Intact inner speech use in autism spectrum disorder: Evidence from a short-term memory task

Article *in* Journal of Child Psychology and Psychiatry · February 2008

DOI: 10.1111/j.1469-7610.2007.01836.x · Source: PubMed


David Williams
University of Kent


Francesca Happe
King's College London


Chris Jarrold
University of Bristol

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

Intact Inner Speech use in Autism Spectrum Disorder: Evidence from a Short-Term Memory Task

David Williams & Francesca Happé
Institute of Psychiatry, Kings College London

Christopher Jarrold
University of Bristol

Acknowledgements
This research was supported by a Medical Research Council Studentship awarded to the first author. We would like to warmly acknowledge the support of all the staff, students and parents/guardians who participated in this research, from the following schools: Linden Bridge school, Springhallow school, Sybil Elgar school, Queensmill school, Garratt Park school, Brent Knoll school and Turney school. Finally, we would like to thank two anonymous reviewers for their helpful comments on previous drafts of this manuscript.

Abstract
Background: Inner speech has been linked to higher-order cognitive processes including 'theory of mind', self-awareness and executive functioning, all of which are impaired in Autism Spectrum Disorder (ASD). Individuals with ASD, themselves, report a propensity for visual rather than verbal modes of thinking. This study explored the extent to which children with ASD used inner speech or visual imagery to support recall from short-term memory.
Method: Twenty-five children with ASD and 20 comparison children with moderate learning disabilities completed an immediate serial recall task, in which stimuli consisted of items with either phonologically similar features, visuo-spatially similar features or control items which were neither visuo-spatially nor phonologically similar.
Results: ASD and comparison participants, with verbal mental ages above 7 years, recalled phonologically similar stimuli less well than control stimuli, indicating that both groups were using inner speech to recode visually presented information into a phonological code. In contrast, those participants with verbal mental ages below 7 years, whether with ASD or not, recalled visuo-spatially similar stimuli less well than control stimuli, indicating visual rather than phonological coding. This developmental pattern mirrors that found in typically developing children.
Conclusions: Under experimental conditions, individuals with ASD use inner speech to the same extent as individuals without ASD of a comparable mental age.

Keywords: Autistic disorder, inner speech, phonological processing, visuo-spatial functioning, memory, cognition.

Introduction

According to Baddeley's (1986) working memory model, inner, or subvocal, speech is controlled by the phonological loop which is specialized for the short-term retention of verbally encoded information. This phonological loop has two components: a phonological store, which holds information in a verbal/phonological form for a short period, and an articulatory control process which translates visual representations into a phonological code (to enter the store) and then acts to rehearse this material so as to reduce its decay in the store. This recoding of material into a verbal medium and its subsequent rehearsal is seen as an example of inner speech.

Whilst inner speech within Baddeley's (1986) model is thought to be independent of executive, self-regulative processes, the notion that it plays a key mediating role in higher mental functions was suggested by Vygotsky (1962) and, more recently, validated empirically in studies of executive functioning in adults (Baddeley, Chincotta & Adlam, 2001; Cinan & Tanor, 2002). According to Vygotsky's socio-cultural approach, inner speech represents an internalised form of social-communicative dialogue with others, which serves to guide and regulate one's ongoing behaviour by allowing multiple, alternative perspectives (originally manifested in external dialogue) upon a situation to be adopted.

The implications of Vygotsky's theory are relevant to our understanding of Autism Spectrum Disorder (ASD), a condition diagnosed on the basis of severe impairments in social interaction (DSM-IV, APA, 1994) and which, on the cognitive level, is characterised by deficits in higher mental functions such as executive functioning (e.g., Hill, 2004) and 'theory of mind' (ToM; e.g., Happé, 1995). Theoretically, if the transition to dialogic modes of thinking in inner speech relies on an adequate experience of interpersonal relations, then individuals with ASD would be expected to show limited, if any, subvocal speech for the purpose of problem-solving. Indeed, Russell, Jarrold and Hood (1999) suggest that the specific profile of executive dysfunction in ASD could be explained by an underlying paucity of inner speech use. They argue that children with autism are challenged reliably only by those executive tasks which require the maintenance, in mind, of a novel, arbitrary rule and which also have non-verbal response modes. Russell et al. suggest that such tasks require the ability to engage in self-talk in order to remind oneself of the rule to be followed. A lack of spontaneous inner speech use could therefore explain the difficulties that individuals with ASD have in performing such tasks.

Anecdotal evidence of a reduced propensity for inner speech use in ASD emerges from the autobiographical accounts of high-functioning individuals who report a predominantly visual thinking style (e.g. Grandin, 1991; Hurlburt, Happé & Frith, 1994). The current experiment was designed therefore to assess the performance of individuals with ASD on a task which was amenable to either visual or verbal solutions (cf. Hurlburt et al., 1994). Specifically, the experiment assessed whether participants with ASD internally recode visually presented material into a verbal medium, or whether they, instead, rely solely on visual representations, as might be predicted on the basis of these introspective reports.

