Williams, D. (2010). Theory of own mind in autism: Evidence of a specific deficit in self-awareness? *Autism: International Journal of Research and Practice*, *14*, 474-494.

Theory of *own* mind in autism: Evidence of a specific deficit in self-awareness?

David M. Williams, Durham University, UK.

Abstract

Assuming that self-awareness is not a unitary phenomenon, and that one can be aware of different aspects of self at any one time, it follows that selective impairments in self-awareness can occur. This paper explores the idea that autism involves a particular deficit in awareness of the "psychological self", or "theory of *own* mind". This hypothesised deficit renders individuals with autism spectrum disorder (ASD) at least as impaired at recognising their own mental states as at recognising mental states in other people. This deficit, it is argued, stands in contrast to an apparently typical awareness of the "physical self" amongst people with autism. Theoretical implications of the empirical evidence are discussed.

Introduction

According to the original use of the term, 'theory of mind' (ToM) refers to the ability to attribute mental states, like beliefs and intentions, to self and others in order to explain and predict behaviour (Premack & Woodruff, 1978). According to one, arguably dominant, theoretical position, the ability to recognise mental states in oneself depends on the same psychological mechanism, or process, or core knowledge, as recognising mental states in other people (e.g., Carruthers, 2009; Frith & Happé, 1999; Gopnik, 1993; Hobson, 1990; Leslie, 1987). The implications of this theory for our understanding of autism, a disorder widely acknowledged to involve a diminished understanding of *others*' minds, were the subject of an influential paper by Frith & Happé (1999; but see Hobson, 1990 for an earlier discussion of the *developmental* relation between self-awareness and ToM in autism). In this paper, Frith and Happé argued that people with autism might have as little awareness of their own mental states as they have awareness of the mental states of other people (see also Frith, 2003). This controversial claim has been questioned by several researchers, each of whom has argued that self-awareness (of mental states) is intact in autism, despite a deficit in recognising mental states in others (Goldman, 2006; Nichols & Stich, 2003; McGeer, 2004; Raffman, 1999). According to these philosophers, the ability to recognise mental states in others is either (a) dependent upon direct first-person access to one's own mental states, such that those states can be used to *simulate* (or imagine) the other's mental states (e.g., Goldman), or (b) reliant on an entirely separate underlying system to that employed to represent mental states in self (Nichols & Stich, 2003).

A key aim of the current paper is to evaluate whether autism involves a deficit in theory of *own* mind (as well as theory of others' minds) and to what extent an observed deficit in self-awareness is 'domain-specific'. It is widely assumed amongst both philosophers and psychologists, that the self is not a unitary construct, but consists of various aspects/dimensions each of which can be both (non-reflexively) experienced and (reflexively) known at any one time. One particularly salient distinction is between the physical self and the psychological self (Gillihan & Farah, 2005; Neisser, 1988). Intuitively, experience and knowledge of one's own body seems quite separable from experience and knowledge of one's own mind. For some theorists, however, these two aspects of self-awareness are intimately connected (e.g., Lewis, 2003), with the latter depending in large part on the former (e.g., Russell, 1996). How these issues help to understand autism and how research into self-awareness in autism helps to clarify these issues will be the focus of this paper.

Awareness of the Psychological Self, or "Theory of own Mind"

For typically developed humans, mindreading is a swift (frequently non-conscious) process. We routinely attribute thoughts, feelings, perceptions and so forth to other people to make sense of their behaviour. It is widely acknowledged that, for one reason or another, individuals with ASD show impairments in recognising mental states in other people, particularly epistemic states like belief and knowledge. The question to be addressed in this part of the paper, though, is whether the difficulty in recognising such mental states in others, amongst people with ASD, extends to a similar (or maybe even more profound) difficulty in recognising mental states in *self*.

Recognising own (false) beliefs

Amongst both typically and atypically developing children, the task used most frequently to assess awareness of mental states in others and self has been the "Smarties" (unexpected contents) false belief task (Hogrefe, Wimmer, & Perner, 1986). In this task, participants are shown a Smarties tube and 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 and is then asked (a) what they thought was inside the tube before they looked (Self test question) and (b) what another person, who has not seen the actual contents of the tube, would believe was inside (Other-person test question). Success on tasks like this is often seen as the 'acid-test' of ToM competence because it requires an understanding of behaviour in terms of an individual's mis-representation of reality, as opposed to reality itself.[1] Typically developing children begin to pass these tasks consistently from approximately four years of age and show parallel performance across the Self and Other-person test questions of the Smarties task (Wellman, Cross, & Watson, 2001).

In those (relatively few) studies that have employed the Smarties task and analysed the performance of children with autism on both the Self and Other-person test questions, some striking results have been reported. In keeping with the argument that self-awareness of mental states in *intact* amongst people with autism, two early studies found that children with this disorder found the Self test question significantly easier than the Other-person test question (Leslie & Thaiss, 1992; Perner et al., 1989). In each of these studies, participants with autism were impaired at recognising another's false belief, despite passing the Self test question at similar rates to comparison participants.

If reliable, these results would present a significant challenge to Frith and Happé's (1999) suggestion that children with autism have an impaired theory of own mind. However, in both Perner et al. (1989) and Leslie and Thaiss (1992), participants were asked, in the Self test question, what they had previously *said* was in the Smarties tube, rather than what they had *thought* was inside. It may be that individuals with autism, unlike typically developing children, do not automatically associate what a person (including themselves) *says* with what they *believe*. In other words, children with autism may have been quite able to remember what they had said was in the Smarties tube, especially considering their statement was made only approximately 30 seconds prior to the Self test question being asked. Whether they represented their (false) *belief* about the contents of the tube is a different matter. By directing children with autism to their earlier statement, Perner et al., and Leslie and Thaiss may have inadvertently confounded their results. Indeed, several subsequent studies have observed deficits on the Smarties Self test question amongst children with autism when the question has been phrased in terms of the participant's previous *thought* about what was inside the box, rather than their previous statement (Baron-Cohen, 1991, 1992; Fisher, Happé, & Dunn, 2005; Russell & Hill, 2001; Williams & Happé, 2009a).

