

Error analyses reveal contrasting deficits in “theory of mind”:

Neuropsychological evidence from a 3-option false belief task

Dana Samson, Ian A. Apperly, & Glyn W. Humphreys

University of Birmingham, U.K.

PREPRINT of the following paper:

Samson, D., Apperly, I. A., & Humphreys, G. W. (2007). Error analyses reveal contrasting deficits in "theory of mind": Neuropsychological evidence from a 3-option false belief task. *Neuropsychologia*, 45(11), 2561,2569. <https://doi.org/10.1016/j.neuropsychologia.2007.03.013>

Short title: Error analyses of ToM deficits

Published in *Neuropsychologia*, 45(11), 2561-2569

Address for correspondence:

Dana SAMSON

Behavioural Brain Sciences Centre

School of Psychology – Hills Building

University of Birmingham

Edgbaston

Birmingham, B15 2TT, U.K.

Tel: ++44 (0) 121 414 3661

Fax: ++44 (0) 121 414 4897

E-mail: d.samson@bham.ac.uk

Abstract

Perspective taking is a crucial ability that guides our social interactions. In this study, we show how the specific patterns of errors of brain-damaged patients in perspective taking tasks can help us further understand the factors contributing to perspective taking abilities. Previous work (e.g., Samson, Apperly, Chiavarino, & Humphreys, 2004; Samson, Apperly, Kathirgamanathan, & Humphreys, 2005) distinguished two components of perspective taking: the ability to inhibit our own perspective and the ability to infer someone else's perspective. We assessed these components using a new nonverbal false belief task which provided different response options to detect three types of response strategies that participants might be using: a complete and spared belief reasoning strategy, a reality-based response selection strategy in which participants respond from their own perspective, and a simplified mentalising strategy in which participants avoid responding from their own perspective but rely on inaccurate cues to infer the other person's belief. One patient, with a self-perspective inhibition deficit, almost always used the reality-based response strategy; in contrast, the other patient, with a deficit in taking other perspectives, tended to use the simplified mentalising strategy without necessarily transposing her own perspective. We discuss the extent to which the pattern of performance of both patients could relate to their executive function deficit and how it can inform us on the cognitive and neural components involved in belief reasoning.

Key words: social cognition; theory of mind; perspective taking; false belief; frontal lobe; temporo-parietal junction

1. Introduction

Cognitive neuropsychology has made major contributions towards understanding the architecture of the mind and brain in many areas of cognition including memory, attention, language, vision, action, calculation and music. The cognitive neuropsychological approach attempts to isolate particular cognitive processes by studying patterns of spared and impaired performance in patients with different brain lesions. In addition, analyses of the error patterns in patients can help to specify which cognitive processes are affected by the brain lesion (e.g., locating a deficit at an early stage of word processing when visual errors occur in reading; Ellis & Young, 1997). Until recently, few attempts had been made to study processes involved in social cognition using this approach. However, important insights can be gained by examining how performance on tasks requiring social cognitive processes breaks down after brain lesion, and there is a growing body of evidence using deficits on social reasoning tasks to isolate component operations. To date, though, error analyses have not been used to demonstrate qualitatively different forms of impairment. The study here presents a first attempt to do this, using error patterns to distinguish patients with contrasting deficits in “theory of mind” (ToM) reasoning.

The term ToM is commonly used to describe our ability to use concepts such as beliefs, intentions, desires or emotions in order to predict or explain human behaviour (Premack & Woodruff, 1978). Early neuropsychological studies in this area were concerned with the identification of the brain regions that lead to a ToM impairment. These studies first highlighted the right hemisphere (Happe, Brownell, & Winner, 1999; Siegal, Carrington, & Radel, 1996; Surian & Siegal, 2001) and the frontal lobes (Lough, Gregory, & Hodges, 2001; Lough et al., 2006; Rowe, Bullock, Polkey, & Morris, 2001; Stone, Baron-Cohen, & Knight, 1998) then, more recently the amygdala (Stone, Baron-Cohen, Calder, Keane, & Young, 2003) and the left temporo-parietal junction (Samson et al., 2004). Overall these findings are consistent with neuroimaging studies, which have consistently shown an extensive fronto-

parieto-temporal network, with peak activations especially in the medial prefrontal cortex, the temporo-parietal junction and the temporal poles when healthy adults are engaged in ToM tasks (for a review, see Frith & Frith, 2003; Saxe, Carey, & Kanwisher, 2004). The neuropsychological data indicate that these different brain regions play a necessary role in ToM (but see Bird, Castelli, Malik, Frith, & Husain, 2004).

Neuropsychological studies have also begun to investigate the nature of the processes involved in ToM. Some studies have addressed the issue of domain-specificity: to what extent are the processes involved in ToM specific to the social domain or instead, part of more general language or executive processes also employed outside the social domain (Gregory et al., 2002; Lough et al., 2001; Lough et al., 2006; Rowe et al., 2001; Varley & Siegal, 2000; see also Apperly, Samson, & Humphreys, 2005, for a critical discussion of current findings)? Other work attempts to decompose ToM abilities themselves. For example, there is some suggestion that the integration of emotion with other mental states might be a distinct aspect of ToM that can be specifically affected following lesions to the orbito-frontal cortex (e.g., Lough et al., 2006; Stone et al., 1998). In our own work, we attempt to distinguish between two further component processes concerned with perspective taking, respectively the ability to infer someone else's perspective and the ability to inhibit one's own perspective (Apperly, Samson, Chiavarino, & Humphreys, 2004; Samson et al., 2004; Samson et al., 2005). To further investigate this distinction, we present here a new test of ToM that leads to different classes of error, according to the deficit present in a particular patient.

