Williams, David M., Peng, Cynthia and Wallace, Gregory L. (2016) *Verbal thinking and inner speech use in autism spectrum disorder.* Neuropsychology Review, 26 (4). pp. 394-419. ISSN 1040-7308.

Williams, D. M., Peng, C., & Wallace, G.L. (2016). Verbal thinking and inner speech use in autism spectrum disorder. *Neuropsychology Review*, *26*, 394-419. DOI: 10.1007/s11065-016-9328-y

Verbal thinking and inner speech use in autism spectrum disorder

David M. Williams[a]
University of Kent, UK

Cynthia Peng
George Washington University, USA

Gregory L. Wallace
George Washington University, USA

[a] Corresponding author: School of Psychology, Keynes College, University of Kent, Canterbury, Kent, CT2 7NP, United Kingdom. Email d.m.williams@kent.ac.uk. Tel. +44 (0)1227 827652

Abstract

The extent to which cognition is verbally mediated in neurotypical individuals is the subject of debate in cognitive neuropsychology, as well as philosophy and psychology. Studying "verbal thinking" in developmental/neuropsychological disorders provides a valuable opportunity to inform theory building, as well as clinical practice. In this paper, we provide a comprehensive, critical review of such studies among individuals with autism spectrum disorder (ASD). ASD involves severe social-communication deficits and limitations in cognitive/behavioural flexibility. The prevailing view in the field is that neither cognition nor behaviour is mediated verbally in ASD, and that this contributes to diagnostic features. However, our review suggests that, on the contrary, most studies to date actually find that among people with ASD cognitive task performance is either a) mediated verbally in a typical fashion, or b) not mediated verbally, but at no obvious cost to overall task performance. Overall though, these studies have methodological limitations and thus clear-cut conclusions are not possible at this stage. The aim of the review is to take stock of existing empirical findings, as well as to help develop the directions for future research that will resolve the many outstanding issues in this field.

Keywords: Inner speech; private speech; executive functioning; cognitive flexibility; memory; motor control; autism

The debate about the relation between language and thought is rich and has a long history in philosophy, psychology, and – more recently – cognitive neuropsychology (e.g., Davidson, 1973; Fodor, 1975; Whorf, 1956). In relation to the topic of the current paper, Vygotskian theories are particularly (but certainly not exclusively) relevant, because they focus on the development of this relation and, in particular, because they have been the biggest influence to date on research into verbal thinking in autism spectrum disorder (ASD), which is the focus of the current paper.

According to Vygotskian theory, language and basic forms of thought ("elementary mental functions") develop along parallel trajectories initially before fusing together at around seven years of age among typically developing (TD) children. Vygotsky outlines the developmental trajectory of verbal thinking as follows; from early in life – shortly after language emerges – children employ overt self-talk ("private speech") as a means of guiding their attention and aiding problem solving. At around age seven, the use of private speech declines rapidly, going "underground" to form inner speech. At this point, Vygotsky argues, thinking takes place *in* speech (as opposed to visual imagery, primarily), and allows for uniquely human forms and levels of cognitive and behavioural flexibility, and self-regulation. According to (neo-) Vygotskian theory, verbal thinking has social-communicative origins, emerging initially from the dialogic exchanges that children have with others (especially caregivers) early in life (see Fernyhough, 2008). For example, when caregivers scaffold children's problem solving through verbal instruction and encouragement, an internalisation of the exchange takes place such that when children are required to problem solve on their own they are able to view the problem not only from their own perspective, but also from the (now internal) perspective of the caregiver. This idea is a key one in Vygotsky's broader socio-cultural theory: "Any function in the child's cultural development appears twice, or on two planes. First it appears on the social plane and then on the psychological plane." (1981, p.163) The crucial implication is that Vygotskian theory predicts any impediment to engaging in meaningful dialogic exchanges with others early in development will result in the atypical development of verbal thinking and, as a result, cognitive/behavioural inflexibility and difficulties with self-regulation. This pattern of deficits is strikingly similar to the profile of behavioural impairments that are diagnostic of ASD.

ASD is a developmental disorder characterised by behavioural impairments in social-communication and behavioural flexibility (restricted and repetitive behaviour and interests; American Psychiatric Association, 2013). These limitations in behavioural flexibility are frequently mirrored by a degree of cognitive inflexibility also (e.g., Van Eylen et al., 2011). It is precisely this kind of inflexibility that Vygotskian theory predicts as a *consequence* of a failure to utilise inner speech. Moreover, social-communication impairments are seen in Vygotskian theory as a direct *cause* of a failure to utilise inner speech appropriately. From a theoretical perspective, therefore, we have a potentially unifying link between the core behavioural features of ASD via a *developmental* pathway; social communication impairments lead to a relative lack of inner speech use, which exacerbates and/or contributes to limited behavioural and cognitive flexibility.

On the cognitive level, a diminished propensity to employ inner/private speech could explain the profile of executive dysfunction that is frequently observed in studies of ASD. Executive functioning is an umbrella term that refers to a set of higher-order abilities that are involved in the control of thoughts and actions in the service of goal-directed behaviour. In particular, executive functions are thought to include inhibitory control, working memory, and set-shifting (Miyake et al., 2000), as well as planning (Pennington & Ozonoff, 1996). Russell, Jarrold, and Hood (1999), for example, noted that people with ASD do not show blanket deficits in all aspects of executive functioning, but rather perform poorly on tasks (such as the Tower of London and Card Sorting Tasks; see below) that require the maintenance in mind of novel and arbitrary rules. Their reasoning was that if one has to complete a task according to a rule that one has not employed directly before, then constant self-reminding of the rule in inner speech is required for successful task completion. Without the use of inner speech, "goal neglect" will occur when there are no cues available in the environment to remind the participant how to perform. Taken together, this constellation of social-communication deficits, inflexibility, and executive dysfunction characteristic of and unique to ASD could, when viewed from a (neo-) Vygotskian perspective, have shared associations with potential inner speech difficulties.

It is important to state at the outset our view that there are strong theoretical reasons to suppose that inner speech use is diminished in ASD, and that this diminution might contribute to behavioural and cognitive difficulties experienced by people with this disorder. Additionally, findings from studies of verbal thinking may hold important clinical and practical implications leading to the possible development and implementation of various strategies/techniques that could serve to improve executive functioning in ASD. However, conclusions should not be drawn firmly until the relevant empirical research on the issue has been evaluated rigorously. This is the aim of the current review. There are a number of methodological issues considered throughout the review as they apply to particular studies. These issues are easier to describe in the context of particular studies. However, as a guide for the reader, issues that are discussed repeatedly concern matching of participant groups for baseline characteristics, inappropriate use of Analysis of Covariance (ANCOVA) to "control" for significant pre-existing between-group differences in baseline characteristics, inappropriate interpretations of interaction effects in Analysis of Variance (ANOVA), small sample sizes, and piecemeal approaches to data analysis that increase the risk of making type I errors.

