ORIGINAL PAPER

Theory of Mind in Williams Syndrome Assessed Using a Nonverbal Task

Melanie A. Porter · Max Coltheart · Robyn Langdon

Abstract This study examined Theory of Mind in Williams syndrome (WS) and in normal chronological age-matched and mental age-matched control groups, using a picture sequencing task. This task assesses understanding of pretence, intention and false belief, while controlling for social-script knowledge and physical cause-and-effect reasoning. The task was selected because it is entirely non-verbal, so that the WS individuals could not rely on their good verbal skills when performing the task. Results indicated a specific deficit in understanding of false belief within the WS group. There was also evidence of heterogeneity in the WS group, with the false belief impairment restricted to only a particular subgroup of WS individuals identified originally by Porter, M., & Coltheart, M. (2005). Cognitive heterogeneity in Williams syndrome. *Developmental Neuropsychology, 27*(2), 275–306.

Keywords Williams syndrome · Theory of Mind (ToM) · Nonverbal · Heterogeneity

Introduction

'Theory of Mind' (ToM), or the ability to reason about psychological causality, has been assessed across a wide range of developmental disorders. While ToM and intellectual abilities are often commensurate, dissociations occur within certain disorders. For example, people with autism or schizophrenia typically display deficits in their understanding of mental states such as pretence, intentions and beliefs beyond their general level of intellectual functioning (Baron-Cohen 1993; Langdon et al. 1997); not surprisingly, their day-to-day social functioning is also impaired.

ToM and intellectual abilities may also dissociate in people with Williams syndrome, a rare genetic disorder first identified in 1961 and typically associated with mild to moderate intellectual impairment (Williams et al. 1961). Unlike people with autism or schizophrenia, people with Williams syndrome (WS) are sometimes said to show strengths in their day-to-day social functions, so one might expect that ToM abilities in WS are above their general intellectual abilities. The literature remains mixed with respect to ToM abilities in WS however, with some research suggesting that ToM abilities may be above general intellectual abilities (e.g. Karmiloff-Smith et al. 1995; Tager-Flusberg et al. 1998) and other research suggesting they might be below (Tager-Flusberg and Sullivan 2000). Yet other research indicates that ToM abilities in WS are commensurate with their general level of intellectual functioning (Sullivan and Tager-Flusberg 1999).

Three factors that seem to contribute to these variable findings are: (1) The choice of ToM measures; (2) The choice of comparison groups and (3) Heterogeneity within WS. These three issues will be discussed in detail below.

Choice of ToM Measures

ToM measures used within the WS literature have typically relied heavily on verbal skills (verbal comprehension, expressive language, narration and use of mental state language). Tager-Flusberg and Sullivan (2000), for example, read short stories to children with WS (aged 4–8 years) and asked them questions in order to assess their understanding of desire, emotion and belief. Tager-Flusberg and Sullivan found that the verbal responses of their group of individuals

M. A. Porter (✉) · M. Coltheart · R. Langdon
Macquarie Centre for Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia
e-mail: mporter@maccs.mq.edu.au

with WS indicated ToM abilities equivalent to those of children with non-specific intellectual disability and children with Prader-Willi syndrome or PWS,[1] another developmental disorder associated with intellectual impairment. Similarly, Sullivan and Tager-Flusberg (1999) investigated ToM in WS, PWS, and children with non-specific intellectual disability (aged 8–17 years) using classic story vignettes that are common in the autism literature. Sullivan and Tager-Flusberg found that all three groups performed similarly in their ability to answer second order belief questions and in their use of second-order justifications to explain a critical protagonist's behavior (e.g. "John did that because John thinks that Mary thinks"). These studies suggest that ToM abilities in people with WS develop similarly to individuals with other forms of intellectual impairment.

The difficulty here is that people with WS typically display strengths in their general verbal abilities compared to their overall level of intellectual functioning (e.g. see Howlin et al. 1998; Ypsilanti et al. 2005). WS individuals may therefore have relied on their good verbal skills (and possibly their good use of mental state language) to enhance their performance on the verbal ToM tasks compared to intellectually impaired controls.

Indeed there is some evidence to suggest impaired ToM abilities in WS compared to other intellectually impaired groups, at least on false belief tasks. Tager-Flusberg and Sullivan (2000), for example, also used the classic change in location and change in contents tasks[2] on their 4–8 year old

participants with WS. While an ANOVA revealed no significant main effects or interactions, the percentage of children who passed ignorance or false belief questions in these tasks was significantly higher for the PWS and non-specific mental retardation groups compared to the WS group. Similarly, Tager-Flusberg et al. (1997) had found only 43% of their sample of WS children passed two trials of a change in location task compared to 60% of their PWS control group, and argued that this suggested some impairment in false belief understanding in WS. Using these same tasks, however, Karmiloff-Smith et al. (1995) found that 94% of their 18 WS participants (aged 9–23 years) passed these tasks; their study did not include a control group. It might be important, therefore, to focus more specifically on false belief understanding in WS, which might be commensurate with or below general intellectual functioning.