These latter studies confirm that children with autism are impaired at recognising their own as well as others' false beliefs, supporting Frith and Happé's (1999) contention that self-awareness of mental states is diminished in autism. However, even these latter studies may over-estimate the extent to which children

with autism can represent their own false beliefs, given that in each study participants with autism were asked to explicitly state their (false) belief prior to being asked the Self test question. Williams & Happé (2009a) removed this potential confound by pretending to have cut a finger and asking participants to fetch them a plaster (band-aid). Participants were confronted by three different boxes, only one of which was a plasters box (which actually contained birthday cake candles). By choosing to open the plasters box (thereby discovering the candles), rather than either of the other containers, participants clearly demonstrated their (false) belief that the plasters box contained plasters. Having not verbalised this belief, however, success on the subsequent Self test question could not reflect any kind of compensatory 'memory for statement' strategy, but rather self-awareness of their false belief. Results from this experiment were that children with autism were unique as a group in finding the Self test question from this task significantly *more* difficult than the Other-person test question. Amongst both young (3- to 5-year-old) typically developing and learning disabled (intellectually impaired) comparison participants, performance on the Self test question was equivalent to performance on the Other-person test question. At least with respect to false beliefs, therefore, children with autism appear at least as impaired, and probably more impaired, at recognising their own mental states as at recognising mental states in others (see below for further discussion).

Recognising own intentions

Prior to 2009, only two studies had specifically explored awareness of own intentions amongst children with autism. In a study by Phillips, Baron-Cohen, and Rutter (1998), children with autism and well-matched comparison participants were engaged in a rigged shooting game, in which they shot at an intended target (some of which contained a prize) using an electronic gun. After each turn, participants were asked whether or not the fallen target was the one they intended to hit. Phillips et al. found that comparison participants were able to correctly report their prior intention to hit a particular target, regardless of whether their intention had been fulfilled and/or whether the fallen target yielded a prize or not. In contrast, participants with autism tended to report a mis-hit as intentional when it resulted in a prize, but as unintentional when it did not. This suggests that children with autism confuse desires and intentions. When a mistake (hitting the wrong target) resulted in a desire (for a prize) being satisfied, then the action was judged as intentional. When a mistake did not satisfy such a desire, however, then the action was judged as unintentional. As such, Phillips et al.'s results suggest that children with autism are impaired at representing their own intentions.

However, Russell and Hill (2001, Experiment 2) failed to replicate Phillips et al. (1998). One potential reason for this is that different instructions were given to participants in each study. In the Phillips et al. study *children were asked to choose which target* they intended to hit. Children were then given a coloured card, corresponding to the colour of their chosen target, in order to "remind" them which target they had chosen. In Russell and Hill's study, *the experimenter chose the target* for the child, again providing a reminder in the form of a coloured card. By choosing the target for the child, however, the pragmatics of the task may have been altered in relevant ways. When participants were subsequently asked the test questions, involving what they had meant to hit, there may have been some confusion, given that

the intention to hit the designated target was as much a product of the experimenter's desire as the deliberate will of the participant.

In another experiment, Russell and Hill (2001, Experiment 3) gave participants with autism, as well as age- and ability-matched comparison participants, a "transparent intentions" task. In this task, participants were asked to complete a drawing on a transparency (like those used with overhead projectors) of, for example, a boy with a missing ear. Unknown to the child, a second transparency with a different, unfinished drawing (e.g., of a cup with a missing handle) was laid in precise alignment on top of the first transparency (with the drawing of the boy), which they intended to complete. Therefore, the child ended up unintentionally finishing off the top drawing (i.e., drawing a handle on a cup) rather than the bottom drawing. When their mistake was revealed, children were asked both what they had *meant* to draw and what they had *thought* they were drawing during their action. Russell et al. also included an Other-person condition in which children observed a glove puppet performing the same actions and making the same errors.

Results from Russell and Hill (2001, Experiment 3) indicated that, in all but one condition, children with autism performed as well as comparison participants on the transparent intentions task. This led Russell and Hill to conclude that, overall, children with autism were not significantly impaired at recognising their own intentions (or others') intentions. However, there are several potential concerns about the design of their experiment, which suggest that the results should be interpreted cautiously. For example, Russell and Hill adopted a repeated measures design in which the Self and Other conditions of the task were counterbalanced across participants. This somewhat counterintuitive design resulted in some children receiving the Other-person condition first. When subsequently undertaking the Self condition of the task, these children must have approached the procedure in the knowledge that they were being tricked, because they had already observed the crucial experimental manipulation (that there were really two transparencies even though there only appeared to be one) in the Other-person condition. The whole nature of the task appears to be quite different for children under these conditions and could well have confounded results (for a discussion of other potential confounds, see Williams & Happé, 2009b). Indeed, when Williams and Happé (2009b, Experiment 2) eliminated these potential confounds, participants with autism were significantly less likely than age- and ability-matched comparison participants to correctly report mistaken actions (by self or others) as unintended. Indeed, the poor performance of children with autism on the transparent intentions task was significantly associated with their comparable difficulties on a series of false belief tasks (although not independently of verbal intelligence).

Finally, Williams & Happé (2009b, Experiment 1) assessed the extent to which children with autism represent their own intentions, using a different methodology. Participants with and without autism were given a 'knee-jerk task', in which a knee reflex was elicited in the participant, who was then asked if they had intended to move their leg. Perner (1991) speculated that to correctly recognise one's own reflexes as unintentional, one requires a (meta-representational) ToM. In support of this, Lang and Perner (2002) found that the performance of young typically developing children on a false belief task was highly correlated with performance on the knee-jerk task, even after the influences of age and verbal ability were controlled. In their study, Williams and Happé found that children with autism performed significantly less well than age- and ability-matched comparison participants on this knee-jerk task. Relative

to these learning disabled comparison participants, the children with autism were much more likely to report incorrectly that their reflex movement had been under their intentional control. Furthermore, their performance on this task was associated significantly with their performance on false belief tasks, independent of the effects of age and verbal ability. These results support the view that individuals with autism have diminished awareness of their own intentions and that this diminution is directly related to their impairments in recognising others' mental states.