There is good evidence that, in typical development, this recoding does not occur until approximately 7 years of age. An early study by Flavell, Beach and Chinsky (1966), for example, analysed the lip movements of 5-, 7- and 10-year-olds for signs of subvocal rehearsal during a pictorial memory task and found such indications only in the two older groups. A number of studies have also shown that young children's memory for visually presented material is unaffected by the phonological properties of the items. For instance, it is well established that the spoken duration of the items in verbal short-term memory tasks affects recall in adults, longer items (e.g. 'bicycle') being significantly less well remembered than short items (e.g. 'cat') (Baddeley, Thomson & Buchanan, 1975). Baddeley et al. interpreted this 'word length effect' (WLE) as a reflection of covert rehearsal processes.

Longer words are thought to be more difficult to recall because they tend to take longer to subvocally rehearse and are therefore subject to greater decay in the phonological store. In other words, if inner speech (in the form of subvocally labeling pictures and then covertly rehearsing these labels) is being used then a WLE on recall for visually presented stimuli should be evident. Hitch, Halliday, Dodd, and Littler (1989) however, only found reliable WLEs among 8-year-olds, and not 6-year-olds when to-be-remembered items were presented pictorially, suggesting that the younger group were not recoding the picture names into a phonological form, and that this recoding comes on line around 7 years of age (see also Conrad, 1971; Hitch, Halliday, Schaafstal, & Schraagen, 1988).

Before approximately age 7 years, children's thought processes appear to be dominated by visual representations. For example, Hitch, Woodin, and Baker (1989) found that 5-year-olds' serial recall performance for pictorially presented material was detrimentally affected by the visual similarity of the items to be recalled, but not their phonological similarity. In contrast, 11-year-olds showed the opposite pattern, recalling few items with phonologically similar names (e.g., cat, bat, hat), whilst remembering well items with similar visuo-spatial structures (e.g., a pen and a knife, presented at the same angle of orientation). The 'phonological similarity effect' (PSE) shown by these older children indicates a verbal representation of items in memory, since only information which has been translated from a visual to a verbal code could be affected by a manipulation of phonological structure. The 'visuo-spatial similarity effect' (VSE) shown by the young children indicates that, unlike the older children, items were represented only visually, having not undergone any active recoding into verbal form. Whether or not the shift to phonological encoding strategies around age 7 years is an 'all-or-nothing' one, young children certainly appear to be restricted to representing items visually in short term memory (e.g., Brown, 1977; Hayes & Schulze, 1977; Hitch, Halliday, Schaafstal, & Schraagen, 1988).

No study to date has assessed the effect of phonological and visual similarity on serial recall in ASD, although two published studies have assessed the WLE in ASD. Russell, Jarrold & Henry (1996) found a WLE of the typical magnitude in their sample, whilst Whitehouse, Mayberry & Durkin (2006, Experiment 2) found a reduced WLE in children with ASD, concluding from this that individuals 'with autism do not recruit inner speech in order to assist their performance in tasks involving pictorial memory…' (p.8). One difficulty with Whitehouse et al.'s Experiment 2, and hence the conclusion that children with ASD do not use inner speech, is the interpretation of the WLE itself. As the authors note, the suggestion that WLE is the reflection of covert articulatory rehearsal processes is controversial. It may be that rather than resulting from rehearsal processes during the retention period, WLEs are driven by *overt* articulation processes during recall (Cowan et al., 1992; Henry, 1991). That is, the more time it takes to articulate words *at recall* the more time the information in the recall list must be held in the phonological store, and the more it is subject to potential decay. Such output effects would have little to do with the rehearsal strategy (inner speech) during the retention phase, therefore.

Because of these uncertainties regarding the basis of the WLE, the most widely accepted evidence for the verbal encoding of information in memory research is, as noted earlier, a demonstration of the PSE. Unlike the WLE, the PSE cannot be explained in terms of a disruption to output processes (i.e. overt articulation during recall) because items are typically matched for syllable length, and therefore spoken duration. Given the fact that the PSE, unlike the WLE, is an uncontroversial indicator of inner speech use, we decided to adopt the methodology used by Hitch, Woodin and Baker (1989) in their study of typically developing (TD) children. A group of children with ASD and a matched comparison group undertook a serial recall task involving lists of items which were either visually similar, phonologically similar or neither visually nor phonologically similar. If children with ASD do not spontaneously recruit inner speech to mediate this task, then they should exhibit a lesser PSE than comparison children. If children with ASD do not encode information verbally then it remains an open question whether they encode it visuo-spatially, like TD children below approximately 7 years of age (cf. Whitehouse et al., 2006, p. 8). The inclusion of a condition involving sets of visuo-spatially similar items provides a direct test of this possibility since, if children with ASD are using a visuo-spatial encoding strategy, their recall of items with similar visuo-spatial properties should be detrimentally affected.