If, as suggested earlier, individuals with WS use their superior verbal abilities to 'bootstrap' their performance on verbal ToM tasks, how do these individuals perform when the verbal demands are low? Tager-Flusberg et al. (1998) asked participants to simply select the correct labels to match photographs of complex mental state expressions depicted in the eye region of a face and found that their group of WS adults still performed significantly better than a group of adults with PWS. We would suggest, however, that this particular task is more akin to an emotion perception task than a classic ToM task. And again, WS individuals may have used their superior vocabularies to pass this task. In any case, research also exists to suggest that emotion perception abilities in WS are commensurate with their general level of intellect (Plesa-Skwerer et al. 2006; Porter 2004).

Overall, the results thus far remain equivocal as to whether ToM abilities in WS are commensurate with or above or below general intellectual functions. It is possible that the use of predominately verbal tasks has masked ToM impairment in WS, as these tasks allow individuals with WS to use their good verbal skills to their advantage when performing these tasks. Given these concerns, we have chosen to assess ToM in WS using a non-verbal task.

The Dilemma of Choosing a Comparison Group

That ToM impairment in WS might have been masked in previous research also seems particularly plausible as experimental and control groups, when matched on verbal abilities, have typically been matched using a receptive vocabulary task rather than measures of verbal intellectual ability, expressive language or verbal comprehension. According to Shaked and Yirmiya (2004)'s research, these and related matching choices may have biased the results.

In a meta-analytic study on research investigating ToM abilities in autism, Shaked and Yirmiya (2004) found that variables such as chronological age of the participants,

---

[1] According to Tager-Flusberg (1999), people with PWS are characterized by salient food-related characteristics including hyperphagia (overeating) and assiduous food seeking behaviors. PWS involves either a deletion on chromosome 15 derived from the father or maternal uniparental disomy of chromosome 15. Physical features are apparent including facial features, short stature, and small hands and feet. People with PWS typically show higher intelligence than that seen in other genetic disorders, with a mean IQ around 70. There is a hint of uneven academic performance within PWS with strengths in reading and a weakness in arithmetic. There is no universal PWS cognitive profile, though cognitively some authors suggest strengths in spatial-perceptual organization and visual processing and a weakness in short term memory.

[2] The change in location task runs as follows: Participant (P) is shown two boxes. Experimenter 1 (E1) hides an object in Box A, Experimenter 2 (E2) leaves the room, E1 moves object to Box B and asks P three questions: "When E2 comes back where will E2 look for the object?", "Where is the object really?", and "Where does E2 think it is?". The change in contents task runs as follows: P is shown a packet of M & Ms (or a local equivalent) and asked what is inside. Most Ps give the reply: "M & Ms.". E1 then shows P that this packet does not actually contain M & Ms, but small pencils. P is then asked what another child, a classmate, who has not yet seen the contents, will reply when shown the packet. P's task is (1) to differentiate between the mental representation she or he has of the actual contents of the packet (pencils) and the mental representation that the other child will have of its contents as a result of expectancies from world knowledge (M & Ms), and (2) to predict that the other child's response will be based on her or his mental state (what she or he thinks it contains) and not on the actual contents known to P.

matching procedures and choice of clinical comparison group greatly affected the outcome of a study. More specifically, larger effect sizes were reportedly obtained when: children were older rather than younger; participants were matched individually on mental age or chronological age (rather than simply matched based on non-significant group differences); and when clinical comparison groups were more homogeneous (such as Down syndrome) rather than heterogeneous (such as non-specific intellectual disability).

Turning first to the issue of age, the majority of researchers within the WS literature have studied different age ranges (e.g. 9–23 years in one study and 4–8 years in another), without considering the general influence of mental age or chronological age on task performance and without using both mental age and chronological age matched controls as a comparison. Turning further towards the issue of comparison groups, typical control groups for assessing ToM abilities in WS have been individuals with PWS and individuals with non-specific mental retardation. The use of people with PWS may be particularly misleading as people with PWS typically display higher levels of intellectual functioning than are seen in other developmental disorders. Furthermore, the groups of individuals with PWS and non-specific mental retardation were likely to have been quite heterogeneous. PWS for example is said to be a heterogeneous disorder in terms of intellectual strengths and weaknesses, with no particularly common cognitive profile. The implication here is that variability within the comparison clinical groups might have obscured ToM difficulties in WS.

In an attempt to balance these various concerns, we chose to use normal chronological age matched and normal mental age matched controls rather than choosing a clinical control group (although we mention some comparisons to individuals with Down syndrome in the discussion). In addition, we opted not to restrict the mental age or chronological age of our WS sample, especially since Shaked and Yirmiya (2004) suggest that imposing such restrictions may in fact substantially influence one's results on ToM measures. This latter decision was also based on the fact that WS is a very rare disorder and that restricting our age range would substantially reduce our sample size. Even with these considerations, within-groups variability might still be a concern, particularly in our clinical WS group.