Recognising own knowledge/ignorance

Although several studies have explored the extent to which children with autism represent others' states of knowledge (see Lind & Bowler, in press), only two studies have explored awareness of own knowledge. In a study by Perner et al. (1989), each participant was paired with a confederate and shown a series of boxes. The experimenter informed both parties that each box contained a different object. Then, for each box in turn, the experimenter allowed either the participant or the confederate look inside. After one of the two parties had been shown the contents of the box, the participant was asked the critical test questions regarding whether a) they knew what was inside the box (Self test question) and b) the confederate knew what was inside (Other-person test question). Of course, the correct answer to each of these questions depended on who had been given visual access to the contents. If only the participant had seen inside, then they would know the contents, but the confederate would be ignorant. Conversely, if only the confederate had seen inside, then they would know the contents, but the participant would be ignorant.

Perner et al. (1989) found that participants with autism were significantly impaired on each of these test questions. Unlike comparison participants who were quite accurate at judging their own and the confederate's state of knowledge or ignorance, children with autism tended to over-estimate the knowledge of both self and other, answering that a person knew what was inside the box even though they had not had visual access to the contents.

Unlike Perner et al. (1989), Kazak, Collis, and Lewis (1997) did not find significant differences between participants with and without autism on a task assessing understanding of the distinction between knowing and guessing. However, this was mainly due to the poor performance of comparison participants, rather than the successful performance of the participants with autism. Both participants with Down's syndrome and typically developing 4-year-old participants performed at levels significantly below chance. Such poor performance is out of keeping with their typically good performance on standard tests of ToM, including those assessing awareness of states of knowledge (e.g., Baron-Cohen & Goodhart, 1994). As such, it is debateable whether the task used by Kazak et al. was suitable for assessing this kind of understanding.

Autobiographical accounts: a controversial indicator of theory of own mind

In their discussion of self-consciousness/theory of own mind in autism, Frith and Happé (1999) relied heavily on the first-person accounts of high-functioning adults with autism. The intelligent and eloquent autobiographies provided by these individuals have undoubtedly given us a unique insight into 'what it is like' to have

autism. Frith and Happé used these reports to highlight what, to them, seemed to indicate an atypical form of inner experience amongst people with autism. However, these self-reports have been used by others as evidence that self-awareness of mental states is *intact* in autism. Nichols and Stich (2003, p.187) give three examples of autobiographical comments from adults with Asperger syndrome, of which two are reproduced below:

> "I didn't talk until I was almost five, you know. Before I started talking I noticed a lot of things, and now when I tell my mother she is amazed I remember them…" (reported in Dewey, 1991, p.204).
>
> "… One day my mother wanted me to wear a hat when we were in the car. I logically thought to myself that the only way I could tell her that I did not want to wear the hat was to scream and throw it on the car floor" (Grandin, 1984, p.145).

In response to these descriptions, Nichols and Stich (2003, p.187) ague that "if these recollections are accurate, then these individuals must have been aware of their own mental states even though, at the time in question, they could not reliably attribute beliefs to others." Similarly, McGeer (2004, p.239) argues that, "it seems downright implausible to suggest that any group of people who can write as vividly and consistently about their own sensory experiences could really be missing a mechanism that purportedly makes such experiences available to them as experiencing subjects."

So, do the autobiographies of really show that individuals with autism have a good grasp of their own mental states? Well, there are several issues to be borne in mind when considering this issue. Firstly, the impairment in reading others' minds amongst individuals with autism is not taken by any researcher in the field as being an all-or-nothing deficit. Certainly, these autobiographers have some understanding of the minds of other people, even if this understanding is diminished relative to their typical peers, and/or if this understanding has been acquired through an atypical developmental process and is based on compensatory mechanisms (e.g., Hermelin & O'Connor, 1985). Therefore, the fact that high-functioning adults with autism can report on their own mental states at all is not a challenge for the theorist who maintains that recognising one's own mental states relies on the same mechanism/process as reading others' minds.

The second and more controversial issue concerns the fact that self-reports offered by individuals with autism only challenge the notion that this disorder involves a diminished theory of own mind *if those reports are accurate* (Happé, 1991). When considering self-reported memories, it is essential to distinguish between those memories which are *episodically* remembered and those which are *semantically* known. Episodic memory is associated with consciously remembering *personally-experienced* events, whereas semantic memory is concerned with *factual* information. Caution must be exercised when attributing to people with autism memory processes of the *episodic* kind. I may *know* that a particular event has happened to me in the past, and hence report details of the event quite accurately, without actually *remembering* the event. Only this latter kind of 'remembering' is thought to involve the kind of self-awareness, or 'autonoetic' (self-knowing) consciousness, at issue here (e.g., Perner, 2000).

In fact, there is substantial evidence to indicate a selective diminution in episodic remembering and increased reliance on semantic knowing amongst individuals with autism, including high-functioning adults (e.g., Bowler, Gardiner, & Gaigg, 2007; Lind & Bowler, in press). Perhaps this selective difficulty explains a remarkable finding from a study of self-understanding amongst children and adolescents with autism, by Lee and Hobson (1998). Lee and Hobson found that a substantial proportion of the participants with autism, but no comparison child, spontaneously described 'recollections' of their own birth. Importantly, Hobson (personal communication) reports that these recollections were as vivid and descriptive as their other personal recollections. The phenomenon of childhood amnesia, which concerns the fact that the vast majority of typical adults do not recall events which happened prior their fourth birthday, suggests that any memory of one's birth (or of early life, *per se*) is most likely the result of confabulation.