The detailed study of error patterns is not possible in many commonly-used ToM tasks. Take the classic "false" belief reasoning task. Typically, in a false belief scenario the main character is initially aware of a certain state of the world (e.g., the character sees that a chocolate bar is put in a cupboard). This state then changes, but the character is unaware of the change (e.g., the character is outside the

room when his mum puts the chocolate bar in the fridge). This creates a discrepancy between the participant's own-perspective (the participant knows that the chocolate bar is now in the fridge) and the perspective of the character in the scenario (the character thinks the chocolate is still in the cupboard). The participant is then asked a question such as, where does the character think the chocolate is – In the fridge or in the cupboard? The binary choice only leaves two possible interpretations: either the participant can reason about beliefs and thus chooses the correct response (the character thinks the chocolate is in the cupboard), or, alternatively, the participant cannot take someone else's perspective and thus chooses the reality-based response (the character thinks the chocolate is in the fridge). In such tasks it is often assumed that difficulties in perspective taking automatically lead to choices of the reality-based response. However, this may not be true. Previously, we have highlighted two types of perspective taking deficits: some patients have difficulties in inhibiting their own perspective (Samson et al., 2005) whereas other patients have specific difficulties in inferring someone else's perspective (Samson et al., 2004). These separate deficits were associated with lesions to different brain areas (respectively the right inferior frontal gyrus and the left temporo-parietal junction), suggesting that self-perspective inhibition and other-perspective taking are distinct functional and neural processes. On this account, we expect that one of the two processes can be selectively affected by brain damage, leaving the other relatively (or even totally) spared. It follows that patients with a problem in self-perspective inhibition, but not other-perspective taking per se, should be able to take someone else's perspective when there is a low requirement for self-perspective inhibition (that is indeed what we observed, Samson et al., 2005). Conversely, patients with a deficit specific to taking another person's perspective might still be able to inhibit their own perspective. Importantly, this means that such patients might appreciate that it is incorrect to respond from their own perspective, but nonetheless fail to infer the correct content of the other person's perspective. For these patients, errors would not be necessarily egocentric or reality-based but could reflect the use of residual or simplified mentalising

strategies. Different patients could use different strategies (or residual knowledge), but we focused here on one strategy in particular to devise our new false belief task.

Previously, we have observed that patients with other-perspective taking deficits performed well in visual perspective taking tasks provided the task could be carried out by drawing a line of sight from the other person's eyes (an ability which is also acquired quite early in children before they accurately reason about beliefs, Flavell, 1992). For example, the patients could easily state what a given person could or could not see, depending on where the person was looking and depending on whether there was something obstructing the view (e.g., a newspaper that the person was holding in front of her eyes). What stimulus a person can see is a critical piece of information for inferring what that person thinks about the physical world. If we see a visitor looking at a chocolate box, it would be correct to infer that the person thinks there are chocolates inside (even if we know ourselves that we have put teabags inside the box earlier on). Our observations suggested that some of our patients with other-perspective taking deficits were able to make such inferences but were unable to make more complex belief inferences, especially when these inferences required integration of a sequence of events over time (e.g. inferring that the visitor would think there are teabags in the chocolate box because he or she saw *earlier* on that the chocolates were replaced by teabags). This suggested to us that, when inferring someone else's belief, patients could fall back on simpler mentalising processing (e.g. visual perspective taking), using what the person *currently* sees as the content of what the person thinks. To directly test this hypothesis, we created a new video and animation-based false belief task (with low language demands). The task was based on a classic change of content false belief paradigm (where a character is unaware that the content of a box has changed). However, rather than using "neutral" boxes, we employed boxes that have a predictable content (e.g., a pizza box, in which people would expect a pizza). In this way, we could detect if patients use their residual visual perspective taking ability and use as cue what the person is looking at (e.g. seeing a pizza box) to infer what she is

thinking is inside the box (e.g. a pizza) instead of taking into account the full sequence of events (i.e. what the character has or has not seen been previously placed inside the pizza box). The videos showed a person putting a first object in the box (e.g., a passport in the pizza box) and then later changing the content of the box (e.g., replacing the passport by scissors). A second character could see what was first put in the box and then either saw (true belief condition) or did not see the change of content (false belief condition). At the end of the video, the participant was asked what the second character thinks is inside the box. Unlike the classic false belief task, there were three response options: for example, the character will think there is a pizza in the pizza box, the character will think there is a passport in the pizza box or the character will think there are scissors in the pizza box. Correct belief reasoning would consist in taking into account what the character has or has not seen previously to attribute the correct content to the character's belief (e.g., the character will think there is a passport in the box, in the false belief condition). We predicted that participants with a self-perspective inhibition deficit would give a reality-based response, i.e., choosing as belief content what they know is now in the box (e.g., scissors). In contrast, participants with other-perspective taking deficit might give an appearance-based response (e.g., the character sees a pizza box so he/she will think there is a pizza inside) using a simpler mentalising strategy rather than taking into account the full sequence of events.

We present the results of two brain-damaged patients on this new false belief task. The two patients were chosen because they demonstrated in our previous studies a clear pattern of self-perspective inhibition deficit on the one hand (Samson et al., 2005) and other-perspective taking deficit on the other hand (Samson et al., 2004).

2. Methods

2.1. Participants

Patient WBA. WBA is a right-handed man with a degree in law who was 59 years old at the time of testing (2005). As we reported previously (Samson et al., 2005), he suffered a right hemisphere stroke in 2001 which affected the right middle and inferior frontal gyri as well the right superior temporal gyrus. As a consequence of his stroke, WBA suffered a left side weakness, language as well as attention and executive deficits (see Table 1 for WBA's results on general neuropsychological tests). His language impairment was characterised by grammatical processing difficulties (mainly difficulties in understanding and using function words and deictic words). His executive impairment was mainly characterised by difficulties in inhibition and cognitive flexibility. At the time of testing, WBA lived independently and had returned to part-time work.