Heterogeneity in WS

While many developmental disorders are heterogeneous, WS appears to be a particularly heterogeneous disorder with cognitive, social, genetic and physical characteristics of WS varying considerably from one individual with WS to another (Borg et al. 1995; Porter and Coltheart 2005, 2006). The literature on social abilities within WS seems to reflect this variability. Some literature and anecdotal reports suggest good social skills (Karmiloff-Smith et al. 1995), a hyper-social nature (Jones et al. 2000; Doyle et al. 2004) and good conversational language (e.g. see Bellugi et al. 1999) in WS. On the other hand, other literature and other anecdotal reports indicate that people with WS often display inappropriate or invasive social behavior (Udwin 1990), poor conversational abilities, and pragmatic language difficulties (Laws and Bishop 2004). Furthermore, peer relationships in WS can often be superficial, with many WS individuals failing to develop appropriate social relationships with their peers (Davies et al. 1998).

Similarly, people with WS vary considerably in their cognitive strengths and weaknesses. For example, using the Woodcock-Johnson test of cognitive ability—revised (WJ-R COG, Woodcock and Johnson 1989, 1990), a standardized measure of verbal skills, processing speed, long-term memory, short-term memory, visual processing, auditory processing, fluid reasoning and comprehension-knowledge, Porter and Coltheart (2005) found extreme cognitive variability in WS, with individuals not only differing in degree of impairment, but also displaying distinct patterns of strength and weakness. Porter and Coltheart therefore challenged the notion of a "syndrome-specific" (Howlin et al. 1998, p. 183) WS cognitive profile. While some individuals displayed the characteristic WS profile of a strength in verbal skills and a weakness in nonverbal abilities, other individuals displayed a strength in nonverbal skills and a weakness in verbal abilities.

Porter and Coltheart (2005) suggested that subgroups might exist within WS and have since found evidence to support their claims. In a cohort of 31 WS participants, Porter and Coltheart (2005) found evidence for up to eight distinct WS subgroups. The two largest subgroups (known as Subgroup 1 and Subgroup 4) each included nine individuals. Subgroups were defined according to similarities in their cognitive profile on the WJ-R COG. Subgroup 1 displayed significant strengths in long-term memory and auditory processing and significant weaknesses in speed of processing and oral language, while Subgroup 4 displayed significant strengths in verbal comprehension, oral language, long-term memory and auditory processing and a significant weakness in speed of processing.

Subsequent research indicated further differences with double dissociations of function across these two groups in terms of perception, attention and spatial construction abilities (Porter and Coltheart 2006) and differences in social-emotion skills (Porter 2004, & paper in preparation), providing further validation that these two subgroups are inherently different and that differences across subgroups appear to be consistent across a variety of measures. In more detail, Subgroup 1 displayed a perceptual integration deficit, but good spatial construction abilities, response inhibition and emotion perception abilities, whereas

Subgroup 4 showed good perceptual integration skills, but poor spatial construction abilities, poor response inhibition and poor emotion perception abilities.

Overall, there is clear evidence to suggest cognitive and social heterogeneity in WS. This heterogeneity does not appear to reflect differences in degree of impairment, but rather distinct patterns of strength and weakness or spared and impaired abilities. This heterogeneity may help to explain the equivocal findings within the literature in relation to ToM and social abilities in WS.

Since the identification of these two subgroups arose from a study with relatively few participants ($N = 31$) measured on relatively many variables (seven), a cross-validation with a new set of WS participants would be beneficial in order to see whether these same two subtypes emerge when the same seven variables are used. We are engaged in such a study, but it will take some time to attain a sufficiently large sample of new WS participants.

There is another way, however, in which the subtyping can be further validated and this is to use new variables rather than new participants—that is, to show that the same subtype structure still emerges when data on new variables is collected from the original participants.

Summary of Aims

The main aim of the present study, then, was to explore ToM functioning in WS using a nonverbal task which assesses understanding of false beliefs, as well as pretence and intentions. A nonverbal task was used so that individuals with WS were unable to use their good verbal skills to their advantage. People with WS were compared to normal chronological age matched controls (NCA) and normal mental age matched controls (NMA) in order to avoid matching dilemmas outlined in Shaked and Yirmiya (2004, see above). We also used a wide age range so as not to bias our results in a particular direction (Shaked and Yirmiya 2004). The second aim was to explore whether ToM abilities may differ for WS individuals in two previously determined subgroups of WS, namely Subgroup 1 and Subgroup 4 from Porter and Coltheart (2005), with the possibility that one group might display intact and the other impaired ToM abilities, thus providing further validation that these subtypes are indeed distinct.

Method

Participants

Williams Syndrome

Thirty of the original 31 WS individuals studied by Porter and Coltheart (2005) took part in the study. These WS individuals (14 male and 16 female) were recruited through the Williams Syndrome Association New South Wales and the Williams Syndrome Association South Australia. All participants were diagnosed independently by a minimum of two professionals (cardiologists, ophthalmologists, geneticists, pediatricians) based on a combination of unique facial, physical and behavioral characteristics associated with the syndrome (McKusick 1988; Morris and Sigman 1988).