On the other hand, Lyons and Fitzgerald (2005) argue that early autobiographical episodic memories reported by some people with autism are accurate and should be considered a savant talent. They suggest that a strong bias toward local perceptual/cognitive processing, as predicted by weak central coherence theory (Happé & Frith, 2006), and associated neural under-connectivity (Just et al., 2004), could support precociously early memory development. There are several reasons to be suspicious of Lyon and Fitzgerald's suggestion, however. Schacter and Tulving (1994, p.28, emphasis added) provide a classic definition of episodic remembering as consisting of "multifeature representations in which *numerous kinds of information* – spatial, temporal, contextual, and so forth – are *bound together* with the individual's awareness of personal experiences in subjective time." It is counter-intuitive to suggest that weak central coherence/functional under-connectivity would enhance one's ability to bind together such features into a coherent representation. Quite the opposite, a difficulty with binding together diverse elements of events could plausibly account for the widely acknowledged deficit in episodic memory amongst individuals with autism (Bowler, Gaigg, & Lind, in press)

The point here is that, without empirical confirmation, we need to be cautious about self-reported memories of early mental state understanding made by individuals with a disorder that is known to be associated with impairments in theory of mind *and* episodic memory. In this regard, Bruck, London, Landa, and Goodman (2007) assessed the ability of children with and without ASD to recall events from their lives (with reliability of reports being established through parental report)One important study, which may bear upon this issue, was conducted by Bruck, London, Landa, and Goodman (2007). In this study, children with and without autism were asked to recall events from their life (with reliability of reports being established through parental report) and also details of a staged event that the children had witnessed several days previously. Bruck et al. found that, for each event that participants were asked about, children with autism produced significantly fewer utterances that were confirmed by their parents, indicating reduced levels of episodic remembering. However, in general, children with autism did not produce significantly more utterances that were "unconfirmed" by or "inconsistent" with parental report. This led Brick et al. to conclude that children with autism were no more likely than children without autism to produce confabulated memories.

In response to this conclusion, however, two points need to be made. Firstly, even if Bruck et al.'s (2007) findings are accurate, this would not show that individuals

with autism can accurately recall the *mental states* that they experienced early in life. Secondly, there were, in fact, several indicators that children with autism were more prone to confabulation than comparison participants. Hence, children with autism were significantly more likely than comparison to incorrectly report that they had experienced a "silly event", such as "help[ing] a lady find a monkey in the park" (p.79). Indeed, 31% of participants with autism, but only 8% of comparison participants falsely claimed to have experienced at least one such an event. Similarly, participants with autism were significantly more likely than comparison participants to falsely report details from the staged event that not actually occurred. Bruck et al. interpret the latter finding in terms of reduced episodic memory in autism, such that participants with autism "could not consistently remember which events had occurred and which ones had not" (p.90). Therefore, Bruck et al. seem to acknowledge the possibility that, in at least some circumstances, a reduction in episodic memory could result in increased levels of confabulation.

The most important thing to note from all of this is that we should be cautious about any individual's memories that include details of skills or capacities that the individual would not have possessed at the time of the original event, even if they possess those skills at the time of 'remembering'/reporting. As such, the reports by individuals with autism of early self-awareness of mental states should not be considered *prima facie* evidence against the idea that autism involves a diminished theory of own mind.

Awareness of the Bodily Self

Recognising one's own body

For many, the litmus test of bodily self-awareness is the mirror self recognition task (e.g., Amsterdam, 1972). In the classic procedure, the experimenter surreptitiously places coloured, odourless pigment on the nose of the participant. The individual is then shown their reflection in a mirror, and their reactions to the image are assessed. Touching the nose is usually used as the criteria for ascribing objective, reflexive awareness of one's own body. The reasoning here is that the individual must possess knowledge both of how their face typically looks and of how this differs from their current appearance – they must have a 'body schema or own body representation that is mapped onto what is seen in the mirror' (Rochat, 2003, p.726). Mirror self-recognition has been used to assess objective self-awareness (of own body) in animals (e.g., Gallup, 1982), and in human infants, where it is first observed between the ages of 15 and 24 months of age (e.g., Lewis & Brooks-Gunn, 1979).

According to Lewis (2003), mirror self-recognition has the same representational underpinnings as required for awareness of the psychological self. According to Lewis, mirror self-recognition demonstrates a level of self-awareness that "involves explicit consciousness and self-referential behaviour. It is based on the mental state of me and allows for the capacity to reflect on one's self and to reflect on what one *knows*. This mental state is a *meta*representation." (p.118, emphasis added) In support of this suggestion, Lewis points to findings by Lewis and Ramsey (2004) that show mirror self-recognition to be closely associated with personal pronoun use and pretend play amongst young typically developing children.

For various reasons, both personal pronoun use and pretend play have been taken as indicators of awareness of the psychological self. According to Bruner (1975), the appropriate use of personal pronouns indicates awareness of the respective roles that each partner has in discourse settings. Because pronouns, unlike proper names, constantly shift according to who is speaking – I am 'me' to me and you are 'you' to me, but I am 'you' to you and you are 'me' to you – only 'non-egocentric' who are aware of themselves *in relation to* others individuals can use them appropriately (Bates, 1980). According to Leslie (1987), pretend play involves an awareness of one's own mental state of pretending (e.g., "this banana is a telephone") as separate ('decoupled') from one's representation of reality ("this banana is a banana"). Leslie argues that such awareness is necessary to avoid the problem of representational abuse, whereby children could become confused when the use of object does not conform to the object's literal meaning.