In false belief tasks, WBA made errors only when the demands in self-perspective inhibition were high (see Table 1). As reported previously, this pattern was observed not only in the case of belief reasoning, but also when WBA was asked to infer someone else's desire, emotion or visual experience (for the full results, see Samson et al., 2005). We thus concluded that WBA had a selective deficit in self-perspective inhibition. For the purposes of the current study, WBA was presented with two additional visual perspective tasks. In the level 1 visual perspective taking task, WBA was shown photographs of a girl sitting in a room with circles pinned on the walls. The girl could face the left or the right side of the room (and thus could not see any circle pinned on the opposite side of the room) and her view could be clear or obstructed (because she held her hands or a newspaper in front her eyes). WBA was asked how many circles the girl could see. He had no difficulty performing this task. In the level 2 visual perspective task, WBA was presented with displays consisting of a circle and a rectangle printed on a sheet of paper. From the patient's perspective, on any given display the circle

could be aligned with the rectangle in a horizontal (either to the left or to the right) or vertical fashion (either above or below). The examiner sat 45 degrees to the left of the patient. The patient was asked to infer how the examiner saw the display (pointing to a multiple choice response sheet that depicted all four possible configurations of the circle and rectangle). Because of the 45 degrees difference between the examiner's and the patient's position, what appears as horizontally aligned for the patient is actually vertically aligned from the examiner's perspective and vice versa. WBA made errors on almost every single trial and his errors consisted in choosing the display as seen from his own perspective. Similarly to the pattern observed in children (Flavell, Everett, Croft, & Flavell, 1981), the level 2 visual perspective task was much harder for WBA than the level 1 task.

Patient PF. PF is a right-handed female care worker with basic secondary education (leaving school at the age of 16 following completion of the O-levels in the British educational system) and who was 56 years old at the time of testing (2005). As reported previously (Apperly et al., 2004; Samson et al., 2004), PF suffered a left hemisphere stroke in 1996 which affected the left temporo-parietal junction (angular and supramarginal gyri as well as the left superior temporal gyrus). As shown in Table 1, PF's main cognitive impairments consisted in a visuo-constructive apraxia as well as an impairment in attention and executive function (inhibition and flexibility). At the time of testing, PF also lived independently and had returned to a part-time job.

As previously reported (Samson et al., 2004), PF had difficulties inferring someone else's false belief even though the demands on self-perspective inhibition were low (see Table 1). We thus concluded that PF had a deficit affecting the processes involved in other-perspective taking rather than self-perspective inhibition. To our surprise, when presenting the false belief task with high self-perspective inhibition demands, PF was in fact perfectly able to infer someone else's false belief. One critical difference between the two tasks (other than the demands in self-perspective inhibition) is the wording

of the test question. In the low self-perspective inhibition condition, i.e. the one in which PF performed poorly, the question consisted in asking where a given object was located. On false belief trials, to correctly locate the object, the participant had to infer that a character has a false belief but nothing in the wording of the question directed the attention to the character or his belief. In contrast, in the high self-perspective inhibition condition (i.e., the task that PF performed well), the question consisted in asking which box the character would open first to find a given object. In this case, the question explicitly mentioned that the character's perspective needed to be taken. We will refer back to that profile of performance in the discussion section. PF was also presented with the level 1 and level 2 visual perspective taking tasks described above. PF was very accurate in the level 1 visual perspective task but the results from the level 2 visual perspective task were more difficult to interpret because PF suffered from visuo-spatial processing difficulties (visuo-constructive apraxia). She made a high proportion of errors and many errors consisted in left/right or above/below inversions. In fact, even when reporting her own visual perspective, PF made errors in selecting the correct response.

Control participants. Eight control participants were selected from our pool of control participants based on their age and educational level. Five of the control subjects (four males and one female), had the same educational profile as patient PF. Their age ranged between 52 and 64 (mean age: 58) and they all left school at the age of 16 (following completion of the O-levels in the British educational system). The three remaining control subjects (two females and one male) were matched to WBA's educational profile. Their age ranged between 54 and 64 (mean age: 59.7) with one participant leaving school at 18 (following completion of the A-levels in the British educational system) and the other participants completing a university degree.

Insert Table 1 about here

2.2. The Nosy Neighbour Task

The task consisted of nonverbal video clips inserted within powerpoint animations. Participants were told that they would see on the computer screen the window of a house. Through the window, they would see what the occupier is doing with different boxes. Participants were also told that they would see neighbours passing by the window. Each time a neighbour would stop at the window, the neighbour would see what happened inside the house. In contrast, when the neighbour walked away from the window, the neighbour could not see what happened inside the house. Each trial was divided into three scenes (see Table 2 for an example): a first scene where the occupier placed an object (e.g., a pizza) in a particular container (e.g., a pizza box), a second scene where the occupier changed the content of the container (e.g., replacing the pizza by a passport), and a third scene where only the closed container could be seen. On each trial, a neighbour passed by, stopped and looked through a window to witness the first scene. On the true belief trials, the neighbour remained in front of the window until the end of scene 2, so that the neighbour knew that the content of the box had changed. Then on half of true belief trials, the neighbour moved away and on the other half of trials the neighbour remained in front of the window in scene 3. On the false belief trials, the neighbour walked away at the end of scene 1. The neighbour remained absent during scene 2 so he or she could not see the change of content. On half of the false belief trials, the neighbour stayed away from the window in scene 3 and for the other half of these trials the neighbour returned and stopped in front of the window in scene 3. This variability in the presence and absence of the neighbour in scene 3, both for the true and false belief trials, was introduced to reduce systematic regularities in the task. Neighbours were cartoon-like figures (adapted from <http://www.barrysclipart.com>) that appeared or disappeared from the screen using powerpoint animations. Only one neighbour appeared per trial, each time a different one.

For the true and false belief trials, two types of videos were presented. For half the videos, the initial object put in the container was the most likely content for the container (e.g., a pizza in a pizza box) and this object was then replaced by an object that would not usually be placed in the container (e.g., a passport in a pizza box). This is later referred to as the predictable object condition (see Table 2). For the other half of the videos, both the initial and the second objects were unusual contents for the container (e.g., a passport and then a pair of scissors in a pizza box). This is later referred to as the non-predictable object condition. At the end of each animation the examiner asked “What does the person now think is inside the container?”. Participants were given three choices: the most likely content of the box (pizza) and two other unusual objects (passport and scissors). Response options were depicted by showing the neighbour with three thought bubbles, each bubble showing a photograph of a possible content of the box.