Normal Controls

Controls were recruited through state schools in New South Wales and through Macquarie University, Australia. For the chronological age matched controls (NCA) 30 typically developing individuals were chosen to individually match WS participants on chronological age and gender. Similarly, 30 mental-age matched healthy controls (NMA) were chosen to individually match WS participants on mental age and gender. Controls were screened for a history of developmental delay or cognitive or intellectual impairments.

The top of Table 1 shows mean chronological and mental ages for each group (WS, NCA and NMA). For the Williams individuals, mental ages were obtained using the Woodcock-Johnson test of cognitive ability - revised (WJ-R COG, Woodcock and Johnson 1989, 1990). All subtests from the WJ-R COG were administered according to standardized instructions described in the WJ-R COG Examiner's manual (Woodcock and Mather 1989b, 1990b). Individual subtests included measures of short-term memory/attention, expressive and receptive language abilities, conceptual reasoning and visual perception.

As expected, analyses revealed no significant difference in chronological age between the WS group and the NCA group ($F (1, 58) = 0.10$, $p = .90$) and no significant difference in mental age between the WS group and the NMA group ($F (1, 58) = 0.08$, $p = 0.80$).

Materials

The Picture Sequencing Task

The picture sequencing task used in the present study was designed by Langdon et al. (1997) to test ability to reason about psychological causality in people with schizophrenia. The task was based on an earlier Baron-Cohen et al. (1986) study used to assess ToM in children with autism, Down syndrome and normal preschool children. There are mental state stories, as well as social script and mechanical stories. In addition to the critical false-belief stories, stories were also designed to assess understanding of pretence,

Table 1 Mean mental age and chronological age for each group

	Chronological age	Mental age
Williams syndrome	17.02 (9.53), 5.33–43.67	5.65 (1.31), 3.58–9.33
Normal CA matched controls	17.74 (10.05), 5.58–44.58	n/a
Normal MA matched controls	5.62 (1.26), 3.67–9.50	n/a
WS Subgroup 1	23.20 (8.59), 9.41–34.25	6.50 (0.84), 5.33–8.08
Normal MA matched controls for Subgroup 1	6.44 (0.86), 5.08–8.00	n/a
WS Subgroup 4	17.15 (11.60), 6.00–43.67	5.41 (0.88), 4.00–6.67
Normal MA matched controls for Subgroup 4	5.42 (0.82), 4.08–6.67	n/a

Note: Mean (s.d.), range; MA = mental age, CA = chronological age, scores are in years

unrealized goals and intention (see below). The idea is that if a person has a selective ToM deficit, they should experience difficulty sequencing some or all of the mental state stories, but not the social script or mechanical stories.

Twenty-four stories were depicted in 4-card picture sequences (21 cm × 15 cm) using a simple black and white cartoon style. Illustrations of similar picture sequences are published in Leslie and Frith (1990). There were eight story types, with three examples per story type: (a) mechanical 1: Objects interacting casually with each other, (b) mechanical 2: A person and objects interacting casually, (c) social script 1: A single person acting out an everyday social routine, (d) social script 2: More than one person interacting in everyday social routines, (e) Pretence: One or more persons involved in pretend play, (f) Unrealized-goal: Goal directed activity, where the goal is not achieved, (g) Intention: Goal-directed activity, where initial actions could be associated with multiple goals, (h) False-belief: A person, unaware of an event in a story, acts on a false belief.

Design and Procedure

The picture sequencing task was administered in accordance with Langdon et al. (1997). After two practice items, stories were presented in a random order. Cards for each story (four in total) were placed in a random order presented face down. Participants were instructed to turn the cards over and to put them in order so they told a story 'that made sense'. There was no time limit. Order of cards was recorded for each story.

Data Scoring

The picture sequencing task was scored using Langdon et al.'s (1997) criterion. In short, each sequence scored two points if the first card was positioned correctly, two points if the last card was positioned correctly, and one point each

for correct second and third card placement. Scores ranged form zero to six. A score of zero was also provided when a participant failed to produce a sequence. Scores were averaged across the three examples of a story type (range 0–6).

Results

Table 2 shows the mean position score for WS and normal control groups across the eight story types. Table 2 indicates that the picture sequencing task was appropriate for the mental age range of our WS group, with no apparent floor or ceiling effects. Greenhouse-Geisser corrections are reported due to violation of sphericity. An ANOVA with Group as the between-subjects factor (3 levels: WS, NCA and NMA) and Story Type as the within-subjects factor (8 levels: social script 1, social script 2, unrealized goal, false belief, pretence, intention, mechanical 1, and mechanical 2) indicated a significant main effect for Group (F (2, 87) = 20.412, $p < .001$), a significant main effect for Story Type (F (6, 484) = 22.977, $p < .001$) and a significant interaction of Group by Story Type (F (11, 484) = 3.789, $p < .001$).

