Theory of Mind and Self-Consciousness: What Is It Like to Be Autistic?

UTA FRITH AND FRANCESCA HAPPÉ

Abstract: Autism provides a model for exploring the nature of self-consciousness: self-consciousness requires the ability to reflect on mental states, and autism is a disorder with a specific impairment in the neurocognitive mechanism underlying this ability. Experimental studies of normal and abnormal development suggest that the abilities to attribute mental states to self and to others are closely related. Thus inability to pass standard ‘theory of mind’ tests, which refer to others’ false beliefs, may imply lack of self-consciousness. Individuals who persistently fail these tests may, in the extreme, be unable to reflect on their intentions or to anticipate their own actions. In contrast, individuals with high-functioning autism or Asperger syndrome often possess a late-acquired, explicit theory of mind, which appears to be the result of effortful learning. An experimental study with three people with Asperger syndrome suggested that level of performance on standard theory of mind tasks was strongly related to the ability to engage in introspection. Qualitative differences in the introspections of high-functioning people with autism are also reflected in autobiographical accounts which may give a glimpse of what it is like to be autistic.

1. Introduction and Overview

A widely accepted theory is that the core symptoms of autism are due to a deficient neuro-cognitive mechanism which underpins the normal ability to develop a ‘theory of mind’: the ability to attribute mental states and predict behaviour accordingly. The empirical basis of this theory comes from extensive tests of (false) belief attribution, which are simple for normally developing children but are failed by children with autism. Here we want to extend the idea of a lack of theory of other minds, which is the ability standardly tested, towards the notion of a lack of theory of own mind. Taken to its logical conclusion, the inability to ‘attribute mental states to self and others’, i.e. ‘theory of mind’ (Premack and Woodruff, 1978), is the same as not having introspective awareness. This is not a new idea, but still one that we have shied away from, since it seemed potentially pernicious to attribute

We would like to thank Diana Raffman, Martin Frith and Clare Sainsbury for their helpful comments on this paper.

Address for correspondence: Uta Frith, Institute of Cognitive Neuroscience, University College London, Alexandra House, 17 Queen Square, London WC1N 3AR, UK.

Email: u.frith@ucl.ac.uk.
impaired self-consciousness to those who are handicapped and not able to present their own side of things.

Fortunately, there are some talented individuals with autism who have been able to speak up, and some of their recollections of childhood experiences are surprisingly in line with this notion. They have provided spontaneous reports which may tell us something about the inner life of less able individuals with autism who never achieve this competence. Like other children with autism, and unlike normally developing children, these individuals very probably did not have an intuitive ‘theory of mind’ during their early years. However, there is no doubt that, as adults, these individuals can reflect on their own and on others’ mental states. We will present some examples which illustrate how candidly and lucidly they write about their inner life. We will also present some examples of experimentally induced introspections. We propose that both sources of data hint at a different kind of self-consciousness. These individuals appear to arrive at an explicit theory of other minds by a slow and painstaking learning process, just as they appear to arrive at self-consciousness through a long and tortuous route.

First, however, we need to present some general background facts about theory of mind in normal and abnormal development, and some facts about autism. On the way, we will touch briefly on current ideas about the neuro-cognitive basis of the ability to attribute mental states to self and others.

2. Theory of Mind in Normal Development

In the last ten years developmental psychology has been greatly affected by new ideas about the awareness of own and other minds. ‘Theory of mind’ refers to the ability to attribute independent mental states to self and others in order to predict and explain behaviour (Premack and Woodruff, 1978). This ability appears to be a prerequisite for normal social interaction: in everyday life we make sense of each other’s behaviour by appeal to a belief-desire psychology (see, for example, Astington, Harris and Olson, 1988; Byrne and Whiten, 1988; Carruthers and Smith, 1996). For instance, it is trivially easy to explain why John will carry his umbrella with him: it is because he believes it will rain and he wants to stay dry. Attribution of mental states is vital for everyday social interaction (e.g. cooperation, lying, keeping secrets). Theory of mind ability is also fundamental to communication and possibly also to the acquisition of language (Bloom, 1997; Frith and Happé, 1994): We make sense of utterances by reference to what the speaker intends to convey rather than what is literally said (Sperber and Wilson, 1986).

The cognitive processes which underlie the development of Theory of Mind (ToM) are still a matter of debate. The field can be divided into those who favour a more general explanation for ToM (e.g. simulation, general theory building), and those who argue for the necessity of a dedicated cognitive mechanism (for debate see, for example, Goldman, 1993; Gopnik, 1993; and chapters in Carruthers and Smith, 1996). Evidence in favour of a dedi-
cated, innately specified cognitive mechanism underlying ToM is strong (Leslie, 1987; Baron-Cohen, 1994). It includes the relative lack of normal individual difference or cross-cultural variation, the rapidity of acquisition in early childhood, and the case of autism, a biologically based disorder characterized by selective ToM impairment. The existence of a hard-wired proto-mechanism is, of course, not sufficient for a full belief-desire psychology to develop. Other cognitive processes, especially those concerned with inferences, will be involved. Furthermore, triggering input will be required, and at least some experience of social interactions will be necessary.

