P.F., D.B., and R.H.) had a lesion to the left TPJ, in each case involving the superior temporal and the angular gyri. Patients who showed difficulty with both false belief and control trials all had lesions to the frontal lobe, either unilaterally left (D.S.), right (P.W.), or bilaterally (F.K. and G.A.). In contrast, the lesion sites of the unimpaired group of patients were more diverse, including frontal, temporal, parietal, and subcortical regions (see Table 2).

DISCUSSION

By reducing the incidental language and executive demands of first-order false-belief tasks, we were able to test more severely impaired patients than was possible with standard methods, including patients with aphasia. Five patients who did not respond significantly above chance on control questions of the story-based tasks were able to pass the control trials of the video-based task. Using a relatively simple belief-reasoning task enabled us to create closely matched control conditions. This approach makes it harder to explain ''pure'' belief-reasoning errors in terms of difficulty with incidental task demands. We believe that tasks of this kind will significantly extend the utility of a neuropsychological approach to the study of reasoning about mental states.

As reported by Samson et al. (2004), three patients (D.B., P.F., and R.H.) were not above chance on false-belief trials but made no errors on any other trials. Strikingly, all three patients had lesions involving the left TPJ. As far as we are aware, this evidence is the first from a study of patients that concurs with the neuroimaging data highlighting the importance of this region (e.g., Hooker et al., 2003; Saxe & Kanwisher, 2003; Gallagher et al., 2000; Calvert et al., 1997; Rizzolatti et al., 1996) in addition to the frontal lobes. The common lesion site in these three patients was the left superior temporal and the left angular gyri. Interestingly, lesions encroaching on the left superior temporal gyrus not involving the left angular gyrus such as in the case of P.H., or conversely, lesions to the left angular gyrus not extending into the superior temporal gyrus, such as in the case of M.H., were not sufficient to produce a similar pattern of ''pure'' false-belief-reasoning deficit. Although we show here the importance of the left TPJ, we do not exclude, at this stage, that lesions to the right TPJ would produce a similar pattern of deficit (it happened that none of the patients in our sample had lesions to the right TPJ). However, in the light of neuroimaging studies that usually show bilateral TPJ activation (e.g., Saxe & Kanwisher, 2003; Gallagher et al., 2000), our findings do suggest that unilateral lesions to the TPJ are sufficient to disrupt belief reasoning.

We previously argued that the ''pure'' belief-reasoning deficit in D.B., P.F., and R.H. cannot be due to difficulty processing low-level social cues because both our control and false-belief trials required patients to compre-

Table 1. Patients' and Controls' Mean Score (% Correct Responses Unless Otherwise Stated) on the Independent Executive Function and Language Measures

	Patients with a Pure False-Belief-Reasoning Impairment			*Patients Making Errors for Both False-Belief and Control Trials*				*Patients Unimpaired for the False-Belief Trials*					*Controls [Mean (Range)]*
	P.F.	*R.H.*	*D.B.*	*D.S.*	*F.K.*	*G.A.*	*P.W.*	*W.B.A.*	*J.B.*	*C.N.*	*M.H.*	*P.H.*	
Executive function													
Working memory: manipulation (%)	94	**Impaired**[a]	**73**	94	**23**	**81**	**86**	**79**	97	100	94	**Impaired**[a]	99 (88–100)
Working memory: resistance to interference (%)	39	**Impaired**[a]	**23**	52	**15**	44	47	90	47	54	81	**Impaired**[a]	71 (31–100)
Working memory: updating (%)	**33**	**Impaired**[a]	**50**	**31**	**29**	85	86	67	78	71	**56**	**Impaired**[a]	86 (67–100)
Inhibition: stimulus selection (cost)	0.25	0.33	0.48	0.36	**12.79**	**3.8**	0.20	**0.81**	0.30	0.39	0.50	0.30	0.18 (0.02–0.53)
Inhibition: response selection (cost)	**3.93**	0.20	0.40	0.35	**8.31**	**0.85**	**1.20**	0.48	**1.21**	0.22	**1.65**	**1.49**	0.30 (0.09–0.63)
Shifting: focus of attention (cost)	1.72	**2.46**	1.56	1.65	**16.72**	0	1.55	**2.47**	**4.91**	0.64	0.12	**1.78**	0.97 (0.44–1.75)
Shifting: arithmetical operation (cost)	**3.59**	1.23	1.53	0.91	–	0.63	2.16	1.23	0.89	0.39	–	1.3	0.87 (0.05–2.20)
Brixton (%)[b]	50	**39**	67	67	**35**	61	**35**	**37**	57	70	65	78	Impaired if < 42
Language													
Written synonym matching (%)	84	**39**	**50**	66	**48**	**44**	75	93	–	88	86	**59**	
Sentence/picture matching (%)[c]	**78**	**58**	**77**	–	**43**	90	**73**	**65**	–	92	90	**67**	–

