

# Profiles of executive function in parents and siblings of individuals with autism spectrum disorders

D. Wong<sup>\*,†</sup>, M. Maybery<sup>†</sup>, D. V. M. Bishop<sup>†,‡</sup>,  
A. Maley<sup>†,§</sup> and J. Hallmayer<sup>§,¶,\*\*</sup>

<sup>†</sup>School of Psychology, University of Western Australia, Crawley, Australia, <sup>‡</sup>Department of Experimental Psychology, University of Oxford, Oxford, UK, <sup>§</sup>Centre for Clinical Research in Neuropsychiatry, University of Western Australia, <sup>¶</sup>School of Psychiatry and Clinical Neurosciences, University of Western Australia, Crawley, Australia, and <sup>\*\*</sup>Department of Psychiatry & Behavioral Science, Stanford University, Stanford, CA, USA

\*Corresponding author: Dr D. Wong, School of Psychology, University of Western Australia, 35 Stirling Highway, Crawley WA 6009, Australia. E-mail: danaw@graduate.uwa.edu.au

**Delineation of a cognitive endophenotype for autism is useful both for exploring the genetic mechanisms underlying the disorder and for identifying which cognitive traits may be primary to it. This study investigated whether first-degree relatives of individuals with autism spectrum disorders (ASDs) demonstrate a specific profile of performance on a range of components of executive function (EF), to determine whether EF deficits represent possible endophenotypes for autism. Parents and siblings of ASD and control probands were tested on EF tasks measuring planning, set-shifting, inhibition and generativity. ASD parents showed poorer performance than control parents on a test of ideational fluency or generativity, and ASD fathers demonstrated a weakness in set-shifting to a previously irrelevant dimension. ASD siblings revealed a mild reduction in ideational fluency and a weakness in non-verbal generativity when compared with control siblings. Neither ASD parents nor siblings displayed significant difficulties with planning or inhibition. These results indicated that the broad autism phenotype may not be characterized primarily by impairments in planning and cognitive flexibility, as had been previously proposed. Weaknesses in generativity emerged as stronger potential endophenotypes in this study, suggesting that this aspect of EF should play a central role in cognitive theories of autism. However, discrepancies in the EF profile demonstrated by parents and siblings suggest that factors related to age or parental responsibility may affect the precise pattern of deficits observed.**

**Keywords:** Autism, broad autism phenotype, cognitive endophenotype, executive function, generativity, inhibition, planning, set-shifting

Received 7 November 2005, revised 28 November 2005, accepted for publication 29 November 2005

Twin concordance studies have now firmly established the genetic basis for autism (Bailey *et al.* 1995; Folstein & Rutter 1977; Steffenburg *et al.* 1989), with heritability estimated to be greater than 90% (Bailey *et al.* 1995). An elevated recurrence risk for autism has been observed in siblings, averaging around 2.2% across studies (Szatmari *et al.* 1998) compared with a population base rate of around 0.1% (Fombonne 2003). An increased rate of other autism spectrum disorders (ASDs) such as Asperger syndrome in twins and other relatives of individuals with autism has also been reported (Bailey *et al.* 1995; Bolton *et al.* 1994; Le Couteur *et al.* 1996). However, variability in phenotypic expression within monozygotic (MZ) twin pairs (e.g. Le Couteur *et al.* 1996; MacLean *et al.* 1999), along with the rapid decrease in risk rates from MZ twins to dizygotic twins and siblings to more distant relatives (DeLong & Dwyer 1988), indicates that the genetic mechanisms of autism are not Mendelian in nature and are likely to involve epistatic interactions among several genes (Pickles *et al.* 1995; see also Risch *et al.* 1999).

Numerous studies have demonstrated that milder forms of autistic symptomatology, which do not meet criteria for an ASD diagnosis, are frequently exhibited in relatives of individuals with ASDs (Bishop *et al.* 2004; Bolton *et al.* 1994; Landa *et al.* 1992; Le Couteur *et al.* 1996; Piven *et al.* 1994), giving rise to the notion of a spectrum of autistic traits or 'broad phenotype' of autism. Exploration of the broad phenotype has been a useful method of identifying possible genetic mechanisms (e.g. whether traits are contributed by both parents) and of increasing the power to identify genes linked with autism by increasing the number of 'affected' individuals available for analysis (Folstein *et al.* 1998). Studying the characteristics of the broad phenotype is also a useful method of identifying which traits may be central to autism, as a core characteristic should show increased prevalence in individuals who carry the autism genotype, including those with a milder variant which does not meet criteria for an ASD diagnosis (Bailey *et al.* 1996).

However, defining the boundaries of the broad phenotype has proven to be a challenging task. Family history interview methods have been useful in documenting the presence of social and communicative impairments and repetitive behaviours in relatives of autistic probands (Bailey *et al.* 1995; Bolton *et al.* 1994; Le Couteur *et al.* 1996), but individual behavioural signs suffer from the problem of low diagnostic specificity (Bailey *et al.* 1998) and are therefore of limited utility as unique markers for the broad autism phenotype. Researchers have therefore concurrently searched for a

more basic subclinical marker of the autism genotype – or ‘endophenotype’ (see Gottesman & Gould 2003) – at the level of cognition.

The search for a cognitive endophenotype for autism began with investigations of IQ and academic ability in first-degree relatives, but failures to find consistent patterns of difference led to a narrowing in focus to the identification of specific cognitive deficits as potential endophenotypes, driven by current cognitive theories of autism. Current research on core cognitive deficits in autism has identified impairment in executive function (EF) as one of a number of promising candidates (Hill 2004). EF is an umbrella term covering a number of related but distinct high-level cognitive capacities which help guide and control purposeful behaviour towards attainment of a goal (e.g. Welsh & Pennington 1988). These capacities include planning, set-shifting or cognitive flexibility, inhibition, working memory, generativity, strategy formation and self-monitoring. EFs are often considered to be synonymous with the functions of the prefrontal cortex, even though EFs can be intact when the frontal lobes are damaged (e.g. Eslinger & Damasio 1985) and non-frontal damage can produce EF deficits (e.g. Anderson *et al.* 1991).

A number of studies have now shown impairment of individuals with autism compared with age and IQ-matched controls on tasks tapping a range of EF components, including cognitive flexibility or attentional shifting (Goldstein *et al.* 2001; Hughes & Russell 1993; Hughes *et al.* 1994; Ozonoff & Jensen 1999; Ozonoff & McEvoy 1994; Ozonoff *et al.* 2004), planning (Hughes 1996; Hughes *et al.* 1994; Ozonoff & Jensen 1999; Ozonoff & McEvoy 1994; Ozonoff *et al.* 2004) and generativity (Boucher 1988; Craig & Baron-Cohen 1999; Turner 1999; Wong *et al.* 2003). These EF deficits do not appear to be attributable to impairments in more basic attentional processes, such as sustained or selective attention or basic attentional capacity (e.g. Garcia-Villamisar & Della Sala 2002; Goldstein *et al.* 2001). These findings resulted in the hypothesis that EF deficits may be primary in autism (e.g. Hughes & Russell 1993; Ozonoff *et al.* 1991; Russell 1997) and can explain a number of aspects of autistic symptomatology including repetitive behaviours (Turner 1997), impairments in social interaction (e.g. Berger *et al.* 2003) and communicative abnormalities (Bishop & Norbury 2005).

Critics of this hypothesis have noted that EF impairments are seen in other disorders, such as attention-deficit hyperactivity disorder (ADHD) (Geurts *et al.* 2004; Sergeant *et al.* 2002), schizophrenia (Hoff & Kremen 2003) and Tourette’s syndrome (Channon *et al.* 1992) and are therefore not unique to autism. There are two ways of viewing such findings. Proponents of the hypothesis that EF deficits are primary to autism have contended that different disorders are characterized by different profiles of impairment on the various components of EF. Ozonoff and colleagues (Ozonoff 1997; Ozonoff & Jensen 1999) have specifically proposed that autism is characterized by deficits in cognitive flexibility and planning but spared inhibitory capacity (whereas children

with ADHD, for example, have intact flexibility but impaired inhibition). This need to distinguish EF profiles specific to certain disorders has placed added importance on designing and choosing EF tasks which are relatively ‘pure’ and where underlying processes are not easily confounded. Another possibility is that the same EF deficits may have different consequences for behaviour in different disorders, depending on which other cognitive deficits co-occur (e.g. autism may result when EF deficits coincide with other deficits in theory of mind, face processing etc.).

Investigating EF components in relatives of individuals with ASDs is another strategy for determining which aspects of EF play a central role in the aetiology of autism (Hughes *et al.* 1999). Dawson and colleagues have argued that EF is one of six candidate broad phenotype measures worthy of focus in genetic studies of autism (Dawson *et al.* 2002). In addition, there is now some evidence that EF deficits are heritable (Anokhin *et al.* 2003; Coolidge *et al.* 2004), consistent with the notion that executive dysfunction may characterize the broad autism phenotype. Studies of EF in relatives of individuals with autism have so far concentrated on measures of planning and set-shifting, domains in which probands with autism have consistently been found to be impaired. Significantly poorer performance by parents of individuals with autism compared with control parents (including parents of children with learning disabilities and Down syndrome) on Tower tasks (i.e. the Towers of Hanoi and London and the Stockings of Cambridge test from the CANTAB battery) was found by Hughes *et al.* (1997) and Piven and Palmer (1997), and the same result in siblings was obtained by Hughes *et al.* (1999) and Ozonoff *et al.* (1993). In Hughes *et al.*’s (1997) study, the difference in planning ability was restricted to fathers only, and in both of the studies by Hughes *et al.* (1997, 1999), a planning deficit was restricted to a subset of the relatives of autistic probands, with group differences only emerging clearly when the proportions of participants showing a deficit were compared. Findings of no group differences on the Wisconsin card-sorting test in parents (Szatmari *et al.* 1993) or siblings (Ozonoff *et al.* 1993) were initially suggestive of intact cognitive flexibility in relatives of individuals with autism. However, two subsequent studies using the intradimensional, extradimensional (IDED) set-shifting task found that a subset of both parents and siblings of autistic probands demonstrated difficulties with the extra-dimensional shift stage of the task (Hughes *et al.* 1997, 1999). The two studies by Hughes *et al.* (1997, 1999) also incorporated working memory measures, but results were variable across parents and siblings and suggested that a working memory deficit is not as reliable or unique a characteristic of the broad phenotype as problems with planning and set-shifting.