2.1 Tests of Theory of Mind

Despite the sometimes tortuous debate concerning the nature of ToM, paradigms for testing ToM are extremely simple. The litmus test for ToM has been the ability to attribute false beliefs to others, where prediction and explanation of action cannot be based simply on one’s own convictions or the state of the world. In other words, in order to pass the test, the particular beliefs held by another must be considered and held separate from one’s own knowledge. Because of its relatively high verbal and executive task demands, this test can only be given to children from age 3 or 4 onwards. This is not to say that below that age children are not implicitly aware of others’ mental states.

The two most frequently used false-belief tasks, both due to Wimmer and Perner (1983), are location change and content change tasks. In the Sally-Anne task, a location change task, Sally has a box and Anne has a basket. Sally puts her marble into her box. Then she goes out for a walk. While she is out, naughty Anne takes the marble from the box and puts it into her own basket. Now Sally comes back and wants to play with her marble. Where will she look for the marble—where does she think the marble is? The answer that seems obvious to a 4-year-old child is: Sally will look inside her box! This is where Sally thinks the marble is. Younger children have some difficulty with this task and often point to the basket, indicating that Sally will look where the marble really is.

In the Smarties task (Hogrefe, Wimmer and Perner, 1986), a content change task, the experimenter shows a well known sweet container, a tube, to the child and asks: ‘What is in here?’ The child answers ‘Smarties’, and this is exactly the answer an adult would give too. The experimenter reveals that the tube contains a small pencil. The child may well look disappointed indicating that he or she did indeed think that there would be sweets inside. The tube is closed again. Now the experimenter says: ‘Your friend John is going to come in now. He hasn’t seen this tube. When John comes in, I’ll show him this tube just like this and ask: ‘What’s in here?’ What will John say?’ The average 4-year-old will answer that John will say ‘Smarties’. After all, John has not seen the contents of the tube. Younger children have trouble with this task and claim that John will say ‘a pencil’. Furthermore, when
asked what they themselves at first thought was in the tube, children under 4 will typically assert: ‘a pencil’.

Children normally pass these tasks between 3 and 4 years and tend to pass many tasks of a similar kind (e.g. appearance-reality tests, deception) all at the same age (Gopnik and Astington, 1988). In everyday life, examples that involve understanding that another person may have a belief that is different from reality can be found at even earlier ages (Dunn et al., 1991). There are plenty of signs that even in infancy the young child is capable of tracking another’s intention (e.g. Gergely et al., 1995). Interactions involving teasing in children as young as 8 months may reflect an awareness of another person’s mistaken expectation (Reddy, 1991). At 18 months children show clear signs of sensitivity to others’ intentions; for example, they infer and imitate an intended incomplete action, but not an unintended action (Meltzoff, 1995). The early appearance of ToM and its presence even in learning-disabled individuals (other than those with autism) are all consistent with the notion of a computationally specialized processing mechanism.

To date there have been three functional imaging studies attempting to locate brain systems underlying the ToM mechanism (Baron-Cohen et al., 1994; Goel et al., 1995; Fletcher et al., 1995). Each of these has located frontal brain regions which are significantly more active during tasks considered to require ToM, than during control tasks. The different patterns of specific brain activation found were surprisingly circumscribed, considering the higher-order nature of the cognitive process targeted. This lends credence to the hypothesis that in normal development ToM does not emerge simply from general reasoning processes but involves some dedicated brain circuitry.

2.2 Theory of Own Mind and Theory of Other Minds

A theoretically important question for philosophers and psychologists is whether the same cognitive mechanism required for attributing thoughts and feelings to others is also necessary for attributing mental states to self. At first glance the two attributions seem entirely different: own mental states do not have to be inferred through observation like those of others, and they may be less likely to be erroneous. However, even though the input channels by which the relevant information is received may well be different, a crucial part of the process is to distinguish mental states, be they first-person or other people’s, from representations of the physical world. For example, it is necessary to distinguish the representation of physical reality (‘there is a pencil in the tube’) from the representation of belief (I thought ‘there are sweets in the tube’, or John thinks ‘there are sweets in the tube’). It seems plausible that the mechanism that keeps (second-order) representations of mental states separate from (first-order) representations of physical states is the same for self and other attribution (Leslie, 1987). Even if the appreciation of others’ mental states results in representations that are more error prone
than the representations of own mental states, this difference becomes trivial if one is unable to represent mental states at all.

As far as we can see, there is little evidence from the developmental literature to suggest that mental states are attributed to self before they are attributed to others. For instance, children do not systematically pass the self-question in the Smarties test (‘What did you think was in the tube?’) before passing the other’s belief question (‘What will John think . . .?’). Relevant studies are summarized by Gopnik and Meltzoff (1994), who conclude that when children are able to report their own mental states they are also able to report the mental states of others. Conversely, when they cannot report and understand the psychological states of others, they do not report those states of themselves. If there is a common representational mechanism for attributing mental states to self and to others, then these findings make sense. It would follow that we can test for the presence of self-consciousness using standard ToM tasks. In other words, tests which tap children’s ability to attribute mental states to others may also tap their insight into their own beliefs, feelings and knowledge.