Although it is not at all clear why recognising one's *body* in a mirror should constitute evidence for the additional recognition of one's *mental* states, Lewis' (2003) theory is directly testable, particularly by exploring the case of self-awareness in autism. The extensive evidence, described above, that children with autism have diminished awareness of their own mental states should lead to the prediction that they will similarly show diminished awareness of their own physical selves, as demonstrated by a deficit in mirror self-recognition. To date, four studies have explored mirror self-recognition amongst children with autism (Dawson & McKissick, 1984; Ferrari & Matthews, 1983; Neuman & Hill, 1978; Spiker & Ricks, 1984). These studies have yielded a consistent pattern of results, with an average of 74% of children with autism successfully recognising their own mirror image, across the four studies.

The difficulty with drawing absolute conclusions from these studies is that none included age- and mental age-matched comparison children with whom the recognition performance of the children with autism could be compared. Nonetheless, from the available evidence it can be reasonably assumed (although not categorically concluded) that mirror self-recognition is present amongst children with autism who reach an appropriate developmental level. Hence, those children with autism in Ferrari and Matthew's study who failed the mirror self-recognition task had an average mental age of 22 months, which is within the developmental window when typically developing children fail the task (i.e., up to 24 months). This led Ferrari and Matthews (p.322) to conclude that "even when autistic children fail to recognise their self-images, this failure cannot be taken as evidence for a syndrome-specific deficit, but instead is a reflection of a general developmental delay". Supporting this conclusion, Dawson and McKissick found that the only two (of 15) children with autism who failed mirror self-recognition in their study, were also the only two children who failed to display stage V/VI object permanence. From this, they argued that "when an impairment in self-recognition is exhibited by an autistic child, it is a function of global mental retardation, not a specific feature of autism." (p.392)

What is so striking about the results of each of these studies is how successful mirror self-recognition stands in direct contrast to the severe social and intellectual deficits observed in the children with autism. The fact that severely learning disabled children with autism are capable of mirror self-recognition (*at least* from a mental age of 38 months, which was the average mental age of the self-recognisers in Ferrari and Matthew's 1983 study) is remarkable, and out of keeping with their severe ToM difficulties.

Two studies have also assessed children with autism on an extension of the mirror self-recognition paradigm. In the typical version of the delayed video self-recognition paradigm, the experimenter is filmed surreptitiously placing a large sticker on the participant's head, during a distractor task. The sticker remains on the participant's head for a period of three minutes, after which time the participant views the original video recording of the sticker placement. Reaching up to remove the sticker from one's head after viewing this recording is taken to indicate the possession of a temporally extended self-representation. The logic here is that only if the participant recognises that the individual in the recording of this earlier event is the same individual watching the recording in the present will they recognise that the sticker is on their head *here and now*, and hence reach up to remove it. In typical development, this task is passed from around 4 years of age (Povinelli, Landau, & Perilloux, 1996). However, highlighting that this task measures awareness of one's temporally extended physical self and not one's extended *thinking*, success on this task is not significantly associated with performance on traditional ToM tasks (Suddendorf, 1999; Zelazo, Sommerville, & Nichols, 1999).

Lind and Bowler (2009) found that, amongst participants who displayed live video self-recognition (which 95% of all participants did), 93% (n = 25) of participants with autism and 100% (n = 30) of age- and ability-matched comparison participants passed the delayed video self-recognition task. These results are also reported by Nielsen, Suddendorf, & Dissanayake (2006) who used the same task with high-functioning children with autism. It should be noted that the average mental age of the participants with autism in Lind and Bowler's study was 6.18 years, which is some two years above the chronological age at which typically developing children pass the delayed self-recognition task. Hence, like with studies of mirror self-recognition, absolute conclusions about whether delayed video self-recognition is impaired amongst children with autism are not possible to draw. However, two important points need to be made, in this regard. Firstly, Lind (personal communication) has confirmed that amongst the six children with autism from Lind and Bowler's study who had verbal mental ages below 4.5 years (and who passed live video self-recognition), five passed the delayed video self-recognition task. Hence, the vast majority of children with autism succeed on this task by the time they reach the developmental level at which typically developing children recognise their delayed video image. Secondly, what is particularly striking about Lind and Bowler's findings is that successful performance amongst the children with autism on the delayed video task stood in contrast to both their poor performance on an unexpected contents false belief task and to their significantly reduced propensity to use personal pronouns ('me') when labelling their video image. Therefore, contrary to Lewis' (2003) theory that awareness of the physical self and awareness of the psychological self each depend on the same underlying representational system, these findings suggest that each is underpinned by its own dedicated system, only one of which is impaired in autism. Individuals with autism appear to possess a coherent representation of their own bodies (even across time), but nonetheless fail to recognise aspects of their psychological selves.

Recognising own agency

Another important aspect of physical self-awareness concerns the ability to monitor one's own actions. Russell and Hill (2001, p.317) define action monitoring as, "the mechanisms that ensure that agents know, without self-observation, (a) for which changes in perceptual input they are responsible and (b) what they are currently engaged in doing". Effective action monitoring therefore allows an individual to distinguish between 'self-caused' and 'world-caused' changes in experience and hence, in Russell's (1996) theory, gives rise to an experience of *agency*. This process is widely thought to involve generating a visual ('efference') copy of a motor intention for action and then monitoring this copy in relation to the sensory consequences of one's action (e.g., Wolpert, Ghahramani, & Jordan, 1995).

Recently, Williams and Happé (2009c) explored action-monitoring amongst children with autism, using a task based on that used by Russell and Hill (2001) who found no significant differences between children with and without autism. In this task, participants needed to judge which one of several coloured squares on a computer screen was under their intentional control (through movements of the mouse) and which 'distractor' squares were under the control of the computer. Hence, every time participants moved the mouse, all of the squares began moving, but only one of the squares moved *directly* in accordance with the participant's movements. All of the other squares moved randomly, being controlled directly by the computer. So as to ensure that participants could not solve this task through self-observation (i.e., by visually comparing the movements of their hand with the movements of the squares on the screen), their hand was located inside a box while they were moving the mouse. Therefore, recognising which square the participant was responsible for moving required them to monitor their own efference copy and compare this to the visual input from the computer screen (i.e., to detect their own agency).