Specifically, this study explored whether children with ASD spontaneously recruit inner speech to mediate their recall from short-term memory, or whether they instead rely on the use of visuo-spatial representations as suggested by Whitehouse et al. (2006) and as might be predicted both on the basis of the introspective reports of individuals with ASD (Grandin, 1991; Hurlburt et al., 1994) and on theoretical accounts of the typical development of internalised speech (Vygotsky, 1962).

Method

Participants

Ethical approval for this research was obtained from the joint South London and Maudsley NHS Trust/Institute of Psychiatry Research Ethics Committee. Twenty-five children with ASD and 20 comparison children participated in this study, after parents/guardians had given written, informed consent for their children to be included. The participants in the ASD group had received formal diagnoses, by a trained psychiatrist or pediatrician, of autistic disorder (n = 22), Asperger's disorder (n = 2) or pervasive developmental disorder not otherwise specified (PDD-NOS; n = 1) according to established criteria (DSM-IV; APA, 1994). All participants in this group attended specialist autism schools, which required a diagnosis of autism, Asperger's syndrome or PDD-NOS for entry into the school. The comparison group consisted of 18 children with generalised learning disability, who were attending specialist schools for pupils with learning difficulties, and two TD children who were recruited for comparison with those (two) children in the ASD group who achieved full-scale IQ scores of over 100. Children with a primary diagnosis of attention deficit hyperactivity disorder (ADHD) or who had a primary specific language impairment (SLI) were not included in the study, since it was felt that individuals with ADHD would have difficulty conforming to the task and that individuals with SLI would not form a fair comparison group, given their established difficulties with phonological processing (see Leonard, 2000). Verbal and non-verbal abilities were assessed by an appropriate measure for the chronological age of each participant. Thus, 23 (out of 25) children with ASD and 18 (out of 20) comparison children received a short form of the Wechsler Intelligence Scale for Children – Third Edition UK (WISC-III; Wechsler, 1991), consisting of Information, Vocabulary, Picture Completion and Block Design subtests. The IQ estimate gained from this short form has high reliability (Sattler, 1992). The verbal and non-verbal abilities of the four remaining children (two from each group) were assessed by the Wechsler Preschool and Primary Scale of Intelligence – Revised (Wechsler, 1990). Due to limited child availability, two comparison participants only received assessment of verbal abilities by the WISC-III. Following Field (2005) all effect sizes in this study are reported in terms of r. The groups were matched on chronological age and verbal mental age (VMA): chronological age, $t(43) = 0.12$, $p = .91$, $r = .05$; VMA, $t(43) = 0.63$, $p = .53$, $r = .12$; verbal IQ, $t(43) = 0.90$, $p = .37$, $r = .14$; performance IQ, $t(41) = 0.37$, $p = .71$, $r = .09$; and full-scale IQ, $t(41) = 0.62$, $p = .54$, $r = .12$.[1] Participant details are reported in Table 1.

Apparatus and Stimuli

Stimuli were 24 pictures, similar to those used by Hitch, Woodin and Baker (1989), divided into three sets: those with phonologically similar labels (bat, cat, hat, mat, map, rat, tap, cap), those with visually similar appearances, all presented at the same angle, (key, spoon, pen, knife, tie, bone, spade) and control items with no phonological or visual similarity (drum, shoe, fork, bell, leaf, bird, lock, net). All items were one syllable in length and matched for word frequency as indexed by Kucera-Francis (1967) and Thorndike-Lorge (1944) counts, and for imageability and concreteness as reported in the MRC Psycholinguistic Database (Coltheart, 1981). A multivariate analysis of these four measures across the three stimulus types revealed a non-significant main effect of stimulus type using Wilks' criterion, $F(8, 34) = 0.76$, $p = .64$, confirming the adequacy of this matching.[2] Fifteen of the 24 pictures were drawn from Snodgrass and Vanderwart's (1980) standardized set. Nine of the pictures (tap, rat, key, spade, cap, fork, mat, net, map) were not available from Snodgrass and Vanderwart's set and so were either hand drawn or selected from Microsoft Clipart so as to match as closely as possible the style of Snodgrass and Vanderwart's pictures. Figure 1 provides examples of the stimuli used in each condition.

Pictures were mounted in rows on black card and covered individually by coloured squares of card which could be lifted to reveal the pictures underneath.

Design and Procedure

Participants took part in each condition (phonological, control, visual), with the order of presentation counterbalanced. Short-term memory for the materials in each condition was assessed using an incremental span procedure. Items were organized into rows, the sequence length of which varied from two to eight pictures per row. There were four trials (i.e. rows) at each sequence length. Items in each trial were revealed in turn for approximately one second. After presentation of the last item in each trial, the experimenter pointed to the cover of the first picture and said 'which one was here?' and, having been offered an answer, continued down the row pointing to each picture cover in turn. Each trial was considered to have been successfully completed if all items were recalled in correct order. If at least one of the four trials at a given sequence length had been successfully completed then the participant was given another set of (four) trials at a greater sequence length. When none of the trials at a given sequence length were successfully completed, the participant moved onto the next condition. Each condition began with three-item sequences (i.e., four trials of three items). If none of the trials were successfully completed at this sequence length then participants were given a set of trials with two-item sequences. All conditions were completed in one session, which lasted approximately 20-25 minutes.