These arguments are in line with Leslie’s account (1987, 1988), the first and most explicit description of the computational underpinnings required for ToM. Leslie suggests that understanding of pretence in infancy demonstrates the availability of a special form of representation for mental states as mental states. Second-order or metarepresentations are kept separate from first-order or primary representations of real states of affairs. Hence, the child observing his mother playfully using a banana as a telephone does not get confused about the normal use of phones and bananas. Leslie specifies metarepresentation as a four-part relation between agent (mother), propositional attitude (pretends), target in the world (banana) and proposition (it is a telephone). A metarepresentation represents the attitude an agent takes to a description of a particular aspect of reality. In contrast, a primary representation describes a particular aspect of reality. Leslie postulates that metarepresentation is necessary for attribution of any mental state, including (false) belief, and is necessary equally for self and other attribution.

Whether or not Leslie’s general theory of ToM is accepted, it seems clear that underlying our social understanding must be representations which capture who is thinking what and in what sense they are thinking it. Are they believing it, desiring it, hoping it, fearing it? Without in some way tagging a representation with an agent marker, own and other’s beliefs would be confused. Without marking an attitude, beliefs, desires, and pretence would be confused. The implication for self-consciousness is that without metarepresentation one would not know one’s own propositional attitude to the information in one’s head. Could this be the case in autism?

3. Autism – A Disorder of ‘Theory of Mind’

It is arguable whether we would ever have thought of such a thing as a neurologically specified theory of mind (ToM) mechanism, let alone a cir-
cumscribed brain system underlying this mechanism, were it not for the fact that individuals with autism appear to lack the ability to attribute mental states. Autism is a developmental disorder with a genetic basis (Bailey, Phillips and Rutter, 1996) and a prevalence of 0.1 to 0.3% (Wing, 1993). It is diagnosed on the basis of early emerging qualitative abnormalities in social interaction, communication and imagination (with restricted interests and activities). One striking feature of young children with autism is the lack of pretend play (Wulff, 1985). It was this observation which originally suggested that, on Leslie’s account, metarepresentation and hence ToM might be impaired in autism. From this observation it was predicted that false-belief attribution might also be impaired, despite sufficient verbal and nonverbal ability to follow the Sally-Anne task (Baron-Cohen, Leslie and Frith, 1985).

There now exists ample experimental evidence that individuals with autism have difficulty in conceptualizing mental states, and thus fail to attribute (false) beliefs to others (Baron-Cohen, Tager-Flusberg and Cohen, 1993). This failure is extremely specific and cannot be reduced to a failure in more general cognitive processes. This is seen in a number of contrasts between assets and deficits in otherwise very similar behaviour, distinguished only by the necessity to attribute mental states. Thus, understanding of deception is impaired (where manipulation of another’s mental state is necessary), while understanding of sabotage (manipulation of physical means of access) is not (Sodian and Frith, 1992). Similar ‘fine cuts’ between intact and impaired behaviour have been reviewed by Frith and Happé (1994). There is also some preliminary evidence from brain imaging that the ToM-dedicated area in prefrontal cortex is not activated when people with autism attempt theory of mind tasks (Happé et al., 1996).

Consistent with our hypothesis that understanding of mental states relies on a dedicated cognitive system which can be selectively impaired, some individuals with autism have good general cognitive capacities (as reflected in high IQ scores) alongside their ToM deficits. On the other hand, non-autistic children with mental handicap perform relatively well on false-belief tasks. The notion that ToM is impaired in autism makes good clinical sense and explains the very specific pattern of impaired (e.g. keeping secrets) and intact abilities (e.g. keeping house rules) in everyday life (Frith, Happé and Siddons, 1994).

### 3.1 Asperger Syndrome—An Unusually Acquired Theory of Mind?

Apparently inconsistent with the hypothesis that autistic disorder is a consequence of a failure in the ToM mechanism, is the well documented existence of able individuals with autism who pass tests of mental state attribution. This minority within the autistic spectrum, who are now often given the label Asperger syndrome, also have higher social and communication abilities than those without ToM ability. Individuals with Asperger syndrome tend to develop better social understanding as they get older, when they
show clear evidence of passing ToM tests (Bowler, 1992). So far, no single case has been reported of a child with autism/Asperger syndrome who showed good ability to attribute mental states in standard tests at the early age that young children normally show such ability. It is possible that this group has a less severe impairment in the underlying neurocognitive mechanism, which may manifest itself as slowness and developmental delay in ToM.

There is reason to believe that the understanding of mental states developed by these individuals is rather different from the effortless automatic ToM of the normal preschooler. First, they require much higher verbal ability to pass ToM tasks than do normal children, and do so at later ages (typically in adolescent, not preschool, years). Second, even as adults they are prone to making tell-tale slips in mental state attribution (Happe´, 1994, 1995). For instance, they might mistake a joke for a lie, and often find it hard to distinguish, for example, sarcasm from outright deception. Third, their approach to social tasks has been said to resemble slow, conscious calculation. They appear to do better with written than spoken communication, where the fast to and fro of mental state appraisal is avoided. As mentioned above, a first brain imaging study presenting ToM tasks to such individuals supports the idea that different brain processes underlie the late developed social understanding in Asperger syndrome (Happe´ et al., 1996).