This aspect of the task was similar to the task employed by Russell and Hill (2001), except that Williams and Happé's (2009c) version was much more incremental in terms of difficulty (i.e., the number of distractor squares and the degree to which the target square could move relative to the distractor squares), so as to avoid both floor and ceiling levels of performance. In Russell and Hill's study, the majority of participants performed at either floor or ceiling levels, leaving such little variation in performance that a difference between the groups in action-monitoring ability would have been difficult to detect even if it existed.

There was another way in which the study design employed by Williams and Happé (2009c) critically differed from the design implemented by Russell and Hill (2001): In Williams and Happe's study, participants also completed an Other-person condition in which their hand was located on a mouse, as in the Self condition, but this time the experimenter (having placed his hand in the other end of the box) actually moved the mouse. In this condition, there is no efference copy of a motor intention for action and so, for an individual with an intact sense of their own agency, this condition should be significantly more challenging than the Self condition.

In keeping with the idea that individuals with autism have a typical awareness of their own physical selves, Williams and Happé (2009c) found children with autism performed almost identically to age- and ability-matched comparison participants in terms of levels and patterns of performance. Participants showed neither floor nor ceiling effects on the task and found the Other-person condition significantly more difficult than the Self condition. Unpublished data from Williams and Happé's study

also confirms that the participants with autism showed impaired ToM. Relative to comparison participants, significantly more participants with autism failed a composite measure of false belief understanding (comprising two false belief tasks), $\chi^2(1, N = 32) = 4.57$, $p = .03$, $\phi = .38$. Interestingly, participants in whom ToM was impaired performed somewhat better (though non-significantly) on the Self condition of the experimental task than participants who passed the ToM tasks. Whereas ToM passers succeeded on an average of 10.96 (*SD* = 7.79) trials in the Self condition, ToM failers managed an average of 14.00 (*SD* = 10.66) successful trials. Although this difference did not approach statistical significance, it highlights that difficulties with ToM do not have a negative impact on one's ability to monitor one's own actions/agency (see also David et al., 2008).

The results are also in keeping with those from an early study by Frith and Hermelin (1969; but see Hermelin & O'Connor, 1975) who found that children with autism tended to make *better* use of motor feedback (i.e., monitoring an efference copy), and rely *less* on visual feedback, than TD or learning disabled comparison participants. In Frith and Hermelin's study, participants with and without autism were equally fast at moving a stylus along the grooves of a curved 'track' which had been cut into Perspex. However, when visual access to the track was prevented in a second condition, participants with autism completed the task significantly *faster* than comparison participants. From this finding, Frith and Hermelin (p.162) concluded that the strategies used by participants with autism 'were those based on immediate feedback from kinaesthetic or motor cues', consistent with enhanced, rather than diminished, action-monitoring abilities.

Theoretical Issues and Implications

The above evidence lends substantial support to the idea that whilst individuals with autism have difficulty in reflecting on their own psychological selves, their awareness (both objective/reflexive and subjective/non-reflexive) of their physical selves remains remarkably unimpaired. A number of important theoretical considerations follow from such a suggestion and from the findings on which it is based.

The relation between physical and psychological self-awareness

Overall, the pattern of findings described above provides strong support for a taxonomy of self-awareness that clearly distinguishes between physical and psychological aspects of self (e.g., Gillihan & Farah, 2005). Developmentally speaking, the case of autism seems also to demonstrate that one can have a diminished awareness of the psychological self without any obvious implications for awareness of the physical self. Therefore, contrary to the view expressed by Lewis and Ramsey (2004), for example, reflexive awareness of the physical self does not depend on the kind of meta-representation required for reflexive awareness of one's own mind. Vice versa, awareness of the physical self is not *sufficient* for the development of awareness of the psychological self, although such awareness may be *necessary* (cf. Russell, 1996). In an important sense, therefore, the development of these aspects of self-awareness appears *domain-specific*.

The relation between theory of own mind and theory of others' mind

Another important implication from the above findings is that becoming aware of one's own psychological states (at least, one's own propositional attitudes) appears to depend on the same underlying functional system/process as that employed when one is aware of others' psychological states. This is contrary to a prominent alternative approach to characterising self-other awareness in autism, which states that first-person *access* to mental states is intact, but that the process of adopting one's mental states in a *simulation* of another person's mental states is impaired (e.g., Goldman, 2006). If this were true, people with autism should be perfectly able to report their own mental states, despite showing impairment in recognising others' mental states. As Goldman (p.223) highlights, "no simulation theorist claims that simulation itself is used in first-person attribution…" As such, it seems the best explanation of the particular self-awareness (and other-awareness) deficits seen amongst individuals with autism implicates a single theory of mind mechanism, which is responsible for detecting mental states in both self and others, and which is impaired amongst people with this disorder (cf. Carruthers, 2009).

One key finding by Williams and Happé (2009a) arguably provides important information about how this mechanism (mal-) functions in ASD. As discussed above, Williams and Happé found that children with ASD were unique as a group in finding it significantly more difficult to detect and report their own false belief than to detect and report a false belief held by another person. At least in some circumstances, then, self-awareness of mental states is *more* impaired amongst children with ASD than is awareness of mental states in others. This is clearly contrary to simulationist explanations of self-awareness in ASD, but also appears somewhat out of keeping with the view that a concept of belief (underpinned by a single theory of mind mechanism) is acquired for self and others in parallel.