Empirical studies of verbal thinking in ASD

Table 1 summarises the experimental techniques used frequently to assess verbal thinking. In theory, these types of approaches and techniques can be applied to nearly any cognitive-experimental task, but studies of verbal thinking in ASD have focussed primarily on five domains of cognition, namely planning, task-switching, set-shifting, short-term/working memory, and motor control. Table 2 summarises studies of verbal thinking in ASD, but should be read in the context of the discussion of each study we provide in the text of the paper.

Planning

The ability to plan one's future actions is important for everyday functioning and is considered an executive function by many (Pennington & Ozonoff, 1996). It is well-established that people with ASD tend to show diminished performance in everyday planning and on cognitive-experimental, laboratory-administered measures of planning ability (e.g., Williams & Jarrold, 2013). For example, in the classic Tower of Hanoi/Tower of London task, participants are required to move coloured discs, one disc at a time, across three (or more) pegs. The aim of the task is to move from an initial configuration of pegs (the start-state) to a pre-specified end state in as few moves as possible. The fewer the moves a person takes to reach the goal/end state, the better their planning ability. Russell et al. (1999, p.104) identify the Tower of London task as one of those executive function tasks that requires the maintenance in mind of novel, arbitrary rules and, thus, depends on the utilisation of inner speech for success: "frontal tasks, such as the Tower of London… in which the means-end task requires the following of arbitrarily chosen rules such as 'move from that configuration of coloured balls to this configuration'".

Crucially, three studies have explored the effect of completing the Tower of London task (i.e., planning performance) under conditions of concurrent "articulatory suppression" among individuals with ASD. Articulatory suppression involves repeating a simple word or phrase over and over while performing the primary task (in this case, the Tower of London). Under conditions of articulatory suppression, inner speech use should not be possible, because verbal resources are exhausted by an overt speech stream that is irrelevant to completion of the primary task. Thus, if planning *is* verbally mediated, then articulatory suppression should detrimentally affect planning performance on the Tower of London task. If planning is *not* verbally mediated, then articulatory suppression should have relatively little impact on Tower of London task performance. In relation to ASD, more specific predictions can be made. If inner speech use is diminished in ASD *and* if this diminution contributes significantly to planning difficulties among people with this disorder, then a) Tower of London task performance should be significantly poorer among participants with ASD than comparison participants when completing the task in silent conditions (because comparison participants will support their planning through the spontaneous and beneficial use of inner speech, whereas participants with ASD will not), and; b) between-group differences in Tower of London performance should *not* be apparent when the task is completed under articulatory suppression (because removing the capacity of comparison individuals to employ inner speech should reduce their planning performance to the level of participants with ASD).

Wallace, Silvers, Martin, and Kenworthy (2009) had 28 adolescents with ASD and 25 age-, VIQ-, PIQ-, and sex-matched neurotypical comparison participants complete the Tower of London in silence, as well as under concurrent articulatory suppression. The dependent variable was the number of *extra* moves taken to complete Tower of London trials over and above the minimum number of moves in which completion would be possible. In an ANOVA, the main effects of Group and Condition were significant (indicating diminished planning performance *across* conditions among participants with ASD, and that articulatory suppression negatively affected planning performance in *both* groups). However, the crucial interaction between Group and Condition was not (p = .30) and was associated with a small effect size only (Cohen's d = 0.29). On the basis of this overarching analysis, the conventional approach is to conclude that inner speech was employed by both participant groups to mediate the task (i.e., both groups were affected by articulatory suppression) and that participants with ASD merely showed a general planning deficit that could not be explained by a diminished propensity to employ inner speech. However, Wallace et al. performed additional *post-hoc t*-tests on the data. Statistically, it is not advisable or valid to break down non-significant interactions, in part because it significantly increases the risks of producing type I errors as a result of conducting multiple (unnecessary) comparisons. Nonetheless, Wallace et al. reasoned that, given a (relatively) small sample size (and consequent risk of making a type II error) and, given their *a priori* predictions regarding the pattern of results, conducting additional post hoc tests was acceptable, on balance. The additional tests indicated that, within-participants, planning performance was significantly less affected by articulatory suppression among participants with ASD than among comparison participants. Moreover, between-group differences in planning performance were evident in the silent condition of the task only. The pattern of results yielded by the post-hoc tests is entirely consistent with the notion that planning is less verbally mediated in ASD than in typical development *and* that this difference explains planning deficits among people with ASD. We urge caution in concluding this firmly, however, given that the interaction effect was not significant in the overarching ANOVA. Even if one accepts the grounds suggested by Wallace et al. for conducting the additional post hoc tests, the interpretation of the tests themselves is limited by the non-significant interaction effect. Specifically, given the results from the post-hoc tests, it may be legitimate to conclude that articulatory suppression *significantly* negatively affected planning performance in one group (neurotypical group), but not the other group (ASD group). However, it is not possible to conclude, further, that articulatory suppression negatively affected comparison participants significantly *more* than it affected participants with ASD (see Gelman & Stern, 2006) – only a significant interaction effect would allow this latter conclusion to be drawn (and even then only with appropriate caution, given that significant interaction effects can emerge as a result of measurement issues, rather than substantive differences between groups; see Strauss, 2001). Nonetheless, Wallace et al.'s findings are potentially important.

Williams, Bowler, and Jarrold (2012) also explored the role of inner speech use in planning in ASD. In their relatively small-scale study, 15 adults with ASD, and 16 age-, VIQ-, PIQ-, and sex-matched neurotypical comparison participants completed a Tower task in silence and under concurrent articulatory suppression. The dependent variable was similar to the one used by Wallace et al. (2009), namely the total number of moves taken to complete the Tower trials on average. Unlike Wallace et al., Williams et al. did observe a significant interaction effect; neurotypical comparison participants were significantly more negatively affected by the imposition of articulatory suppression than were participants with ASD. At the individual level, almost 90% of comparison participants were negatively affected by articulatory suppression, whereas only 40% of participants with ASD were.