Follow-up tests were undertaken to explore the significant Group by Story Type interaction. Analyses revealed that the NCA group performed significantly better than the WS group and the NMA group on all story types ($p < .01$ for all comparisons). Further analyses revealed that the WS group performed significantly below the NMA group on the false belief stories (F (1, 58) = 7.783, $p < .01$), yet these groups performed similarly on all other story types ($p > .07$ for all other comparisons). This suggests that individuals with WS display a specific impairment in understanding false belief compared to normal mental age matched controls and, therefore, compared to their general level of intellectual functioning. Results also suggest that understanding of pretense and intent, at least as assessed using the picture sequencing task, are commensurate with intellectual abilities in WS.

Table 2 Mean position score by group on the picture sequencing task

	M1	M2	S1	S2	P	U	FB	I
WS	3.41 (1.58)	3.51 (1.79)	4.48 (1.68)	4.16 (1.77)	2.96 (1.51)	3.61 (1.92)	2.42 (1.64)	4.03 (1.66)
NCA	5.81 (0.41)	5.66 (0.76)	5.84 (0.36)	5.58 (0.89)	5.08 (1.59)	5.53 (0.89)	5.02 (1.16)	5.02 (1.16)
NMA	4.03 (1.81)	4.33 (1.57)	4.79 (1.69)	4.59 (1.68)	3.00 (1.61)	4.21 (1.65)	3.58 (1.61)	3.59 (1.61)
WS Subgroup 1	3.80 (1.61)	4.22 (1.63)	5.11 (1.67)	5.22 (1.31)	3.81 (1.53)	4.41 (2.26)	3.31 (1.63)	4.85 (1.19)
WS Subgroup 4	2.85 (1.49)	2.52 (1.78)	3.93 (1.69)	3.00 (1.67)	2.67 (1.72)	2.94 (1.62)	1.45 (1.03)	3.48 (1.84)

Note: Mean (s.d.); M1 = mechanical 1; M2 = mechanical 2; S1 = social script 1; S2 = social script 2; P = pretence; U = unrealized-goal; FB = false belief; I = intention

Heterogeneity

Upon close examination of individual scores, results suggest heterogeneity in ToM abilities within WS, as well as some heterogeneity in social script knowledge. In order to further explore this heterogeneity, we compared ToM abilities in Subgroup 1 and Subgroup 4 from Porter and Colthart (2005).

Subgroup 1 and Subgroup 4 did not differ in terms of overall chronological age (F (1, 16) = 1.84, $p > .1$), but Subgroup 1 displayed a significantly higher mental age, overall, compared to Subgroup 4 (F (1, 16) = 7.26, $p < .05$). The bottom of Table 2 shows the estimated marginal means (controlling for mental age) for WS Subgroup 1 and WS Subgroup 4 on each of the eight story types. These subgroups did not differ on individual subtests from the WJ-R COG assessing short-term memory/attention, conceptual reasoning or visual perception ($p > .1$ for all comparisons). Analyses revealed that WS Subgroup 1 performed significantly better than Subgroup 4 in sequencing the false belief stories (F (1, 16) = 8.478, $p = .01$) and social script 2 stories (($(1, 16) = 9.873$, $p = .01$), while differences on all other story types failed to reach significance ($p > 0.06$). These differences remained significant when mental age was covaried. An analysis with group as a between-subjects factor (2 levels: WS Subgroup 1 and WS Subgroup 4), Story Type as the within-subjects factor (2 levels: false belief, social script 2) and mental age as a covariate indicated a significant main effect for Group (F (1, 15) = 6.06, $p < .05$) and no significant interaction (F (1, 15) = 0.05, $p > .1$); results remained similar when chronological age was also entered as a covariate.

Effect sizes for these subgroup differences were very large, 1.4 for false belief and 1.5 for social script 2, with less than 30% overlap in distributions (see http://web.uccs.edu.au/lbecker/psy590/es.htm). This indicates that the two subgroups display distinct patterns of performance on false belief and social script 2 story sequences.

We next entered false belief and social script 2 scores into a backward stepwise logistic regression analysis with subgroup membership as the dependent variable. The overall model was highly significant (χ^2 (2, N = 18) = 11.50, –2 Log Likelihood = 13.45, $R^2 = 0.63$, $p < .01$) with false belief and social script 2 scores together accurately predicting subgroup membership in 83.3 percent of cases. Both false belief ($B = –0.97$, SEB = 0.63, $\beta = .38$, Change in –2 Log Likelihood = 4.16, Wald = 2.34, $R^2 = 0.36$, $p < .05$) and social script 2 ($B = –0.80$, SEB = 0.45, $\beta = .45$, Change in –2 Log Likelihood = 4.45, Wald = 3.14, $R^2 = 0.41$, $p < .05$) scores remained a significant predictor of subgroup membership in the presence of the other variable. Results indicate that both variables provide a unique contribution when it comes to predicting subgroup membership.