The evidence to date suggests that individuals with Asperger syndrome are part of the autism spectrum: for instance, the two conditions tend to be found in the same families, and one and the same individual may show a change in clinical picture from one to the other over time (Wing, 1997). We will therefore treat autism and Asperger syndrome as similar in some essential respects. In particular, we will consider whether individuals with Asperger syndrome can provide an exceptional window into the mind of the individual with autism.

4. Does Impaired Theory of Mind Imply Impaired Self-Consciousness?

While it has been easy for many people to accept that the devastating social and communicative handicaps of autism may result from a failure to attribute mental states to others, the notion that this mind-blindness might also apply to the child’s own mind has scarcely been acknowledged (but see Carruthers, 1996). We have argued above that if the mechanism which underlies the computation of mental states is dysfunctional, then self-knowledge is likely to be impaired just as is the knowledge of other minds. The logical extension of the ToM deficit account of autism is that individuals with autism may know as little about their own minds as about the minds of other people. This is not to say that these individuals lack mental states, but that in an important sense they are unable to reflect on their mental states. Simply put, they lack the cognitive machinery to represent their thoughts and feelings as thoughts and feelings. Likewise, although they are
able to observe the behaviour and emotional expressions of other people, they are still unable to make sense of their behaviour by attribution of mental states.

What would a mind without introspective awareness be like? Perhaps it would contain only first-order representations of events and experiences. First-order representations can be seen as relatively close to freshly processed perceptual information (description of objects and events: what it is and where it is). They could be tagged for being true or false, and tagged for when the event happened. However, representations in such a mind would not be tagged for propositional attitude (whether they are a thought, an ongoing experience or a memory), and they might all be held of equal factual status. There would be no modification of representations according to whether they were derived from another person’s opinion or from own opinion. If so, this would explain, for instance, why people with autism can suddenly adopt another person’s suggestion that they previously rejected, without any acknowledgement of the reversal. One important consequence of impaired self-consciousness might be impaired understanding of one’s own actions. In other words, without self-awareness, an individual might not know how she is going to act until she acted, nor why she acted as she did. This is different from the usual experience of actions, where we take access to our own imagined or true motivations for granted. It is easy to believe that we know what we are going to do before we do it. On the other hand, if thoughts, utterances or motor actions ‘just happen’, we presume that they are involuntary actions. A person who lacks self-consciousness may be unable to distinguish between her own willed and involuntary actions (C. D. Frith, 1992). These examples suffice to suggest that the notion of impaired introspective awareness can lead to some extraordinarily radical yet testable predictions.

4.1 Some Experimental Evidence about Impaired Self-Attribution of Mental States

While the (in)ability to attribute mental states to others has been studied extensively in children with autism, there is scarcely any work on the ability to attribute mental states to self. However, some evidence is available to indicate that the latter may be just as impaired as the former. For instance, Perner et al. (1989) asked children with autism about what they knew and what the experimenter knew concerning the contents of a box, depending on whether they were allowed to look inside. On some occasions only the child was allowed to look inside, and on other occasions only the experimenter. While it may seem obvious that only the person who has looked inside the box will know what it contains, this was not at all obvious to children with autism. Of the 23 autistic children tested (mean age 13 years, mean verbal mental age 6 years), only 13 showed this basic understanding: answering the ‘know’ question correctly (i.e. in line with the ‘see’ question). All the children who were able to answer correctly for their own knowledge
could also answer correctly for the experimenter’s knowledge. There was no significant difference between numbers of children succeeding on questions concerning self and questions concerning other; that is, no sign of an advantage for reporting own knowledge.

A similar finding is reported by Kazak, Collis and Lewis (1997) who asked young people with autism whether they knew or only guessed what was in a box, having on some trials seen inside. The same questions were asked for an experimenter who sometimes saw into the box. The results showed no superiority in judging own knowledge versus judging other’s knowledge in any of the experimental groups (autism, 4- and 5-year-olds, people with Down’s syndrome). With the exception of those people with very high verbal level, the autism participants found the know/guess task extremely hard, and succeeded (nonsignificantly) more often when judging the experimenter’s knowledge than when reporting their own.

In an experiment investigating how well children understand and keep track of their own intentions, Phillips (1993) used a target shooting game which was rigged so that hits were actually controlled by the experimenter. Children were first asked to name which of four targets they would aim to hit, when one of the targets contained a hidden reward. After shooting, they were asked again which target they had meant to hit. Normal 5-year-olds were able to say, for instance, that they had meant to shoot the red target, but hit the blue one instead, even when it turned out—luckily—that the blue target contained the reward. Mental age matched children with autism were much less able to report their initial intention correctly. While it is often reported that children with autism have a basic understanding of their own and other people’s desires (Tan and Harris, 1991), this finding suggests that this understanding may be rather fragile. In particular, the results suggest that the children with autism did not conceptualize their own intentions as intentions. To speculate, if intentions are not available as metarepresentations, clearly distinct from primary representations, then they may be confused or conflated with other primary representations, such as the representation of outcome (presence or absence of a reward).