The findings of Williams et al. (2012), taken in conjunction with the findings of the post hoc tests conducted by Wallace et al. (2009), may well suggest that planning is less verbally mediated among individuals with ASD than among neurotypical individuals. However, implicit in the rationale for both studies was the premise that a failure to utilise inner speech could explain previously-reported difficulties with planning among people with ASD. In that case, one would expect to see a specific pattern of performance across conditions. First, participants with ASD should manifest a significant deficit, relative to comparison participants, in the silent condition of the task. Second, planning performance should be equivalent among ASD and comparison participants in the articulatory suppression condition, because the performance of comparison participants will be reduced to the level of ASD participants once their advantage of using inner speech is removed. In fact, however, there were no between-group differences in planning performance in the silent condition of Williams et al.'s experiment (the effect size - Cohen's *d* - associated with the between-group difference in planning in the silent condition was only 0.26), whereas participants with ASD showed significantly *better* planning than comparison participants in the articulatory suppression condition of the experiment (with an associated effect size of 0.78). In other words, it might be possible to conclude from Williams et al.'s results that people with ASD rely less on inner speech to mediate planning than do neurotypical individuals, but it is not immediately apparent that one can conclude, further, that a failure to utilise inner speech contributes to difficulties with planning in this population. It is important to note that participant groups in Williams et al.'s study were closely matched for age, and verbal, and non-verbal IQ (all between-group differences being non-significant and associated with only a small effect size, all ds $\leq$ 0.38).

The findings of Williams et al. (2012), in particular, raise the question of how planning was mediated by participants with ASD *if* inner speech is not employed by people with this disorder. One obvious suggestion is that planning is visually mediated among individuals with ASD. This suggestion seems plausible on the basis of the anecdotal evidence that people with ASD have a tendency to be "visual thinkers" or to "think in pictures" (e.g., Grandin, 1995; Hurlburt et al., 1994; Kunda & Goel, 2011). Tentative evidence for the suggestion comes from Williams et al.'s post-hoc analyses that revealed planning ability to be associated significantly with performance on the Block Design subtest of the Wechsler Scales among individuals with ASD only (ASD: r = .64; comparison: r = .03). The Block Design subtest is considered to be a good measure of visuo-spatial ability/perception and, thus, the fact it was associated with planning among participants with ASD only suggests that people with this disorder tend to utilise visual processing for the purposes of planning, whereas neurotypical individuals clearly utilise inner speech for that purpose. However, correlational evidence is relatively limited for the purpose of proving hypotheses. What both the studies by Wallace et al. (2009) and Williams et al. (2012) lacked is a further secondary task condition (other than articulatory suppression) to test the hypothesis. Specifically, dual task studies of inner speech use among neurotypical individuals have often employed three secondary tasks; silence, a verbal processing task designed to disrupt verbal processing/inner speech (e.g., articulatory suppression), and a task that is equated for general difficulty with the verbal processing task, but which is designed to disrupt a capacity other than inner speech use (e.g., foot tapping; Emerson & Miyake, 2003). If planning is visually mediated among individuals with ASD, but verbally mediated among neurotypical individuals, then the use of an additional secondary task that selectively disrupts visuo-spatial processing should produce a particular pattern of results; completing the Tower of London task under conditions of articulatory suppression should negatively affect planning efficiency significantly more among neurotypical individuals than among individuals with ASD. Conversely, completing the Tower of London task under conditions of visuo-spatial suppression should negatively affect planning efficiency significantly more among individuals with ASD than among neurotypical individuals. Holland and Low's (2010) study was designed to test this idea.

Holland and Low (2010, Experiment 3) had 13 participants with ASD and 13 neurotypical comparison participants complete a Tower task under silence, articulatory suppression, and visuo-spatial suppression. The visuo-spatial suppression secondary task involved repeated tapping of four blocks (situated among a number of other blocks) in a particular sequence (like a corsi-blocks task; Corsi, 1972). In this experiment, the dependent variable was the time taken to complete trials (rather than number of moves/excess number of moves taken, which is a more traditional measure of planning ability). The predicted interaction effect emerged. This reflected the facts that a) articulatory suppression negatively affected Tower performance among comparison participants but not participants with ASD; b) visuo-spatial suppression affected Tower performance negatively among both groups of participants equally. Indeed, visuo-spatial suppression had a greater negative effect on planning than articulatory suppression among neurotypical participants. This latter finding is difficult to square with the suggestion that planning is primarily verbally mediated in typical development. Holland and Low argue that the finding that visuo-spatial suppression and articulatory suppression *both* affected planning performance in comparison participants is evidence that dual (verbal and visual) codes are utilised for planning in typical development. While this is possible, it is not in keeping with the findings of studies of inner speech use in typical development, which often report a "trade off" between visual and verbal mediation. That is, studies of, for example, working memory tend to show that only when inner speech use is not possible do individuals switch to/prefer to employ visual strategies to mediate the task (e.g., Hale et al., 1996; Hitch, Woodin, & Baker, 1989; Ford & Silber, 1994). Regardless, there are good reasons to be very cautious about over-interpreting the results of Holland and Low's study.

First, the study design employed by Holland and Low (2010) was unusual. It appears that only a single Tower puzzle (which required a minimum of seven moves to complete) was used across all three secondary task conditions. What is more striking is that it appears one and the same trial was used in each condition. That is, participants completed the *same* trial in silence, under articulatory suppression, and under visuospatial suppression. It is very difficult to interpret results under these circumstances. The actual Tower task employed by Holland and Low comprised nine trials in total, which increased incrementally in complexity (i.e., in the minimum number of moves taken to complete each trial). It seems that participants actually completed all of these trials under each secondary task condition, but Holland and Low analysed the data from only one trial (that took seven moves to complete). The trials that took more than seven moves to complete were deemed too “relationally complex” and the trials that took fewer than seven moves were considered as practice trials. Holland and Low’s rationale for this was that “the overall performance of [children with autism] on a [Tower] task that included all nine trials may mask or dilute any effects of the experimental suppression conditions employed” (p.15). However, apart from this being a highly unusual strategy for analysing data from a Tower task, such a selective approach to data analysis runs a significant risk of producing chance results.

Second, while the inclusion by Holland and Low (2010) of an additional secondary task (other than articulatory suppression) represents a methodological advance on previous studies, the particular task used may not have been fit for purpose. The spatial tapping task used was designed to disrupt visuo-spatial processing selectively, but it may have disrupted other functions directly relevant for planning too. Specifically, participants had to keep track of a relatively complex and arbitrary sequence of (four repeated) movements. It is well-established that keeping track of arbitrary sequences draws on executive resources, as well as (in the current context) visuo-spatial resources. Given that planning is considered an executive ability, demanding executive resources in the secondary task would represent a confound, therefore. Moreover, the secondary task necessarily involved *co-ordinating* movements according to rules, so tapped motor skills significantly. This is not a problem in itself, but it is a significant concern when one considers that motor control is a demand of the Tower task. In effect, participants had to make complex movements with one hand (to complete the secondary task), while simultaneously moving discs across pegs (to complete the primary task). These directly competing motor demands would surely confound results, especially given that the dependent variable analysed was time taken to complete trials (rather than the more standard variant of number of moves taken to complete trials). Either of these potential difficulties could explain the unexpected finding that visuo-spatial suppression negatively affected performance of *both* groups of participants to a significantly greater degree than did articulatory suppression.