In order to investigate ToM abilities in WS Subgroup 1 and WS Subgroup 4 further, we compared the performance of individuals in each subgroup with normal controls matched individually to each WS participant on mental age. Mean picture sequencing scores for these subgroups of normal mental age matched control groups were similar to the NMA scores in Table 2 and are not reproduced here.

WS Subgroup 1 versus Mental Age Matched Controls

An ANOVA with Group as the between-subjects factor (2 levels: WS Subgroup 1 and NMA for WS Subgroup 1) and Story Type as the within-subjects factor (8 levels: social script 1, social script 2, unrealized goal, false belief, pretence, intention, mechanical 1, and mechanical 2) indicated no significant main effect for Group (F (1, 16) = 1.21, $p > .1$), a significant main effect for Story Type (F (4, 69) = 3.97, $p < .01$) and no significant interaction (F (16, 69) = 2.05, $p = .09$). This suggests that ToM abilities in Subgroup 1 are commensurate with their general level of intellect or their overall mental age.

WS Subgroup 4 versus Mental Age Matched Controls

In contrast, an ANOVA with Group as the between-subjects factor (2 levels: WS Subgroup 4 and NMA for WS Subgroup 4) and Story Type as the within-subjects factor

(8 levels: social script 1, social script 2, unrealized goal, false belief, pretence, intention, mechanical 1, and mechanical 2) indicated no significant main effect for Group (F (1, 16) = 1.94, $p > .1$), a significant main effect for Story Type (F (5, 78) = 7.45, $p < .001$) and a significant Group by Story Type interaction (F (5, 78) = 2.797, $p < .05$).

Follow-up analyses to explore the significant Group by Story Type interaction revealed that WS Subgroup 4 performed significantly below their normal mental age matched control group in sequencing false belief stories (F (1, 16) = 6.289, $p = .02$), but that the two groups performed similarly on all other story types ($p > .06$ for all other comparisons).

Summary

Results suggest a specific impairment in social understanding (assessed using social script and false belief picture sequences) within WS Subgroup 4, but not WS Subgroup 1, providing further validation that subgroups exist within WS.

General Discussion

The first aim of the present study was to explore ToM functioning in WS using a nonverbal picture sequencing task. A nonverbal task was used so that individuals with WS were unable to use their good verbal skills to their advantage. People with WS were compared to normal chronological age matched controls and normal mental age matched controls, as Shaked and Yirmiya (2004) found that the use of clinical comparison groups lead to equivocal and biased findings within the ToM literature. We also used a wide age range so as not to bias our results in a particular direction (Shaked and Yirmiya 2004). The second aim was to explore whether ToM abilities in WS were variable and, more specifically, to compare ToM abilities in two previously defined subgroups of WS. Differences in ToM abilities between these two subgroups would provide further validation that these groups are, in fact, distinct rather than simply differing in general degree of impairment.

In relation to our first aim, results suggested impaired understanding of false belief in the WS group. On average, the WS group performed significantly below normal mental age matched controls in sequencing false belief stories, but performed similarly to this group on all other story types including understanding of intention, unrealized goals and pretence, as well as understanding of social script knowledge and physical cause and effect reasoning. This suggests that, on average, false belief understanding in WS

is significantly below the level of their overall mental age or significantly below the level expected on the basis of their general intellectual abilities.

Our finding of a deficit in false belief understanding within WS is consistent with Tager-Flusberg et al. (1997), but is inconsistent with other studies which found false belief understanding to be intact in WS (e.g. Karmiloff-Smith et al. 1995). These previous studies used restricted age ranges (adults or children) and neither used normal mental age matched controls for comparison. Sample characteristics and choice of control group may account for the inconsistent findings within the literature on ToM abilities in WS.

Heterogeneity

The second aim of the paper was to explore whether ToM abilities varied amongst individuals with WS and, more specifically, whether ToM abilities differed for Subgroup 1 and Subgroup 4 from Porter and Coltheart (2005). We found that the deficit in understanding false belief in WS was specific to a subset of individuals, namely those in WS Subgroup 4, providing further validation that these subgroups are in fact distinct; we also found differences in understanding of social script knowledge between these two subgroups.

ToM Abilities

Individuals in Subgroup 4 display a strength in verbal skills compared to their overall mental age (see Porter and Coltheart 2005), so using a nonverbal task meant that these individuals were unable to use their good verbal skills to pass the ToM task; the picture sequencing task was perhaps a more sensitive measure of ToM abilities within this group.

At this stage, it is difficult to determine whether these results suggest a specific deficit in understanding false belief or a more generalized difficulty with representational understanding of mind. Although Subgroup 4 performed well on pretense, unrealized goal and intention stories, which may initially suggest a specific difficulty in understanding false belief, these particular stories, although designed to assess understanding of mind as a representational medium, may not in fact do so. That is, the unrealized goal and intention stories might, perhaps, be understood using a simple desire-goal folk psychology and the pretense stories might require only an appreciation of 'substitution pretense' (one object being used in the place of another rather than one object representing another: e.g. see Perner 1991). Pretense, unrealized goal and intention

stories may, therefore, be sequenced correctly irrespective of whether a person possesses a representational ToM or not. Either way, results indicate a theory of mind deficit in Subgroup 4 which may or may not be specific to understanding false belief.