Participants were tested individually in a quiet room in their school. They were first shown and asked to name each picture from the task. Any failures to produce the designated name for each picture were corrected. This was done in order to ensure that the correct labels, which had been matched for syllable length, were being used. If, for example, a child had consistently used an incorrect/alternative label (e.g., 'baseball bat', for the item 'bat') in a particular condition, then any findings would be confounded by uncontrolled word length effects.

Following the naming phase, participants completed an unrelated task (as part of a battery of tasks) which took approximately five minutes to complete. They were then given a practice trial of the experimental task, consisting of four two-item sequences. If, during this practice trial or at any other point in the test, the participants overtly articulated the labels of pictures during the sequence presentation they were instructed to remain silent until all the pictures in a trial had been seen.

Data Scoring and Analysis

Hitch, Woodin, and Baker (1989) did not use a span procedure in their experiment, but rather presented individuals with a set of trials of a fixed list length. They then coded recall in terms of number of items correctly recalled across each condition. However, such an approach is not appropriate in a span procedure of the form employed here, because it risks giving undue weight to trials at longer as opposed to shorter list lengths. Rather, we adopted the preferred method in the literature of scoring span performance (Conway et al., 2005); namely totaling the number of trials on which serial recall was completely correct in each condition[3].

Given that TD children below approximately 7 years of age characteristically display a VSE, indicating that they mediate this kind of task not by phonological recoding but by relying on visuo-spatial representations of items, we investigated whether experimental and comparison participants at the equivalent developmental level (i.e., with mental ages below 7 years) showed this same pattern of performance in our study. To explore this issue we subdivided participants according to their VMA and compared the recall performance of participants with VMAs over 7 years (high VMA group) to that of participants with VMAs under 7 years (low VMA group). There were 32 participants (18 ASD and 14 comparison) in the high VMA group and 13 participants (7 ASD and 6 comparison) in the low VMA group.

Data were therefore analysed using a 2 × 2 × 3 repeated-measures ANOVA, with diagnostic group (ASD/comparison) and VMA group (under 7 years/over 7 years) as the between-participants factors and with condition (phonological, visuo-spatial and control) as the within-participants variable.

Results

The ANOVA indicated that the main effect of condition was not significant, showing that participants' recall performance was not reliably affected, overall, by the type of stimulus, $F(2, 82) = 1.92$, $p = .15$, $r = .15$. The main effect of diagnostic group was also not significant, indicating that participants with ASD showed the same level of recall performance, across conditions, as did comparison participants, $F(1, 41) = 0.91$, $p = .35$, $r = .15$. The main effect of VMA group was significant, however, reflecting the superior overall performance of the high VMA group, $F(1, 41) = 6.42$, $p = .02$, $r = .37$.

There was no significant interaction between condition and diagnostic group, indicating that participants with ASD showed the same pattern of recall performance as did comparison participants, $F(2, 82) = 0.43$, $p = .65$, $r = .07$. There was, however, a significant interaction between condition and VMA group, $F(2, 82) = 7.54$, $p = .001$, $r = .29$. To break down this interaction, within-participants contrasts were performed comparing the recall of phonologically similar and visuo-spatially similar stimuli to control stimuli, by each VMA group independently. The first contrast revealed that, for those participants with VMAs above 7 years (high VMA group), the recall of phonologically similar stimuli was significantly lower than the recall of control stimuli, $F(1, 31) = 15.74$, $p < .001$, $r = .58$. The recall of visuo-spatially similar stimuli did not differ significantly from the recall of control stimuli in these participants, however, $F(1, 31) < 0.01$, $p > .99$, $r < .01$. Figure 2 illustrates the performance of ASD and comparison participants, with VMAs over 7 years, on the experimental task. The second contrast revealed the opposite pattern of performance by participants with VMAs below 7 years (low VMA group). In these participants, the recall of visuo-spatially similar stimuli was significantly poorer than the recall of control stimuli, $F(1, 12) = 4.59$, $p = .05$, $r = .53$, whilst the recall of phonologically similar stimuli did not differ significantly from the recall of control stimuli, $F(1, 12) = 0.18$, $p = .68$, $r = .12$. Figure 3 shows the performance of participants, collapsed across diagnostic groups, with VMAs over and under 7 years.

Next, between-participants contrasts were performed comparing the recall performance of each VMA group (high/low) on each of the three stimulus conditions. These contrasts revealed that, compared to participants in the low VMA group, participants in the high VMA group were significantly superior in their recall of control, $F(1, 43) = 5.88$, $p = .02$, $r = .35$

and visual, $F(1, 43) = 11.85$, $p = .001$, $r = .46$, stimuli. The two VMA groups of participants did not differ significantly, however, in their recall of phonological stimuli $F(1, 43) = 0.27$, $p = .61$, $r = .08$.

