Williams, D., & Happé, F. (2009a). 'What did I say?' versus 'What did I think?': Attributing false beliefs to self amongst children with and without autism. *Journal of Autism and Developmental Disorders, 39*, 865-873.

'What did I *say*?' versus 'What did I *think*?': Attributing false beliefs to self amongst children with and without autism.

David M Williams & Francesca Happé,

Institute of Psychiatry, Kings College London

Abstract

The task used most widely to assess recognition of false belief in self *and* others is the 'Smarties' unexpected contents task. Amongst individuals with and without autism, the Self and Other-person test questions of this task are of an equivalent level of difficulty. However, a potential confound with this task may allow the Self test question to be passed without false belief competence. Three groups of participants (with autism, developmental disability and typical development) undertook a new unexpected contents task which did not suffer from this confound. The main finding was that participants with autism performed significantly less well on the Self test question than the Other-person test question on this new task. Individuals with autism may have greater difficulty representing their own beliefs than the beliefs of other people.

Keywords: Autism Spectrum Disorder, Theory of Mind, False Belief, Unexpected Contents Task, Self-Awareness

Author Note

This study was supported by a Medical Research Council Studentship, awarded to the first author. The manuscript was prepared during an Economics and Social Research Council Fellowship awarded to the first author. Our thanks go to Peter Hobson and Sophie Lind and for their helpful comments on an early draft of the paper.

Correspondence concerning this article should be addressed to: David Williams, School of Psychology, Keynes College, University of Kent, Canterbury, Kent, CT2 7NP. Tel. +44 (0)1227 827652. Email: d.m.williams@kent.ac.uk;

'What did I say*?' versus 'What did I* think*?': Attributing false beliefs to self amongst children with and without autism.*

An individual is said to possess a 'Theory of Mind' (ToM) if they attribute mental states both to others and to self in order to explain and predict behaviour (Premack & Woodruff, 1986). The majority of research on ToM has focused on 'reading other minds', with attribution of mental states to self relatively neglected in the literature. The task used most widely to assess mental state understanding in self *and* others has been the unexpected contents false belief task (Perner, Leekam, & Wimmer, 1987). In the original and best known version of this task, a participant is shown a Smarties tube and is asked what they think is inside. Having responded that they believe there are Smarties/sweets inside, the participant is shown that the tube actually contains a pencil. The tube is then resealed with the pencil inside and the participant is asked two 'false belief' test questions. The 'Other-person' test question requires the participant to predict what another person, who has not yet seen the actual contents of the Smarties tube, would believe was inside (before they were allowed to look). This test question is designed to assess the participant's awareness of another person's false belief. Participants are also asked what they, themselves, thought the tube contained before they were allowed to look inside. This is the 'Self' test question and is thought to assess the participant's awareness of their own prior false belief.

There is substantial evidence that, in typical development, the Self and Other-person test questions from the Smarties task are of an equivalent level of difficulty. Wellman, Cross and Watson's (2001, p.665) meta-analysis of false belief task performance in typically developing (TD) children concluded that:

> The essential age trajectory for tasks requiring judgements of someone else's false belief is paralleled by an identical age trajectory for children's judgements of their own false beliefs. Young children, for example, are just as incorrect at attributing a false belief to themselves as they are at attributing it to others.

At the cognitive level of description, autism spectrum disorder (ASD) is characterised by marked deficits in ToM (e.g., Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998). Those studies that have incorporated tasks designed to assess awareness of mental states in self *and* others have tended to find similar patterns of performance to those observed amongst TD children. On the Smarties task, individuals with ASD tend to perform as poorly on the Self test question as they do on the Other-person test question (Baron-Cohen, 1991, 1992; Fisher, Happé & Dunn, 2006; Russell & Hill, 2001). However, a potential shortcoming of the Smarties Self task may provide individuals with ASD the opportunity for success despite lacking false belief competence. The difficulty with the Self task is that it involves an initial demonstration in which participants are required to *state* their (false) belief about the contents of the tube before the actual contents are revealed and before the critical test questions are asked. In this instance, it is possible that the Self question of the Smarties task could be answered correctly merely by recalling what was previously *said* rather than what was previously *believed*.

This sort of 'memory for statement' strategy could be used by individuals with ASD to 'hack out' a solution to the Self question on the Smarties task even when they lack a grasp of their own (or others') mental states. Several researchers have cautioned that success on false belief tasks amongst children with ASD may not

reflect the same underlying cognitive processes that operate in individuals who do not have ASD (e.g., Bloom & German, 2000; Happé, 1995). As Leslie and Roth (1993) note, it is important that surface level behaviour is distinguished from the mental architecture underlying behaviour.