We used 8 different types of containers: an egg box, a bin (trash can) and a pizza box as well as 5 additional boxes or packs labelled with a well known brand in the UK - a Cornflakes box, a chocolate box, a beer box, a crisp pack (potato chip pack) and a tea box. Each container was used four times, twice in a predictable object condition (once in a true belief and once in a false belief scenario) and twice in a non-predictable object condition (once in a true belief and once in a false belief scenario). For each container, a unique set of objects was used (the set consisted of the predictable object and two unrelated objects, one of which was presented twice – see Table 2 for an example). There were thus 32 trials in total. The 32 trials were presented in two blocks of 16 trials each. The number of true and false belief trials as well as the number of predictable and non-predictable object trials was balanced across the two blocks. A given container was not presented more than twice in a block. The order of the trials was pseudo-randomised to avoid more than 3 consecutive trials of the same type (true/false belief or predictable/non-predictable content). Both patients saw the trials in the same order and had a few

minutes break in between the two blocks in which an informal chat took place (the task itself was not discussed during the break).

For all true belief trials, the correct belief content could be inferred either by genuinely reasoning about beliefs or by transposing what the participant knew to be the latest content of the box (a ‘reality-based’ response). However, choosing the most likely content of the box, i.e., giving an appearance-based response, would always lead to an incorrect response. For half of the false belief trials (i.e., the ones in the predictable object condition), the correct belief content could be inferred either by genuinely reasoning about beliefs or by giving an appearance-based response. However, giving a reality-based response would lead to an (egocentric) error. For the other half of the false belief trials (those in the non-predictable object condition), only genuine belief reasoning could generate the correct answer (see Table 2). This led us to make the following predictions:

- (1) Participants with spared belief reasoning abilities would rely on genuine belief reasoning and would give correct responses for all types of trials.
- (2) Participants with a self-perspective inhibition deficit, like in the case of patient WBA, would rely on a reality-based strategy and would make errors on all types of false belief trials but would be accurate on the true belief trials. Moreover, errors should consist in choosing the reality response rather than the appearance response.
- (3) Participants with a spared self-perspective inhibition ability but who nevertheless have difficulty inferring someone else’s perspective – as in the case of patient PF – could rely on an appearance-based strategy. In this case, they would make errors on the true belief trials as well as the false belief trials in the non-predictable object condition. Errors should consist in choosing the appearance response rather than the reality response.

Before the task was presented a pre-test ensured that the participants knew what would be the most likely content of each container used in the task.

Insert Table 2 about here

3. Results

Given the three-choice option, the probability of generating a correct response by chance on one trial is 0.33. For all conditions, the level at or above which we can be confident that the participant did not achieve correct responses simply by guessing is 6/8 (since at least 6, i.e. 75%, correct response are required for the score to be statistically significantly above chance, the probability to get 6/8 correct = 0.0187 by one-tailed binomial test). The task was presented to all the control participants. Six of the controls made no errors and the two other controls (both from the lower educational background group matched to PF) made one error each (one participant wrongly attribute a false belief on a true belief trial and the other participant made a reality-based response on a false belief trial). None of the control participants chose any appearance-based response. The response distributions for PF and WBA are shown in Figure 1.

WBA made no errors in the true belief condition but he failed to perform significantly above chance level in the two false belief conditions: he performed almost at floor giving 2/8 and 0/8 correct responses on false trials in the first and second block respectively. All his errors consisted in choosing the reality-based response. WBA was thus not simply guessing but instead responded according to his own perspective both in the true belief condition (leading to the correct response) and the false belief condition (leading to an incorrect response). This profile is exactly the one predicted in case of a self-perspective inhibition deficit.

As can be seen in Figure 1, PF's score was not significantly above chance level in two conditions: the true belief – predictable object condition (TB1) and the false belief – non-predictable object condition (FB2). Quite noticeably, all her errors were made in the first block. PF showed a quite different pattern of performance to WBA in several ways. Firstly, unlike WBA, PF made errors in the true belief conditions. Secondly, PF made errors on false belief trials but only in the non-predictable object condition. Thirdly, unlike WBA, all but one error committed by PF consisted in giving an appearance-based response. In fact, overall in the first block, PF gave 4/8 appearance-based responses in the true belief condition (all leading to an incorrect response, three errors were made in the predictable object condition – TB1 – and one error was made in the non-predictable object condition – TB2) and 7/8 appearance-based responses in the false condition (4 of which led to the correct response in the predictable object condition – FB1 – and 3 leading to the incorrect response in the non predictable object condition – FB2). Although appearance-based responses were more likely to be given when the predictable object was used in the scenarios (7 responses), this type of response also happened when the predictable object was not used (4 responses). Furthermore, whether or not the character was present in the last scene did not influence the likelihood of choosing the appearance-based option (in total 6 appearance-based responses were given when the character was no longer in front of the box at the time the belief question was asked and 5 appearance-based responses were given when the character was still in front of the box at the time the belief question was asked). Overall, PF's profile of a high proportion of appearance-based responses, leading to errors both on true and false belief trials, is consistent with the one we predicted for someone with a deficit in other-perspective taking who used a simplified mentalising strategy (possibly relying on residual visual perspective abilities). Quite strikingly, however, all errors made by PF were made in the first block. In the second block, PF seemed to rely on genuine belief reasoning, making no errors in the true or false belief conditions. Note that the second block was presented in the same session as the first block and no changes were

made to the instructions. Also worth noting is the fact that no feedback was provided during the task. So it is unlikely that PF developed a superficial strategy to solve the task other than integrating what the character had or had not seen previously to infer the character's belief content.