Finally, neither the two-way interaction between diagnostic group and VMA group, $F(1, 41) = 0.07$, $p = .80$, $r = .04$, nor the three-way interaction between diagnostic group, VMA group and condition, $F(2, 82) = 0.52$, $p = .60$, $r = .08$, was significant, indicating that participants with ASD and comparison participants were similar in terms of overall levels of performance and patterns of performance across conditions, regardless of which VMA group they were in.

Discussion

The results of this experiment clearly show that children with ASD do not differ from the comparison individuals with moderate learning difficulties in their ability to spontaneously recruit inner speech to mediate the experimental short-term memory task. For children with VMAs over 7 years, in both the experimental and comparison groups, phonologically similar pictorial stimuli were significantly less well remembered than control stimuli (see Figure 2). Indeed, the effect size (r) for the contrast between phonological and control conditions, in those participants in the high VMA group, was .58, a magnitude considered large (Cohen, 1992). As noted above, in short-term memory research, such a PSE is arguably the 'gold standard' for demonstrating intact verbal recoding of visual information. Finding a typical PSE in children with ASD therefore provides the best evidence yet of unimpaired phonological loop functioning in this disorder.

Our results also provided some evidence regarding the development of inner speech in children with ASD and children with moderate learning disability. As highlighted above, TD children do not begin to spontaneously recode visual information into a verbal medium until roughly 7 years of age (Flavell et al., 1966; Hitch, Halliday et al., 1989). Before this, visual information is represented in terms of its visuo-spatial properties. Exactly this pattern of performance was observed among the participants in our sample when they were divided around a developmental level of 7 years mental age. Those participants with VMAs above 7 years clearly recalled phonologically similar items least well, indicating a verbal representation of items. In contrast, participants with VMAs below 7 years showed no detrimental effects of phonological similarity, relative to control condition performance. Instead, these individuals showed significantly poorer performance on visually confusable items (see Figure 3). Consequently, these patterns of performance are remarkably similar to those shown by TD 5- and 11-year-olds in Hitch, Woodin and Baker's (1989) study suggesting that among individuals with learning disability, whether with autism or not, it is developmental level rather than chronological age that predicts short-term memory competence (cf. Jarrold, Baddeley, & Hewes, 2000).

These findings runs counter to those of Whitehouse et al. (2006), who found a reduced word length effect in children with ASD. As noted above, however, the controversy regarding the suggestion that such length effects reflect covert articulatory processes, leads us to believe that the phonological similarity effect, which was not reduced in children with ASD in this study, provides a purer measure of inner speech processes. One further difference between the current study and that of Whitehouse et al. is in the type of comparison group employed. Whereas the comparison group assessed in this study comprised individuals matched with the ASD group on chronological and verbal mental age (and hence verbal IQ), Whitehouse et al. included only a (TD) comparison group matched with their ASD group on VMA, but not chronological age. It could, therefore, have been the verbal IQ decrement shown by the participants with ASD, rather than their diagnostic status per se, in their study that was driving performance differences between groups on experimental tasks (cf. Henry, 2001).

The current findings also run counter to self-reported propensity of individuals with autism for a tendency toward visual thinking. The finding of a predominantly verbal solution to the task in participants with ASD, despite the fact that visual mediation was equally applicable, seems to rule out the possibility that visual thinking is the only problem-solving strategy available to people with this disorder. One speculative explanation is that people with ASD actually do implement inner speech in their everyday lives but have difficulty (meta)representing and, as a result, reporting their own internal worlds as consisting of such dialogue. Perhaps people with autism, given their ToM difficulties, come to see thoughts as synonymous with ‘pictures in the head’ and hence report their own thoughts as consisting in such images even when, in reality, they consist in linguistic forms. This suggestion is compatible with the finding of Flavell, Green, Flavell and Grossman (1997) that young TD children, at an age before they consistently pass ToM tasks, are unable to distinguish their visual from verbal thought processes.

Perhaps the idea of an impoverished self-concept in ASD, rather than a lack of inner speech use, could also partially explain the contrast between the performance of participants on our memory task and their typically poor performance on executive functioning tasks which also seem to require inner speech use (Russell et al., 1999). This argument revolves around the idea that the nature or quality of inner speech may vary according to the robustness of one’s self-concept, whereby (inner) language can serve as a problem-solving device, even in individuals with diminished self-awareness, without being truly dialogic in nature. Inner speech in this case could be said to be ‘for oneself’ rather than ‘to oneself’ and perhaps it is the latter type of internal dialogue which is required to remind oneself of which rule to follow in executive functioning tasks. Our study did not allow an assessment of the quality of the inner speech used by participants with ASD. What is clear from our study, however, is that the *mechanism* for engaging in inner speech is unimpaired in ASD and that children with this disorder are quite capable of implementing verbal representations to mediate complex tasks.