The idea that individuals with ASD employ a compensatory 'memory for statement' strategy to succeed on the Self question of the Smarties task is somewhat suggested by evidence that their performance on this question is dramatically improved when they are asked specifically about their original *statement* about the contents of the box than when they are asked about their original *belief*. In studies by Perner, Frith, Leslie, and Leekam (1989), and Leslie and Thaiss (1992), children with ASD and TD children were given a standard task in which they were initially asked to state what they believed was inside the box, and were subsequently asked the critical Self and Other-person test questions. Now, in the studies of false belief understanding cited above (e.g., Fisher et al., 2006) participants were asked what they *thought* was inside the box before they looked (Self question) and what another person will *think* is inside the box before they look (Other-person question). However, in the studies by Perner et al. and Leslie and Thaiss, the Self and Other-person test questions were phrased subtly differently, participants being asked what they *said* was inside the box before they looked (Self question) and what another person will *say* is inside before they are allowed to look (Other-person question). This apparently simple change in the wording of the test questions resulted in children with ASD, but not TD participants, performing significantly less well on the Other-person test question than on the Self test question.

The results of Perner et al. (1989) and Leslie and Thaiss (1992) highlight the possibility that individuals with autism are able to recall what they stated was inside the familiar container without necessarily grasping this statement as a reflection of their belief about the box's contents. As Leslie and Thaiss (1992, p.239, original emphasis) suggest, individuals with ASD may "not relate the *uttering* of a sentence to the speaker's underlying propositional attitude: for example they will not connect assertion with belief". Supporting this argument, Roth and Leslie (1991) found that in a modified false belief task TD children, but not children with ASD, reported that a speaker believes what they say.

Given these findings, it seems a reasonable concern that asking children with ASD to state what they believe to be in the Smarties tube before asking them the false belief test question might lead to artificially inflated levels of performance, in the absence of true false belief competence. On this basis, the results from previous studies employing the Smarties task amongst children with ASD might be questioned. Whereas these studies have observed either parallel performance across the Self and Other-person test questions (e.g., Fisher et al., 2006) or an advantage on the Self question over the Other-person question (e.g., Perner et al., 1989), perhaps individuals with ASD would have *greater* difficulty representing their own false beliefs if the task demands of the Self and Other questions were better equated.

This latter prediction might strike the reader as unusual given that most theories of ToM suggest either that a concept of belief is acquired for self and others in parallel (Gopnik, 1993; Gopnik & Meltzoff, 1994; Leslie, 1987), or that knowledge of one's own beliefs emerges prior to, and forms the basis of, knowledge of others' beliefs (Goldman, 1993; Harris, 1992). However, if one takes the competence-performance framework seriously, there are grounds to believe that reasoning about beliefs in self might engage different processes to those involved in reasoning about beliefs in others (German & Leslie, 2000). The competence-performance framework

is usually invoked to support the argument that typically developing children prior to age 4 years have false belief competence, but that performance factors (e.g., executive limitations) prevent them from expressing this competence in false belief tasks (Leslie & Thaiss, 1992). However, this argument cuts both ways: if it is assumed that many children with ASD have diminished false belief competence but 'hack out' solutions to false belief tasks using compensatory (performance) strategies (Happé, 1995), then different processing routes for self and other could theoretically be seen amongst these individuals also. Indeed, for an individual with only partial knowledge of beliefs, it may be easier to identify a false belief in another person, whose behaviour is readily *observable*, than in oneself whose behaviour is less visually accessible.

In an attempt to test this hypothesis, we devised an unexpected contents task – the 'Plasters' (band-aid) task – in which participants did not verbalise their beliefs about the contents of a familiar container before they were asked the false belief test questions. In this modified task, a plasters box and two other, unrelated containers were placed in a convenient location within reaching distance of the participant, but out of the experimenter's reach. The experimenter pretended that he had cut his finger and so asked the participant if s/he could get him a plaster (pointing in the direction of the containers). The participant obligingly attempted to fetch a plaster but unexpectedly found that the plasters box contained birthday cake candles. Once the participant had discovered the actual contents of the box, they were asked the Self and Other-person test questions, following the standard procedure of the traditional Smarties task.

The crucial manipulation in the Plasters task is that participants never explicitly verbalise their belief that the plasters box contains plasters although, by *selecting* this box when asked to help the experimenter, participants demonstrated their false belief unambiguously. Therefore, success on the Self test question of this Plasters task is not possible through a simple 'memory for statement' strategy since no statement was ever made. Rather, an individual must recognise and recall their false belief. It is important to note that the Other-person test question in this Plasters task was of the same form as in the traditional Smarties task. The aim of the current study was not to adapt the unexpected contents task *per se*, but to better equate the demands of the Self and Other-person aspects of the task.