A note on group matching; Holland and Low's (2010) participant groups were not closely matched for baseline characteristics. In particular, participants were matched for verbal mental age (VMA; ASD: M = 11;5, SD = 5;3; comparison: M = 11;3, SD = 4;0), but not for chronological age (ASD: M = 10;9, SD = 2;4; comparison: M = 9;4, SD = 1;7). Holland and Low report that the between-group difference in age was non-significant ("p > .05", p. 374). However, the difference was associated with a large F value of 3.17 and, according to our calculations (from the means and SDs provided), a large effect size (Cohen's d = 0.84). This means that not only were groups not equated for chronological age, but also that they could not have been matched for verbal IQ. Verbal IQ is a function of age and mental age. Hence, if verbal mental age was equivalent in each group, yet chronological age was higher among ASD participants than among comparison participants, then participants with ASD must have had substantially lower verbal IQs. Using the formula verbal mental age ÷ chronological age = VIQ, ASD participants in Holland and Low's study had a mean VIQ of *approximately* 106, whereas comparison participants had a mean VIQ of approximately 121.

Why is it important to match for baseline characteristics in studies of inner speech use? If participant groups are not matched for these baseline variables, then any between-group differences in experimental measures could be attributable entirely to differences in the baseline variables, rather than to differences in diagnostic group status (see Mervis & Klein-Tasman, 2004). It is not reported whether age or verbal abilities were associated significantly with the key measures of inner speech use in Holland and Low's (2006) Experiment 3 (although, as we note below on p.16, age was moderately correlated with measures of inner speech use in a different experiment reported in the same 2010 paper). Nonetheless, the failure to match groups closely for baseline abilities represents a potential confound in the experiment discussed here.

Summary

In all three studies of the role of inner speech use in planning, it was reported that articulatory suppression did not show a significant detrimental effect on planning performance among participants with ASD (unlike among comparison participants). According to the logic of the dual-task design, this would suggest that planning is not verbally-mediated among people with ASD. However, we argue that this conclusion should be deemed tentative at best. One of the three studies (Holland & Low, 2010) suffered from significant methodological limitations that, in our view, prevent any meaningful conclusions being drawn from the results. In Wallace et al. (2009), the lack of an interaction effect that even approached significance in their overarching analysis significantly limits the validity of their conclusion that people with ASD relied on inner speech significantly less than comparison participants did. Finally, the study by Williams et al. (2012) did provide positive evidence that people with ASD relied on inner speech to mediate planning significantly less than comparison participants did. However, there was no evidence that this atypicality had a negative influence on the planning ability of people with ASD. Instead, participants with ASD performed significantly better than comparison participants on the planning task in the articulatory suppression condition.

Task switching

The ability to switch flexibly between different, often competing, tasks is important for everyday functioning, but an ability that appears to be diminished among individuals with ASD (e.g., Mackinlay, Charman, & Karmiloff-Smith, 2006). From a Vygotskian perspective, inner speech should facilitate flexible switching by providing an ideal means of self-cuing. In effect, inner speech allows one to remind oneself what one's task goals are and to cue oneself to switch between different tasks (or mind sets) when necessary. Modern cognitive science has provided empirical support for this notion. Among neurotypical adults, switching between tasks (e.g., responding to the left side of a display on one trial, then the right side of a display on the following trial, in alternating fashion) results in lower task accuracy and speed of task completion than repeating the same task on all trials (e.g., responding only to the left side of a display on each trial). The difference in accuracy on/speed of completion between task-switch trials and task-repeat trials is known as the "switch cost". The larger the switch cost, the more difficult a person finds switching flexibly between tasks. Crucially, several studies have shown that, among neurotypical individuals, articulatory suppression–but *not* equivalently difficult secondary tasks (e.g., foot tapping) that do not tap verbal resources–significantly increases switch costs (e.g., Baddeley et al., 2001; Baldo et al., 2005; Emerson & Miyake, 2003; Miyake et al., 2004). In other words, preventing inner speech use affects people's ability to switch between tasks in a relatively specific manner, supporting the notion that inner speech plays a major role in facilitating task switching.

Two studies have investigated the effect of articulatory suppression on the task switching performance of individuals with ASD. The authors of each study made the prediction that concurrent articulatory suppression would not significantly increase the switch cost among participants with ASD because participants with this disorder would not tend to mediate the task verbally anyway. Whitehouse et al. (2006, Experiment 3) compared the performance of 23 children with ASD, and 23 verbal ability- and sex-matched comparison participants on an arithmetic switching task. In a "task-repeat" condition, participants were required to add one number to another number on all trials. In contrast, in a "task-switch" condition, participants were required to add one number to another number on one trial, then subtract one number from another on the next trial, in alternating fashion. Participants completed each condition once under articulatory suppression and once in silence. Whitehouse et al. found the predicted interaction between group (ASD/comparison), condition (task-repeat/task-switch), and verbal output (silence/articulatory suppression) in an ANCOVA (non-verbal ability as a covariate). This interaction reflected that articulatory suppression negatively affected performance in the task-switch condition among comparison participants only. In other words, articulatory suppression increased the switch cost for neurotypical participants only, suggesting that task switching is not mediated verbally among participants with ASD.

Whitehouse et al.'s (2006) study was seminal in the field, and the finding that participants with ASD were unaffected by articulatory suppression on the switching task is potentially revelatory in offering a possible explanation for the task switching difficulties that people with this disorder are often reported to have (e.g., Van Eylen et al., 2011). However, the interpretation of findings is not straightforward. The authors went to admirable lengths to rule out alternative interpretations of their finding. However, some important issues should be considered.

First, the source of the significant interaction effect was not entirely the one predicted by the authors. Just as Williams et al. (2012), and Holland and Low (2010, Experiment 3) report, individuals with ASD did not show significant deficits in baseline executive task performance (i.e., when the primary task was completed in silence). Rather, interaction effects in all three studies, including Whitehouse et al., were driven by the significantly *superior* performance of individuals with ASD versus comparison participants in the articulatory suppression conditions of the respective tasks. Thus, it is not clear that any of these studies show inner speech *impairments*, as such.