Scores outside the normal range are in **bold**. Except for the Brixton for which we considered the published norms, all executive tasks were presented to a group of 16 controls (ages 46–68). For the inhibition and shifting task, the cost was calculated as the RT divided by the number of correct responses in the executive condition minus the RT divided by the number of correct responses in the control (nonexecutive) condition.

[a]The patients' digit span was too low (2) for them to be tested on the working memory tasks that require to recall at least 3 digits.

[b]Burgess and Shallice (1997).

[c]PALPA 55 (Kay, Lesser, & Colheart, 1992).

Table 2. Patients' Characteristics and Lesion Description

Patient	*Sex/Age/ Handedness*	*Main Lesion Site*	*Major Clinical Symptoms*	*Etiology*	*Years Post-Onset*
C.N.	M/47/R	Bilateral medial temporal lobes (more pronounced on left)	Mild amnesia	Herpes simplex encephalitis	10
D.B.	M/68/R	Left parietal inferior (angular gyrus), superior, and middle temporal gyri	Aphasia	Stroke	6
D.S.	M/70/R	Left inferior, middle and superior frontal gyri	Right hemiplegia, aphasia	Stroke	14
F.K.	M/35/R	Bilateral superior and medial frontal regions, bilateral superior and medial temporal gyri, bilateral lateral occipital gyri	Agnosia, aphasia, dysexecutive syndrome	Anoxia	14
G.A.	M/49/R	Bilateral medial and anterior temporal lobes, extending into left medial frontal region	Aphasia, amnesia, dysexecutive syndrome	Herpes simplex encephalitis	13
J.B.	F/58/R	Left thalamus and ischemic change related to anterior horns of lateral ventricles	Right hemiplegia	Stroke	2
W.B.A.	M/58/R	Right inferior and middle frontal gyri, right superior temporal gyrus	Aphasia	Stroke	3
M.H.	M/50/R	Left angular and supramarginal gyri, lentiform nucleus	Right extinction, optic ataxia	Anoxia	10
P.F.	F/55/R	Left inferior parietal (angular and supramarginal gyri) and superior temporal gyri	Right extinction, dysgraphia	Stroke	8
P.H.	M/31/R	Left medial and superior temporal, left inferior and middle frontal gyri	Right hemiplegia, aphasia	Stroke	5
P.W.	M/72/R	Right inferior and middle frontal gyri, right superior temporal gyrus	Left hemiplegia, dysexecutive syndrome	Stroke	4
R.H.	M/70/L	Left inferior parietal (angular and supramarginal gyrus) and superior temporal gyrus	Right neglect, aphasia	Stroke	8

M = male; F = female; R = right; L = left.

hend exactly the same social cues (i.e., the character's pointing, see Samson et al., 2004). This highlights the role of the TPJ not only in low-level social processing but also in high-level social reasoning. In the current article, we show that the patients' performance could not be uniquely linked to an impairment in a specific component of executive function. Although there is evidence that these patients have executive function problems, for each measure on which one of these patients is impaired, there is an example of another patient who is more impaired but who does not make belief-reasoning errors. It remains possible that the level of impairment on a particular measure is not vital, but that belief-reasoning errors will occur if there are certain combinations of impairment on independent tasks, or a certain overall level of impairment. This issue remains open and needs to be addressed with a larger sample of patients. However, on the current data, there is little evidence

that belief-reasoning errors in these three patients are due to a deficit in some component of executive process of belief reasoning.