An anonymous reviewer of the manuscript kindly brought to our attention an early study of false belief understanding amongst TD children by Bartsch and Wellman (1989) in which a task with similarities to the Plasters task was employed. In Bartsch and Wellman's study, participants discovered that a familiar box (in one case, a plasters box) did not contain its usual contents, the typical contents instead being located in a nearby, unmarked box. Participants were then asked, for instance, (a) where a puppet who needed a plaster would look for the item (prediction question) and; (b) why a puppet who needed a plaster would look in the plasters box (explanation question). There are notable similarities between Bartsch and Wellman's experimental procedures and those involved in the current Plasters task. However, the focus in Bartsch and Wellman's study was upon the relative difficulty of predicting versus explaining another's false belief, not upon the relative difficulty of recognising false beliefs in self versus others. As such, participants in their study were not asked the critical test question about own false beliefs, which was pivotal to the current investigation. Therefore, Bartsch and Wellman's results aren't considered here, although the similarity in procedures to those employed in the current study is noted.

The Plasters task was completed by a group of individuals with ASD, a comparison group of individuals with developmental disability (DD), equated for age and verbal ability, and a group of young TD children who were of an age at which most would be on the cusp of passing traditional false belief tasks. All of the TD children, as well as a sub-sample of children with ASD, were also given a traditional Smarties task.

Two specific predictions were made about participants' performance *within* each of the Plasters and Smarties tasks. Firstly, within the new Plasters task, it was predicted that participants with ASD would be unique in performing significantly *less* well on the Self test question than the Other-person test question. Parallel performance across the two test questions was predicted amongst both DD and TD participants. Secondly, within the original Smarties task, parallel performance across the Self and Other-person test questions was predicted amongst all participants, in line with previous findings using the same methodology.

Two specific predictions were also made about participants' performance *across* the Plasters and Smarties tasks. Firstly, it was predicted that participants with ASD would be unique in performing significantly less well on the Plasters Self question than the Smarties Self question. Equivalent levels of performance across the Smarties Self and Plasters Self questions were predicted amongst TD participants. Secondly, given that the Other-person aspect of the Plasters task was identical to the Other-person aspect of the Smarties task, equivalent levels of performance across the Other-person test question from each measure were expected amongst all participants.

Method

Participants

Ethical approval for this research was obtained from the joint South London and Maudsley NHS Trust/Institute of Psychiatry Research Ethics Committee. Fifty-two children with ASD, 25 children with DD and 41 TD children participated in the study, after parents/guardians had given written, informed consent for their children to be included. The participants in the ASD group had received formal diagnoses, by a trained psychiatrist or pediatrician, of autistic disorder (n = 47), Asperger's disorder (n = 3) or atypical autism/pervasive developmental disorder not otherwise specified (PDD-NOS; n = 2) according to established criteria (DSM-IV-TR, American Psychiatric Association, 2000; ICD-10, World Health Organisation, 1993). All participants in this group attended specialist autism schools, which required a diagnosis of autism, Asperger's syndrome or atypical autism/PDD-NOS for entry into the school. The DD group consisted of 25 children with general learning disability of unknown origin who attended UK schools for children with special educational needs. Any DD participant who was described in their statement of special educational needs or by their head teacher as having any social-communication difficulties was excluded from the study. This allows confidence that comparison participants in the current study did not have autism-related symptoms. Finally, 41 TD participants were recruited from mainstream nurseries and primary schools in the Greater London area.

Background assessments

Baseline verbal abilities were assessed by an appropriate measure for the developmental level of the participant. The verbal abilities of 26 (out of 52) children

with ASD and 18 (out of 25) comparison children were determined by performance on the Vocabulary and Information subtests of the Wechsler Intelligence Scale for Children – Third Edition UK (WISC-III; Wechsler, 1991). The verbal IQ estimate gained from this short form has high reliability (Sattler, 1992). Because the lowest test age-equivalent offered by the WISC-III is 6 years 2 months, the verbal mental age (VMA) of any participant who fell below this level on either of the verbal subtests could not be calculated. Under these circumstances, participants were administered the British Picture Vocabulary Scale – Second Edition (BPVS; Dunn et al., 1997), which offers test age-equivalents down to 2 years 11 months. In this instance, the verbal abilities of 26 (out of 52) children with ASD and 7 (out of 25) DD children were assessed with the BPVS. Of the 52 children with ASD who completed the Plasters task, a sub-sample of 19 also received a traditional ('Smarties') unexpected contents task, as did each of the 41 TD participants. DD participants received the Plasters task only.