Social Abilities

In addition to heterogeneity in ToM abilities, results also suggest variability in social script knowledge within WS. Subgroup 4 performed significantly below the level of Subgroup 1 when sequencing social script 2 stories. Social script 2 stories differ from social script 1 stories in that social script 2 stories display an interaction between two people, whereas social script 1 stories involve only one person carrying out a routine activity. Social script 2 stories are therefore likely to be more sensitive to difficulties in day-to-day social interactions.

Although there was no significant difference between Subgroup 4 and their mental age matched control group in terms of sequencing social script 2 stories, this result nevertheless suggests more general social deficits in Subgroup 4 that extend beyond their ToM impairment. In line with this idea, Porter (2004) also found that emotion perception abilities were significantly worse in Subgroup 4 when compared to Subgroup 1 and when compared to normal mental age matched controls and individuals with Down syndrome.

Results suggesting ToM impairments and possibly general difficulties with social understanding in Subgroup 4 are consistent with results from a separate qualitative study (in preparation) where we administered the Child Behavior Checklist (Achenbach and Rescorla 2001) to parents of this cohort, a questionnaire where parents (or guardians) are requested to rate their child's behavioral, emotional and social well-being using a 3-point scale (0 = Not True, 1 = Somewhat or Sometimes True and 2 = Very True or Often True). We found a higher proportion of ratings suggesting social difficulties in day-to-day functioning for individuals in Subgroup 4 compared to individuals in Subgroup 1 (38% of ratings for Subgroup 4 versus 13% for Subgroup 1 were in the clinical range of significance).

General Sequencing Abilities, Developmental Delay, Mental Age and Chronological Age: Potential Confounds?

ToM and social differences between Subgroups 1 and 4 could not be accounted for by general sequencing abilities, chronological age, short-term memory, nonverbal reasoning or visual perceptual abilities, nor could they be accounted for by differences in mental age. First, subgroups displayed similar general sequencing abilities on mechanical or control stories. Second, there were no significant differences in chronological age, short-term memory, nonverbal reasoning or visual perceptual abilities between subgroups and, third, the above differences remained significant when mental age (and chronological age) were covaried.

Additional preliminary analyses also suggest that Subgroup 4's difficulties with false belief understanding and social script knowledge cannot be attributed to greater developmental delay or intellectual disability. We compared participants from Subgroup 4 to a group of individuals with Down Syndrome matched individually to members of WS Subgroup 4 on gender, chronological age and mental age and found that the Down syndrome group performed significantly better in sequencing both false belief ($F(1, 16) = 11.82$, $p < .01$, mean = 5.00, s.d. = 0.90 for the DS group) and social script 2 ($F(1, 16) = 10.05$, $p < .0101$, mean = 3.11, s.d. = 1.03 for the DS group) stories. A similar analysis between members from Subgroup 1 and a group of Down syndrome individuals indicated no significant differences between groups on either the false belief ($F(1, 16) = 0.03$, $p > .1$, mean = 3.41, s.d. = 0.66 for the DS group) or the social script 2 ($F(1, 16) = 1.44$, $p > .1$, mean = 4.44, s.d. = 1.44 for the DS group) stories.

Summary

Our research, using a non-verbal task to assess ToM, indicates impaired false belief understanding in one select group of WS individuals, providing further evidence of distinct subgroups within the syndrome who differ in *kind* rather than *degree* of impairment. There are some indications that the deficits in social understanding in this subgroup go beyond an impaired ToM. Future studies include recruiting a new cohort of WS individuals and attempting to replicate the various patterns of strength and weakness displayed by our two existing subgroups; this is an equally important way to validate these subgroups. Finally, our findings accord with Shaked and Yirmiya (2004) in highlighting how task selection, choice of comparison group and heterogeneity can greatly affect the outcomes of a study.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Baron-Cohen, S. (1993). From attention-goal psychology to belief-desire psychology: The development of a theory of mind, and its dysfunction. In D. J. Cohen (Ed.), *Understanding other minds* (pp. 59–83). New York: Oxford University Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioral, and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4*, 113–125.

Bellugi, U., Lichtenberger, L., Mills, D., Galaburda, A., & Korenberg, J. R. (1999). Bridging cognition, the brain and molecular genetics: Evidence from Williams syndrome. *Trends in Neuroscience, 22*(5), 197–207.

Borg, I., Delhanty, J. D., & Baraitserer, M. (1995). Detection of hemizygosity at the elastin locus by FISH analysis as a diagnostic test in both classical and atypical cases of Williams syndrome. *Journal of Medical Genetics, 32*(9), 692–696.

Davies, M., Udwin, O., & Howlin, P. (1998). Adults with Williams syndrome. Preliminary study of social, emotional and behavioral difficulties. *The British Journal of Psychiatry, 172*, 273–276.