Insert Figure 1 about here

4. Discussion

The results of our new 3-option false belief task (the Nosy Neighbour Task) support the view that self-perspective inhibition and other-perspective taking abilities rely on distinct functional and neural mechanisms. We have now three lines of neuropsychological evidence in favour of this hypothesis. Firstly, as we documented previously, patients with a deficit in one or the other mechanisms are differentially sensitive to the demands in self-perspective inhibition of a task. Patients with a deficit in self-perspective inhibition have difficulties in the tasks that place high demands on self-perspective inhibition but they can correctly take someone else's perspective when the self-perspective inhibition demands are kept low (Samson et al., 2005). In contrast, patients with a deficit in other-perspective taking have difficulties even when the demands on self-perspective inhibition are kept low (Samson et al., 2004). Secondly, patients showing these different patterns have lesions in different brain areas. For the sample of patients we reported, a deficit in self-perspective inhibition is associated with lesions to right inferior frontal lobe (Samson et al., 2005) whereas a deficit in other-perspective taking was associated with lesions in the area of the left temporo-parietal junction (Samson et al., 2004). Third, in this paper, we show that the two types of patients make qualitatively distinct types of errors in a false belief task. When asked to attribute content to someone else's perspective, patient WBA with self-perspective inhibition problems transposed his own perspective; in contrast, patient PF, with other-perspective taking problems, successfully rejected the egocentric response even when she was unable

to provide the correct response. The pattern observed in WBA is entirely consistent with the pattern of “realist” errors most commonly observed in false belief tasks. The pattern observed in PF is more unusual, and is discussed in more detail below.

4.1. The origin of PF’s appearance-based responses

The appearance-based response option in the Nosy Neighbour Task was designed to detect the reliance on a simplified mentalising strategy which would use what the person sees at the time the belief question is asked to infer what the person is thinking (rather than relying on the more accurate and sophisticated processing that would also take into account what the person has or has not seen in the past). However, there could be other reasons why PF chose the appearance-based response so often.

Firstly, it could be that PF’s appearance-based responses resulted from difficulties in keeping track of the sequence of events. This does not seem very plausible. In other false belief tasks, PF could keep track of the movements of critical objects (Samson et al., 2004). Moreover, using this account it is difficult to explain why PF always chose the appearance-based option rather than the object last seen (or a mixture of reality-based and appearance-based incorrect responses that would reflect real confusion as to which object was presented when). It also seems unlikely that PF simply forgot whether the character had or had not seen the change of content. On true belief trials, PF gave appearance-based and incorrect responses even when the character remained on the screen, in full view of the patient, during the complete sequence of events.

Secondly, it could be also argued that PF’s appearance-based responses resulted from a misunderstanding of the task. In principle, the pre-test in which PF was asked what was the most likely content of the boxes could have mislead her to believe that we asked the same question in the test-

phase. However, it is important to highlight that, in the actual test phase, for each item, the question was made explicit (i.e., “What does the character now think is in the box”, and PF does not have a language impairment that would impede her understanding of this instruction). Moreover, the response choices clearly showed the character with thought bubbles in order to avoid any confusion as to the goal of the task (this format of response was completely different from the format of responses in the pre-test). Perhaps, the strongest evidence against the hypothesis that PF was confused as to the purpose of the task is the fact that, on the very first trial, which was a true belief trial, PF gave a correct response which was *not* an appearance-based response. If she had misunderstood the task and perseverated from the pre-test phase, we would have expected an appearance-based response from the earliest trial.

Finally, it is possible that whenever PF didn’t know how to attribute the character’s belief content, the appearance-based response was the most salient option available to her but *not* because she used residual visual perspective taking abilities. Probably the strongest evidence that PF was not using visual perspective in its simplest form is the observation that PF made appearance-based responses even when the character had left the scene at the time of the belief question. (However, it is still possible that PF reasoned about what the character would see if he or she would return to the scene.) Alternatively, it could be that PF suffered a *general* interference from the appearance of the box in her reasoning. The developmental literature shows that young children, at an age where they cannot pass false belief tasks, are easily misled in their reasoning by the appearance of objects (Flavell, 1993), deciding, for example, that a sponge that looks like a rock is really a rock and not a sponge (even though the children had touched the item and felt that it was a sponge). The extent to which such a general interference effect would be linked to an executive function deficit is unclear. Both PF and WBA are impaired on standard attention and executive function tests and yet WBA was not influenced at all by the appearance of the box. The difference might be due to subtle differences in the

contribution of the frontal and parietal areas to executive control that are not captured by standard neuropsychological tests. Recent data has shown, for example, that the left parietal areas (damaged in the case of PF) might play a crucial and specific role in ignoring salient but yet irrelevant visual information (Mevorach, Humphreys, & Shalev, 2006b; Mevorach, Humphreys, & Shalev, 2006a). Whatever the reason for which she chose the appearance-based response (either because she was relying on visual perspective or because she suffered more general interference from the appearance of the box in her reasoning), the crucial finding is that, when PF was inaccurate in inferring what the character thought, she did not make reality-based or egocentric responses. This provides a striking contrast to patient WBA.

4.2. The origin of PF's improvement on the second block

Interestingly, all errors made by PF were made in the first block, whereas later she appeared to use genuine belief reasoning to solve the task. PF thus had the knowledge about how beliefs are constructed but she had difficulties accessing and/or using that information in the first part of the task. It is unclear what triggered this change in her performance, because no external information was given to PF between the two blocks. (Since the collection of the data on the Nosy Neighbour Task, PF has performed other variations of false belief tasks as part of other research projects and a similar profile of improvement from the first to the second block has been observed). As we mentioned in the case report, it also seems that PF's performance is influenced by how directly the test question mentions that someone else's perspective needs to be taken (performing better with more direct belief questions). To us, this highlights the fact that many false belief tasks (and other theory of mind tasks, and of course, many everyday instances of theory of mind) not only require the ability to ascribe a mental state, but also to work out when to ascribe the mental state and what elements in the environment need to be attended to in order to infer the content of that mental state (e.g., in the case of

belief reasoning, whether the person could have an expectation about what is in a box; whether the person has seen or been told the box's content; and such knowledge is presumably part of our semantic knowledge of how the mind works). We suggest that PF has difficulty with accessing and using such knowledge. One reason for this might be that working out when to take someone else's perspective, and what factors determine the content of someone's mental state, is a demanding problem that puts pressure on PF's limited executive processes. Another possibility is that PF's lesion has affected an access route to and/or from what we could term "social semantic" knowledge. Several studies have shown that the temporo-parietal junction plays an important role in processing gaze and other socially meaningful cues (for a review, see Allison, Puce, & McCarthy, 2000). This brain area must have efficient connections to higher order social processes so that activation generated by the perception of social cues can feed into those areas dealing with higher level social processing, including social semantic knowledge. Either way, it seems that, once more complete knowledge has been accessed, it can be used successfully to perform the remaining part of the task.