Second, a subsequent reanalysis of the data from Whitehouse et al. (2006, Experiment 3) suggests that the failure to employ inner speech use among participants with ASD was somewhat apparent, rather than real. When Lidstone, Fernyhough, Meins, and Whitehouse (2009) focussed on individual-level, rather than only group-level, data, it became clear that task-switching performance was substantially negatively affected by articulatory suppression among the majority of participants with ASD (60%). In other words, only a sub-sample of participants with ASD in Whitehouse et al.'s study were not relying on inner speech to mediate the task-switching measure. Importantly, this sub-sample of participants whose task performance was unaffected by articulatory suppression had a mean verbal mental age of only 7 years 9 months (*SD* = 1 year; 4 months). Given that participants below approximately seven years of age would not be expected to utilise inner speech to support cognitive task performance, it is not necessarily indicative of inner speech *impairments* that this sub-sample of participants was unaffected by articulatory suppression (see Williams & Jarrold, 2010).

Holland and Low (2010, Experiment 1) also explored the effect of articulatory suppression on the same task switching paradigm employed by Whitehouse et al. (2006) among 13 children with ASD and 13 neurotypical comparison participants. The results reported were remarkably similar to those reported by Whitehouse et al. Switch costs were significantly negatively affected by articulatory suppression among neurotypical participants only, which led the authors to conclude that individuals with ASD do not employ inner speech to mediate task switching. Moreover, just as reported by Whitehouse et al., differences between the groups in task switching performance were driven by *superior* performance among participants with ASD in the articulatory suppression condition, rather than diminished performance under silent conditions.

In a second experiment, Holland and Low (Experiment 2) had the same participants who completed Experiment 1 complete further arithmetic trials (task-switch and task-repeat) under the same conditions of "visuo-spatial suppression" as in their study of planning (their Experiment 3, discussed above). Rather than combine all of the data from experiments 1 and 2, and then conduct a 2 (Group: ASD/comparison) x 2 (trial type: switch/repeat) x 3 (Condition: silent/articulatory suppression/visuo-spatial suppression) ANOVA, Holland and Low took a more piecemeal approach to data analysis. First, they *partially* combined data from experiments 1 and 2 in order to perform a 2 x 2 x *2* (rather than 2 x 2 x 3) ANOVA, excluding data from the articulatory suppression condition. In other words, they compared the effect on task-switch and task-repeat trials of a) visuo-spatial suppression and b) silence among each participant group. The results demonstrated that participants from *both* diagnostic groups were significantly slower to complete *both* types of trial in the visuo-spatial suppression condition than in the silent condition. It is important to clarify that this means that switch costs per se were not selectively affected by visuo-spatial suppression (relative to silence), but that task performance as a whole was negatively affected. Second, Holland and Low combined data from experiments 1 and 2 in a different way and conducted a 2 (Group: ASD/comparison) x 2 (Condition: articulatory suppression/visuo-spatial suppression) ANOVA on task-*switch* trials only. Data from task-repeat trials and all data from the silent secondary task condition were excluded from this second analysis. The results revealed a significant interaction effect, reflecting the fact that participants with ASD were less negatively affected by articulatory suppression than by visuo-spatial suppression, whereas comparison participants were equally negatively affected by each type of suppression. Third, Holland and Low performed the same ANOVA as in their second analysis, but used performance in task-*repeat* trials, rather than task switch trials as the dependent variable. Results were that visuo-spatial suppression and articulatory suppression equally negatively affected performance on task-repeat trials among both ASD and comparison participants. Holland and Low conclude from these three analyses that their "findings can be parsimoniously regarded as evidence for [children with autism] demonstrating impoverished representational elaborations in the context of service to executive control, whereby task-switching would otherwise typically afford inner speech in addition to visuospatial resources" (pp.13-14).

One difficulty with the conclusion drawn by Holland and Low (2010) is that the analyses on which the conclusion was based were not themselves parsimonious. As discussed in relation to Wallace et al. (2009), failure to find overarching interaction effects limits the conclusions that one can draw from the data. Aside from this issue, several potential limitations with Holland and Low's experiments (discussed in relation to their Experiment 3, above) prevent clear interpretation of experimental results, in our view. As noted above, participant groups were not closely matched for age (or verbal IQ). This is critical, because Holland and Low report that the association between age and the key measure of inner speech use (the extent to which articulatory suppression negatively increased the latency switch cost) in Experiment 1 was moderate and negative (r = -.37) among participants with ASD. Remarkably, Williams et al. (2012) found an almost identical correlation between age and the size of the articulatory suppression effect on planning among their sample of adults with ASD (r = -.36). Thus, in both studies, the older participants with ASD were, the *less* their executive functioning performance was negatively affected by articulatory suppression. The fact that, in Holland and Low's study, participants with ASD were substantially older than comparison participants to start with means that between group differences in inner speech use could be partially, or totally, the result of differences between the groups in chronological age. Moreover, the ambiguity about whether the visuo-spatial suppression task selectively disrupted visuo-spatial processing, or whether it also tapped executive functioning, makes it difficult to interpret the results from their Experiment 2 (discussed above on p.12).

Summary

On the basis of our review, it seems difficult to conclude with any conviction that participants with ASD are atypical in recruitment of inner speech to mediate task switching. It may be the case that individuals with ASD employ inner speech for this purpose to a lesser extent than neurotypical participants (although the analysis provided by Lidstone et al. does not support this possibility). However, the potential confounds inherent in both existing studies make it difficult to conclude this with certainty. Even if these confounds are ignored/are irrelevant and the experimental results stand, it is not clear what explanatory power the results have. The fact that in Whitehouse et al., and Holland and Low the only *between*-group differences in task switching performance reflected *superior* performance among participants with ASD (when completing the task under articulatory suppression) makes it difficult to conclude that failing to employ inner speech could be considered an impairment. If anything, it could be argued that the reverse is true, namely that neurotypical individuals have a deficit in performing multiple tasks (i.e., task switching while talking out loud).