The participant characteristics of each group, including the 'Smarties sub-sample' of 19 ASD participants, are presented in Table 1. Statistical analyses showed that the ASD and DD groups were adequately equated on all variables: CA: $t(75) = -1.80$, $p = .08$, $r = .21$; VMA: $t(75) = -0.72$, $p = .47$, $r = .08$; VIQ: $t(75) = 1.72$, $p = .09$, $r = .20$.

(Table 1 about here)

Given that some ASD and DD participants received the WISC-III whilst others received the BPVS, independent *t*-tests were conducted on each sub-sample to ensure adequacy of matching in each case. ASD and DD participants who received the WISC-III were equated adequately for CA and VMA (all *t*s < 0.81, all *p*s > .42). However, in this WISC-III sub-sample, participants with ASD had significantly higher VIQs ($M = 79.12$, $SD = 17.83$) than DD participants ($M = 69.39$, $SD = 12.00$) ($t = 2.02$, $p = .05$, $r = .30$). ASD and DD participants who received the BPVS were adequately equated for CA, VMA and VIQ (all *t*s < 0.63, all *p*s > .53).

Design and Procedures

Amongst those participants who received both the new Plasters task and the traditional Smarties task, the order of task completion was counterbalanced.

Standard Smarties task. The participant was shown a Smarties tube, with which all children in UK schools are familiar. They were asked "What is in here?" All participants said "Smarties" or "Chocolate". The tube was then opened to reveal that it really contained a pencil. The pencil was placed back inside the tube which was then resealed. The participant was then asked the Self test question ("*Before you looked in the tube, what did you think was inside?*"), the Other-person test question ("*Later on I am going to show this tube to your teacher. He/she hasn't seen inside here though. What will he/she think is in there before he/she looks inside?*") and a reality control question ("*What's inside the tube, really?*"). These questions were asked in a fixed order across all participants. Only participants who passed the reality control question were included in the study.

Plasters task. Three containers (an empty 'Pringles' crisps tube, an empty mints box, and a plasters box which contained birthday cake candles) were placed in a

convenient location within reaching distance of the participant, but out of the experimenter's reach. The experimenter pretended that he had a cut on his little finger and asked the participant to get him a plaster: "*Oh, I have got this little cut on my finger* [experimenter points to his finger without showing it directly, up close, to the participant]. *Could you get me a plaster?* [Experimenter points in the direction of the three containers]". The participant then picked up the plasters box and opened it to find birthday cake candles inside. If the participant did not open the box spontaneously the experimenter requested: "*Could you just get one out for me?*". Once the participant had discovered the real contents of the plasters box, the experimenter replaced the candles and resealed the box. The participant was then asked the Self test question ("*Before you looked in the box, what did you think was inside?*"), the Other-person test question ("*Later on I am going to show this box to your teacher. He/she hasn't seen inside here though. What will he/she think is in there before he/she looks inside?*") and a reality control question ("*What's inside the box, really?*"). These questions were asked in a fixed order across all participants. Only participants who passed the reality control question were included in the study.

Results

Preliminary Between-Subjects analysis

On the Plasters task, 27/52 (51.9%) participants with ASD passed the Self test question compared to 23/25 (92.0%) participants with DD. This between-group difference was significant, $\chi^2(1) = 11.91$, $p = .001$, $\phi = .39$. Also, 36/52 (69.2%) participants with ASD passed the Plasters Other-person test question compared to 22/25 (88.0%) participants with DD. This difference was also significant $\chi^2(1) = 3.20$, p (one-tailed) $< .04$, $\phi = .20$.

Within-subjects analyses

In order to assess the relative difficulty of the Plasters and Smarties test questions, both within and between tasks, patterns of within-participant performance were explored using McNemar tests.

Plasters task: self versus other-person. Table 2 shows the contingency between performance on the Plasters Self and Other-person test questions by ASD, DD, and TD participants. Participants who performed inconsistently across the two questions, passing only one of the two, are highlighted in bold.

To reiterate, the critical prediction tested through this analysis that previous findings of parallel performance amongst participants with ASD across the Self and Other-person test questions of the Smarties task may have been confounded. Because children with ASD could plausibly pass the Self question of the traditional task, in the absence of false belief competence, by recalling their *statement* that the tube contained Smarties, their performance on the Self question could have been artificially inflated. Therefore, it was predicted that on the Plasters task (where no initial statement of the participant's belief is made) participants with ASD would perform significantly less well on the Self test question than the Other-person test question.

On the other hand, it was speculated that individuals who do not have ASD rely rarely on a compensatory strategy of recalling their previous statement on the

Smarties task. Hence, their performance across the Plasters Self and Other-person test questions should be equivalent, in line with their performance on the traditional task.

The results were that amongst participants with ASD, 11/52 (21.2%) failed the Plasters Self question despite passing the Plasters Other-person question, and 2/52 (3.8%) participants showed the opposite pattern of performance. Amongst participants with ASD, the Plasters Self question was significantly more difficult than the Plasters Other question, McNemar's $p = .02$.