Even though evidence of a significant generativity impairment in autism has been found in several studies (Boucher 1988; Lewis & Boucher 1988; Turner 1999), and the apparent sparing of inhibitory capacity has been proposed to

distinguish the EF profile of individuals with autism from other disorders, these components of EF have not been well studied in relatives. Hughes *et al.* (1999) included a verbal generativity task (word fluency) in their battery with siblings and found both a significant group difference overall in the number of words generated and a higher proportion of 'low fluency' participants in the autism sibling group, indicating that a weakness in generativity may be an important characteristic of the broad phenotype. As yet, inhibition has not been tested in parents and siblings of individuals with autism. Furthermore, the interaction between inhibition and working memory has received some attention in modelling EF development (Beveridge *et al.* 2002; Roberts & Pennington 1996), and Russell (1997) has specifically proposed that children with autism only demonstrate EF deficits on tasks requiring both capacities; however, this interaction is yet to be directly examined either in probands with autism or in their relatives.

In this study, a large sample of both parents and siblings of individuals with ASDs were tested on measures of inhibition (and the inhibition/working memory interaction) and generativity in addition to measures of planning and set-shifting. The aims of the study were (1) to identify potential endophenotypes for autism by specifying the EF profile which characterizes the broad autism phenotype and (2) to determine which types of EF deficits may be core characteristics of autism (as core characteristics should occur in first-degree relatives of individuals with an ASD to a greater degree than in control samples but to a lesser degree than in individuals who express the ASD phenotype). Emphasis was placed on choosing relatively 'pure' tasks where the target EF component could be isolated for analysis. It was hypothesized that

parents and siblings of individuals with ASDs would show mild impairments in planning, set-shifting, generativity and the inhibition/working memory interaction but not in inhibition.

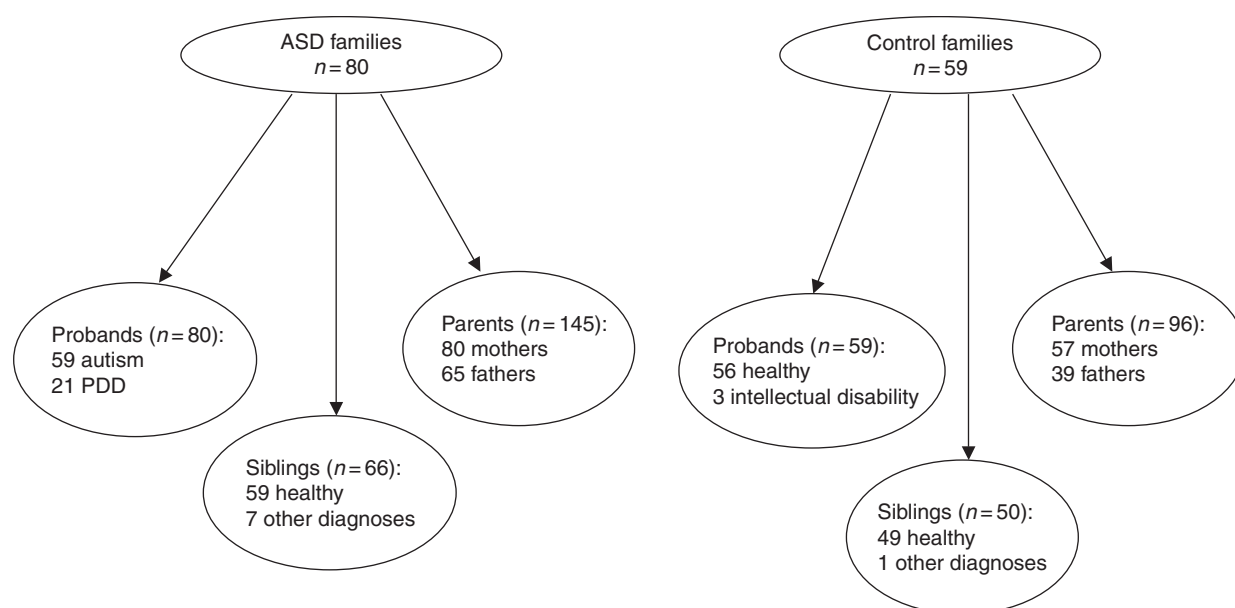
## Materials and methods

### Participants

The data reported here come from the Western Australia Family Study of ASDs (WAFSASD). The study was approved by the Human Research Ethics Committee at the University of Western Australia. Figure 1 shows the composition of the family groups from which our sibling and parent samples were drawn.

### ASD group

Families with at least one child with an ASD ( $n = 80$ ) were recruited through autism centers and support groups. Families were excluded if any family member had a diagnosis of genetic abnormality or other serious organic condition which may have been causally related to autism. Probands (children with an ASD) had received a diagnosis of autism ( $n = 55$ ), Asperger syndrome ( $n = 25$ ) or pervasive developmental disorder not otherwise specified (PDDNOS) ( $n = 11$ ) from a health professional (e.g. paediatrician, psychiatrist). The presence of autistic symptomatology in at least one domain was then verified using the Autism Diagnostic Interview – Revised (ADI-R; Lord *et al.* 1994). The ADI-R was administered in all cases by an accredited trained interviewer, who had passed reliability checks as part of the accreditation process. Fifty-nine probands met standard



**Figure 1: The final sample of autism spectrum disorder (ASD) and control families.**

ADI-R criteria for autism, and 21 probands scored above threshold on one or two of the three symptom domains assessed by the ADI-R (communication, social interaction and repetitive behaviour/restricted interests), with these latter probands being designated as cases of PDD. Table 1 summarizes the number of ASD probands receiving each clinical diagnosis (autism, Asperger syndrome or PDDNOS) and the mean, SD and range of ADI-R scores for each of these groups in each symptom domain. All of the 21 cases designated as PDD (on the basis of the ADI-R) met the ADI-R threshold on either the social or the communication domain and were within 1–3 points on the other. Two cases that were referred but did not score above threshold in any domain of the ADI-R were excluded from the study. For the analyses reported here, families of the autism and PDD cases were treated as a single group.

**Siblings.** There were 108 siblings of the final sample of 80 ASD probands. All siblings who were available and willing to participate were recruited to the project. Autistic symptomatology in siblings was assessed using the Social Communication Questionnaire (SCQ; Berument *et al.* 1999), and the ADI-R was administered to any sibling scoring above 10 (this conservative cutoff point was adopted to ensure that any individuals with mild ASD symptomatology were administered the ADI-R). There were 10 siblings who met ADI-R criteria for autism ( $n = 2$ ) or PDD ( $n = 8$ ), coming from nine families (in one family there were two siblings meeting criteria for an ASD). In these nine families, the child who was referred to the project first was considered the proband. The 10 siblings meeting criteria for an ASD were excluded from all analyses, leaving 98 siblings in the sample. In addition, there were several families with more than one sibling in the sample. To avoid possible confounds due to the effect of family relatedness (i.e. the fact that several of the siblings were genetically related and therefore were not independent participants), we randomly chose one sibling from each family for inclusion in the final ASD sibling sample ( $n = 66$ ). Of these 66 ASD siblings, there were seven with non-ASD clinical diagnoses (five with ADHD, one with dyspraxia and one with dyslexia).

**Parents.** Of the probands meeting ADI-R criteria for autism or PDD, there were 145 parents who participated in the study, including 80 mothers and 65 fathers.

#### Control group

Recruitment of control families was mainly achieved through schools. Families of children with mild intellectual disabilities were also recruited through the Western Australian Disability Services Commission, as controls for families of ASD probands with low performance IQ (PIQ). Exclusion criteria were the same as for the ASD families, with the addition of a known or suspected ASD in any family member. As for siblings in the ASD group, the SCQ was used to screen for symptoms of autism in control probands and the ADI-R administered for any proband scoring above 10. Five participants, all of whom had a mild intellectual disability, met at least partial criteria for autism on the ADI-R, and they and their families were excluded from further analysis, leaving 59 control families.

**Siblings.** Sixty-seven siblings of control probands were recruited to the study. No sibling in this group had clinical diagnoses of ASDs or exceeded cutoff on the SCQ. As was the case for the ASD siblings, when there was more than one sibling in a family, one sibling was randomly chosen for inclusion in the final control sibling sample ( $n = 50$ ). Of this final sample, one control sibling had a non-ASD clinical diagnosis (epilepsy).

**Parents.** There were 96 parents (57 mothers and 39 fathers) of the 59 control probands.

#### Age, IQ and gender comparisons

Descriptive statistics for the groups are presented in Table 2. The ASD and control siblings were matched on chronological age,  $t(114) = 0.48$ ,  $P > 0.1$ , PIQ,  $t(114) = 0.46$ ,  $P > 0.1$  and verbal IQ (VIQ),  $t(114) = 1.26$ ,  $P > 0.1$ . All siblings had a PIQ and VIQ of 60 or above. There was a higher proportion of girls in the control sibling group (70% vs. 47% in the ASD sibling group),  $\chi^2(1, n = 116) = 6.15$ ,  $P < 0.05$ . This is probably because in attempting to select proband samples

**Table 1:** Autism Diagnostic Interview – Revised (ADI-R) scores in each domain for autism spectrum disorder (ASD) probands with clinical diagnoses of autism, Asperger syndrome and PDDNOS

	Autism ( $n = 55$ )	Asperger syndrome ( $n = 14$ )	PDDNOS ( $n = 11$ )
ADI-R social domain			
Mean (SD)	23.29 (4.97)	18.14 (5.95)	17.18 (8.13)
Range	9–30	9–28	7–28
ADI-R communication domain			
Mean (SD)	16.42 (4.57)	13.71 (4.30)	15.82 (4.31)
Range	6–24	7–21	11–23
ADI-R interests domain			
Mean (SD)	5.40 (2.63)	4.71 (2.70)	3.55 (3.17)
Range	0–12	1–9	0–8

**Table 2:** Characteristics of probands, siblings and parents in the autism spectrum disorder (ASD) and control groups

	<i>n</i>	Mean age (years, months)	Male : Female	Mean PIQ	Mean VIQ
Probands					
ASD	80	10, 5	70:10	82.8	75.9
Control	59	9, 0	43:16	101.1	105.3
Siblings					
ASD	66	10, 8	35:31	108.9	104.6
Control	50	10, 4	15:35	107.4	108.0
Parents					
ASD	145	41, 5	65:80	105.1	102.5
Control	96	40, 11	39:57	108.1	104.8

PIQ, performance IQ; VIQ, verbal IQ.

matched on gender, often the male child in the family was selected as a control proband, resulting in a higher proportion of females in the control sibling group. Gender was included as an independent variable in all analyses.