What is the nature of the deficit in these patients? Many authors, from a variety of theoretical perspectives have argued that belief reasoning involves a domain-specific cognitive module (e.g., Sperber, 2000a, 2000b; Segal, 1996; Baron-Cohen & Ring, 1994; Fodor, 1992; Leslie & Thaiss, 1992). The current data are clearly consistent with this hypothesis, but the scope of any such conclusion must be regarded with caution. The most we can conclude on the strength of the findings from our video-based task (and other existing neuropsychological studies of belief reasoning) is that reasoning of the formal complexity of belief reasoning is impaired. Stronger conclusions about modularity or domain specificity require that patients who show a deficit in belief reasoning are significantly less or significantly more impaired on formally similar reasoning tasks that are not about beliefs. Our narrative-based false-belief tasks had the potential to provide evidence on this issue since they included a counterfactual reasoning question, and counterfactual reasoning has important formal similarities to belief reasoning (Peterson & Riggs, 1999). Interestingly, patient P.F. was not above chance on false-belief questions but was above chance on counterfactual and control questions. Further investigation is clearly needed to describe the scope of P.F.'s reasoning impairment more precisely.

When we presented the video task to a larger group of patients with various lesion sites we found no clear evidence for the same pattern of ''pure'' false-belief errors resulting from lesions to another brain area. In particular, four patients with frontal lesions (D.S., F.K., G.A., and P.W.) were not above chance on false-belief trials, but also made errors on working memory control trials. This finding is worthy of examination because a number of authors have used evidence of belief-reasoning difficulties in patients with frontal lesions to argue that frontal regions are specifically involved in belief reasoning (Frith & Frith, 2003; Gallagher & Frith, 2003; Stone et al., 1998).

There are a number of reasons why patients with frontal lesions might fail belief-reasoning tasks, each with different implications for the role of frontal regions. First, it is well known that damage to the frontal lobes can lead to difficulty resisting interference from salient alternative responses (e.g., Stuss et al., 2001; Stuss & Benson, 1986). In existing studies of false-belief reasoning, the participant must resist interference from their own knowledge of the objectively correct answer (e.g., that the chocolate is in the fridge, in the example given in the Introduction) to attribute a false belief (e.g., that the chocolate is in the cupboard). Moreover, because everyday occurrences of social reasoning problems (including belief reasoning) commonly entail setting aside what one knows, thinks, or feels to be the case, problems with this process may be important in explaining some of the social–cognitive impairments that are described in patients with frontal impairments (see e.g., Damasio, Tranel, & Damasio, 1990). However, without removing this confounding factor from experimental tasks, it is difficult to reach strong conclusions about the role of frontal systems in belief reasoning. Our video-based task is unique in neuropsychological studies of belief reasoning because at the point when the participant infers a false belief, they do not know the objectively correct answer themselves. Although there was no possibility of interference from the correct answer, four patients with frontal lesions (D.S., F.K., G.A., and P.W.) made belief-reasoning errors on the video-based task. However, because these patients also made errors on working memory control trials, it appears that their difficulty might lie with meeting other general processing demands of the task, rather than with belief reasoning per se.

It is common in existing studies, as in the current study, to use patients' errors on independent tests of language and executive function to help interpret the basis for errors on belief-reasoning tasks. The logic of this approach is that if there is a correlation between performance on independent tests and false-belief tasks, then belief-reasoning errors are probably due to a more general processing deficit. The absence of such a relationship is (weaker) evidence of functional independence of belief reasoning. It is clear that belief-reasoning tasks make demands on executive function. It is also known that frontal systems are consistently found to be (at least one of) the brain regions sustaining a variety of executive processing such as shifting or working memory (see Duncan & Owen, 2000, for a review). Therefore, the relationship between performance on belief-reasoning tasks and independent tests of executive function is of particular interest for interpreting belief-reasoning errors of patients with frontal lesions. Given that D.S., F.K., G.A., and P.W. made errors on both false belief and working memory control trials, it might have been expected that they would show clear impairment on independent executive function tests. In fact, although these patients did perform poorly on these measures, no single executive measure was specifically affected. Indeed, for each measure on which one of these patients was impaired, there was another patient with the same impairment who did not make belief-reasoning errors. Perhaps, this should not be surprising. Executive function encompasses a variety of different processes. and different tasks can be differently loaded on each process (e.g., Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). It is thus unlikely that any single executive task would tap the same combination of executive processes as a particular false-belief task. Conversely, it is likely that a number distinct executive impairments could give rise to errors on any particular false-belief task. Perhaps, our working memory control trials were more suitable for identifying these patients' processing problems because

they were specifically tailored to the combination of executive demands of the false-belief trials.