Amongst participants with DD, 1/25 (4.0%) failed the Plasters Self question despite passing the Plasters Other-person question, and 2/25 (8.0%) showed the opposite pattern of performance. Amongst participants with DD, the Plasters Self and Other test questions did not differ significantly in difficulty, McNemar's $p = .63$.

Amongst TD participants, 7/41 (17.1%) failed the Plasters Self question despite passing the Plasters Other-person question, and 2/41 (4.9%) showed the opposite pattern of performance. Amongst TD participants, the Plasters Self and Other test questions did not differ significantly in difficulty, McNemar's $p = .18$.

(Table 2 about here)

Smarties task: self versus other-person. Table 3 shows the contingency between performance on the Smarties Self and Other-person test questions by ASD and TD participants. The aim of this analysis was to confirm that participants from each group were typical in finding the two test questions from this traditional task of an equivalent level of difficulty.

Amongst participants with ASD, 4/19 (21.1%) failed the Smarties Self question despite passing the Smarties Other-person question, and 3/19 (15.8%) participants showed the opposite pattern of performance. Amongst participants with ASD, the Smarties Self and Other-person test questions did not differ significantly in difficulty, McNemar's $p > .99$.

Amongst TD participants, 2/41 (4.9%) failed the Smarties Self question despite passing the Smarties Other-person question, and 5/41 (12.2%) showed the opposite pattern of performance. Amongst TD participants, the Smarties Self and Other test questions did not differ significantly in difficulty, McNemar's $p = .45$.

(Table 3 about here)

Plasters self versus Smarties self. Table 4 shows the contingency between performance on the Plasters Self and Smarties Self test questions by ASD and TD participants. To reiterate, the critical prediction tested through this analysis is that individuals with ASD, but not those without ASD, use a compensatory strategy of recalling their own statement to succeed on the Smarties Self question in the absence of false belief competence. Therefore, it was predicted that performance on the Plasters Self question (on which success is not possible through such a strategy) would be significantly poorer than the Smarties Self question amongst participants with ASD only.

Amongst participants with ASD, 4/19 (21.1%) failed the Plasters Self question despite passing the Smarties Self question, and 1/19 (5.3%) participants showed the opposite pattern of performance. Amongst this sub-sample of participants with ASD, the Plasters Self and Smarties Self test questions did not differ significantly in difficulty, McNemar's $p = .45$.

Amongst TD participants, 9/41 (22.0%) failed the Plasters Self question despite passing the Smarties Self question, and 1/41 (2.1%) showed the opposite pattern of performance. Amongst TD participants, the Plasters Self question was significantly more difficult than the Smarties Self question, McNemar's $p = .02$.

To summarise, contrary to predictions, typically developing participants found the Self question of the Smarties task significantly less difficult than the Self question of the Plasters task. Also contrary to predictions, participants with ASD did *not* perform significantly less well on the Plasters Self question than on the Smarties Self question.

(Table 4 about here)

Plasters other-person versus Smarties other-person. Table 5 shows the contingency between performance on the Plasters Other-person and Smarties Other-person test questions by ASD and TD participants. The aim of this analysis was to confirm that the Plasters task was not more difficult than the Smarties task *per se*. Given that the Other-person aspect of the Plasters task was identical to the Other-person aspect of the Smarties task, it was predicted that equivalent levels of performance would be seen across each task amongst all participants.

Amongst participants with ASD, 3/19 (15.8%) failed the Plasters Other-person question despite passing the Smarties Other-person question, and 1/19 (5.3%) participants showed the opposite pattern of performance. Amongst participants with ASD, the Plasters Other-person and Smarties Other-person test questions did not differ significantly in difficulty, McNemar's $p = .63$.

Amongst TD participants, 4/41 (9.8%) failed the Plasters Other-person question despite passing the Smarties Other-person question, and 4/41 (9.8%) showed the opposite pattern of performance. Amongst TD participants, the Plasters Other-person and Smarties Other-person test questions did not differ significantly in difficulty, McNemar's $p > .99$.

(Table 5 about here)

Discussion

This study explored the effects of removing a potential confound from the traditional Smarties false belief task that may have allowed the Self test question (supposedly assessing awareness of one's own false belief) to be passed in the absence of a representational ToM. The format of the new 'Plasters' false belief task did not require participants to verbalise their own (false) belief about the contents of a container prior to being asked the usual Self test question. Therefore, success on the Self question of this task must reflect an awareness of one's own prior *belief*, as opposed to one's prior *statement*.