Parents of ASD and control probands were matched on age,  $t(239) = 0.63$ ,  $P > 0.1$ , PIQ,  $t(239) = 1.36$ ,  $P > 0.1$  and VIQ,  $t(239) = 1.09$ ,  $P > 0.1$ . All parents had a PIQ and VIQ above 60. Although there was a similar ratio of mothers to fathers in the ASD and control groups,  $\chi^2(1, n = 241) = 0.42$ ,  $P > 0.1$ , gender was included as an independent variable (IV) in analyses of group differences because of the possibility that any differences could be limited to one gender. ASD and control parents were also matched in socioeconomic status. This was assessed using education data from mothers and fathers, which were coded using the following system: 1 = up to year 10 of high school; 2 = up to year 12 of high school; 3 = diploma, trade certificate, apprenticeship or other traineeship and 4 = university degree. A  $\chi^2$  analysis revealed comparable education levels for ASD and control parents (using the highest code from each family),  $\chi^2(3, n = 132) = 5.99$ ,  $P > 0.1$ .

## Measures

### IQ measures

Two verbal subtests (Vocabulary and Similarities) and two performance subtests (Picture Completion and Object Assembly) from the Wechsler scales (WPPSI-R, WISC-III or WAIS-III) were used to estimate VIQ and PIQ using pro-rating scaled scores. These subtests were chosen because they load highly on the verbal and performance factors and are the least similar of the Wechsler subtests to cognitive measures in the test battery.

### EF measures

Most of the EF tasks were suitable for both children and adults; however, certain tasks were not appropriate for young children, and one was not suitable for adults. The age range for each EF task is listed below.

**Planning: Tower of London (administered to all ages).** The Tower of London (ToL) is thought to be a measure of planning (Culbertson & Zillmer 1998a; Shallice 1982). Our

version was based on the task outlined by Culbertson and Zillmer (1998b), except that the floor was lowered by including 1- and 2-move items (rather than beginning with 3-move items); there were four items at each level of difficulty instead of three; the instructions provided more encouragement to plan moves in advance in order to obtain a purer index of planning ability that was less affected by a failure to grasp the need to plan, and scores were adjusted (see below) to account for rule violations.

Participants are presented with three wooden posts of descending heights mounted on a wooden base. Three coloured discs (red, black and white) are placed on the posts in a standard starting position. The task is to rearrange the discs on the posts so that the new configuration corresponds to a pattern presented on a 21 × 15-cm stimulus card. The participants are informed that this must be accomplished in the minimum number of moves, which is told to them verbally as well as being written at the top of the stimulus card. They are also told to adhere to the following rules: (1) to use only one hand to move the discs; (2) to move only one disc at a time; (3) to not place disks on the board or table – only on the posts and (4) to not put more discs on a post than it will hold (i.e. 1, 2 and 3 for the three posts). Participants are given one 1-move and two 2-move practice examples, during which feedback is provided.

Items range in difficulty from one to seven moves, with four items at each level of difficulty. Participants aged between 4 and 13 years begin with 1-move items, and participants aged 14 years and over begin with 3-move items (and are given automatic credit for easier items if they complete at least two 3-move items in the minimum number of moves). If the first item at a new level of difficulty is failed (either by breaking the rules or using too many moves), the correct solution is demonstrated. There is a time limit of 120 seconds on each item. Testing is discontinued if participants fail all items at a particular level of difficulty. Remaining items are assumed to have been failed and are assigned the maximum total number of moves (i.e. 20).

Two indices of performance were calculated for the ToL task. First, for each item the adjusted extra moves score was

coded as 0 if the solution was achieved in the minimum number of moves and without rule violation; as 1 if 1–5 additional moves were taken (with two moves added for each rule violation, because breaking a rule often enabled the item to be completed in fewer moves than would otherwise be possible) or as 2 if six or more moves beyond the minimum number were required. These scores were summed for all items. The second ToL index was the mean number of rule violations per block of items administered.

*Set-shifting: IDED set-shifting task (7 years and older).* Developed by Owen *et al.* (1993), this task contains two conditions to distinguish whether impairments in attentional set-shifting are caused by an inability to release attention from a relevant stimulus dimension (perseveration) or an inability to re-engage attention to a previously irrelevant dimension (learned irrelevance). In both conditions, each trial consists of two patterns appearing on a computer touchscreen (their positions randomly varying between four rectangular boxes to the top, bottom, left and right of screen centre), and the participant is required to choose which one is 'correct' according to an unspecified rule, with feedback provided (see Owen *et al.* 1993). Each condition comprises eight stages presented in the same fixed order. These eight stages involve a simple discrimination (SD), in which the participant chooses between two stimuli; a reversal (SDR), in which the participant is required to reverse the rule just learned and respond to the previously incorrect stimulus; then a compound discrimination (CD), in which an additional stimulus dimension is introduced, and another reversal (CDR); then an intradimensional shift (IDS), where new exemplars are introduced but the rule for choosing the correct stimulus remains the same, and a reversal (IDR) and finally an extra-dimensional shift (EDS), which requires a new rule to be adopted, and a reversal (EDR). In the EDS stage of the *Perseveration* condition, the previously irrelevant stimulus dimension (lines) is replaced by a new stimulus dimension which becomes relevant (solidity), and the previously relevant dimension (shape) becomes irrelevant. Thus, participants must shift their attention from a previously relevant to a new stimulus dimension, and failure reflects perseveration to the previously relevant dimension. In the EDS stage of the *Learned Irrelevance* condition, the relevant dimension (colour) is replaced by a previously irrelevant dimension (number), and a new dimension (size) becomes the irrelevant dimension. Hence, participants must shift their attention to a previously irrelevant stimulus dimension, and failure reflects learned irrelevance associated with the previously irrelevant dimension.

Failure to achieve the criterion of six consecutive correct responses within 50 trials at any one stage results in discontinuation of the condition. There is a 1000 milliseconds interval between successive trials. The two conditions are separated by at least 30 min of unrelated tests. In the current

study, the dimensions used in each condition (e.g. shape, solidity) were consistent across participants, and the *Perseveration* condition was always presented first.

The index of performance was the number of *errors to criterion* within each stage of the task. If the test was discontinued because the criterion of six consecutive correct responses was not met within 50 trials, a value of 25 (the value expected with random responding) was assigned for the errors to criterion score for subsequent stages of the task which were not administered.

*Inhibition/working memory: Response Inhibition and Load (RIL) task (7 years and older).* Created by the authors, this non-verbal computerized test of inhibition and its interaction with working memory is presented on a touch screen. In Condition 1 (control), either a pink or green 5-cm stimulus circle appears at the top of the screen for 250 milliseconds, and then two 3.5-cm response circles, one pink and one green, appear simultaneously at the bottom left and right corners of the screen. Participants are instructed to touch the response circle which is the *same colour* as the stimulus circle. Participants have 4 seconds to respond (using their preferred hand) before the next trial begins. An equal number of pink and green stimulus circles are presented, and the order of presentation is random. The response circles also change sides randomly from trial to trial. Performance indices are the percentage of errors, and the median RT for correct trials. Condition 2 (inhibition) is identical to Condition 1 except that the colours of the stimulus and response circles are purple and yellow, and the participant is required to touch the response circle *opposite in colour* to the stimulus circle. In Condition 3 (working memory load), instead of the stimulus being a circle, it is either a square, a triangle or a cross. As in Condition 2, participants must touch the response circle opposite in colour to the stimulus shape (the colours in this condition are orange and grey). However, at random intervals between trials, the three shapes are displayed on the screen, and the participant must touch the shape which was presented in the most recent trial. This occurs for 25% of the trials. The participant must therefore remember the shape of the stimulus as well as inhibit the prepotent tendency to respond to the same colour. Performance indices for the responses to the colour of the stimulus are identical to those in Conditions 1 and 2. For the questions about the shape of the stimulus, performance is measured by the percentage of errors. In each condition, participants perform seven practice trials, during which any errors are corrected, and then 60 critical trials, during which every third error is pointed out and the task rules reiterated. The unfilled intertrial interval is 1000 ms in all conditions.

*Generativity: Pattern Meanings (all ages).* Pattern meanings is a measure of 'ideational fluency' or generativity. The stimuli are five meaningless line drawings, taken from Wallach and Kogan (1965) and also used by

Turner (1999), which were printed on 14.3 × 9.2-cm cards. An additional drawing was used for a practice item. The procedure was identical to that used by Turner (1999), except that participants were given 90 seconds instead of 150 seconds to generate responses for each item. This shorter interval was introduced following pilot testing, in which it was found that participants tended to produce only a very small number of responses in the last minute and would often become restless or inattentive. Participants are told they will be shown some patterns and asked to think of all the things each looks like or could be. They are then shown the practice card and asked 'What could this be?'. Any appropriate response is reinforced and encouragement given to think of other things the pattern looks like. The experimenter also suggests a number of responses (if not already provided by the participant). Participants are told they can rotate the cards. They are then given the test stimuli one at a time in random order and for each one asked 'What could this be?'