In attempting to reconcile this latter view with their intriguing finding, Williams & Happé (2009a) suggested that many individuals with ASD employ alternative/atypical routes to mental state understanding in order to compensate for an impaired theory of mind mechanism. More specifically, a limited competence in recognising mental states, amongst most (or all) individuals with ASD, is supplemented by the application of rule-bound, cognitively acquired heuristics that are used to "solve" theory of mind problems (in both laboratory and real-world settings) (cf. Hermelin & O'Connor, 1985). Now, the available information, to which such compensatory heuristics can be applied, clearly differs in the cases of self and others. For example, one visually perceives the actions of other people and can attribute mental states on the basis of these perceptions. Yet, one rarely visually perceives one's own actions (unless looking at oneself in a mirror or on a video recording). Conversely, one has direct access to one's own somatosensory experiences and inner imagery, for example, but never has access to these in other people. Williams and Happé argued that when the theory of mind mechanism is damaged (as it arguably is in ASD), repeated perception of regularities in others' actions provides an opportunity to learn "behaviour rules" (cf. Povinelli & Vonk, 2004), and thereby attribute mental states to others via compensatory heuristics. Such opportunities are less frequent in the case of one's own behaviour, which one rarely visually perceives. As a result, self-awareness of mental states is most at risk when the theory of mind mechanism, itself, is damaged (see Carruthers, 2009, pp.53-54 for a similar argument).

Regardless of the validity of these hypotheses, the studies described above provide a clear indication that self-awareness of mental states is impaired amongst individuals with ASD. There may, in fact, be other aspects of self-awareness that are also impaired in this disorder (see Hobson, 1990; Loveland, 1993) and these may interact with deficits in theory of own mind in important ways. The puzzle of self-awareness in ASD is no doubt a highly complex one, but a puzzle that, with enough good research and theorising, might just be successfully solved.

References

Amsterdam, B. (1972). Mirror self-image reactions before age two. *Developmental Psychobiology, 5,* 297-305.

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.

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

Bowler, D.M., Gaigg, S., & Lind, S. (in press). Memory in autism: Binding, self and brain. In I.Roth & P. Rezaie (Eds.). *The autistic spectrum: Research reviews.* Milton Keyes: Open University Press.

Bowler, D. M., Gardiner, J. M., & Gaigg, S. B. (2007). Factors affecting conscious awareness in the recollective experience of adults with Asperger's syndrome. *Consciousness and Cognition, 16,* 124-143.

Bremner, A.J., & Mareschal, D. (2004). Reasoning... what reasoning? *Developmental Science*, **7**, 419- 421.

Bruck, M., London, K., Landa, R., & Goodman, J. (2007). Autobiographical memory and suggestibility in children with autism spectrum disorder. *Development and Psychopathology, 19*, 73-95.

Bruner, J. (1975). The ontogenesis of speech acts. *Journal of Child Language, 2*, 1-19.

Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences, 32*, 1-18.

David, N., Gawronski, A., Santos, N.S., Huff, W., Lehnardt, F., Newen, A., & 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*, 593-605.

Dawson, G., & McKissick, F. C. (1984). Self-recognition in autistic children. *Journal of Autism and Developmental Disorders, 14,* 383-394.

Ferrari, M. & Matthews, W. S. (1983). Self-recognition deficits in autism: Syndrome-specific or general developmental delay? *Journal of Autism and Developmental Disorders, 13,* 317-324.

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.

Flavell, J.H., Green, F.L., Flavell, E.R. & Grosmann, J.B. (1997). The development of children's knowledge about inner speech. *Child Development, 68*, 39-47.

Frith, U. (2003). *Autism: Explaining the enigma*. (2nd ed.) Oxford: Blackwell.

Frith, U., & Happe, F. (1999). Theory of mind and self-consciousness: What is it like to be autistic? *Mind and Language, 14,* 1-22.

Frith, U., & Hermelin, B. (1969). The role of visual and motor cues for normal, subnormal and autistic children. *Journal of Child Psychology and Psychiatry, 10*, 153-163.

Gallup, G. G. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology, 2*(3), 237-248.

Gillihan, S.J., & Farah, M.J. (2005). Is self special? A critical review of evidence from experimental psychology and cognitive neuroscience. *Psychological Bulletin, 131*, 76-97.

Goldman, A. (2006). *Simulating minds: the philosophy, psychology, and neuroscience of mindreading*. Oxford University Press.

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

Happé, F.G.E. (1991). The autobiographical writings of three Asperger syndrome adults: problems of interpretation and implications for theory. In U. Frith (Ed.), *Autism and Asperger syndrome* (pp. 207-242). Cambridge: Cambridge University Press.

Happé, F. G. E. & Frith, U. (2006). The weak coherence account: Detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 36,* 5-25.

Hermelin, B., & O'Connor, N. (1975). Location and distance estimates by blind and sighted children. *Quarterly Journal of Experimental Psychology, 27*, 295-301.

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. (1993). *Autism and the development of mind*. Psychology Press: Erlbaum (UK).

Hogrefe, G. J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57,* 567-582.

Hurlburt, R.T. (1990). *Sampling normal and schizophrenic Inner Experience*. Plenum Press: New York.

Hurlburt, R. T., Happé, F. & Frith, U. (1994). Sampling the form of inner experience in three adults with Asperger syndrome. *Psychological Medicine, 24,* 385-395.

Joseph, R.M., Steele, S.D., Meyer, E., & Tager-Flusberg, H. (2005). Self-ordered pointing in children with autism: Failure to use verbal mediation in the service of working memory? *Neuropsychologia, 43*, 1400-1411.

Just, M.A., Cherkassy, V.L., Keller, T.A., & Minshew, N.J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: Evidence of underconnectivity. *Brain, 127*, 1811-1821.

Kazak, S., Collis, G. M., & Lewis, V. (1997). Can young people with autism refer to knowledge states? Evidence from their understanding of 'know' and 'guess'. *Journal of Child Psychology and Psychiatry, 38,* 1001-1009.

Lang, B., & Perner, J. (2002). Understanding of intention and false belief and the development of self-control. *British Journal of Developmental Psychology, 20*, 67-76.

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

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

Lewis, M. (2003). The development of self-consciousness. In J. Roessler and N. Eilan (Ed.), *Agency and Self-Awareness* (pp.275-295). Oxford, UK: Oxford University Press.