Cognitive flexibility

A cognitive construct that bears similarity to task switching is cognitive flexibility. The latter refers to one's ability to shift flexibly between different mental operations and is frequently measured using one variation or other of the Wisconsin Card Sorting Test (WCST). In the WCST, participants are presented with a number of stimulus cards, each of which varies on several dimensions (e.g., colour, shape, and number of items), and they are asked to match their cards to one of an array of presented cards. However, participants are not told how to match (i.e., on what dimension) but they are given feedback on whether each match they make is correct; thus, they must infer the matching rule. Periodically throughout the task, the matching rule is changed. The key dependent measure of cognitive flexibility is the number of perseverative errors made on the task (i.e., the number of times an individual continues to sort by an old rule, after the experimenter has switched the matching criterion). Other measures of performance include the number of categories completed and number of correct responses provided. Multiple studies have observed deficits in WCST performance among individuals with ASD (see Van Eylen et al., 2011), regardless of whether the task was computerised or administered manually (see Williams & Jarrold, 2013)

Russell-Smith, Bronwynn, Comerford, Maybery, and Whitehouse (2014) investigated the performance of 17 children with ASD and 18 neurotypical comparison participants on a modified version of the WCST. Participants completed the task under each of four conditions: silence, articulatory suppression, concurrent mouthing, and overt self-talk. In the mouthing condition, children were required to open and close their mouths in time to a metronome. This condition was included to control for general motor demands of articulatory suppression. In the overt self-talk condition, children were required to verbalise their strategies each time they sorted a response card. This self-talk condition is of high importance, because it offers a further opportunity to investigate the role of inner speech in mental set-shifting/cognitive flexibility. If set-shifting is not verbally mediated among individuals with ASD and if this explains their characteristic impairments in this domain, then encouraging participants to verbalise their strategies in the self-talk condition of the WCST should enhance performance relative to performance in all other conditions. Indeed, this kind of experimental manipulation was suggested by Russell et al. (1999) as a direct test of the hypothesis that a failure to employ inner speech contributes to executive dysfunction in ASD: "encouraging children with autism to encode the rules verbally should improve their performance on executive tasks to a greater degree than the performance of comparison children" (p.111). The logic here is that forcing individuals to employ *overt* self-talk during task performance should ameliorate the executive difficulties that they normally experience as a result of not employing inner speech.

In relation to Russell-Smith et al.'s study, participants with ASD should perform significantly less well than comparison participants in the silent and mouthing conditions of the WCST, because neurotypical individuals will have the advantage offered by the spontaneous use of inner speech in these conditions. However, in the articulatory suppression and self-talk conditions, no between-group differences in performance should be apparent; in the articulatory suppression condition, the performance of comparison participants should be *reduced* to that of ASD participants because the usual advantage they gain from utilising inner speech will not be possible in the suppression condition. In the overt self-talk condition, the performance of participants with ASD should be *enhanced* compared to that of comparison participants, because the forced use of language to mediate the task should mask the failure of individuals with ASD to mediate the task verbally. As it turned out, this was not quite the pattern of results that Russell-Smith et al. observed.

Russell-Smith et al. (2014) subjected their data to a series of 2 (Group: ASD/comparison) x 2 (Verbal ability: high/low) x 4 (condition order: four levels) x 4 (Condition: silence, articulatory suppression, concurrent mouthing, and overt self-talk) ANCOVAs, including age as a covariate (because participant groups were not closely matched for age). The dependent variable in each ANCOVA was the number of categories completed, number of correct responses, and number of perseverative errors, respectively. Broadly, the most important results were reflected in significant group x condition interaction effects for each dependent variable. Russell-Smith et al. broke down these interactions by exploring performance across conditions *within* each diagnostic group. In the main, however, they did not report *between*-group differences in performance. Within-participants analyses revealed broadly that WCST performance was equivalent across conditions among participants with ASD (i.e., they were not affected by any of the manipulations), whereas comparison participants performed a) significantly less well in the articulatory suppression condition than in any of the other conditions, and b) significantly better in the self-talk condition than in any of the other conditions. Russell-Smith et al. concluded that these results constituted "further evidence for a link between inner speech limitations and executive functioning" in ASD (p.1236). However, there are arguably a number of difficulties with the study that prevents such a conclusion from being drawn with any conviction.

First, the groups were not matched closely for age or VIQ (between group differences in these variables approached significance and were associated with moderate effect sizes), which is why they were included either as a covariate (age) or (unusually) as a between-subjects factor (verbal ability) in the ANCOVAs conducted by the authors: "as age and verbal ability were marginally discrepant across the two groups, these two variables were included in the analyses to ensure that these differences were not influencing key results in any meaningful way" (p.1237). In essence, the logic behind this use of ANCOVA lies the question, "if the diagnostic groups did not differ on the covariate(s), would the between-group difference in the dependent variable still exist?" Put another, more specific, way, "would individuals with ASD use less inner speech to mediate task performance if they were the same age as comparison participants?" Although use of ANCOVA to "control for" pre-existing group differences is relatively frequent in ASD research, it is flawed statistically and conceptually, and leads to erroneous conclusions about results (see Miller & Chapman, 2001, for an excellent discussion of this issue; see Joseph, McGrath, & Tager-Flusberg, 2005a for an excellent discussion of this issue in relation to the study of inner speech in ASD). Use of ANCOVA is subject to two assumptions. The first is that the association between the dependent variable and the covariate is equivalent in each of the groups (the "homogeneity of regression slopes" assumption). If the relation between the two variables differs in magnitude or direction across groups, then the subsequent regression model is certain to be inaccurate. This is particularly important in relation to the study of verbal thinking in ASD, because in the few studies that have reported the relevant associations, ASD and comparison groups frequently appear to differ (see Liss et al., 2001; Joseph et al., 2005a). Even if regression slopes were equivalent across diagnostic groups in studies of verbal thinking that have employed ANCOVA, none of these studies would meet the second assumption of this test, namely that groups are equated on the covariate (the "independence of covariate and treatment effect" assumption). Controlling for non-random group differences (e.g., in age) is not possible using ANCOVA, or any other statistical technique. To quote Fleiss and Tanur (1973, pp.513-517, in Miller & Chapman, 2001, p.43),

> "[N]o amount of statistical manipulation can tell one what might have been had certain differences been non-existent… The overwhelming weight of logic is on the side of those who warn that neither the analysis of covariance nor any other statistical technique can undo systematic differences which were out of the investigator's control".

Ignoring the issues concerning use of ANCOVA, Russell-Smith et al. (2014) argue that age discrepancies between groups probably did not confound results because "as the ASD group was slightly older on average than the typically developing group, if age was influencing results, it should be in the direction of masking inner speech deficits in ASD children" (p.1237). Although this argument is intuitively appealing and likely correct in some cases/with regard to some domains of cognition, it does not always hold true and cannot therefore be relied upon as an *a priori* justification for not matching groups [see above discussion of Holland & Low (2010, Experiment 1) and Williams et al. (2012, Experiment 2)].