Scoring procedures were similar to those used by Turner (1999), but an extra 'uninterpretable response' category was added. Each response was therefore classified as belonging to one of the following five scoring categories, and the number of responses in each category was summed across the five test items: (1) correct (a plausible interpretation of the pattern); (2) incorrect (an inappropriate or implausible interpretation of the pattern); (3) repetition (a response which repeats a previous response for the current stimulus or a previous stimulus); (4) redundant (a response that varies minimally from a previous response, e.g. 'two hills', 'two mountains') and (5) uninterpretable (a nonsensical response, which cannot be interpreted as fitting into any of the above categories, e.g. 'up and down'). Because the scoring of Pattern Meanings was more subjective than it was for other tasks in the protocol, interrater reliability was calculated using data from 22 participants (sampled randomly from the ASD and control groups). There was 93.3% agreement between the two raters, and Cohen's kappa was 0.81, indicating excellent interrater reliability.

**Generativity: Stamps task (ages 5–16 years).** This task assesses the spontaneous self-generation of underlying rules in patterns. It is considered a test of 'design fluency' or non-verbal generativity in this research, as it is a non-verbal task requiring participants to produce multiple novel responses. The procedure was based on Frith (1972), with some minor modifications. Participants are provided with four stamps of different shapes and colours and a piece of paper with a row of 16 boxes on it. They are asked to make whatever pattern they like with the stamps, putting one stamp in each box. There are eight trials, four using only two of the stamps and four using all four stamps. If the child does not use all the stamps available during the first eight boxes of a trial, at that point s/he is reminded that there are more stamps available. The two-stamp and four-stamp

trials are presented alternately. The trials are divided into two blocks, separated by at least half an hour, with each block consisting of four trials.

Four types of scores are calculated for each trial:

- 1 Complexity.** Rules are defined as consistently recurring subunits of a fixed number of elements (e.g. the pattern red/green/red/green, etc. has an underlying alternation rule as two elements are repeated). Ratings of complexity are based on the number of elements contained in a subunit. Repetitions of single elements are given the lowest rating of 0, repetitions of two single elements (i.e. alternations) are rated 1 and repetitions of three or four single elements score 2. Also, on two-stamp trials, if the pattern consists of more than just an alternation (e.g. red/red/green/red/red/green), a score of 2 is given, and on four-stamp trials, if the pattern consists of more than just cycling through the three or four elements (e.g. red/red/green/black/black/blue/red/red/green/black/blue/blue), then a score of 3 is given. Finally, the rule must account for at least one-half of the 16-item sequence to receive its score; otherwise the pattern is considered unidentifiable and given a rating of 0.
- 2 Rule adherence.** Each sequence which can be completely accounted for by a single rule (i.e. a repeated subunit or a mirror-reversed pattern) is given a score of 1, other sequences score 0.
- 3 Restriction.** A score of 1 is awarded if the child uses fewer than the maximum number of stamps available on the trial.
- 4 Originality.** Any sequence with an identifiable pattern that occurs only once in all of the trials is considered 'original' and given a score of 1. Other sequences score 0.

Scores are summed across the eight trials to produce overall complexity, rule adherence, restriction and originality scores for each participant.

### Procedure

Informed written consent was obtained from each participant, or from the mother of children who were either under 12 years of age or whose level of understanding of the research was judged to be insufficient to give informed consent. Tests and parental interviews were usually administered at the participants' homes, or at a research center. Testing was often divided into two sessions across different days, in order to prevent fatigue and distractibility.

### Results

#### **Power, data screening and approaches to statistical analysis**

The power of our study to detect a medium-sized effect ( $d = 0.5$ ) at  $\alpha = 0.05$  exceeded 0.8 for both the parent and sibling samples (parents = 0.98; siblings = 0.84; this was calculated using GPower, Faul & Erdfelder 1992). The effect

size chosen for these calculations is comparable with those obtained in similar studies using similar measures (e.g. in Hughes *et al.*'s 1999 study, the effect size of the difference between ASD and control boys on the ToL extra moves score was 0.45). We chose not to use any correction to the alpha level for multiple comparisons. This was because Bonferroni corrections are widely regarded as an over-correction (e.g. Perneger 1998), particularly when measures are intercorrelated and when there are *a priori* predictions for specific measures, which was the case for our study. As there is always a risk that some associations may arise by chance when there are numerous dependent variables, we therefore gave strongest weight to findings that were consistent with prior literature.

Data from all measures were screened for normality and outliers. For variables with distributions that did not depart substantially from normality, outliers falling more than three standard deviations (SD) from the mean of the group (i.e. ASD or control) were trimmed to 3 SD from the mean. A few variables demonstrated highly skewed distributions, and these were either transformed or dichotomized, as detailed below. For those tasks administered only to participants within a certain age range, or for which data were missing, group comparisons of age, PIQ and VIQ were repeated for the limited samples. The samples remained matched on all variables for all tasks.

As previously mentioned, gender was included as an independent variable in all group comparisons on continuous variables. Separate descriptive statistics for males and females are reported only if there were significant group by gender interactions. In the case of dichotomous variables, gender effects were assessed by conducting separate  $\chi^2$  analyses for males and females. The separate analyses are reported only if outcomes were discrepant for the two genders.

In all sibling group comparisons where there were significant group differences, the influence of siblings with (non-ASD) clinical diagnoses (e.g. ADHD) was checked by repeating all group comparisons after excluding these participants from the sample. This affected outcomes in only one case (on the Stamps task complexity score), and reporting is restricted to that case.

### Planning: ToL

#### Parents

One control parent had missing data on the ToL and was not included in analyses. A two-way group  $\times$  gender ANOVA on the adjusted extra moves score revealed no significant difference between ASD and control parents (ASD parents:  $M = 10.10$ ,  $SD = 5.29$ ; control parents:  $M = 10.53$ ,  $SD = 6.57$ ),  $F_{1,236} = 0.02$ ,  $P > 0.1$ . There was a significant main effect of gender such that fathers ( $M = 8.78$ ,  $SD = 5.06$ ) made significantly fewer extra moves than mothers ( $M = 11.36$ ,  $SD = 6.11$ ),  $F_{1,236} = 14.12$ ,

$P < 0.001$ , but the interaction between group and gender was not significant,  $F_{1,236} = 2.40$ ,  $P > 0.1$ . The number of rule violations per block administered was highly skewed and was recoded as a dichotomous variable, where a score of 0 remained as 0 and any higher number of rule violations was coded as 1. A  $\chi^2$  analysis showed that the proportion of high-rule violators did not differ for ASD (21.4%) and control (28.4%) parents,  $\chi^2 (1, n = 240) = 1.55$ ,  $P > 0.1$ .

#### Siblings

A two-way ANOVA comparing the total adjusted extra move scores of the ASD siblings and control siblings revealed no significant main effect of group or gender and no significant interaction; largest  $F_{1,112} = 1.02$ ,  $P > 0.1$  (ASD siblings:  $M = 22.21$ ,  $SD = 8.37$ ; control siblings:  $M = 21.78$ ,  $SD = 8.34$ ). The rule violations variable was also dichotomized for siblings, such that 0–1 violations per block were given a score of 0 and any higher number of violations scored 1. The proportion of high-rule violators in the ASD-sibling group (24.2%) did not differ significantly from the proportion in the control sibling group (26.0%),  $\chi^2 (1, n = 116) = 0.05$ ,  $P > 0.1$ .

### Set-shifting: IDED set-shifting task

All set-shifting variables were highly skewed, and all variables were recoded such that any participant making 0 or 1 errors was assigned a score of 0 (low-error scorers) and any participant making a higher number of errors was given a score of 1 (high-error scorers). The percentage of high-error scorers in parent and sibling groups for the IDS and EDS stages in each task condition is displayed in Table 3.

#### Parents

There were no significant differences between ASD and control parents on any of the 'control' stages of the task in either condition (i.e. the SD, SDR, CD, CDR, IDS and IDR stages), as expected. There was also no significant difference between parent groups on the EDS stage of the perseveration condition,  $\chi^2 (1, n = 236) = 0$ ,  $P > 0.1$ , and differences were non-significant for both fathers and mothers. However, a significantly higher proportion of ASD parents than control parents were high-error scorers on the EDS stage of the learned irrelevance condition,  $\chi^2 (1, n = 235) = 4.52$ ,  $P < 0.05$ . When fathers and mothers were analysed separately, it was revealed that there was no significant difference between ASD and control mothers,  $\chi^2 (1, n = 133) = 0.51$ ,  $P > 0.1$ , but a significantly higher proportion of ASD fathers were high-error scorers compared with control fathers,  $\chi^2 (1, n = 101) = 6.50$ ,  $P < 0.05$ .

#### Siblings

There were no significant sibling group differences overall on any variable. When brothers and sisters were analysed separately, a significant difference was observed in the SD



**Table 3:** Intradimensional, extradimensional (IDED) set-shifting task results: percentage of high-error scorers in each group for each stage of each task condition

	Parents		Siblings	
	ASD	Control	ASD	Control
Perseveration condition				
<i>n</i> *	142	94	50	37
SD stage	25.2	26.9	26.0	24.3
SDR stage	32.2	31.2	40.0	29.7
CD stage	17.5	22.6	28.0	27.0
IDS stage	16.9	20.2	18.0	18.9
EDS stage	29.6	29.8	22.0	27.0
Learned irrelevance condition				
<i>n</i> *	143	92	51	37
SD stage – brothers only	11.9	7.5	15.4	37.5
SD stage – sisters only			24.0†	3.4
SDR stage	16.1	10.8	21.6	21.6
CD stage	21.7	15.2	25.5	21.6
IDS stage	10.5	6.4	23.5	10.8
EDS stage – fathers only	64.1†	37.8	66.7	67.6
EDS stage – mothers only	54.4	48.1		

\*Due to computer malfunction, two autism spectrum disorder (ASD) parents and two control parents had unusable data for the whole set-shifting task; one ASD parent had missing data for the Perseveration condition; two control parents had missing data for the Learned Irrelevance condition, and data for the Perseveration condition from one ASD sibling were invalid and not included in analyses. The task had a restricted age range for siblings.