Our findings demonstrate the need for considerable caution before concluding that a patient has specific belief-reasoning problems. They show that closely matched control trials are a vital addition to independent tests of executive function for interpreting the nature of false-belief-reasoning errors. Without such trials, there is clearly a danger of reaching incorrect conclusions about the relationship between belief reasoning and domain-general cognitive processes. Our findings also show that although we reduced the incidental processing demands of our video-based false-belief task enough to test a wider range of patients than is possible with existing methods, the frontal damage in D.S., F.K., P.W., and G.A. still leads to difficulty with the remaining task demand of maintaining and updating information in working memory. Importantly, however, although this means that the current study cannot be said to have fully separated belief reasoning from more general cognitive demands for these frontal patients, it is clear that this concern applies with much greater force to existing neuropsychological studies on which claims about the importance of frontal systems have been based. Of course, this does not mean that frontal systems have no specific role in belief reasoning. Imaging studies using a variety of methods commonly show activation of frontal systems (e.g., Gallagher et al., 2000; Fletcher et al., 1995), and a number of authors have suggested theoretically interesting roles for frontal systems in shifting perspective or maintaining separation between alternative perspectives (e.g., Frith & Frith, 2003; see also Gallagher & Frith, 2003; Ruby & Decety, 2003). However, we do believe that far stronger evidence is necessary before the conclusion that frontal systems have a specific role in belief reasoning can be secured.

In conclusion, detailed investigations of the nature and scope of the deficit of patients who fail false-belief tasks will help develop cognitive models of ''theory of mind'' as a reasoning domain and help us to understand the relation of this domain to other forms of reasoning and other cognitive processes. Even for patients who show pure belief-reasoning errors, substantially different patterns of impairment are possible. It could turn out that a particular patient's difficulties are quite domain general, extending to include problems that are formally similar to belief reasoning, such as counterfactual thinking and reasoning about nonmental representations (such as words and pictures). At the other extreme, a patient's difficulties could be highly specific to belief reasoning and not extend to formally similar problems within the ''theory of mind'' domain (e.g., problems involving other mental states such as knowledge, desires, and intentions). Equipped with appropriate empirical tools, a neuropsychological approach based both on group and single-case studies has great potential for addressing such questions.

METHODS

Subjects

Patients were recruited based on their lesions affecting the frontal, parietal, and/or temporal lobes. The patients' characteristics are shown in Table 2.

Materials and Procedures

Video-Based False-Belief Task

Participants watched short videos sequences in which a character gives a visual clue to the location of a hidden object by ostensively placing a pink marker on top of one of two boxes. The task principles were explained to the participant at the beginning of each testing session, and comprehension was checked with a number of warm-up trials on which corrective feedback was given, as necessary.

In false-belief trials, the participant sees a man allowing a woman to look inside both boxes, but the participant him- or herself does not see in which box the object is located. The woman leaves the room, and the man swaps the locations of the two boxes. This means that the woman has a false belief about the object's location. The woman returns and gives her clue to the participant by indicating the box where she (falsely) thinks the object is located. At this point, the video was paused and the participant was prompted to point to the box containing the object. To locate the object, participants needed to realize that the woman has a false belief and so has pointed to the wrong location. Participants judged where they thought the object was located, then received feedback by viewing the end section of video-clip where the man opens the boxes and shows them to the camera.

False-belief trials required the participant to process the order of the events in the video, in particular that the woman gives her clue after the boxes have been swapped. To control for this incidental processing demand, working memory control trials reversed the order of clue-giving and box-swapping events. The woman indicates a box before leaving the room, thus, enabling the participant to infer the location of the object. While the woman is absent, the man swaps the locations of the two boxes without opening them to reveal the object's location to the participant. The woman returns but does nothing further. The participant was prompted to point to the box containing the object. Thus, the participant had to use the fact that the boxes had swapped to update his or her knowledge of the object's location and maintain this information until a response was requested.

False-belief trials also required the participant to disengage their attention from the box just indicated by the woman and point to the other box. A participant who lacked the inhibitory control to disengage their attention from the incorrect location would fail the task, whether or not they could reason about beliefs. On

inhibition control trials, the woman leaves the room, then, in full view of the participant, the man moves the object from one box to another. The woman returns and (unwittingly) indicates the box that the participant now knows to be empty. The participant was then invited to point to the box containing the object. As in the false-belief trials, a correct answer required the participant to disengage their attention from the box just indicated by the woman. However, unlike false-belief trials, no belief reasoning was required.