Second, individuals with ASD did not show any impairments on the WCST in the silent condition of the task. On the basis of the only between-subjects analyses conducted by Russell-Smith et al. (2010), the authors reported that ASD and comparison participants did not differ significantly in the number of perseverative errors made, number of categories completed, or number of correct responses made in the silent condition of the WCST (between-group differences were associated with only very small effect sizes, ds ≤ 0.11, as Russell-Smith et al. report on p.1239). Thus, even if individuals with ASD did not utilise inner speech to mediate the task in this baseline condition, it is not clear that this detrimentally affected their task performance. Indeed, based on Russell-Smith et al.'s Figure 2 (p.1239), it is not clear that between group differences were apparent in any of the conditions for any of the dependent variables. Therefore, it is not clear what explanatory power the study has. That is, even if one accepts the validity of the results, it is not clear that diminished inner speech use in ASD can explain the characteristic limitations in cognitive flexibility that are apparent among individuals with ASD.

Third, the finding that Wisconsin Card Sorting Task performance was *not* enhanced by overt labelling among individuals with ASD, whereas it *was* enhanced among neurotypical participants, is contrary to predictions. The rationale for including an overt labelling condition was that, if participants with ASD did not spontaneously employ inner speech to mediate the task, then requiring them to use overt self-talk should improve their WCST performance and bring it closer to the level achieved by neurotypical participants in the silent condition (see Russell et al., 1999, also). This logic has been borne out in studies of inner speech use in other disorders.

Among people with schizophrenia – another disorder in which inner speech limitations have been implicated in established executive difficulties – enforcing overt self-talk *does* appear to improve performance on the WCST (Perry et al., 2001). The fact that enforcing overt self-talk did not improve the WCST performance of individuals with ASD suggests, therefore, that we should be cautious in drawing the conclusion that diminished inner speech use lies at the core of executive difficulties in ASD.

Thus far, the literature on inner speech usage in ASD has been dominated by paradigms utilizing articulatory suppression (versus various control conditions) to interfere with self-talk. However, when working with younger children, a more commonly utilized approach is to measure overt self-talk, termed private speech, often through coding in vivo behaviors. Private speech can be viewed as the developmental precursor to inner speech, particularly given that as Vygotsky documented, when children are asked to complete developmentally appropriate problem solving tasks, they typically progress during the preschool years from overt (often interpretable) words and sentences to progressively more covert (i.e., less interpretable), ranging from muttering to fully internalized, speech. To date, only one study has examined the influence of private speech on cognitive flexibility among individuals with ASD. Winsler et al. (2007) examined private speech use during completion of the WCST by 33 school aged children with ASD, as well as 27 age-matched typically developing (TD) children and 21 children with attention deficit hyperactivity disorder (ADHD)[1]. The authors predicted that children with ASD would use significantly less private speech during completion of the WCST than either of the other diagnostic groups. Moreover, they predicted that the private speech that children with ASD *did* use would be less *task-relevant* than the private speech used by children from each of the other groups. In other words, even when overt self-talk was employed by individuals with ASD, it would not be of the quality or directedness necessary to influence task performance positively. Contrary to expectation, however, no significant differences between the ASD and TD groups emerged in either total amount of private speech used or in the extent to which private speech was task-relevant, according to omnibus tests (all *p*s > .25). Rather, it was the ADHD group that showed significantly less task-relevant private speech than the TD group, which replicated earlier findings of atypical verbal thinking in this disorder (e.g., Winsler et al., 2000). Importantly, the ASD group in Winsler et al.'s study showed significantly poorer WCST performance (in terms of numbers of items correct and number of perseverative errors made) than either of the other groups.

The findings that children with ASD showed clear deficits in cognitive flexibility despite normal use of private speech in Winsler et al.'s (2007) study is not obviously in keeping with the idea that limited self-talk contributes to executive dysfunction in this disorder. However, it is difficult to draw definitive conclusions from Winsler et al.'s study, given that participant groups were not matched for verbal or non-verbal ability (or gender), which could have confounded results. Although the authors attempted to overcome this limitation by controlling for verbal ability in ANCOVAs, this approach does not solve the problem (as discussed immediately above in relation to Russell-Smith et al.'s study).

Summary

Two studies have explored the extent to which cognitive flexibility is verbally mediated in ASD. Russell-Smith et al. (2014) found that articulatory suppression did not interfere with performance on the Wisconsin Card Sorting Task among participants with ASD, in keeping with their prediction that task performance would not be verbally mediated in people with this disorder. However, there were no notable between-group differences in baseline cognitive flexibility, so it is not clear that diminished inner speech use negatively impacted on cognitive flexibility in their sample. In contrast, Winsler et al. (2007) found that Wisconsin Card Sorting Task performance was diminished among children with ASD, despite the typical spontaneous use of task-relevant private speech. Thus, taken together, these studies do not support the view that executive dysfunction is the result of diminished verbal thinking in ASD. However, both studies had methodological issues that limit the extent to which one can draw conclusions with certainty.

Short-term and working memory

In Baddeley and Hitch's influential working memory model, the phonological loop is a "slave" system that is specialised for the short-term retention of novel verbal information. It comprises a short-term store that holds verbal information for a limited time to facilitate transference of the information to long-term memory, and an articulatory rehearsal process that serves to translate visual information into a phonological form and refresh decaying representations in the store. In other words, in Baddeley and Hitch's model, inner speech facilitates memory by allowing internal rehearsal of information that would otherwise be forgotten. Importantly, a distinction is drawn in the field between short-term memory and working memory. Short-term memory refers to the temporary storage of visual or verbal information. Working memory involves both the domain specific (visual/verbal) storage of information, as well as some additional (possibly executive) processing of that information (e.g., Bayliss et al., 2006).

Whitehouse, Maybery, and Durkin (2006, Experiment 2) investigated inner speech use in ASD by investigating the extent to which children with ASD demonstrated a "word length effect" in short-term memory. Among neurotypical adults, the spoken duration of to-be-remembered items is negatively associated with the number of items recalled (e.g., Baddeley, Thomson, and Buchanan, 1975). In particular, pictorial stimuli with long names (e.g., "dinosaur") are harder to recall than those with short names (e.g., "ball"). The explanation for this is that adults spontaneously recode visually-presented information into a verbal medium and then rehearse this information internally; long names take longer to rehearse than short names and, therefore, subject to greater decay in the phonological store. Crucially, the word length effect on serial recall is apparent only from approximately seven years of age onwards in typically developing individuals (Hitch et al., 1989a).