4.3. The link with the patients' executive function abilities

Nothing in the results of the two patients on standardised executive function tests can easily and straightforwardly explain the differential pattern of performance observed for WBA and PF. Both patients suffer a certain degree of executive impairment on most executive tasks that we have administered, yet, in ToM tasks, the two patients are sensitive to different factors (WBA is more likely to make errors when the demands on self-perspective inhibition are high and PF is more likely to make errors when the belief question is indirect) and they produce qualitatively different types of errors. This can be explained in two ways. It is possible that the ToM deficits observed in the two patients are independent of their executive impairment. On this account, the mechanisms of self-perspective inhibition (impaired in the case of WBA) and the mechanisms helping us to attend to the

relevant stimuli in the environment to infer someone else's belief (impaired in the case of PF) might be domain-specific. The patients' associated deficits on executive function tasks would result from the patients' lesion being large enough to affect several independent components. On the other hand, ToM tasks might require specific combinations of executive abilities which are not captured by standardised measures of executive function (see Apperly et al., 2005 for a discussion). Indeed, it seems quite plausible that different ToM components (or tasks) rely on different combinations of executive abilities (e.g., inhibition, working memory, switching of attention). On this account, certain specific combinations of executive abilities might be compromised in one patient but not in the other patient, explaining the different profiles of impairment observed across and within ToM tasks. Disentangling both interpretations will require further investigation into the nature of the various cognitive processes involved in different ToM tasks.

To conclude, the present paper illustrates how error analyses, in a new social cognition task, can inform us on the spared and impaired ToM abilities of patients with acquired brain damage. The pattern of errors of patient WBA shows that lesions to the right prefrontal cortex result in egocentric response biases when the patient is asked to take someone else's perspective. This is in line with our previous research (Samson et al., 2005) suggesting that the right prefrontal cortex is necessary to inhibit our own perspective. With the data from patient PF with lesions to the temporo-parietal junction, we show for the first time that a ToM deficit does not *necessarily* result in egocentricity. PF's pattern of errors suggests that the temporo-parietal junction might play an important role in helping us attending to the relevant elements in the environment when inferring the content of someone else's perspective. The exact nature of these processes, i.e. whether there are domain-specific or more general cognitive processes (e.g., executive function processes), still needs to be investigated. But, importantly, our findings show that neuropsychological investigations can move beyond the basic question of *whether* ToM is impaired by brain damage, to a more detailed consideration of *why* impairment is observed in

patients with different lesions. In this way, neuropsychology can help to prise-apart the multiple processes that are likely to be necessary for the complex task of judging someone else's perspective.

Acknowledgments

We are very grateful to WBA and PF for their patience and willingness to participate in this study. This research was supported by grants from the Stroke Association and the Medical Research Council (UK).

Reference List

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

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

Apperly, I. A., Samson, D., & Humphreys, G. W. (2005). Domain-specificity and theory of mind: evaluating neuropsychological evidence. *Trends in Cognitive Sciences*, 9, 572-577.

Bird, C. M., Castelli, F., Malik, O., Frith, U., & Husain, 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.

Ellis, A. W. & Young, A. W. (1997). Reading: And a composite model for word recognition and production. In A.W.Ellis & A. W. Young (Eds.), *Human cognitive neuropsychology: a textbook with readings* (Augmented edition ed., pp. 191-238). Hove: Psychology Press.

Flavell, J. H. (1992). Perspectives on perspective taking. In H.Beilin & P. Pufall (Eds.), *Piaget's Theory: Prospects and Possibilities* (pp. 107-139). Hillsdale, NJ: Erlbaum.

Flavell, J. H. (1993). The Development of Childrens Understanding of False Belief and the Appearance-Reality Distinction. *International Journal of Psychology*, 28, 595-604.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young Childrens Knowledge About Visual-Perception - Further Evidence for the Level 1-Level 2 Distinction. *Developmental Psychology*, 17, 99-103.

Frith, U. & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London.Series B: Biological Sciences*, 358, 459-473.

Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S. et al. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: theoretical and practical implications. *Brain*, 125, 752-764.

Happe, F., Brownell, H., & Winner, E. (1999). Acquired 'theory of mind' impairments following stroke. *Cognition*, 70, 211-240.

Kay, J., Lesser, R., & Coltheart, M. (1992). *Psycholinguistic Assessment of Language Processing in Aphasia*. Hove: Psychology Press.

Lough, S., Gregory, C., & Hodges, J. R. (2001). Dissociation of social cognition and executive function in frontal variant frontotemporal dementia. *Neurocase*, 7, 123-130.

Lough, S., Kipps, C. M., Treise, C., Watson, P., Blair, J. R., & Hodges, J. R. (2006). Social reasoning, emotion and empathy in frontotemporal dementia. *Neuropsychologia*, 44, 950-958.

Mevorach, C., Humphreys, G. W., & Shalev, L. (2006a). Effects of saliency, not global dominance, in patients with left parietal damage. *Neuropsychologia*, 44, 307-319.

Mevorach, C., Humphreys, G. W., & Shalev, L. (2006b). Opposite biases in salience-based selection for the left and right posterior parietal cortex. *Nature Neuroscience*, 9, 740-742.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 4, 515-526.

Qureshi, K. N. & Hodkinson, H. M. (1974). Evaluation of a ten-question mental test in the institutionalized elderly. *Age and Ageing*, 3, 152-157.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The test of everyday attention*. Cambridge: The Thames Valley Test Company.