True-belief filler trials were designed to guard against participants passing the false-belief trials by adopting the strategy of always pointing to the opposite box from that indicated by the woman. The woman leaves the room, but the man does not swap the boxes, meaning that the woman's belief about the object's location remains true, and the woman returns and indicates (accurately) the box where the object is located. To answer correctly, the participant had to point to the same box indicated by the woman. Although it was possible that participants were inferring the woman's belief on these trials, we did not regard this as a reliable index of belief-reasoning ability because it is also possible to make a correct response simply by pointing to wherever the woman indicates, without inferring her belief. The key point in the current study is that correct answers to the true-belief trials required the participant to point to the location indicated by the woman, while correct answers to false-belief trials required the participant to point to the opposite location. Thus, if participants performed well on true-belief trials, we could be confident that good performance on false-belief trials reflected genuine belief reasoning, not a superficial strategy of pointing to the opposite box from that indicated by the woman.

On clue confirmation filler trials, the woman indicates a box before leaving the room. The man opens this box to reveal the object, providing a very salient reminder that the woman is acting in good faith. The man moves the object to the second box. The woman returns to the room. The participant was then prompted to respond. Interpolation of both types of filler trial with experimental trials meant that experimental trials did not appear in any regular pattern and were not repeated in long sequences.

There were a total of 12 video trials of each type. The videos were presented on a standard desktop computer using PowerPoint software. Video presentation was controlled manually by the experimenter, enabling the time allowed for responding and the rate of progress to the next video to be adapted to the needs of the participant. The participant responded nonverbally by pointing to one of the two boxes on the screen, and this response was recorded by the experimenter. Each testing session lasted approximately 20 min, and sessions were typically held at 1- to 2-week intervals. Each testing session included three trials of each type, presented in a pseudorandom order designed to avoid runs of more than two trials of the same type. For each trial type overall and across trial types within each session, half of the correct responses were the box on the right and half were the box on the left.

Story-Based False-Belief Tasks

We created 12 narrative-based false-belief tasks. Story-based tasks are more typical of the stimuli that have been used in previous studies of neurological patients (e.g., Rowe, Bullock, Polkey, & Morris, 2001; Stone et al., 1998); however, because they were only required reasoning about a single first-order mental state, they were substantially simpler than many of the studies reviewed in the introduction. The tasks were based around simple six-line stories followed by four questions. For example,

> Jeremy is eating out at a restaurant. Inside the restaurant, Jeremy hangs his coat on the stand by the door and leaves his bag underneath. The waitress shows Jeremy to his table and tells him about today's special dishes. When she comes back, the waitress notices Jeremy's bag beneath the coat stand by the door. She decides that it is unsafe for the bag to stay by there, as it would be easy for someone to steal. Leaving the coat on the coat stand, she locks the bag in the store-cupboard.

False-belief question: Where does Jeremy think the bag is? On the coat stand or in the store-cupboard?
Counterfactual question: What if the waitress had not noticed the bag? Where would the bag be? On the coat stand or in the store-cupboard?
Memory control question: Where was the bag at the beginning? On the coat stand or in the store-cupboard?
Reality control question: Where is the bag now? On the coat stand or in the store-cupboard?

As is common in the developmental and neuropsychological literatures (e.g., Rowe et al., 2001; Wellman, Cross, & Watson, 2001; Stone et al., 1998), we asked memory and reality control questions to check that participants could recall two facts that were crucial to accurate attribution of a false belief. As a more accurate comparison for the formal reasoning demands posed by making a belief inference, we also included a counterfactual question (Riggs et al., 1998). The false belief and counterfactual questions were always asked before the control questions. False-belief questions occurred equally often before and after counterfactual questions and reality control questions occurred equally often before and after memory control questions. The order of the two-alternative forced choice (e.g., ''On the coat stand or in the store-cupboard?'') was varied so that, within each block, the correct answer was equally often the first and second items mentioned.

Participants completed the 12 false-belief tasks over three sessions of around 20-min duration and separated by 1–2 weeks. Before the false-belief stories, participants

completed two warm-up trials using stories of a similar length, but with simpler factual questions.