Lewis, M., & Brooks-Gunn, J. (1979). *Social cognition and the acquisition of self.* New York: Plenum.

Lewis, M., & Ramsey, D. (2004). Development of self-recognition, personal pronoun use, and pretend play during the 2nd year. *Child Development, 75,* 1821-1831.

Lidstone, J.S.M., Fernyhough, C., Meins, E., & Whitehouse, A.J.O. (2009). Inner speech impairment in autism is associated with greater nonverbal than verbal skills. *Journal of Autism and Developmental Disorders*, in press.

Lind, S. E. & Bowler, D. M. (in press). Impaired performance on see-know tasks amongst children with autism: Evidence of specific difficulties with theory of mind or domain-general task factors? *Journal of Autism and Developmental Disorders*.

Lind, S. E. & Bowler, D. M. (2009). Delayed self-recognition in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 39,* 643-650.

Loveland, K. A. (1993). Autism, affordances, and the self. In U. Neisser (Ed.), The perceived self: Ecological and interpersonal sources of self-knowledge (pp.237-253). New York: Cambridge University Press.

Lyons, V., & Fitzgerald, M. (2005). Early memory and autism. *Journal of Autism and Developmental Disorders, 35,* 683.

McGeer, V. (2004). Autistic self-awareness. *Philosophy, Psychiatry and Psychology, 11*, 235-251.

Nichols, S. and Stich, S. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford University Press.

Neuman, C. J., & Hill, S. D. (1978). Self-recognition and stimulus preference in autistic children. *Developmental Psychobiology, 11,* 571-578.

Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology, 1,* 35-59.

Nielsen, M., Suddendorf, T., & Dissanayake, C. (2006). Imitation and self-recognition in autism. In S. Rogers & J. W. Williams (Eds.), *Imitation and the development of the social mind: Autism and typical development* (pp. 138–156). New York: Guilford Press.

Perner, J. (1991). *Understanding the Representational Mind.* Cambridge, MA: MIT Press.

Perner, J. (2000). Memory and theory of mind. In E. Tulving & F.I.M. Craik (Eds.), *The Oxford handbook of memory* (pp. 297-312). Oxford: Oxford University Press.

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.

Philips, W., Baron-Cohen, S., & Rutter, M. (1998). Understanding intention in normal development and autism. *British Journal of Developmental Psychology, 16,* 337-348.

Perner, J., & Ruffman, T., (2005). Infants' insight into the mind: How deep? *Science, 308*, 214-216.

Povinelli, D. J., Landau, K. R., & Perilloux, H. K. (1996). Self-recognition in young children using delayed versus live feedback: Evidence of a developmental asynchrony. *Child Development*, *67*, 1540–1554.

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

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

Raffman, D. (1999). What autism may tell us about self-awareness: a commentary on Frith and Happe. *Mind & Language, 14*(1), 23-31.

Rochat, P. (2003). Five levels of self-awareness as they unfold early in life. *Consciousness and Cognition, 12,* 717-731.

Russell, J. (1996). *Agency: Its role in mental development.* Hove, U.K.: Erlbaum (U.K.), Taylor and Francis.

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

Schacter, D.L., & Tulving, E. (1994). What are the memory systems of 1994? In D.L. Schacter & E. Tulving (Eds.). *Memory systems 1994* (pp.1-38). Cambridge, MA: MIT Press.

Suddendorf, T. (1999). Children's understanding of the relation between delayed video representation and current reality: A test for self-awareness? *Journal of Experimental Child Psychology, 72*, 157–176.

Vygotsky, L. S. (1987). Thinking and speech. *The collected works of Lev Vygotsky* (Vol. I). New York: Plenum (Original work published in 1934).

Williams, D. M. & 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

Williams, D. M. & Happé, F. (2009b). Representing intentions in self and others: Studies of autism and typical development. *Developmental Science,* in press.

Williams, D. M. & Happé, F. (2009c). Pre-Conceptual aspects of self-awareness in autism spectrum disorder: The case of action-monitoring. *Journal of Autism and Developmental Disorders, 39,* 251-259.

Williams, D., Happé, F., & Jarrold, C. (2008). Intact inner speech in children with autism spectrum disorders: evidence from a short-term memory task. *Journal of Child Psychology and Psychiatry*, *48*, 51-58.

Winsler, A., Abar, B., Feder, M. A., Schunn, C. D., & Rubio, D. A. (2007). Private speech and executive functioning among high-functioning children with autistic spectrum disorders. *Journal of Autism and Developmental Disorders, 37*, 1617–1635.

Whitehouse, A.J.O., Maybery, M.T., & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry, 47,* 857-866.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science, 269*, 1880-1882.

Zelazo, P. D., Sommerville, J. A., & Nichols, S. (1999). Age-related changes in children's use of external representations. *Developmental Psychology, 35*, 1059–1071.

Footnotes

1. It should be noted that use of the false belief task to assess ToM amongst typically developing children has been much maligned in recent times, with critics arguing both that success on these tasks involves skills over and above ToM, and that false belief tasks assess only a fraction of ToM skills (e.g., Bloom & German, 2000). A defence of the false belief task as a measure of an important aspect of ToM is beyond the scope of this paper (but see, for example, Bremner & Slater, 2004; Perner & Ruffman, 2005). The important thing to note, however, is that even the most ardent critics of the false belief task agree that it is an important tool for understanding ToM amongst children with developmental disorders (particularly autism), and that it can be used reliably in studies assessing "the factors that improve or diminish performance on the task" (Bloom & German, p.30).

Acknowledgements

This paper was prepared during a City University Research Fellowship awarded to the author.

I am very grateful to Dr Sophie Lind for kindly allowing me access to her data. I am also grateful to two anonymous reviewers for their helpful comments on a draft of the manuscript. Finally, I would like to thank Professor Peter Hobson for the many interesting discussions we have had about self-awareness in autism, which have influenced my thinking on the subject and, correspondingly, the content of this paper.