†Significant difference between ASD and control groups at  $P < 0.05$ .

stage (the first 'control' stage) of the learned irrelevance condition such that there was a higher proportion of high-error scorers among sisters of ASD probands than among control sisters,  $\chi^2(1, n = 54) = 5.03$ ,  $P < 0.05$ . However, this difference was not predicted and does not reflect a difference in set-shifting ability between the two sister groups and, therefore, may well represent a chance result given the large number of comparisons. There was no significant difference between the brother groups on this variable, and no other differences in the results from brothers and sisters on other variables.

### **Inhibition/working memory: RIL task**

For all RIL conditions, variables representing the percentage of errors made in each condition were highly skewed, with many participants making a low percentage of errors. Instead of using these separate error scores for each condition, three error difference scores were calculated: (1) an *inhibition* error score – the difference between Conditions 2 (inhibition) and 1 (control); (2) a *load* error score – the difference between Condition 3 (working memory load) and Condition 2; (3) an *inhibition + load* error score – the difference between Conditions 3 and 1. These difference scores were normally distributed.

Three RT difference scores were also calculated: an *inhibition* RT score, a *load* RT score and an *inhibition + load* RT

score. The percentage of errors made in choosing the most recently displayed shape in Condition 3 was labelled the *shape error score* (a measure of working memory ability under conditions requiring inhibitory control). Table 4 displays the means and SDs of each group for all error and RT difference scores and the shape error score.

### *Parents*

There were five participants with outlying scores (one ASD parent and four control parents) on one or more of the three error difference scores, nine with outliers on the RT difference scores (five ASD parents and four control parents) and four outliers on the shape error score (two ASD parents and two control parents), all of which were trimmed. On two-way ANOVAS, there were no significant main effects of group or gender and no significant interactions for any RIL task variables.

### *Siblings*

There were three siblings (two ASD and one control) with outliers on one or more of the error difference scores and three siblings (two ASD and one control) with outliers on the RT difference scores. There were no significant main effects of group or gender and no significant group–gender interactions on the error difference scores, RT difference scores or the shape error score.

**Table 4:** Response Inhibition and Load (RIL) task results: means (and SDs) of each group

	Parents		Siblings	
	ASD ( $n^* = 141$ )	Control ( $n^* = 94$ )	ASD ( $n^* = 50$ )	Control ( $n^* = 37$ )
Error difference scores				
Inhibition	0.21 (1.12)	0.38 (1.34)	0.70 (2.97)	1.17 (3.33)
Load	0.88 (2.39)	1.50 (3.23)	0.50 (3.26)	0.68 (3.48)
Inhibition + load	1.10 (2.19)	1.86 (3.27)	1.10 (3.59)	1.89 (3.63)
RT difference scores				
Inhibition	81.89 (98.46)	79.41 (93.32)	175.89 (164.29)	148.84 (141.37)
Load	300.69 (182.39)	313.41 (191.95)	138.11 (145.69)	191.32 (142.22)
Inhibition + load	384.51 (209.33)	393.06 (213.69)	313.57 (222.36)	339.39 (210.10)
Working memory measure				
Shape error score	6.95 (9.72)	6.25 (7.90)	13.73 (14.78)	9.73 (13.11)

\*Because of computer problems, four autism spectrum disorder (ASD) parents, two control parents and one ASD sibling had missing data and were not included in analyses. This task had a restricted age range for siblings.

### Generativity: Pattern Meanings

For ease and brevity of analysis, a 'total responses' variable was created where the number of correct and error responses (including all types of errors) were summed to give a measure of overall response generativity. The number of correct and error responses were also calculated separately. The error responses variable was significantly skewed for both parents and siblings and was transformed using a logarithm equation.

#### Parents

One control parent had missing data. There was one control parent and two ASD parents with an outlier on the total responses variable, and one control parent and four ASD parents with an outlier on the proportion correct variable, all of which were trimmed. In a two-way group  $\times$  gender ANOVA, ASD parents were found to produce significantly fewer responses overall than control parents,  $F_{1,236} = 15.52$ ,  $P < 0.001$  (ASD parents:  $M = 30.69$ ,  $SD = 10.62$ ; control parents:  $M = 36.27$ ,  $SD = 11.06$ ). There was no effect of gender and no significant group by gender interaction, largest  $F_{1,236} = 2.42$ ,  $P > 0.1$ . ASD parents were found to produce both significantly fewer correct responses,  $F_{1,236} = 14.86$ ,  $P < 0.001$ , and significantly fewer error responses,  $F_{1,236} = 5.54$ ,  $P < 0.05$ , than control parents.

#### Siblings

ASD siblings produced significantly fewer responses overall ( $M = 32.84$ ,  $SD = 12.18$ ) than control siblings ( $M = 37.12$ ,  $SD = 16.90$ ),  $F_{1,112} = 4.18$ ,  $P < 0.05$ . There was no significant effect of gender on this variable and no significant group by gender interaction, largest  $F_{1,112} = 2.42$ ,  $P > 0.1$ . However, there was no significant difference in either the number of correct responses produced,  $F_{1,112} = 0.001$ ,  $P > 0.1$ , or the number of error responses,  $F_{1,112} = 3.40$ ,  $P = 0.07$ .

### Generativity: Stamps task (siblings only)

Both the rule adherence and restriction scores demonstrated highly skewed distributions and were recoded as dichotomous variables. For rule adherence, a score between 0 and 6 inclusive was coded as 0 and a score of 7 or 8 was coded as 1. For restriction, a score of 0 was left as 0 and a score between 1 and 8 inclusive was coded as 1. The complexity and originality scores were approximately normally distributed. Means and SDs for the latter two variables and the proportion of low scorers for the former two variables, along with the significance of group comparisons for all scores, are presented in Table 5. There was a significant group difference on the complexity score,  $F_{1,105} = 5.70$ ,  $P < 0.05$ , indicating that the ASD siblings produced less complex patterns than did the control siblings. There were no significant main effects or interactions involving gender for this variable, largest  $F_{1,105} = 0.004$ ,  $P > 0.1$ . However, when participants with non-ASD clinical diagnoses were excluded, the group difference in the complexity score became non-significant,  $F_{1,97} = 2.73$ ,  $P > 0.1$ . No effects were significant for the originality scores, largest  $F_{1,105} = 0.50$ ,  $P > 0.1$ .  $\chi^2$  analyses revealed that there was

**Table 5:** Stamps task results: means (and SDs) of each group [or the percentage of low scorers for dichotomous variables]

	Siblings	
	ASD ( $n^* = 60$ )	Control ( $n^* = 49$ )
Complexity score	18.92 (2.77)†	20.51 (3.82)
Originality score	4.70 (2.99)	4.71 (2.57)
Restriction score	[95.0%]	[87.8%]
Rule adherence score	[16.9%]	[22.9%]

\*This task had a restricted age range.

†Significant difference between autism spectrum disorder (ASD) and control groups at  $P < 0.05$ .

no significant group difference in the percentage of low scorers on the restriction score,  $\chi^2(1, n = 109) = 1.87$ ,  $P > 0.1$ , or the rule adherence score,  $\chi^2(1, n = 109) = 0.60$ ,  $P > 0.1$ .

### **Additional analyses**

A number of extra analyses were conducted to evaluate the effects of additional factors which could have influenced the pattern of results.

#### *Assortative mating*

The possible effect of assortative mating (whereby the independence of parents from the same family may be reduced) was assessed by correlating the test scores of mothers and fathers from the same families. There were significant mother-father correlations on three variables: ToL rule violations ( $r = 0.23$ ,  $P < 0.05$ ), RIL task inhibition error score ( $r = 0.25$ ,  $P < 0.05$ ) and the RIL task shape error score ( $r = 0.25$ ,  $P < 0.05$ ). When mothers and fathers were analysed separately for these variables, there were no changes in the outcome of the previous analyses.

#### *Inclusion of relatives of PDD probands*

To assess whether relatives of probands with partial symptomatology (i.e. those we classified as PDD on the basis of the ADI-R) showed a different pattern of results to relatives of probands with autism (and therefore may have 'weakened' group differences in the overall sample), comparisons between relatives of autistic and PDD probands were conducted. There was a significant difference between 'autism parents' and 'PDD parents' on one variable only, such that a significantly higher proportion of autism parents (35.2%,  $n = 105$ ) than PDD parents (13.5%,  $n = 37$ ) were high-error scorers on the EDS stage of the perseveration condition of the IDED set-shifting task,  $\chi^2(1, n = 142) = 6.20$ ,  $P < 0.05$ . However, when PDD parents were removed from the parent sample, the difference between autism parents and control parents on that variable remained non-significant,  $\chi^2(1, n = 198) = 0.87$ ,  $P > 0.1$ . For siblings, there was a significant difference between 'autism siblings' and 'PDD siblings' again on one variable only, but on this occasion PDD siblings ( $n = 16$ ,  $M = 26.50$ ,  $SD = 9.24$ ) produced significantly fewer responses overall on the Pattern Meanings task than autism siblings ( $n = 50$ ,  $M = 33.12$ ,  $SD = 11.34$ ),  $F_{1,62} = 4.48$ ,  $P < 0.05$ . When PDD siblings were removed from the sample, the difference between autism siblings and control siblings on this total responses variable became only marginally significant,  $F_{1,96} = 3.62$ ,  $P = 0.06$ . Hence, the inclusion of PDD parents in the sample made no difference to the outcomes of group comparisons, and the inclusion of PDD siblings made a minor difference on one variable measuring overall response generativity.

#### *Restrictive selection of sibling sample*

To evaluate whether the process of selecting the sibling sample (i.e. eliminating all siblings with ASDs and including only one sibling per family) resulted in weaker group differences, we conducted all group comparisons on the original expanded sibling sample. There were no differences in the outcomes of group comparisons, except that the difference between ASD and control siblings on the Pattern Meanings total responses variable did not reach significance in the broader sample.

#### *Age range of siblings*

To check that age variability was not masking genuine group differences in siblings, we re-ran all analyses on sibling groups with age as a covariate, and the pattern of results was unchanged.