Independent Tests of Executive Functions and Language

Executive Function Tasks

Working memory tasks. In the digit manipulation task, participants were presented with 12 sequences of either 3 or 4 orally presented digits (1–9). The length of the sequence was determined by the participant's basic digit span (if a participant had a digit span of 4 or less, the sequence was 3 digits long; if a participant had a digit span above 4, the sequence was 4 digits long). Participants were then asked to report the digits in ascending order. In the resistance to interference task, participants were presented with 12 similar sequences of digits and asked to recall the digits in the same order as they were presented. However, before recalling the sequence, they were presented with an interference task, requiring them, on five consecutive occasions, to name the day that follows a particular day of the week (e.g., What comes after Wednesday?). In the updating task, participants were presented with 12 trials consisting of a sequence of digits of unpredictable length. They were asked to remember the last digits of the sequence in the same order as they were presented (i.e., the 3 or 4 last digits depending on the span).

Inhibition tasks. In both tasks, a trial consisted of 10 items centrally presented on an A4 sheet (with 8 trials per condition). In the stimulus selection task, the items consisted of one or two hands raising either one or two fingers. Participants were asked to cross out the hands with two fingers raised irrespective of the number of hands presented. For all items in the compatible or baseline condition, the number of fingers raised was the same as the number of hands presented (i.e., ✌✌ or ☝). For all items in the incompatible or executive condition, the number of fingers raised was different to the number of hands presented (i.e., ✌ or ☝☝). In the mixed condition, both types of items were presented. This latter condition was not taken into account in the analyses but was aimed to discourage the participants from using strategies such as basing the response on the number of hands instead of the number of fingers. In the response selection task, the items consisted of a hand raising either one or two fingers. Participants were asked to say aloud "one" ± or "two," depending on the item presented. In the congruent or baseline condition, the participants had to say the number of fingers raised on each hand (if ☝, say "one"; if ✌, say "two"). In the incongruent or executive condition, the participants had to say the opposite number of fingers (if ☝, say "two"; if ✌, say "one").

Shifting tasks. Both tasks consisted of three lists of 30 items presented in a central column on an A4 sheet. In the alternation of focus of attention task, each item consisted of a pair of one number (1–9) and one letter (A–Z). With the first list, participants were asked to cross out the numbers. With the second list, they were asked to cross out the letters. With the third list, they were asked to cross out number and letter stimuli in an alternating way (e.g., first line, crossing out the number, next line, crossing out the letter, next line, crossing out the number, etc.). In the alternation of arithmetical operation task, the items consisted of a column of numbers (2–9). With the first list, participants were asked to add 1 to each number presented. With the second list, they were asked to take away 1 from each number presented. With the third list, they were asked to add 1 and take away 1 in an alternating way.

Acknowledgments

This research was supported by grants from the Leverhulme Trust, the MRC, and the Stroke Association. We are very grateful to all the participants for their kind participation.

Reprint requests should be sent to Ian Apperly, School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK, or via e-mail: i.a.apperly@bham.ac.uk.

REFERENCES

Allison, T., Puce, A., & McCarthy, G. (2000). Social perception from visual cues: Role of the STS region. *Trends in Cognitive Sciences, 4,* 267–278.

Astington, J. W., Harris, P. L., & Olson, D. R. (1998). *Developing theories of mind* (1st ed.). Cambridge: Cambridge University Press.

Baron-Cohen, S., & Ring, H. (1994). A model of the mindreading system: Neuropsychological and neurobiological perspectives. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and development* (pp. 183–202). Hove: Erlbaum.

Baron-Cohen, S., Tager-Flusberg, H., & Cohen D. J. (2001). *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd ed.). New York: Oxford University Press.

Bird, C. M., Castelli, F., Malik, O., Frith, U., & Hussain, M. (2004). The impact of extensive medial frontal lobe damage on "theory of mind" and cognition. *Brain, 127,* 914–928.

Burgess, P. W., & Shallice, T. (1997). *The Hayling and Brixton Tests.* Bury St Edmunds: Thames Valley Test Company.

Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development, 70,* 381–395.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., Iverson, S. D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science, 276,* 593–596.

Channon, S., & Crawford, S. (2000). The effects of anterior lesions on performance on a story comprehension test: Left anterior impairment on a theory of mind-type task. *Neuropsychologia, 38,* 1006–1017.

Damasio, A. R., Tranel, D., & Damasio, H. (1990). Individuals with sociopathic behaviour caused by frontal damage fail to respond autonomically to social stimuli. *Behavioural Brain Research, 41,* 81–94.