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.

Samson, D., Apperly, I. A., Chiavarino, C., & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, 7, 499-500.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: a case of a selective deficit in inhibiting self-perspective. *Brain*, 128, 1102-1111.

Saxe, R., Carey, S., & Kanwisher, N. (2004). Understanding other minds: Linking developmental psychology and functional neuroimaging. *Annual Review of Psychology*, 55, 87-124.

Siegal, M., Carrington, J., & Radel, M. (1996). Theory of mind and pragmatic understanding following right hemisphere damage. *Brain and Language*, 53, 40-50.

Stone, V. E., Baron-Cohen, S., Calder, A., Keane, J., & Young, A. (2003). Acquired theory of mind impairments in individuals with bilateral amygdala lesions. *Neuropsychologia*, 41, 209-220.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, 10, 640-656.

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. (2000). Evidence for cognition without grammar from causal reasoning and 'theory of mind' in an agrammatic aphasic patient. *Current Biology*, 10, 723-726.

Wilson, B., Cockburn, J., & Halligan, P. (1987). *The Behavioural Inattention Test*. Bury St Edmunds: Thames Valley Test Company.

Table 1. Patients' general cognitive profile (number correct responses unless otherwise stated) when first referred to us (2002-2004) and at the time of testing for the current study (2005-2006).

	WBA		PF	
	2002-2004	2005-2006	2002-2004	2005-2006
Orientation and Long-Term Memory				
Orientation in time and space (AMT ^a)	-	10/10	-	10/10
Immediate recall of story (BUCS ^b)	12/15	8/15	-	7/15
Delayed recall of story (BUCS ^b)	14.5/15	10/15	-	7/15
Attention and Executive Function				
Spatial selective attention (Star Cancellation Test ^c)	-	49/54	-	51/54
Auditory sustained attention (Elevator Counting Test ^d)	4/7	3/7	7/7	6/7
Auditory selective attention (Elevator Counting with Distraction ^d)	-	3/10	-	0/10
Inhibition – stimulus selection (error cost ^e)	0.25	-	0	-
Inhibition – response selection (error cost ^e)	0.50	-	1.56	-
Verbal Inhibition (scaled score – Hayling Test ^f)	-	13	-	17
Rule detection and rule shifting (Brixton Test ^f)	20/54	37/54	27/54	-
Language				
Sentence/picture matching (PALPA 55 ^g)	39/60	-	47/60	-
Sentence reading (BUCS ^b)	-	42/42	-	40/42
Sentence construction (BUCS ^b)	-	4/8	-	8/8
Picture naming (BUCS ^b)	12/14	11/14	-	12/14
Praxis				
Multi-step object use (BUCS ^b)	-	11/12	-	11/12
Gesture production (BUCS ^b)	-	11/12	-	11/12
Gesture recognition (BUCS ^b)	-	4/6	-	5/6
Imitation of meaningless actions (BUCS ^b)	-	11/12	-	5/12
Figure copy (BUCS ^b)	-	42/47	-	21/47
Theory of Mind (ToM)				
False belief with indirect question and low self-perspective inhibition demands ^h				
False belief items	11/12	-	2/12	-
Control items	9/12	-	12/12	-
False belief with direct question and high self-perspective inhibition demands ^h				
False belief items	1/12	-	11/12	-
Control items	12/12	-	11/12	-

Visual Perspective – level 1				
Egocentric errors (max=36)	-	1	-	0
Other error (max=36)	-	2	-	1
Visual Perspective – level 2				
Egocentric errors (max=36)	-	32	-	7
Other error (max=36)	-	0	-	13

Impaired performance is highlighted in bold. The diagnosis is based on published norms (except for ^b, ^f and the ToM tasks). For ^b, the patients' performance was compared to the performance of 22 age-matched control participants (with a cut-off score at 2 SD below the controls' average score); for ^f, the patients' performance was compared to the performance of 16 age-matched control participants (with a cut-off score at 2 SD below the controls' average score); for the false belief tasks, control participants performed at ceiling.

^a Abbreviated Mental Test (Qureshi & Hodkinson, 1974)

^b Taken from the Birmingham University Cognitive Screen (Humphreys, Samson, Kumar, Bickerton, & Riddoch, in preparation)

^c Taken from the Behavioural Inattention Test (Wilson, Cockburn, & Halligan, 1987)

^d Taken from the Test of Everyday Attention (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994)

^e Unpublished tests. In the stimulus selection task, participants were asked to attend to one stimulus dimension (i.e., the number of fingers raised on a hand) and ignore the distracting stimulus dimension (the number of hands raised). The error cost was calculated by subtracting the mean number of errors in the "baseline" condition (when the number of fingers raised was the same as the number of hands raised) from the mean number of errors in the "inhibition" condition (when the number of fingers raised on a hand was not the same as the number of hands raised). In the response selection task, participants were asked to produce a number depending on the number of fingers raised on a hand (one or two). In the baseline condition, they had to produce the number matching the number of fingers raised whereas in the "inhibition" condition, they had to produce the opposite number. The error cost was calculated in the same way as for the stimulus selection task.

^f Taken from the Hayling and Brixton Test (Burgess & Shallice, 1997)

^g Taken from the Psycholinguistic Assessment of Language Processing in Aphasia (Kay, Lesser, & Coltheart, 1992)

^h The tasks are fully described in previous published work (Apperly et al., 2004; Samson et al., 2005)

Table 2. Accuracy of response expected in the different conditions of the Nosy Neighbour task as a function of the strategy used.