Some indirect signs of impaired self-consciousness come from a source other than ToM studies. The observed executive function deficits in autism (problems in planning and monitoring goal-directed action; Pennington and Ozonoff, 1996) also suggest an inability to reflect on own mental states. Carruthers (1996) and Perner (1998) have each argued that these impairments may result from an inability to represent one’s own intended and imagined future behaviours. For example, planning ahead in a task such as the Tower of Hanoi, may require metarepresentation of possible (not actual) moves and of the desired (but not yet realized) end goal state. In the normal case, monitoring of performance and correction of errors can occur even in the absence of external feedback, because of access to own action intentions—access which may be impaired in autism. It has also been argued, however, that executive function deficits may be the primary cause of impaired theory of mind (Russell, 1996).
5. Some Implications of Impaired Self-Consciousness for Everyday Behaviour

While the possibility of impaired self-consciousness in autism has been raised before (e.g. Frith, 1989), the many consequences of such a hypothesis have not been drawn out in any detail. We only want to mention a few speculative possibilities. It is important to note that action without in-depth conscious reflection is by no means poorer than consciously performed action: quite the reverse may be the case. Indeed, performance based on information that is not consciously available may be more accurate than verbal report (see Milner and Goodale, 1995). For example, subjects whose explicit verbal response reflects the fact that they are subject to a visual illusion, nonetheless are not ‘fooled’ as far as their implicit motor coordination is concerned, i.e. when asked to reach for the distorted object (Aglioti, DeSouza and Goodale, 1995). Intriguingly, people with autism appear to succumb less to such visual illusions, even in their verbal report (Happeé, 1996). It is possible that their superior performance in verbal response is not contaminated by conscious reflection.

A human being with a normal capacity for learning can perform many learned actions automatically without conscious control. This individual would run into problems only in novel situations or when action schemata are in conflict, and in this case may react catastrophically, for instance by freezing or by violent outbursts. Similarly, such an individual may not recognize when a goal has been achieved and may repeat the same stimulus-driven action over and over again. Freezing, inexplicable outbursts of emotion, and repetitive action, are characteristic of low-functioning children with autism, the group most likely to fail standard ToM tests, and to exhibit an apparent impairment of self-consciousness. It is striking that the actions they perform well are stimulus driven and of a routine kind, while higher-order non-routine behaviour is often severely impaired (chapters in Russell, 1998; Turner, 1996). When informants rated the everyday behaviour of children with autism, they gave high scores for a whole range of learned skills of a routinized nature, while giving very low ratings for acting flexibly and imaginatively (Frith, Happeé and Siddons, 1994).

If low-functioning individuals with autism are unable to reflect on their inner experiences, then they would be unable to develop over time the richly connected semantic and experiential associations which normally pervade our reflective consciousness. Observations by parents suggest that the awareness of sensations and experiences may be peculiar in children with autism. Anecdotal reports of abnormal sensory and pain experiences are on occasion quite extreme (see below in the autobiographical examples). One anecdotal example is the case of a young girl with autism who was found to have suffered acute appendicitis, but had not complained of pain and, when asked how she felt, did not report anything wrong. Abnormal response to heat and cold, as well as hypo- and hyper-sensitivity to sound, light or touch are frequently reported (Cesaroni and Garber, 1991; chapters in Schopler and Mes-
bov, 1992). Such responses might be expected if there was an inability to reflect on inner experiential states. Of course, normal pain perception is greatly affected by attribution and expectation. These individuals might feel immediate pain in the same way as everyone else, but would not be able to attribute to themselves the emotional significance that normally accompanies pain. This might explain why they do not complain about it. We may speculate that the self-conscious person reflects not only on the pain but also on the experience of pain. This person is feeling ‘misery’ in addition to feeling pain.

5.1 Implications for Intervention

The skillful teacher or parent can help the child with autism to have access to his or her mental states and to tag them as mental states (beliefs, desires) by making them concrete (pictures, words) and explicit for examination. If the person with autism can judge their own inner states only by their actions, it might be important to teach behaviours which express for oneself what one thinks and feels. There are reports that keeping a diary about events that happened, or practising self-talk in the same way, can be very helpful. There is also some evidence for the benefit of concrete images in teaching insight into mental states (Swettenham et al., 1996).

6. Does Unusually Acquired Theory of Mind Imply Unusual Self-Consciousness?

We now turn to the case of Asperger syndrome as a particularly important source of insight into the question of self-consciousness and theory of mind. Asperger syndrome individuals are distinguished by their ability to pass false-belief tasks and their use of an explicit ToM. If ToM and self-consciousness are inextricably linked, as we have argued, then the fact that these individuals can in some cases attribute mental states to others should mean that they also exhibit the ability to reflect on their own mental states. Moreover, the degree of the two abilities should be related.

6.1 Evidence for Abnormal Self-Consciousness from an Experimental Study of Introspection

Given the willingness of able people with autism to talk about themselves, it is surprising that so far hardly any studies on introspection and self-consciousness have been conducted. We have attempted a preliminary study with the help of three high-functioning adults with Asperger syndrome. Their introspection was explored using a descriptive experience sampling method devised by Hurlburt (1990). Previously used with other individuals, the method involves use of a beeper which sounds at random intervals and signals to the subject to ‘freeze’ the moment and make brief notes about current mental contents for later interview discussion. Participants normally
report the following singly or in combination: verbal inner experience, visual images, feelings (located in the body), and unsymbolized thinking (thoughts without words or pictures associated with them). Our study showed marked qualitative differences in introspection in people with Asperger syndrome. We will not describe the study in detail here and the interested reader is referred to Hurlburt, Happé and Frith (1994). Here we provide some previously unpublished examples which are merely intended to give the flavour of the introspection of the three people we studied.