Correspondence to
David Williams, Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, De Crespigny Park, London, SE5 8AF, UK; Tel: 0207 848 0692; Email: d.williams@iop.kcl.ac.uk.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*. 4th ed. (DSM-IV). Washington, DC: American Psychiatric Association.

Baddeley, A.D. (1986). *Working memory.* Oxford, England: Oxford University Press.

Baddeley, A.D, Thompson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14,* 575-589.

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General, 130,* 641-657.

Brown, R.M. (1977). An examination of the visual and verbal coding processes in preschool children. *Child Development, 48,* 38-45.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *The Quarterly Journal of Experimental Psychology*, 33A, 497-505.

Cinan, S., & Tanor, O.O. (2002). An attempt to discriminate different types of executive functions in the Wisconsin Card Sorting Test. *Memory, 10,* 277-289.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Conrad, R. (1971). The chronology of the development of covert speech in children. *Developmental Psychology, 5*, 398-405.

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D. Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review, 12*, 769-786.

Cowan, N., Day, L., Saults, J.S., Keller, T.A., Johnson, T., & Flores, L. (1992). The role of verbal output in the effects of word length on immediate memory. *Journal of memory and language, 31,* 74-79.

Field, A.P. (2005). *Discovering statistics using SPSS* (2nd ed.). Sage Publications.

Flavell, J.H., Beach, D.R. & Chinsky, J.M. (1966). Spontaneous verbal rehearsal in a memory task as a function to age. *Child Development, 37*, 283-299.

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

Grandin, T. (1991). Harvard Mental Health Letter: Overcoming autism: A first person account.

Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66,* 843-855.

Hayes, D.S. & Schulze, S.A. (1977). Visual encoding in preschoolers' serial retention. *Child Development, 48,* 1066-1070.

Henry, L.A. (1991). The effects of word length and phonemic similarity in young children's short-term memory. *Quarterly Journal of Experimental Psychology, 43A,* 35-52.

Henry, L.A. (2001). How does the severity of a learning disability affect working memory performance? *Memory, 9,* 233-247.

Hill, E. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences, 8,* 26-32.

Hitch, G.J., Halliday, M.S., Dodd, A., & Littler, J.E. (1989). Development of rehearsal in short-term memory: Differences between spoken and pictorial stimuli. *British Journal of Developmental Psychology, 7,* 347-362.

Hitch, G.J., Halliday, M.S., Schaafstal, A.M., & Heffernan, T. M. (1991). Speech, 'inner speech', and the development of short-term memory: Effects of picture-labelling on recall. *Journal of Experimental Child Psychology, 51,* 220-234.

Hitch, G.J., Halliday, M.S., Schaafstal, A.M., & Schraagen, J.M.C. (1988). Visual working-memory in young children. *Memory and Cognition, 16,* 120-132.

Hitch, G.J., Woodin, M.E., & Baker, S. (1989). Visual and phonological components of working memory in children. *Memory and Cognition, 17,* 175-185.

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

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

Jarrold, C., Baddeley, A.D., & Hewes, A.K. (2000). Verbal short-term memory deficits in down syndrome: A consequence of problems in rehearsal? *Journal of Child Psychology and Psychiatry, 41,* 233-244.

Jarrold, C., & Brock, J. (2004). To match or not to match? Methodological issues in autism related research. *Journal of Autism and Developmental Disorders, 34,* 81-86.

Kucera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Providence RI: Brown University Press.

Leonard, L.B. (2000). *Children with Specific Language Impairment*. MIT Press

Russell, J., Jarrold, C., & Henry, L. (1996). Working memory in children with autism and moderate learning difficulties. *Journal of Child Psychology and Psychiatry, 37,* 673-686.

Russell, J., Jarrold, C., & Hood, B. (1999). Two intact executive capacities in children with autism: implications for the core executive dysfunctions in the disorder. *Journal of Autism and Developmental Disorders, 29,* 103-112.

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

Snodgrass, J.G., & Vanderwart, M. (1980). A standardised set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174-215.

Thorndike, E.L., & Lorge, I. (1944). *The teacher's word book of 30,000 words*. New York: Teacher's College, Columbia University.

Vygotsky, L. (1962). *Thought and language.* Cambridge: MIT Press.
Wechsler, D. (1992). *Wechsler Intelligence Scale for Children*. (3rd ed.), London, UK: The Psychological Corporation
Wechsler, D. (1990). *Wechsler Preschool and Primary Scale of Intelligence – Revised (UK Edition).* Sidcup, Kent: The Psychological Corporation
Whitehouse, A.J.O., Mayberry, M.T., & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry, 47*, 857-866.