Doyle, T. F., Bellugi, U., Korenberg, J. R., & Graham, J. (2004). "Everybody in the world is my friend": Hypersociability in young children with Williams syndrome. *American Journal on Medical Genetics, 124A*, 263–273.

Howlin, P., Davies, M., & Udwin, O. (1998). Cognitive functioning in adults with Williams syndrome. *Journal of Child Psychology and Psychiatry, 39*(2), 183–189.

Jones, W., Bellugi, U., Lai, Z., Chiles, M., Reilly, J., Lincoln, A., & Adolphs, R. (2000). Hypersociability in Williams syndrome. *Journal of Cognitive Neuroscience, 12*(3), 30–46.

Karmiloff-Smith, A., Klima, E., Bellugi, U., Grant, J., & Baron-Cohen, S. (1995). Is there a social module? Language, face processing, and theory of mind in individuals with Williams syndrome. *Journal of Cognitive Neuroscience, 7*(2), 196–208.

Langdon, R., Michie, P. T., Ward, P. B., McConaghy, N., Catts, S. V., & Coltheart, M. (1997). Defective self and/or other mentalising in Schizophrenia: A cognitive neuropsychological approach. *Cognitive Neuropsychiatry, 2*(3), 167–193.

Laws, G., & Bishop, D.V. (2004). Pragmatic language impairment and social deficits in Williams syndrome: a comparison with Down's syndrome and specific language impairment. *International Journal of Language & Communication Disorders, 39*(1), 45–64.

Leslie, A. M., & Frith, U. (1990). Prospects for a cognitive neuropsychology of autism: Hobson's choice. *Psychological Review, 97*, 122–131.

McKusick, V. (1988). *Mendelian inheritance in man: Catalogues of autosomal dominant, autosomal recessive and X-linked phenotypes*. Baltimore: John Hopkins Press.

Morris, P., & Sigman, M. (1988). Natural history of Williams syndrome: Physical characteristics. *Journal of Pediatrics, 113*, 318–326.

Perner, J. (1991). *Understanding the representational mind*. Cambridge MA: MIT Press.

Plesa-Skwerer, D., Verbalis, A., Schofield, C., Faja, S., & Tager-Flusberg, H. (2006). Social-perceptual abilities in adolescents and adults with Williams syndrome. *Cognitive Neuropsychology, 23*(2), 338–349.

Porter, M. A. (2004). *PhD Thesis: The cognitive neuropsychology of Williams syndrome*. Sydney, Macquarie University.

Porter, M., & Coltheart, M. (2005). Cognitive heterogeneity in Williams syndrome. *Developmental Neuropsychology, 27*(2), 275–306.

Porter, M. A., & Coltheart, M. (2006). Global and local processing in Williams syndrome, autism, and Down syndrome: Perception, attention and construction. *Developmental Neuropsychology, 30*(3), 771–790.

Shaked, M., & Yirmiya, N. (2004). Matching procedures in autism research: Evidence from meta-analytic studies. *Journal of Autism and Developmental Disorders, 34*(1), 35–40.

Sullivan, K., & Tager-Flusberg, H. (1999). Second-order belief attribution in Williams syndrome: Intact or impaired. *American Journal on Mental Retardation, 104*(6), 523–532.

Tager-Flusberg, H. (1999). *Neurodevelopmental disorders*. Cambridge, Massachusetts: MIT press.

Tager-Flusberg, H., Boshart, J., & Baron-Cohen, S. (1998). Reading the windows to the soul: Evidence of domain specific sparing in Williams syndrome. *Journal of Cognitive Neuroscience, 10*(5), 631–639.

Tager-Flusberg, H., & Sullivan, K. (2000). A componential view of theory of mind: Evidence from Williams syndrome. *Cognition, 76*, 59–89.

Tager-Flusberg, H., Sullivan, K., & Boshart, J. (1997). Executive functions and performance on false belief tasks. *Developmental Neuropsychology, 13*(4), 487–493.

Udwin, O. (1990). A survey of adults with Williams syndrome and ideopathic infantile hypercalcaemia. *Developmental Medicine & Child Neurology, 32*(2), 129–141.

Ypsilanti, A., Grouios, G., Alevriadou, A., & Tsapkini, K. (2005). Expressive and receptive vocabulary in children with Williams and Down syndromes. *Journal of Intellectual Disability Research, 49*(5), 353–364.

Williams, J., Barrat-Boyes, B., & Lowe, J. (1961). Supravalvar aortic stenosis. *Circulation, 24*, 1311–1381.

Woodcock R. W., & Johnson M. B. (Eds.) (1989, 1990). *Woodcock–Johnson psycho-educational battery—revised*. Itasca, IL: Riverside Publishing.

Woodcock, R. W., & Mather, N. (1989b,1990b). WJ-R tests of cognitive ability—standard and supplemental batteries: Examiner's manual, *Woodcock–Johnson psycho-educational battery—revised*. Itasca, IL: Riverside Publishing.