### **Effect sizes of group comparisons**

Table 6 presents the effect sizes of all significant group differences, alongside the results of previous studies on ASD proband, parent and sibling samples on measures of the various EF components. According to Cohen's (1988) system for classifying effect sizes, the differences between ASD and control fathers on the IDED set-shifting task and the difference between parent groups on the Pattern Meanings total responses variable are both classified as medium effects. The difference between sibling groups on the Pattern Meanings total responses variable is classified as a small effect, and the difference on the Stamps task complexity score is classified as a small-medium effect. The difference between ASD and control sisters on the first control stage of the IDED set-shifting task is classified as a medium effect but is not shown in the table as it did not represent a meaningful difference in set-shifting performance.

## **Discussion**

ASD parents and ASD siblings both demonstrated mild EF impairments on selected tasks, but the pattern of performance displayed by the two groups was different. ASD parents showed poorer performance than control parents on a test of ideational fluency/generativity, and ASD fathers (but not mothers) demonstrated a weakness in set-shifting to a previously irrelevant dimension. There were no significant differences between ASD and control parents on measures of planning, inhibition or the interaction between inhibition and working memory. ASD siblings revealed a weakness in both ideational fluency and non-verbal generativity. There were no sibling group differences on tests of planning, set-shifting, inhibition or the interaction between inhibition and working memory. None of the significant (or non-significant) differences between parent or sibling groups were attributable to the effects of assortative mating, the

**Table 6:** Previous findings in autism spectrum disorder (ASD) proband, parent and sibling samples on measures of the various executive function (EF) components, and effect sizes of significant group differences in the current study

	Probands	Parents	Siblings
Set-shifting	Impaired	Intact (WCST) <sup>†</sup> /impaired (IDED) <sup>‡</sup> IDED (EDS stage errors): fathers $d = 0.52$	Intact (WCST) <sup>§</sup> /impaired (IDED) <sup>§</sup> IDED: no difference*
Planning	Impaired	Impaired (fathers only)** ToL: no difference	Impaired <sup>††</sup> ToL: no difference
Generativity	Impaired	Not done Pattern Meanings (total responses): $d = 0.51$	Impaired <sup>‡‡</sup> Pattern Meanings (total responses): $d = 0.29$ Stamps task (complexity score): $d = 0.48$
Inhibition	Intact	Not done RIL task: no difference	Not done RIL task: no difference

\*There was a significant difference between ASD and control sisters in one of the control stages of this task, but this did not reflect a difference in set-shifting ability.

<sup>†</sup>Szatmari *et al.* (1993).

<sup>§</sup>Hughes *et al.* (1999).

<sup>‡</sup>Ozonoff *et al.* (1993).

\*\*Hughes *et al.* (1997), Piven and Palmer (1997).

<sup>†††</sup>Hughes *et al.* (1999), Ozonoff *et al.* (1993).

inclusion of relatives of PDD probands in the sample, the restrictive selection of the sibling sample or the variability in age in the sibling sample (except for the difference between ASD and control siblings on the Pattern Meanings total responses variable, which became non-significant when PDD siblings were removed from the sample and when the expanded/unrestricted sibling sample was used, as discussed below).

These results differ from and expand upon the results of previous studies in several important ways, as summarized in Table 6. Firstly, contrary to our predictions, which were based on the findings of Hughes *et al.* (1997, 1999), Ozonoff *et al.* (1993) and Piven and Palmer (1997), no significant group differences were found on the ToL (a measure of planning ability) in either the parents or the siblings of individuals with ASDs. The sample size was much larger for the current study than for any of the previous studies, and our power to detect small-moderate effects was adequate, ruling out power as an explanation for this discrepancy. Of note, the differences on Tower tasks in previous studies were restricted to a subset of the relatives of autistic probands. Hughes *et al.* (1999) did not find overall group differences in their sibling groups on their continuous measures of performance but found differences only when particular variables were dichotomized and the proportions of siblings classified as passers or failers were compared. When the current ToL data were re-analysed using a dichotomous pass/fail performance criterion (such that a 'failer' was anyone who completed less than 50% of the problems in the minimum number of moves), again no group differences were revealed. It therefore does not appear that existing group differences were merely 'hidden' by the methods of analysis used in this study. One alternative explanation is that the

administration of the ToL in this study differed slightly from most other studies in that forward planning was actively encouraged, which may have reduced the size of group differences. This is because if participants are not encouraged to plan, their performance may be confounded by weak generativity (i.e. difficulty producing the strategy to plan the solution in advance). This explanation is consistent with the finding of a generativity deficit in ASD parents, in that prompting and encouragement may have compensated for difficulties in generating a strategy to solve the task. Hence, the current results suggest that a planning impairment is not a strong or reliable endophenotype for autism and that results of previous studies may have been attributable to ASD parents' difficulties with generativity rather than planning.

The poorer performance of ASD fathers on a test of set-shifting is consistent with Hughes *et al.*'s (1997) findings, although there were no gender differences on the equivalent 'errors to criterion' measure in that study. However, the version of the set-shifting task used in Hughes *et al.*'s study did not distinguish between shifting from a previously relevant dimension and shifting to a previously irrelevant dimension. This study showed that ASD fathers' difficulty with set-shifting was specific to the latter ability to shift to a previously irrelevant dimension. However, contrary to prediction, ASD siblings did not show deficits in either of the extradimensional shift stages. Although this contrasts with the findings of Hughes *et al.* (1999), it is interesting that Ozonoff *et al.* (1993) did not find any evidence of a deficit in cognitive flexibility in siblings using the WCST. Together, these results suggest that ASD siblings do not reliably show deficits in attentional shifting. Again, however, it is possible that the task instructions (which included the fact that a shift

would take place) and the structured nature of the task design may have masked subtle impairments in this domain. Sisters of ASD probands also performed significantly more poorly than control sisters on the simple first stage of the IDEd set-shifting task (in the learned irrelevance condition). As there were no significant differences observed in latter stages of the task which involve shifting set and the difference in this control stage was not predicted, this probably reflects a chance result given the large number of comparisons, rather than a deficit in a component of EF.

As predicted, neither parents nor siblings of ASD probands showed a deficit in inhibitory control. These results are consistent with the results of proband studies, which have generally found that inhibition is intact in autism (although see Wong *et al.* 2003, whose findings suggest that this sparing of inhibition may be specific to non-verbal tasks). No previous published studies have examined inhibition in the broad autism phenotype, even though its purported sparing in autism is crucial in distinguishing the EF profile of autism from other disorders such as ADHD. We also found that even when a working memory load was added, ASD parents and siblings showed intact performance, indicating that there was no disproportionate difficulty dealing with the interaction between inhibition and working memory (even though such a difficulty has been proposed to occur in autism). It therefore appears that neither impairments in (non-verbal) inhibition or in the inhibition/working memory interaction are plausible endophenotypes for autism.

In contrast, a generativity deficit emerged as a stronger feature of the broad autism phenotype, particularly in ASD parents. Both mothers and fathers of ASD probands generated significantly fewer responses than control parents on the Pattern Meanings task, with the size of the difference classifying as a medium effect. ASD parents produced fewer responses in all categories – both correct and incorrect – suggesting an overall reduction in response generation. ASD siblings also generated fewer total responses than control siblings. However, unlike the parents, this effect was not strong enough for significant differences to emerge when correct and error responses were analysed separately (see below for a discussion of the different findings in parents and siblings). In addition, the difference on the total responses variable became non-significant when PDD siblings were removed from the sample and when the unrestricted sibling sample (including siblings with ASDs and all siblings per family) was used, suggesting that the overall group difference was not reliable and may have been attributable to extreme scores in a small subgroup. The lack of a robust group difference on this task was somewhat surprising given Hughes *et al.*'s (1999) finding of a weakness in ASD siblings on a word fluency task. Indeed, Turner (1999) found that autistic probands actually showed more striking deficits on ideational fluency tasks (including the Pattern Meanings task) than on the word fluency task used by Hughes *et al.* However, while only 90 seconds were allowed to generate

responses in this study, Hughes *et al.* allowed 120 seconds, which may have been the extra time needed to reveal more robust generativity difficulties in siblings of ASD probands.

Nevertheless, ASD siblings showed a significant weakness in non-verbal generativity, generating less complex responses than control siblings on the Stamps task. Although the group difference on this variable became non-significant when individuals with non-ASD clinical diagnoses were excluded, this does not necessarily suggest that a non-verbal generativity deficit in the broad phenotype is an unimportant artefact of pathology unrelated to ASDs. Rather, it may reflect the possibility that the broad phenotype is itself characterized by higher rates of non-ASD diagnoses (10.6% of ASD siblings in this study compared with 2.0% of control siblings), and these individuals also display a more abnormal cognitive profile.

Hence, the results of this study indicate that weaknesses in generativity were present in both parents and siblings of ASD probands, albeit to different degrees and on different measures, but difficulties with set-shifting were present only in fathers. *A priori*, we might expect parents to show fewer deficits than siblings, because parents are by definition 'selected' for parenthood in that their social and communicative functioning must be sufficient for partnership. The fact that parents appeared more impaired than their children on the ideational fluency and set-shifting tasks is therefore surprising. However, several other studies of cognitive functioning in the broad autism phenotype have also found that group differences occur more frequently in parents than in siblings (e.g. Happé *et al.* 2001; Hughes *et al.* 1999). Hughes *et al.* (1999) proposed that the use of computerized tasks may favour young participants over parents, but this does not account for differences on non-computerized tasks (e.g. the Pattern Meanings task in the current study). It is possible that, as Happé *et al.* (2001) suggested, certain tasks may not be sufficiently sensitive in younger children, or that genetically determined cognitive weaknesses may only emerge at a certain age or become more pronounced with age. This could be the case for certain aspects of EF which develop relatively late, such as generativity, but cannot explain the parent-sibling discrepancy in set-shifting which typically develops earlier in life. Furthermore, the sample of siblings in this study included a significant proportion of individuals over the age of 10 years. An alternative explanation for parent-sibling discrepancies could be that there is a greater range or variability in cognitive and/or behavioural phenotypic expression in young siblings. In addition, it may be that psychiatric conditions such as depression and anxiety (see, for example, Piven & Palmer 1999) may underlie or amplify cognitive difficulties in parents, while not having as great an effect on siblings (who are not as burdened with parental responsibility and/or may not have yet developed significant mood disturbances). This would suggest that EF and other cognitive difficulties displayed by ASD parents may not be core endophenotypes, but rather (at least partially) secondary

consequences of emotional and developmental factors. These possibilities warrant further investigation given the recurring trend across studies.