Our three volunteers were young men, all of normal IQ, and all from that minority of the autism spectrum which passes false-belief tasks like the Sally-Anne and Smarties test. Nevertheless they could be ranked by theory of mind task performance. Robert passed complex double bluff type tasks (second-order false belief; Perner and Wimmer, 1985), Nelson passed only some such tasks, and Peter passed only first-order tasks. We expected that this difference in ability to attribute mental states to others would be reflected in the participants’ ability to reflect on their own mental states. They all enjoyed participating, but only Robert quickly took to the idea of reporting his inner states, while Nelson was able to do this after four sessions, and Peter never satisfied our criteria for understanding the instructions. Instead, Peter persisted over five sessions in recording his ongoing physical action, but never any inner mental state.

**Sample 16.** Robert was on the train to London and had just seen from the train window a suspicious looking car parked or abandoned on the road near the track. He was trying to memorize where the car was so that he could come back and find it when he got off the train, in order to get its registration number (which he did in fact later succeed in doing). At the moment of the beep Robert was seeing an image of the car in accurate colour, about the size of a normal photo (4 × 5 inches). It seemed to be inside Robert’s head, ‘in front of my eyes’. The car in this image was seen just behind a low wall, with some garages in the background behind it; the parts of the car obscured by the wall were not seen. In fact, Robert said that everything in the image was seen exactly as he had seen it in real life.

**Sample 4.** Robert was at home watching football on TV with his mother. He hated football and was ‘thinking about how to get around watching it’. At the moment of the beep, he was seeing an image of himself sitting on the floor playing with his mother’s cat. Robert and the cat were seen from a perspective above them, looking down. He could see the top of his head and the cat on the floor in front of him. The image was in full accurate colour and motion: he could see the pyjamas he was wearing, the cat in motion, swiping his paw across Robert’s hand, the carpet on which they were sitting, and the furniture, although this was in natural indistinctness, being in the periphery.
Robert’s samples illustrate the very concrete nature of the mental images he reports.

Nelson

Sample 1. Nelson was looking at some trees. He could not clarify whether that ‘looking at’ involved real or imaginary seeing. Furthermore, he said he was thinking ‘It’s a lovely picture’ and he was both saying those words and seeing them in black capital letters written in his head. Very little of this description seemed believable. It had apparently never occurred to him to think or talk about his inner experience, and his eagerness to please led him to provide answers to our questions even though he had not made the required initial observations . . . By the fifth sampling day, however, Nelson had become quite convincingly clear about what our questions meant. He was now differentiating in writing immediately after the beep between ‘actual’ and ‘visualized’ phenomena, and his accounts contained convincing detail . . . Nelson used the phrase ‘the shape of my thoughts’ repeatedly throughout the interviews . . . being able to talk about the shape of his thoughts cut down on the work of communication by 50%, he said. He continued intermittently to marvel at the importance of understanding the shape of his thoughts . . .

Sample 4. Nelson was watching a TV nature documentary about water boatmen on a pond. He was seeing on the screen how the insects move, from left to right across the screen, and he was also ‘picturizing’ a classroom scene from when he was younger, when he had watched a similar TV programme. In the image, he could see the TV, a group of children sitting in front of it on the floor, and the teacher standing behind the children. The perspective was from the back of the teacher over her shoulder, past the group of children, towards the TV. The image was in colour and movement: he could see the brown of the water and the colour of the clothing the teacher and children were wearing.

Nelson’s samples too illustrate a visually determined report of inner experience. At first he seemed genuinely puzzled by our request to think and talk about inner experience, and made sense of the request only gradually.

Peter

Sample 10. Peter was doing the cleaning in his group home. He was ‘thinking about’ taking the top off the disinfectant bottle, but that apparently meant he was in the act of unscrewing it. He was seeing the bottle as he unscrewed it: this was apparently an actual seeing of the real bottle, even though he referred to it initially as ‘a mental picture’. Peter often used the expression ‘thinking about’ and the expression ‘it was on my mind’. When asked to explain what these
expressions meant he consistently indicated that he was in the process of doing something.

Although Peter was unable to tell us about his past inner experience using the bleeper method, it was possible to discuss with him current ongoing inner experience during interviews. These experiences, like those of Robert and Nelson, were visually based. In addition, they also give a brief glimpse into the sometimes strange interests and preoccupations of people with Asperger syndrome.

Peter was interested in the moon, since his childhood, and had developed a scheme to number the phases from 1 to 32. He had drawn a chart of these phases, and as he told us about the chart, we interrupted him to ask about his inner experience ‘right now’. He was seeing an image of the chart which he said was an accurate reconstruction of the chart, except that the actual chart also had lunar eclipses marked. He could not remember where on the actual chart those eclipse diagrams were located, and his image simply omitted them. When we sketched a crescent, asking him to tell us the number for that phase, he became absorbed for a few seconds. During this time we could not get his attention and he did not respond to our questions. He eventually mentioned a number and told us in answer to our question why he gave this number: ‘the number seems to suit the phase’.