Footnotes

1. Given that performance and, hence full-scale, IQ data were not available for two comparison participants we wanted to rule out the possibility that our groups could have become significantly different from each other on these ability measures if these individuals had been assessed. To ensure this, we arbitrarily assigned the two comparison participants with the minimum performance IQ score possible (45 points) and re-analysed the data. Group comparisons on these measures yielded the following results: performance IQ, $t(44) = 0.79$, $p = .44$, $r = .11$; full-scale IQ, $t(44) = 0.88$, $p = .38$, $r = .13$. The effect sizes for both of these comparisons were small (Cohen, 1992) and thus, both groups would still have been well matched even if these two comparison participants had extremely low performance IQs.

2. Imageability and concreteness scores were not available for the item *cap*, and consequently this item is not included in the multivariate analysis reported in the text. However, an analysis of the two indices of frequency that included this item again showed a non-significant multivariate effect of stimulus type using Wilks' criterion, $F(4, 40) = 0.974$, $p = .43$.

3. Whilst Conway et al. (2005) argue that totaling the numbers of trials on which serial recall was completely correct in each condition is a superior method for scoring span performance than simply calculating 'span' in terms of the longest length list on which any trial was successfully completed, they also suggest an alternative method, namely 'partial-credit scoring'. Here, participants are given a score of one for every successful trial, plus a proportional score for every unsuccessful trial that corresponds to the proportion of items within each that were recalled in the correct position. Item data were unavailable for four of our participants with ASD and so we could not additionally analyse the data using this scoring method. However, within the remaining sample ($n = 41$), the correlations between participant performance, in each stimulus condition, using the partial-credit scoring method and the trial-scoring method (adopted here) were very high, confirming the validity of our scoring method: control condition, $r = .95$; phonological condition, $r = .89$; visual condition, $r = .96$.

Table 1
Participant details

	ASD (n = 25, 3 females)		Comparison (n = 20, 5 females)	
	Mean (*range*)	*SD*	Mean (*range*)	*SD*
Age (years; months)	12;3 (*4;10 – 15;9*)	3;1	12;1 (*5;1 – 15;8*)	2;8
Verbal mental age (years; months)	8;9 (*4;6 – 12;8*)	2;3	8;4 (*4;10 – 12;4*)	2;1
Verbal IQ	77.16 (*51 – 114*)	15.25	73.20 (*51 – 100*)	13.84
Performance IQ	76.84 (*45 - 134*)	20.27	74.39[a] (*45 – 120*)	22.53
Full scale IQ	74.84 (*47 – 107*)	15.99	71.56[a] (*44 – 111*)	18.57

[a] based on data from 18 out of 20 comparison participants

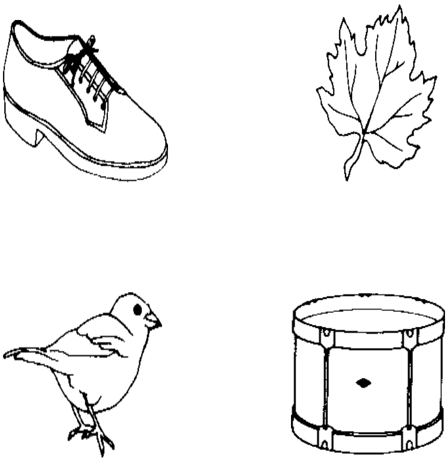
Figure Captions

Figure 1. Examples of experimental materials

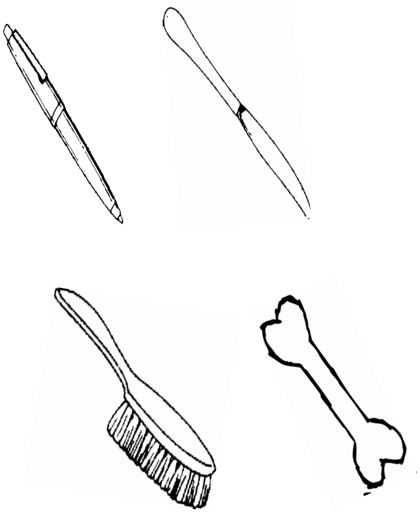
Figure 2. Mean number of trials recalled by ASD and comparison participants with VMAs over 7 years. (Error bars represent one *SE* of the mean).

Figure 3. Number of trials completed in each condition by participants, collapsed across diagnostic groups, with VMAs under and over 7 years. (Error bars represent one *SE* of the mean).