The relationship between the EF profiles displayed by the first-degree relatives in this study and the ASD probands themselves is also relevant in determining the plausibility of the EF deficits identified as potential endophenotypes. While proband-relative correlations on EF task performances in the WAFSASD sample are not covered in the current paper (although they will be addressed in a forthcoming paper from our group), it is relevant to note that the probands in our study demonstrated significant generativity deficits on both ideational and design fluency tasks, with these impairments having the largest effect sizes of all the EF components measured (Wong *et al.* 2003; D. Wong, M. Maybery, D. Bishop, A. Maley & J. Hallmayer, manuscript in preparation). While the ASD probands also demonstrated other EF deficits which were not revealed in their first-degree relatives (e.g. in planning), this finding of a strong generativity deficit in probands with ASDs is at least consistent with the possibility of a generativity-based endophenotype for autism.

The results of this study do not allow us to make direct inferences regarding the specificity or uniqueness of the EF profile that characterizes the broad phenotype of autism as compared with other disorders characterized by a purportedly different profile of EF deficits, such as ADHD. It is nevertheless noteworthy that the evidence so far appears to suggest that if there is a cognitive endophenotype for ADHD, it is distinguished primarily by impairments in inhibitory control (e.g. Slaats-Willemse *et al.* 2003, 2005). Interestingly, parents of children with ADHD have also been found to perform more poorly than control parents on a word fluency task tapping verbal generativity; however, their pattern of performance was characterized by a higher number of rule breaks and error responses rather than difficulties with generating correct novel responses, consistent with a possible inhibitory deficit (Casey *et al.* 2000). Thus, the evidence to date is broadly consistent with the notion that different EF profiles may characterize the broad phenotype of different developmental disorders. However, it may be that the EF deficits that characterize autism and its broader phenotype are not specific to autism but that autism results if these EF deficits occur in combination with other deficits (e.g. in theory of mind, face processing, imitation, etc.). Further studies comparing the EF profiles of relatives of individuals with autism and other developmental disorders would be instructive in determining whether a generativity deficit is specific to the broad autism phenotype.

## Conclusions

Our results suggest that the broad autism phenotype is not characterized by across-the-board EF deficits but rather a specific profile of strength and weakness on various EF

components. This study is the first we are aware of to investigate inhibitory capacity in the broad autism phenotype, with results supporting the prediction of intact performance in this domain. However, our results indicated that the cognitive endophenotype for autism may not be characterized primarily by impairments in planning and cognitive flexibility, as had been previously proposed. No deficits in planning ability were found in either ASD parents or siblings, and although differences in set-shifting performance were identified between ASD and control parents, these were restricted to fathers, and there were no differences between sibling groups. While low task sensitivity cannot be ruled out as an explanation for the lack of differences in these EF components, it is nevertheless significant that weaknesses in generativity emerged more strongly as characteristics of the broad autism phenotype in this study. Our results raise the question of the extent to which deficits in other EF domains identified in previous studies (e.g. planning difficulties) may actually reflect a generativity impairment, given that they were not found in this study when steps were taken to minimize the need for participants to spontaneously generate a strategy for addressing task requirements.

Although our results require replication and are limited by the different profiles displayed by parents and siblings, they nevertheless suggest that impairments in generativity may play a central role in terms of cognitive theories of autism, particularly given that generativity deficits were the most prominent of all EF components measured in our proband sample (Wong *et al.* 2003; D. Wong, M. Maybery, D. Bishop, A. Maley & J. Hallmayer, manuscript in preparation). Previous studies have also provided evidence of significant generativity impairments in children with autism using a range of tasks (e.g. Turner 1999). It is interesting to note that generativity tasks are generally the most unstructured of EF tasks, as they require the participant to produce novel responses, as opposed to reacting to stimuli presented as part of a structured task. They also rely more on internally controlled mechanisms than on external constraints. It would therefore be interesting to examine whether the difficulties displayed by ASD parents and siblings on generativity tasks are due to a specific problem with generativity, or whether impairments in other EF domains would also emerge on unstructured, internally driven tasks which are more representative of many real-life situations (i.e. have higher ecological validity). However, the search for ecological validity often comes at the expense of task purity, and the challenge is therefore to develop ecologically valid tasks which have in-built control conditions that allow isolation of the target ability. The development of such tasks may be a fruitful advancement in the search to define the broad autism phenotype. In addition, it would be interesting to investigate whether the presence of a generativity deficit in first-degree relatives of ASD probands is associated with corresponding behavioural features, such as insistence on sameness (Turner 1999) or poor conversational ability (Bishop &

Norbury 2005), and whether there is any causal association with psychiatric conditions such as depression or anxiety (which could plausibly affect performance on generativity tasks).

## References

- Anderson, S.W., Damasio, H., Jones, R. & Tranel, D. (1991) Wisconsin Card Sorting Test performance as a measure of frontal lobe damage. *J Clin Exp Neuropsychol* **13**, 909–922.
- Anokhin, A.P., Heath, A.C. & Ralano, A. (2003) Genetic influences on frontal brain function: WCST performance in twins. *Neuroreport* **14**, 1975–1978.
- Bailey, A., Le Couteur, A., Gottesman, I. & Bolton, P. (1995) Autism as a strongly genetic disorder: evidence from a British twin study. *Psychol Med* **25**, 63–77.
- Bailey, A., Phillips, W. & Rutter, M. (1996) Autism: towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. *J Child Psychol Psychiatry* **37**, 89–126.
- Bailey, A., Palferman, S., Heavey, L. & Le Couteur, A. (1998) Autism: the phenotype in relatives. *J Autism Dev Disord* **28**, 369–392.
- Berger, H.J., Aerts, F.H., van Spaendonck, K.P., Cools, A.R. & Teunisse, J.-P. (2003) Central coherence and cognitive shifting in relation to social improvement in high-functioning young adults with autism. *J Clin Exp Neuropsychol* **25**, 502–511.
- Berument, S.K., Rutter, M., Lord, C., Pickles, A. & Bailey, A. (1999) Autism screening questionnaire: diagnostic validity. *Br J Psychiatry* **175**, 444–451.
- Beveridge, M., Jarrold, C. & Pettit, E. (2002) An experimental approach to executive fingerprinting in young children. *Infant Child Dev* **11**, 107–123.
- Bishop, D.V.M. & Norbury, C.F. (2005) Executive functions in children with communication impairments, in relation to autistic symptomatology. I: generativity. *Autism* **9**, 7–27.
- Bishop, D.V.M., Maybery, M., Maley, A., Wong, D., Hill, W. & Hallmayer, J. (2004) Using self-report to identify the broad phenotype in parents of children with autistic spectrum disorders: a study using the Autism-Spectrum Quotient. *J Child Psychol Psychiatry* **45**, 1431–1436.
- Bolton, P., Macdonald, H., Pickles, A., Rios, P., Goode, S., Crowson, M., Bailey, A. & Rutter, M. (1994) A case-control family history study of autism. *J Child Psychol Psychiatry* **35**, 877–900.
- Boucher, J. (1988) Word fluency in high-functioning autistic children. *J Autism Dev Disord* **18**, 637–645.
- Casey, M.B., Cohen, M., Schuerholz, L.J., Singer, H.S. & Denckla, M.B. (2000) Language-based cognitive functioning in parents of offspring with ADHD comorbid for Tourette syndrome or learning disabilities. *Dev Neuropsychol* **17**, 85–110.
- Channon, S., Flynn, D. & Robertson, M.M. (1992) Attentional deficits in Gilles de la Tourette syndrome. *Neuropsychiatry, Neuropsychol Behav Neurol* **5**, 170–177.
- Cohen, J. (1988) *Statistical Power Analysis for the Behavioural Sciences*, 2nd edn. Lawrence Erlbaum, Hillsdale, NJ.
- Coolidge, F.L., Thede, L.L. & Jang, K.L. (2004) Are personality disorders psychological manifestations of executive function deficits? Bivariate heritability evidence from a twin study. *Behav Genet* **34**, 75–84.
- Craig, J. & Baron-Cohen, S. (1999) Creativity and imagination in autism and Asperger syndrome. *J Autism Dev Disord* **29**, 319–326.
- Culbertson, W.C. & Zillmer, E.A. (1998a) The construct validity of the Tower of London<sup>DX</sup> as a measure of the executive functioning of ADHD children. *Assessment* **5**, 215–226.
- Culbertson, W.C. & Zillmer, E.A. (1998b) The Tower of London<sup>DX</sup>: a standardized approach to assessing executive functioning in children. *Arch Clin Neuropsychol* **13**, 285–301.
- Dawson, G., Webb, S., Schellenberg, G.D., Dager, S., Friedman, S., Aylward, E. & Richards, T. (2002) Defining the broader phenotype of autism: genetic, brain, and behavioral perspectives. *Dev Psychopathol* **14**, 581–611.
- DeLong, G. & Dwyer, J.T. (1988) Correlation of family history with specific autistic subgroups: Asperger's syndrome and bipolar affective disease. *J Autism Dev Disord* **18**, 593–600.
- Eslinger, P.J. & Damasio, A.R. (1985) Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR. *Neurology* **35**, 1731–1741.
- Faul, F. & Erdfelder, E. (1992) *GPOWER: a Priori, Post-Hoc, and Compromise Power Analyses for MS-DOS*. Bonn University, Department of Psychology, Bonn, FRG [Computer program].
- Folstein, S. & Rutter, M. (1977) Infantile autism: a genetic study of 21 twin pairs. *J Child Psychol Psychiatry* **18**, 297–321.
- Folstein, S.E., Bisson, E., Santangelo, S.L. & Piven, J. (1998) Finding specific genes that cause autism: a combination of approaches will be needed to maximize power. *J Autism Dev Disord* **28**, 439–445.
- Fombonne, E. (2003) Epidemiological surveys of autism and other pervasive developmental disorders: an update. *J Autism Dev Disord* **33**, 365–382.
- Frith, U. (1972) Cognitive mechanisms in autism: experiments with color and tone sequence production. *J Autism Child Schizophr* **2**, 160–173.
- Garcia-Villamisar, D. & Della Sala, S. (2002) Dual-task performance in adults with autism. *Cognitive Neuropsychiatry* **7**, 63–74.
- Geurts, H.M., Verté, S., Oosterlaan, J., Roeyers, H. & Sergeant, J.A. (2004) How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *J Child Psychol Psychiatry* **45**, 836–854.
- Goldstein, G., Johnson, C.R. & Minshew, N.J. (2001) Attentional processes in autism. *J Autism Dev Disord* **31**, 433–440.
- Gottesman, I.I. & Gould, T.D. (2003) The endophenotype concept in psychiatry: etymology and strategic intentions. *Am J Psychiatry* **160**, 636–645.
- Happé, F., Briskman, J. & Frith, U. (2001) Exploring the cognitive phenotype of autism: weak 'central coherence' in parents and siblings of children with autism: I. Experimental tests. *J Child Psychol Psychiatry* **42**, 299–307.
- Hill, E. (2004) Evaluating the theory of executive dysfunction in autism. *Dev Rev* **24**, 189–233.
- Hoff, A.L. & Kremen, W.S. (2003) Neuropsychology in schizophrenia: an update. *Current Opinion Psychiatry* **16**, 149–155.
- Hughes, C. (1996) Brief report: planning problems in autism at the level of motor control. *J Autism Dev Disord* **26**, 99–107.
- Hughes, C. & Russell, J. (1993) Autistic children's difficulty with mental disengagement from an object: its implications for theories of autism. *Dev Psychol* **29**, 498–510.
- Hughes, C., Russell, J. & Robbins, T.W. (1994) Evidence for executive dysfunction in autism. *Neuropsychologia* **32**, 477–492.
- Hughes, C., Leboyer, M. & Bouvard, M. (1997) Executive function in parents of children with autism. *Psychol Med* **27**, 209–220.
- Hughes, C., Plumet, M.-H. & Leboyer, M. (1999) Towards a cognitive phenotype for autism: increased prevalence of executive dysfunction and superior spatial span amongst siblings of children with autism. *J Child Psychol Psychiatry* **40**, 705–718.