What is striking about the reports our three participants gave is that they lack variation in form of inner experience: all three described visual inner experience with content often very close to what they had actually experienced, rather than anything fictitious or imaginary. For Peter, at least, to be asked to imagine (i.e. mentally picture with his eyes closed) something counterfactual (e.g. the interviewer standing up, when in fact she was sitting down) was unpleasant and uncomfortable—he would not comply. None of the three described verbal or unsymbolized thinking, despite our occasionally leading questions in this respect. It is not clear of course whether these volunteers in fact experienced only visual images, or whether ‘pictures in the head’ was a metaphor which allowed them to reflect on and report inner states. We cannot know for sure, but we surmise that they have mental states that they cannot report, especially in the case of Peter. Their ToM ability (but not their verbal IQ) was closely linked to their ability to introspect—best in Robert and least in Peter. This strengthens our hypothesis that self-awareness, like other awareness, is dependent on ToM.

6.2 Evidence for Abnormal Self-Consciousness from Autobiographical Accounts

Some people with high functioning autism/Asperger syndrome manifest their astonishing ability to reflect on their own mental states in autobiographical accounts. These typically include both memories of their early childhood, when we would expect that they did not have an awareness of self, and accounts of their present experiences. The following excerpts, chosen as fairly typical, give a flavour of these accounts.
From Donna Williams (1994)

Autism had been there before I’d ever known a want of my own, so that my first ‘wants’ were copies of those seen in others (a lot of which came from TV). Autism had been there before I’d learned how to use my own muscles, so that every facial expression or pose was a cartoon reflection identification . . . Like someone sleep-walking or sleep-talking, I imitated the sounds and movements of others—an involuntary compulsive impressionist. (p. 3)

There are two ways to be a nobody nowhere. One is to be frozen and unable to do anything spontaneously for yourself. The other is to be able to do anything based on stored mirrored repertoires without any personal self-awareness yet being otherwise virtually unable to do anything complex with awareness. (p. 42)

I was sick to death of my attention wandering onto the reflection of every element of light and colour, the tracing of every patterned shape and the vibration of noise as it bounced off the walls. I used to love it. It had always come to rescue me and take me away from an incomprehensible world, where, once having given up fighting for meaning, my senses would stop torturing me as they climbed down from overload to an entertaining, secure, and hypnotic level of hyper. This was the beautiful side of autism. (pp. 112–13)

Autism makes me hear other people’s words but be unable to know what the words mean. Or autism lets me speak my own words without knowing what I am saying or even thinking. (p. 233)

From Jolliffe, Lansdown and Robinson (1992)

When I was very young I can remember that speech seemed to be of no more significance than any other sound . . . I began to understand a few single words by their appearance on paper. (p. 13)

It was ages before I realised that people speaking might be demanding my attention. But I sometimes got annoyed once I realized that I was expected to attend to what other people were saying because my quietness was being disturbed. I began to start to understand more than just a single word at a time when I realised that speech was sometimes directed at me.

. . . I was once asked by somebody why I repeat the same phrase exactly as I had heard it and in a similar voice to the person who said it. I had no answer at the time. Being asked to write this paper I have had to think more deeply about myself than I have ever had to before. I now know there are several answers. First, you have to work so hard in order to understand speech, that when the words do eventually go into your brain they seem to become imprinted in the way you hear them. Second, because trying to speak is quite an effort . . . Third . . . you seem to believe that the voice of the person used to say the words is the way you too have to say them. You do not seem to be aware
that the words can be put across using all different kinds of voices and that there are alternative ways of expressing things. (p. 14)

From Temple Grandin (1984)
I also liked to sit for hours humming to myself and twirling objects or dribbling sand through my hands at the beach. I remember studying the sand intently as if I was a scientist looking at a specimen under the microscope. I remember minutely observing how the sand flowed, or how long a jar lid would spin when propelled at different speeds. (p. 156)

From Temple Grandin (1991)
I think almost entirely in visual terms. When I read a sentence I immediately see an image of the subjects described. ‘Freedom’ suggests an image of the Declaration of Independence, and ‘truth’, the scales of justice. If someone says ‘cat’ I see in my mind’s eye cats I have known or read about. I also visualize abstract ideas such as relations with people. My original image for getting along with other people was a sliding glass door that must be approached gently, lest it shatter. (p. 16)

From Temple Grandin (1992)
As a child I was hyperactive, but I did not feel ‘nervous’ until I reached puberty . . . The feeling was like a constant feeling of stage fright all the time . . . I had a pounding heart, sweaty palms and restless movements. The ‘nerves’ were almost like hypersensitivity rather than anxiety. It was like my brain was running at 200 miles an hour, instead of 60 miles an hour . . . At a carnival I discovered that riding on the Rotor Ride provided temporary relief. Intense pressure and vestibular stimulation calmed my nerves. (p. 111)
I had an odd lack of awareness of my oddities of speech and mannerisms until I looked at videotapes. (p. 113)
As a child, I often talked out loud because it made my thoughts more ‘concrete’ and ‘real’. Today, when I am alone designing, I will talk out loud about the design. Talking activates more brain regions than just thinking. (p. 122)