In line with the main prediction of the study, participants with ASD found it significantly more difficult to report their own prior false belief than to predict the false belief of another person on the new Plasters task (see Table 2). In contrast, participants who did not have ASD, whether they had developmental disability or not, performed consistently across the Plasters Self and Other test questions. This suggests that the majority of individuals who do not have ASD are not restricted to a strategy of recalling their previous statement to succeed on the Self test question of the unexpected contents task. For these individuals, success on the Self test question

of the traditional unexpected contents task largely reflects an accurate representation of their prior false belief.

In accordance with this reasoning, the other predictions of the study were that (a) in keeping with other studies, all participants who undertook the Smarties task would perform equivalently across the Self and Other-person test questions; (b) participants with ASD would find the Self question of the Plasters task significantly harder than the Self question of the Smarties task because they could not rely on a 'memory for statement' strategy to succeed on the Self test question of the Plasters task; (c) all participants would perform equivalently across the Other-person question on the Plasters task and the Other-person question on the Smarties task, since both tasks were structurally equivalent

In line with predictions (a) and (c), amongst ASD and TD participants the Smarties Self and Other-person test questions were of an equivalent level of difficulty, as were the Plasters Other-person and Smarties Other-person questions. Importantly, this latter finding confirms that is finding confirms that the Plasters and Smarties tasks were equated with respect to the demands involved in answering the Other-person test questions.

However, contrary to the final critical prediction of the study, TD children, and *not* participants with ASD, performed significantly less well on the Plasters Self question than on the Smarties Self question. This final result requires some unpacking. With respect to the TD sample, the results were unambiguous; over 20% of the sample succeeded on the Smarties Self question despite failing on the Plasters Self question. It appears that at least a proportion of TD children were supported on the Smarties task by the verbalisation of their (false) belief prior to the Self test question. Although this result was not predicted at the outset of the study, the finding is perhaps understandable in light of previous research. In studies by Mitchell and Lacohée (1991) and Charman and Lynggaard (1998), TD children were presented with a Smarties tube and asked what they thought was inside the box. Having stated that they (falsely) believed that the box contained Smarties, participants were encouraged to select a picture of Smarties and then to post this into a post-box, provided by the experimenter. Subsequently, in each study, participants performed significantly better on the Self question of this 'picture posting' version of the Smarties task than on the Self question of the traditional task (where no picture was posted).

Although the interpretation of Mitchell and Lacohée's (1991), and Charman and Lynggaard's (1998) results are open to debate (see Perner, Baker & Hutton, 1994), the studies highlight that the performance of TD children on the Smarties task can be scaffolded in meaningful ways through manipulations to the task's structure. However, the current results suggest that although TD children can *benefit* from verbalising their false belief in the Smarties task (or, indeed, posting a picture as a representation of their belief), they do not *rely* on it: In the Plasters task the opportunity to pass the Self question though a 'memory for statement' strategy was not available, yet TD participants (unlike participants with ASD) nonetheless displayed parallel performance across the Self and Other-person test questions in line with their typical performance on the Smarties task. If TD participants had been entirely reliant upon such a strategy then, like participants with ASD, they should have displayed an atypical profile of performance on the Plasters task.

Regarding the failure of the current findings to confirm the prediction that participants with ASD would perform significantly less well on the Plasters Self question than the Smarties Self question, two points should be made. Firstly, because

only a sub-sample of (n = 19) participants with ASD received both the Plasters and Smarties tasks, the power of the analysis to detect differences in performance across the tasks was limited. This point is especially poignant when considering the characteristics of this particular sub-sample, who were developmentally relatively immature in terms of age (M = 7.49 years) and verbal ability (M = 5.24 years). As a result, the majority (10/19; see Table 4) performed at floor, failing the Self question on each task, leaving little variation to detect statistical differences. Therefore, the trend toward superior performance on the Smarties Self question than on the Plasters Self question may have become significant given a larger sample of participants. In fact, looking at Table 4, the percentage of participants with ASD who showed the predicted pattern of performance (i.e., failing Plasters Self but passing Smarties Self; 21.1%) was closely similar to the percentage of TD participants (22.0%) who showed this pattern of performance. However, only in the larger sample of TD participants was the result statistically significant.