	True Belief (<i>the neighbour sees the change</i>)		False Belief (<i>the neighbour does NOT see the change</i>)	
	Predictable object (TB1)	Non-predictable object (TB2)	Predictable object (FB1)	Non-predictable object (FB2)
Example of scenarios	Scene 1: A neighbour stops at the window and watches as the occupier puts the predictable object A (<i>e.g., pizza</i>) inside the box (<i>e.g., pizza box</i>). Scene 2: The neighbour stays at the window while the occupier replaces object A (<i>e.g., pizza</i>) by object B (<i>e.g., passport</i>). Scene 3: On half of the trials, the neighbour stays at the window; on the other half of the trials, the neighbour goes away. Then, the belief question is asked.	Scene 1: A neighbour stops at the window and watches as the occupier puts the non-predictable object B (<i>e.g., passport</i>) inside the box (<i>e.g., a pizza box</i>). Scene 2: The neighbour stays at the window while the occupier replaces object B (<i>e.g., passport</i>) by object C (<i>e.g., scissors</i>). Scene 3: On half of the trials, the neighbour stays at the window; on the other half of the trials, the neighbour goes away. Then, the belief question is asked.	Scene 1: A neighbour stops at the window and watches as the occupier puts the predictable object A (<i>e.g., pizza</i>) inside the box (<i>e.g., pizza box</i>). Scene 2: The neighbour goes away and the occupier replaces object A (<i>e.g., pizza</i>) by object B (<i>e.g., passport</i>). Scene 3: On half of the trials, the neighbour stays away; on the other half of the trials, the neighbour comes back in front of the window. Then, the belief question is asked.	Scene 1: A neighbour stops at the window and watches as the occupier puts the non-predictable object B (<i>e.g., passport</i>) inside the box (<i>e.g., a pizza box</i>). Scene 2: The neighbour goes away and the occupier replaces object B (<i>e.g., passport</i>) by object C (<i>e.g., scissors</i>). Scene 3: On half of the trials, the neighbour stays away; on the other half of the trials, the neighbour comes back in front of the window. Then, the belief question is asked.
Genuine belief reasoning	Correct (<i>e.g., passport</i>)	Correct (<i>e.g., scissors</i>)	Correct (<i>e.g., pizza</i>)	Correct (<i>e.g., passport</i>)
Reality-based strategy	Correct (<i>e.g., passport</i>)	Correct (<i>e.g., scissors</i>)	Incorrect reality response (<i>e.g., passport</i>)	Incorrect reality response (<i>e.g., scissors</i>)
Appearance-based strategy	Incorrect appearance response (<i>e.g., pizza</i>)	Incorrect appearance response (<i>e.g., pizza</i>)	Correct (<i>e.g., pizza</i>)	Incorrect appearance response (<i>e.g., pizza</i>)

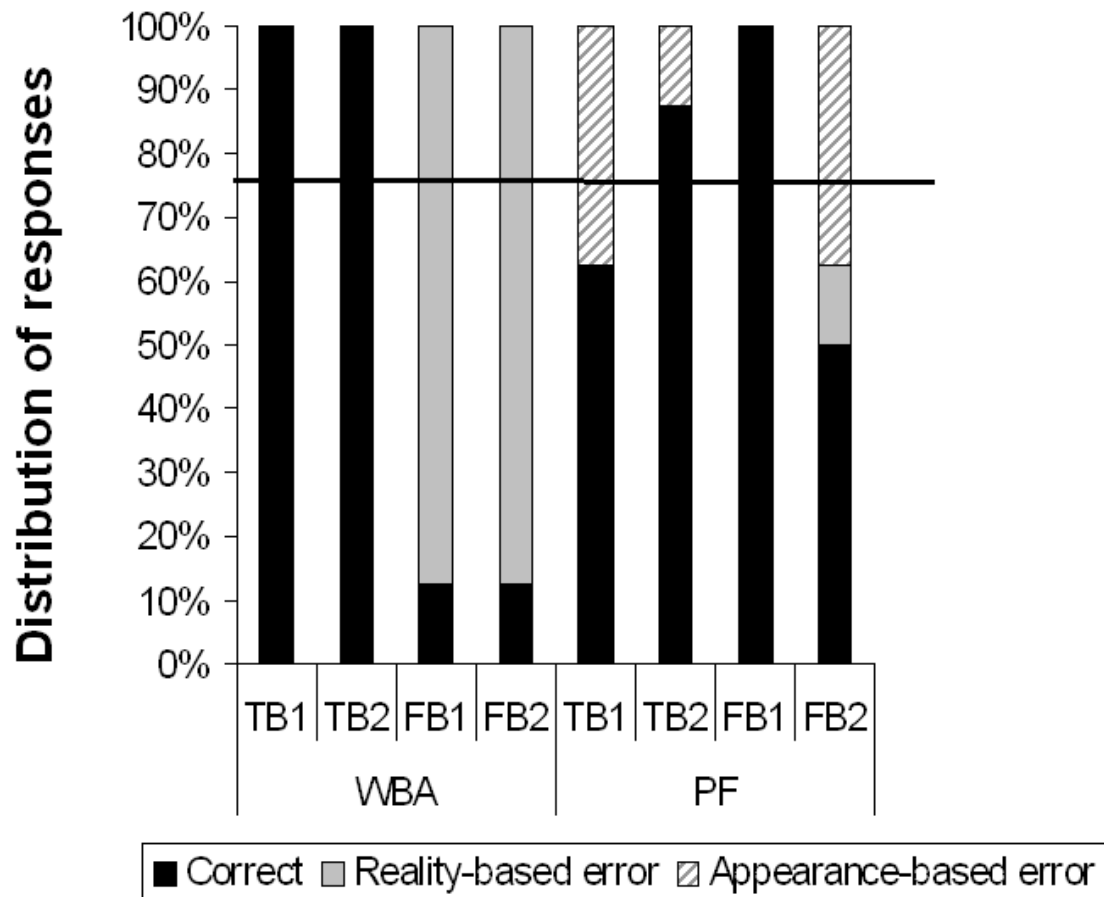


Figure 1. Distribution of responses for WBA and PF in the Nosy Neighbour task for the true belief – predictable object (TB1), true belief – non-predictable object (TB2), false belief – predictable object (FB1) and false belief – non-predictable object (FB2) conditions. For all conditions, 75% correct responses or more is significantly above chance level.