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences, 23,* 475–483.
Fine, C., Lumsden, J., & Blair, R. J. R. (2001). Dissociation between ''theory of mind'' and executive functions in a patient with early left amygdala. *Brain, 124,* 287–298.
Fletcher, P. C., Happé, 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.
Fodor, J. (1992). A theory of the child's theory of mind. *Cognition, 44,* 283–296.
Frith, C. D., & Frith, U. (1999). Interacting minds—A biological basis. *Science, 286,* 1692–1695.
Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London, B, 358,* 459–473.
Gado, M., Hanaway, J., & Frank, R. (1979). Functional anatomy of the cerebral cortex by computed tomography. *Journal of Computer Assisted Tomography, 3,* 1–19.
Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of ''theory of mind.'' *Trends in Cognitive Sciences, 7,* 77–83.
Gallagher, H. L., Happé, 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.
German, T. P., & Nichols, S. (2003). Children's counterfactual inferences about long and short causal chains. *Developmental Science, 6,* 514–523.
Happé, F., Brownell, H., & Winner, E. (1999). Acquired ''theory of mind'' impairments following stroke. *Cognition, 70,* 211–240.
Hooker, C. I., Paller, K. A., Gitelman, D. R., Parrish, T. B., Mesulam, M. M., & Reber, P. J. (2003). Brain networks for analysing eye gaze. *Cognitive Brain Research, 17,* 406–418.
Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic assessment of language processing in aphasia.* Hove: Psychology Press.
Leslie, A., & Thaiss, L. (2003). Domain specificity in conceptual development: Neuropsychological evidence from autism. *Cognition, 43,* 225–251.
Lewis, C., & Mitchell, P. (1994). *Children's early understanding of mind: Origins and development* (1st ed.). Hove: Erlbaum.
Malle, B. F., Moses, L. J., & Baldwin, D. A. (2001). *Intentions and intentionality: Foundations of social cognition*. Cambridge: MIT Press.
Mitchell, P., & Riggs, K. J. (Eds.) (2000). *Children's reasoning and the mind.* Hove: Psychology Press.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex ''frontal lobe'' tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100.
Perner, J., & Wimmer, H. (1985). ''John thinks that Mary thinks that. . .'': Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39,* 437–471.
Peterson, D. M., & Riggs, K. J. (1999). Adaptive modelling and mindreading. *Mind and Language, 14,* 80–112.
Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development, 13,* 73–90.
Rizzolatti, G., Fadiga, L., Matelli, M., Bettinardi, V., Paulesu, E., Perani, D., & Fazio, F. (1996). Localization of grasp representations in humans by PET. 1. Observation versus execution. *Experimental Brain Research, 111,* 246–252.
Rowe, A. D., Bullock, P. R., Polkey, C. E., & Morris, R. G. (2001). ''Theory of mind'' impairments and their relationship to executive functioning following frontal lobe excisions. *Brain, 124,* 600–616.
Ruby, P., & Decety, J. (2003). What you believe versus what you think they believe: A neuroimaging study of conceptual perspective taking. *European Journal of Neuroscience, 17,* 2475–2480.
Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). The left tempero-parietal junction is necessary for representing someone else's belief. *Nature Neuroscience, 7,* 449–500.
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.
Segal, G. (1996). The modularity of theory of mind. In P. Carruthers, J. Boucher, & P. K. Smith (Eds.), *Theories of theories of mind* (pp. 141–157). Cambridge: Cambridge University Press.
Siegal, M., Carrington, J., & Radel, M. (1996). Theory of mind and pragmatic understanding following right hemisphere damage. *Brain and Language, 53,* 40–50.
Sperber, D. (2000a). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (1st ed., pp. 117–138). Oxford: Oxford University Press.
Sperber, D. (Ed.) (2000b). *Metarepresentations: A multidisciplinary perspective.* New York: Oxford University Press.
Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10,* 640–656.
Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes.* New York: Raven Press.
Stuss, D. T., Floden, D., Alexander, M. P., Levine, B., & Katz, D. (2001). Stroop performance in focal lesion patients: Dissociation of processes and frontal lobe lesion location. *Neuropsychologia, 39,* 771–786.
Surian, L., & Siegal, M. (2001). Sources of performance on theory of mind tasks in right hemisphere-damaged patients. *Brain and Language, 78,* 224–232.
Varley, R., Siegal, M., & Want, S. C. (2001). Severe impairment in grammar does not preclude theory of mind. *Neurocase, 7,* 489–493.
Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72,* 655–684.
Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103–128.
Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language, 62,* 89–106.