Theoretically, this study provides the first evidence that individuals with ASD find it relatively more difficult to recognise their own (false) beliefs than the (false) beliefs of other people. The claim here is not that an individual with ASD might possess a coherent conception of other's beliefs but not of their own beliefs. There are persuasive arguments why an individual with theory of mind competence must necessarily recognise mental states in themselves *and* others (Hobson, 1990, 2006; Leslie, 1987; Strawson, 1962). However, individuals with ASD might be atypical in 'solving' theory of mind problems through the application of a kind of rule-bound, cognitively-acquired heuristic, rather than through an affective system/capacity (Hermelin & O'Connor, 1985). Under these circumstances, it may be more or less difficult to apply one's (partial) knowledge (of beliefs) to the case of self or others. Indeed, as speculated above, rule-governed knowledge might be more easily acquired and/or applied to the case of others' mental states than to one's own. After all, there would seem to be many more opportunities to observe the behaviour of others, and thereby learn 'behaviour rules' (Povinelli & Vonk, 2004) to predict their actions, than there are opportunities to observe one's own behaviour and acquire such rules in relation to oneself. This is not to say that there are no modes of first-person experience available to individuals with ASD that could provide relevant information about their own behaviour (Williams & Happé, in press a). It is just that these forms of self-experience appear insufficient for acquiring a theory of (one's own or another's) mind (David et al., 2008).

The results of this study suggest that we still have much to learn about how individuals with ASD come to represent their own mental states (see also Williams & Happé, in press b). On a methodological note, the results also suggest that individuals with autism are able to pass the Self question on traditional unexpected contents tasks without reflecting on their own previous false beliefs. Compared to the Smarties task, the Plasters task may be advantageous not only in terms of removing the identified confound with the traditional Self test question. The task may also provide a more ecologically valid, 'real-world' test of false belief understanding amongst children with and without autism. For these reasons, future studies aiming to assess false belief reasoning in self and other may benefit from implementing the Plasters task or some variation of it.

References

American Psychiatric Association (2000). *Diagnostic and Statistical Manual: 4th Edition Test Revised (DSM IV-TR).* New York: Author.

Baron-Cohen, S. (1991). The development of a theory of mind in autism: deviance and delay. *Psychiatric Clinics of North America, 14*(1), 33-51.

Baron-Cohen, S. (1992). Out of sight or out of mind: another look at deception in autism. *Journal of Child Psychology and Psychiatry, 33*(7), 1141-1155.

Bartsch, K., & Wellman, H. (1989). Young children's attribution of action to beliefs and desires. *Child Development, 60*, 946-964.

Bloom, P., & German, T. (2000). Two reasons to abandon the false belief task as a test of theory of mind. *Cognition, 77*, 25-31.

Charman, T., & Lynggaard, H. (1998). Does a photographic cue facilitate false belief performance in subjects with autism? *Journal of Autism and Developmental Disorders, 28*(1), 33-42.

David, N., Gawronski, A., Santos, N.S., Huff, W., Lehnhardt, F.G., Newen., & Vogeley, K. (2008). Dissociation between key processes of social cognition in autism: Impaired mentalizing but intact sense of agency. *Journal of Autism and Developmental Disorders, 38*(4), 593-605.

Dunn, L.M., Dunn, L.M., Whetton, C., & Burley, L. (1997). *British Picture Vocabulary Scale (2nd Ed.).* NFER-Nelson.

Fisher, N., Happé, F., & Dunn, J. (2005). The relationship between vocabulary, grammar, and false belief task performance in children with autistic spectrum disorders and children with moderate learning difficulties. *Journal of Child Psychology and Psychiatry, 46*, 409-419.

German, T., & Leslie, A.M. (2000). Attending to and learning about mental states. In P. Mitchell & K.J. Riggs (Eds.), *Children's Reasoning and the Mind*. Hove, UK: Psychology Press.

Goldman, A. (1993). The psychology of folk psychology. The *Behavioral and Brain Sciences, 16*, 15-28.

Gopnik, A. (1993). How we know our own minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*, 1-14.

Gopnik, A., & Meltzoff, A.N. (1994). Minds, bodies and persons: Young children's understanding of the self and others as reflected in limitation and theory of mind research. In S.T. Taylor, R.W. Mitchell, & M.L. Boccia (Eds.), *Self-Awareness in Animals and Humans: Developmental Perspectives*. Cambridge: Cambridge University Press.

Harris, P.L. (1992). From simulation to folk psychology: The case for development. *Mind and Language, 7*, 120-144.

Hermelin, B., & O'Connor, N. (1985). Logico-affective states and nonverbal language. In E. Schopler & G. Mesibov (eds.), *Communication Problems in Autism*, pp.283-310. New York: Plenum Press.

Hobson, R. P. (1990). On the origins of self and the case of autism. *Development and Psychopathology, 2*, 163-181.

Hobson, R. P., Chidambi, G., Lee, A., & Meyer, J. (2006). Foundations for self-awareness: An exploration through autism. *Monographs of the Society for Research in Child Development, 71*.

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94*, 412-426.

Leslie, A. M., & Roth, D. (1993). What autism teaches us about metarepresentation. In S. Baron-Cohen, H. Tager-Flusberg, & D. Cohen (Eds), *Understanding Other Minds: Perspectives from Autism* (pp.83-111). Oxford: Oxford University Press.