- Landa, R., Piven, J., Wzorek, M.M., Gayle, J.O., Chase, G.A. & Folstein, S.E. (1992) Social language use in parents of autistic individuals. *Psychol Med* **22**, 245–254.
- Le Couteur, A., Bailey, A., Goode, S., Pickles, A., Robertson, S., Gottesman, I. & Rutter, M. (1996) A broader phenotype of autism: the clinical spectrum in twins. *J Child Psychol Psychiatry* **37**, 785–801.
- Lewis, V. & Boucher, J. (1988) Spontaneous, instructed and elicited play in relatively able autistic children. *Br J Dev Psychol* **6**, 325–339.
- Lord, C., Rutter, M. & Le Couteur, A. (1994) Autism Diagnostic Interview – Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord* **24**, 659–685.
- MacLean, J.E., Szatmari, P., Jones, M.B., Bryson, S.E., Mahoney, W.J., Bartolucci, G. & Tuff, L. (1999) Familial factors influence level of functioning in pervasive developmental disorder. *J Am Acad Child Adolesc Psychiatry* **38**, 746–753.
- Owen, A.M., Roberts, A.C., Hodges, J.R., Summers, B.A., Polkey, C.E. & Robbins, T.W. (1993) Contrasting mechanisms of impaired attentional set-shifting in patients with frontal lobe damage or Parkinson's disease. *Brain* **116**, 1159–1175.
- Ozonoff, S. (1997) Components of executive function in autism and other disorders. In: Russell, J. (ed), *Autism as an Executive Disorder*. Oxford University Press, Oxford, pp. 179–211.
- Ozonoff, S. & Jensen, J. (1999) Brief report: specific executive function profiles in three neurodevelopmental disorders. *J Autism Dev Disord* **29**, 171–177.
- Ozonoff, S. & McEvoy, R.E. (1994) A longitudinal study of executive function and theory of mind development in autism. *Dev Psychopathol* **6**, 415–431.
- Ozonoff, S., Pennington, B.F. & Rogers, S.J. (1991) Executive function deficits in high-functioning autistic individuals: relationship to theory of mind. *J Child Psychol Psychiatry* **32**, 1081–1105.
- Ozonoff, S., Rogers, S.J., Farnham, J.M. & Pennington, B.F. (1993) Can standard measures identify subclinical markers of autism? *J Autism Dev Disord* **23**, 429–441.
- Ozonoff, S., Cook, I., Coon, H., Dawson, G., Joseph, R.M., Klin, A., McMahon, W.M., Minshew, N., Munson, J.A., Pennington, B.F., Rogers, S.J., Spence, M.A., Tager-Flusberg, H., Volkmar, F.R. & Wrathall, D. (2004) Performance on Cambridge Neuropsychological Test Automated Battery subtests sensitive to frontal lobe function in people with autistic disorder: evidence from the Collaborative Programs of Excellence in Autism Network. *J Autism Dev Disord* **34**, 139–150.
- Perneger, T.V. (1998) What is wrong with Bonferroni adjustments. *Br Med J* **136**, 1236–1238.
- Pickles, A., Bolton, P., Macdonald, H., Bailey, A., Le Couteur, A., Sim, C.H. & Rutter, M. (1995) Latent-class analysis of recurrence risks for complex phenotypes with selection and measurement error: a twin and family history study of autism. *Am J Hum Genet* **57**, 717–726.
- Piven, J. & Palmer, P. (1997) Cognitive deficits in parents from multiple-incidence autism families. *J Child Psychol Psychiatry* **38**, 1011–1021.
- Piven, J. & Palmer, P. (1999) Psychiatric disorder and the broad autism phenotype: evidence from a family study of multiple-incidence autism families. *Am J Psychiatry* **156**, 557–563.
- Piven, J., Wzorek, M., Landa, R., Lainhart, J., Bolton, P., Chase, G.A. & Folstein, S. (1994) Personality characteristics of the parents of autistic individuals. *Psychol Med* **24**, 783–795.
- Risch, N., Spiker, D., Lotspeich, L. et al. (1999) A genomic screen of autism: evidence for a multilocus etiology. *Am J Hum Genet* **65**, 493–507.
- Roberts, R.J. & Pennington, B.F. (1996) An interactive framework for examining prefrontal cognitive processes. *Dev Neuropsychol* **12**, 105–126.
- Russell, J. (ed) (1997) *Autism as an Executive Disorder*. Oxford University Press, Oxford.
- Sergeant, J.A., Geurts, H. & Oosterlaan, J. (2002) How specific is a deficit of executive functioning for attention-deficit/hyperactivity disorder? *Behav Brain Res* **130**, 3–28.
- Shallice, T. (1982) Specific impairments in planning. *Philos Trans R Soc Lond B Biol Sci* **298**, 199–209.
- Slaats-Willemse, D., Swaab-Barneveld, H., De Sonnevile, L., van der Meulen, E. & Buitelaar, J. (2003) Deficient response inhibition as a cognitive endophenotype of ADHD. *J Am Acad Child Adolesc Psychiatry* **42**, 1242–1248.
- Slaats-Willemse, D., Swaab-Barneveld, H., De Sonnevile, L. & Buitelaar, J. (2005) Familial clustering of executive functioning in affected sibling pair families with ADHD. *J Am Acad Child Adolesc Psychiatry* **44**, 385–391.
- Steffenburg, S., Gillberg, C., Hellgren, L., Andersson, L. & Gillberg, I., Jakobsson, G. & Bohman, M. (1989) A twin study of autism in Denmark, Finland, Iceland, Norway and Sweden. *J Child Psychol Psychiatry* **30**, 405–416.
- Szatmari, P., Jones, M.B., Tuff, L., Bartolucci, G., Bartolucci, G., Fisman, S. & Mahoney, W. (1993) Lack of cognitive impairment in first-degree relatives of children with pervasive developmental disorders. *J Am Acad Child Adolesc Psychiatry* **32**, 1264–1273.
- Szatmari, P., Jones, M.B., Zwaigenbaum, L. & MacLean, J.E. (1998) Genetics of autism: overview and new directions. *J Autism Dev Disord* **28**, 351–368.
- Turner, M. (1997) Towards an executive dysfunction account of repetitive behaviour in autism. In: Russell, J. (ed), *Autism as an Executive Disorder*. Oxford University Press, Oxford, pp. 57–100.
- Turner, M.A. (1999) Generating novel ideas: fluency performance in high-functioning and learning disabled individuals with autism. *J Child Psychol Psychiatry* **40**, 189–201.
- Wallach, M.A. & Kogan, N. (1965) *Modes of thinking in young children*. Holt, Rinehart & Winston, New York.
- Welsh, M.C. & Pennington, B.F. (1988) Assessing frontal lobe functioning in children: views from developmental psychology. *Dev Neuropsychol* **4**, 199–230.
- Wong, D., Maybery, M., Bishop, D., Maley, A., Hill, W. & Hallmayer, J. (2003) Theory of mind and executive function: primacy and independence in autism. In Katsikitis, M (ed), *Proceedings of the 38th APS Annual Conference*. The Australian Psychological Society, Melbourne, pp. 252–254.

## Acknowledgments

Our thanks to Sarah Davenport, Isabel Fernandez, Kate Fitzpatrick, Wayne Hill, Matt Huitson, Elise Mengler, Sarra Miller, Bronwyn Morgan, Nicole Petterson and Keira Thomson who participated in data collection, scoring, data entry and/or task programming. This study would not have been possible without the support of the families who gave their time generously to participate in this study. This research was funded by an Australian National Health and Medical Research Council grant awarded to Joachim Hallmayer, Murray Maybery and Dorothy Bishop. The work was conducted while Dana Wong was the recipient of an Australian Postgraduate Award and a Jean Rogerson Postgraduate Award.