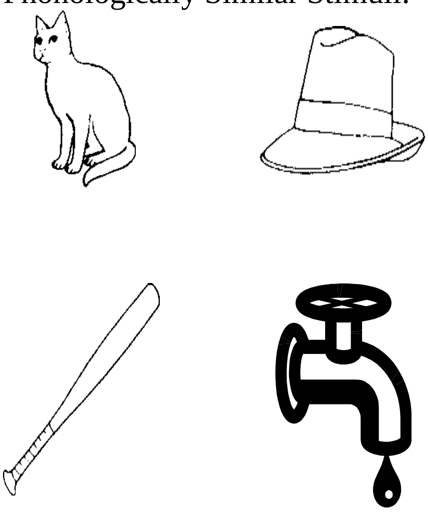
Control Stimuli:

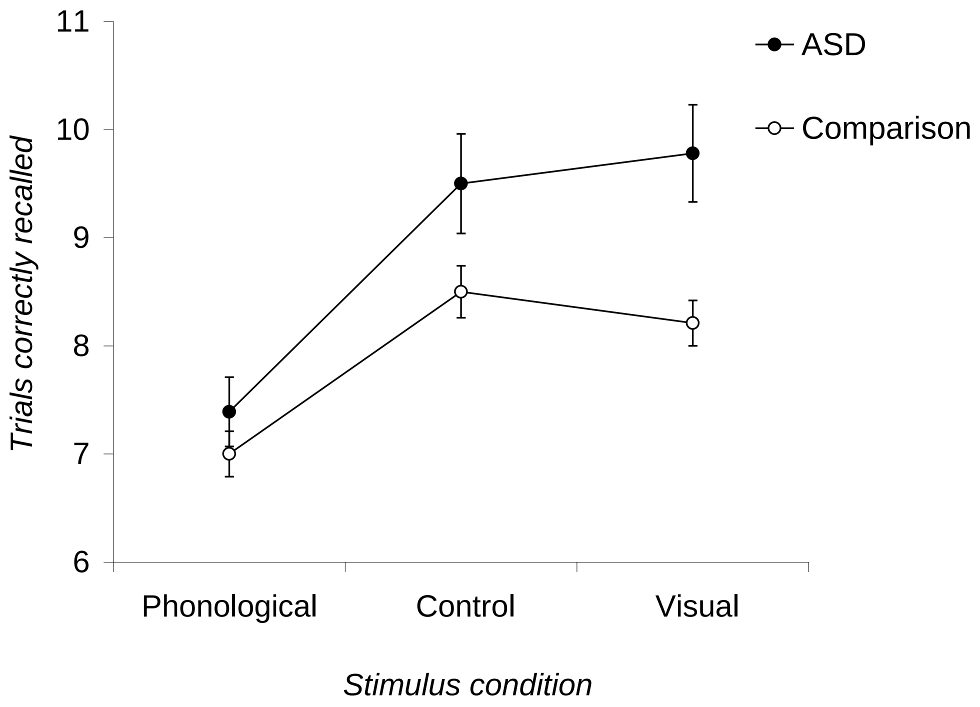


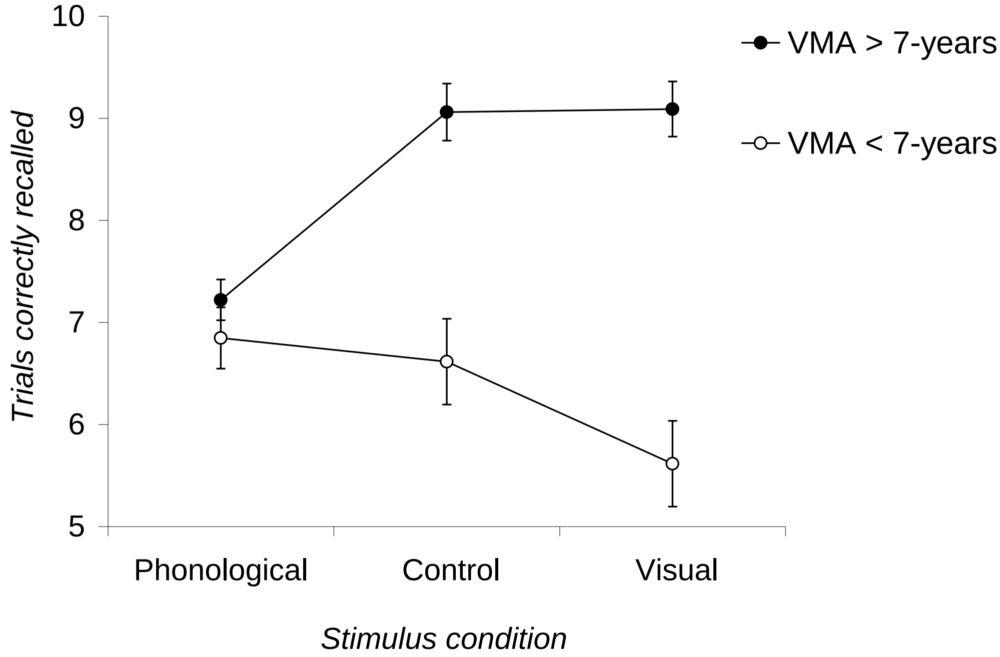
Visuo-spatially Similar Stimuli:



Phonologically Similar Stimuli:




11
10
9
8
7
6
Trials correctly recalled
ASD
Comparison
Phonological
Control
Visual
Stimulus condition


10
9
8
7
6
5
Trials correctly recalled
VMA > 7-years
VMA < 7-years
Phonological
Control
Visual
Stimulus condition