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

Mitchell, P., & Lacohee, H. (1991). Children's early understanding of false belief. *Cognition, 39*(2), 107-127.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Explorations of the autistic child's theory of mind: knowledge, belief, and communication. *Child Development, 60*, 689-700.

Perner, J., Baker, S., & Hutton, D. (1994). Prelief: the conceptual origins of belief and pretence. In C. Lewis & Mitchell, P. (Eds.), *Children's Early Understanding of the Mind: Origins and Development* (pp.261-286). Hove, UK: Erlbaum.

Povinelli, D.J., & Vonk, J. (2004). Do we really need a microscope to explore the chimpanzee's mind? *Mind and Language, 19*, 1-28.

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

Roth, D., & Leslie, A.M. (1991). The recognition of attitude conveyed by utterance: A study of preschool and autistic children. *British Journal of Experimental Psychology, 9*, 315-330.

Russell, J., & Hill, E. L. (2001). Action monitoring and intention reporting in children with autism. *Journal of Child Psychology and Psychiatry, 42*, 317-328.

Sattler, J.M. (1992). *Assessment of children: WISC-III and WPPSI-R supplement*. USA: Jerome M. Sattler Publishers Inc.

Strawson, P. F. (1962, originally 1958). Persons. In V. C. Chappell (Ed.), *The Philosophy of Mind* (pp.127-146). Englewood Cliffs, NJ: Prentice-Hall.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children*. (Third Edition), London, UK: The Psychological Corporation

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development, 72*, 655 – 684.

Williams, D., & Happé, F. (in press a). Pre-conceptual aspects of self-awareness in autism spectrum disorder: The case of action monitoring. *Journal of Autism and Developmental Disorders*.

Williams, D., & Happé, F. (in press b). Representing intentions in self and others: Studies of autism and typical development. *Developmental Science*.

World Health Organisation (1993). *International Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines,* (10^{th} ed.) (ICD-10). Geneva: Author.

Yirmiya, N., Erel, O., Shaked, M., and Solomonica-Levi, D. (1998). Meta-analysis comparing theory of mind abilities in individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin, 124*, 283-307.

Table 1: Participant characteristics: Means and (standard deviations)

	ASD	ASD: Smarties sub-sample	DD	TD
n	52	19	25	41
CA: years	11.15 (3.71)	7.49 (1.76)	12.33 (1.99)	4.47 (0.64)
VMA:years	7.46 (2.76)	5.24 (1.15)	7.91 (2.11)	4.89 (1.24)
VIQ	74.29 (18.52)	81.79 (15.04)	68.48 (11.00)	104.1 (9.76)

Table 2: Contingency between performance on the Plasters Self and Other-person test questions amongst ASD, DD, and TD participants. Inconsistent performers are highlighted in bold.

	Plasters OTHER					
	ASD		DD		TD	
Plasters SELF	Pass	Fail	Pass	Fail	Pass	Fail
Pass	25 (48.1%)	**2 (3.8%)**	21 (84.0%)	**2 (8.0%)**	16 (39.0%)	**2 (4.9%)**
Fail	**11 (21.2%)**	14 (26.9%)	**1 (4.0%)**	1 (4.0%)	**7 (17.1%)**	16 (39.0%)

Table 3: Contingency between performance on the Smarties Self and Other-person test questions amongst ASD and TD participants. Inconsistent performers are highlighted in bold.

	Smarties OTHER			
	ASD		TD	
Smarties SELF	Pass	Fail	Pass	Fail
Pass	5 (26.3%)	**3 (15.8%)**	21 (51.2%)	**5 (12.2%)**
Fail	**4 (21.1%)**	7 (36.8%)	**2 (4.9%)**	13 (31.7%)

Table 4: Contingency between performance on the Smarties Self and the Plasters Self test questions amongst ASD and TD participants. Inconsistent performers are highlighted in bold.

	Smarties SELF			
	ASD		TD	
Plasters SELF	Pass	Fail	Pass	Fail
Pass	4 (21.1%)	**1 (5.3%)**	17 (41.5%)	**1 (2.1%)**
Fail	**4 (21.1%)**	10 (52.6%)	**9 (22.0%)**	14 (34.1%)

Table 5: Contingency between performance on the Smarties Other-person and the Plasters Other-person test questions amongst ASD and TD participants. Inconsistent performers are highlighted in bold.

	Smarties OTHER			
	ASD		TD	
Plasters OTHER	Pass	Fail	Pass	Fail
Pass	6 (31.6%)	**1 (5.3%)**	19 (46.3%)	**4 (9.8%)**
Fail	**3 (15.8%)**	9 (47.4%)	**4 (9.8%)**	14 (34.1%)