From Gunilla Gerland (1997)
I had—and always had had, as long as I could remember—a great fear of jewellery. That terror also included hairclips and metal buttons. I thought they were frightening, detestable, revolting. If I was made to touch jewellery, I felt a sharp whistling metallic noise in my ears,
and my stomach turned over. Like a note falsely electrified, that sound would creep from the base of my spine upwards until it rang in my ears, tumbled down into my throat and settled like nausea into my stomach. (p. 54)

All the time I was growing up, I suffered from an almost constant shudder down my spine . . . It was a constant torture . . . It was like cold steel down my spine. (p. 56)

It confused me totally when someone said that he or she had seen something I had been doing in a different room . . . Not in my wildest fantasies could I imagine that that person had actually been there, in that room, then. It seemed as crazy as if someone had said that the same furniture was in all the rooms I was in, that the chairs and tables followed me wherever I went. (p. 64)

Time and again I was very hurt when people said they knew things about me, things which, according to my view of how it was, they couldn’t know. (p. 65)

I was bullied, usually on my way home from school . . . I was stopped by a fair-haired boy . . . he threw me face down on the ground and sat on my back, and I just let him do it. He filled his hands with snow and rubbed my face hard, several times. I let it happen and just waited. After a while he got up . . . I felt neither flattened nor crushed, only miserable that I didn’t understand all the strange things that happened in the world . . . When I looked up at him, I saw his expression change . . . I saw terror in his face. But I didn’t know where it had come from . . . I looked down at myself. I saw my jacket and trousers were dirty and wet, but I couldn’t see what had frightened him so . . . Not until I got home did I see that my face was criss-crossed and bleeding from thousands of little scratches. The snow he had rubbed into my face had been filled with little bits of gravel, but with my high pain threshold I hadn’t felt it . . . I stood in front of the mirror for a long time, looking at my face. I thought it looked interesting. (p. 99)

My feeling was special in other areas as well, apart from that of pain . . . The vague sense of my body I did have meant that I wasn’t particularly aware whether I was dirty, of how my clothes were sitting. I didn’t feel it. (p. 101)

I also had another problem . . . though it was a problem I never understood until I was an adult. I thought it was the same for everyone. I couldn’t feel that I needed to go to the lavatory, so I had to think out when I needed to go. I didn’t know other people had a signalling system that warned them at intervals before the need to go became urgent. I had no such system. I felt nothing, nothing, nothing . . . (p. 120)

My insensitivity to pain was by now as good as total . . . nothing hurt at all. And yet I felt—my actual feelings were not shut off—because when I was aware that I had injured myself somewhere, I could sense
something, a non-pain, which branched out into my body from the place where the injury was. But the fact was, it didn’t hurt. (p. 157)

The autobiographical writings of these exceptional people suggest a radically different inner experience, both when childhood memories are reported and when present mental states are discussed. However, there are grave limitations with this type of material. While the accounts are intriguing, it might be a mistake to take what is said at face value. This caveat, of course, applies to any autobiographical account. We all are aware of reconstructing our memories of early childhood, sometimes incorporating others’ narrative accounts of early events. In line with the theory presented above, the writers not only have insights into their own mental states, but can also attribute mental states to others. They would certainly pass standard false-belief tasks. Typically in the autobiographical accounts we find relatively little about other people’s feelings or attitudes. Unlike ordinary biographers, they are not constantly wondering about how the reader might see them and their families. They are not interested in making an impression. They are seemingly oblivious to the possibly defamatory effects of what they tell about themselves and their relatives. Thus, harrowing events (in the case of Donna and Gunilla) are reported, while possible reasons for otherwise bizarre behaviour on the part of other people are left extremely vague. However, this may not be so different from autobiographies in general. In this respect the genre of autobiographical writing is the perfect niche for the talented writer with Asperger syndrome. One might say the self is not only central to the world they describe, but it is that world.

7. Conclusion

We have reviewed some of the evidence for a dedicated, circumscribed cognitive mechanism underlying the development of ToM. Self-consciousness, too, we have argued, may be seen as the product of a specific neurocognitive mechanism. If we are right, then it should be possible to find not only cases of good general cognition with impaired self-consciousness, but also of impaired general cognition with good self-consciousness.

If people with autism lack intuitive awareness of own mental states, then they might help us eventually to answer some intriguing questions: What does self-consciousness buy you? What can’t you do without self-consciousness? Why is self-consciousness so important that a minority of people with autism acquire and use it with great and apparently continuous effort? What can these people do that other people with autism can’t? We don’t really know, but they certainly show improved social communication and greater vulnerability to misperception of mental states (e.g. paranoid interpretations).

Autism, then, may hold the key to some of the most interesting questions about self-consciousness. We have argued that a theory of mind mechanism
is vital for the development of self-consciousness, just as it is vital to the understanding of other minds. Autism demonstrates how damaging the lack of ToM is for development of social competence. Autism is a devastating disorder because it disrupts not only understanding of others and their social relationships, but also understanding of self. If our speculations are correct, to the extent that they are testable, then having a theory of other minds, as demonstrated by understanding of false beliefs in others, is also a key criterion for self-consciousness.

Institute of Cognitive Neuroscience
University College London

Social, Genetic and Developmental Psychiatry Research Centre
Institute of Psychiatry, London

References