In Whitehouse et al.'s (2006, Experiment 2) study, children with ASD and neurotypical comparison participants (matched for verbal mental age, reading ability, and sex) completed, under two conditions, a serial recall task involving pictorial stimuli that had one syllable, or three or four syllable names. In one condition, participants were instructed to stay silent during the presentation of items in the study phase of each trial, whereas in the other condition they were instructed to label each picture as it appeared on screen. The central prediction was that there would be a group (ASD/TD) x condition (silent/label) x word length (one syllable/three or four syllable) interaction reflecting a diminished word length effect among participants with ASD in the silent condition *only*. The rationale here was that children with ASD would not spontaneously employ inner speech in the silent condition. Rather, they would process the pictures visually and would, as a result, be relatively unaffected by word length in this condition. However, in the label condition, in which participants overtly named items, a word length effect would be present among all participants. The actual results were complicated to interpret however. Although the overall group x condition x word length interaction was significant, as predicted, a breakdown of the interaction revealed that participants with ASD showed, unexpectedly, a diminished word length effect in the silent *and* label conditions. One explanation for this pattern is that even when overt speech was being employed by individuals with ASD, it was not influencing memory in the same way/to the same extent as it was among the TD comparison participants. If so, this has significant implications for strategies designed to foster inner speech use among people with ASD, because it suggests that training inner speech use in the form of simple labelling will not be sufficient to influence self-regulation in ASD (see below for further consideration of this).

Whitehouse et al.'s (2006, Experiment 2) findings are particularly striking because they contrast with those of an earlier finding by Russell, Jarrold, and Henry (1996) who found that children with ASD showed a typical word length effect in serial recall. In Russell et al.'s (Experiment 1) study, items were spoken words (rather than pictures, in Whitehouse et al.) and the recall phase (rather than the study phase, in Whitehouse et al.) had two conditions. In one condition (verbal recall), participants recalled items verbally immediately after presentation, whereas in the other condition (non-verbal recall), participants were required to point to pictures of the presented spoken items in correct serial order. Results indicated that, relative to children with intellectual impairment and TD children (matched for VMA only), participants with ASD did not show a diminished word length effect in either recall condition. Interestingly, participants with ASD showed a significantly *greater* word length effect than children with intellectual impairment in the non-verbal recall condition, implying enhanced reliance on inner speech to mediate the task among this sample of children with ASD.

Given the conflicting findings regarding the word length effect among people with ASD, Williams, Happé, and Jarrold (2008) explored another manipulation to the structure of items in a serial recall task (other than word length). Arguably the gold standard evidence for the verbal mediation of short-term memory is a demonstration of the phonological similarity effect, rather than the word length effect. Among TD children from the age of seven years onwards, serial recall of pictorial items is negatively affected by the phonological similarity of the items to be recalled; items with similar sounding names (e.g., cat, hat, mat) are less easily recalled than items with dissimilar sounding names (e.g., ball, cow, shoe). In contrast, serial recall among children below approximately age seven years is affected not by the phonological similarity of items to be recalled, but rather by the visual similarity of those items. Hence, pictures with similar appearances (e.g., a pen, a knife, and a tie, all presented in the same orientation) tend to be recalled less well by young children than pictures with dissimilar visual appearances. This implies that the shift from the visual to the verbal mediation of short-term memory takes place at around age seven in typical development. Williams et al. investigated this by comparing serial recall of phonologically similar, visually similar, and control pictures (that were neither phonologically nor visually similar) among a group of children with ASD and a group of age, VIQ, and NVIQ-matched comparison participants. Crucially, participants from both diagnostic groups were divided according to whether their verbal mental age was above or below seven years. Results suggested that short-term memory is mediated entirely typically in ASD; ASD and comparison participants with VMAs over seven years showed a phonological similarity effect of equivalent magnitude, whereas ASD and comparison participants with VMAs below seven years showed a visual similarity effect of equivalent magnitude. This suggests that the short-term memory is mediated in a developmentally typical fashion among individuals with ASD. This is not consistent with the idea that inner speech use is impaired in ASD.

Williams, Bowler, and Jarrold (2012, Experiment 1) also investigated the phonological similarity effect in ASD, but employed articulatory suppression as an additional means of assessing inner speech. In their study, serial recall of phonologically similar and dissimilar (control) pictorial stimuli was assessed among adults with ASD, and age-, VIQ-, and PIQ-matched neurotypical comparison participants. In one condition (silent condition), participants remained silent as they studied the stimuli on each trial. Then, immediately after the study phase, participants attempted to recall the pictures they had seen in the order they had seen them. In the other condition (articulatory suppression condition), participants repeated a simple word at a rate of one word per second throughout the study phase of each trial. Immediately after the study phase was completed, participants stopped repeating the specified word and attempted to recall the pictures in correct serial order. Engaging in articulatory suppression during the study phase of a short-term memory task should prevent the verbal recoding of visually-presented items and their subsequent covert rehearsal (because resources of the phonological loop are already exhausted by articulatory suppression). In the articulatory suppression condition, therefore, participants should be restricted to processing items visually and, thus, should not show a phonological similarity effect. Results indicated no hint of any difference between ASD and comparison participants in either levels or patterns of performance. Both groups showed a large phonological similarity effect in the silent condition, replicating the finding of Williams et al. (2008) and indicating that items were recoded into a verbal form and covertly rehearsed. Moreover, recall was affected negatively among both groups by articulatory suppression and the phonological similarity effect was eradicated in the articulatory suppression condition.

The findings of Williams et al. (2008, & 2012, Experiment 1), following Russell et al. (1996), provide convergent evidence that inner speech *is* used for the purpose of short-term memory in ASD. One potential difficulty with interpreting the results of the former two studies, however, is that participants were required to overtly label all pictorial items prior to beginning the experimental task. This procedure could represent a significant confound, because it may have encouraged individuals with ASD to internally label items when they would not otherwise have done so spontaneously. That is, it may have acted as a cue to employ inner speech among individuals who would not have done so if they had not been cued to. There are two reasons to believe that this approach/procedure did not confound results, however. First, in Williams et al. (2008) neither participants with ASD nor comparison participants who have verbal mental ages *under* seven years manifested a phonological similarity effect despite the fact these participants overtly labelled items in exactly the same manner as all other participants in both Williams et al. (2008) and Williams et al. (2012, Experiment 1). If it was the case that merely overtly labelling items prior to the experiment resulted in an artificial use of inner speech in ASD, then this should have been observed in all participants. Second, the procedure was originally employed in order to ensure that evaluation of the phonological similarity effect was not contaminated by inaccurate labelling of items. That is, if participants with ASD really were using inner speech, but merely used uncommon labels for items (e.g., “sparrow” instead of “bird”) during covert rehearsal, then a phonological similarity effect would not have emerged and a false conclusion would have been drawn that short-term memory was not verbally mediated in ASD. This issue might be relevant when considering a final study of inner speech use for short-term/working memory in ASD by Joseph, Steele, Meyer, and Tager-Flusberg (2005b)

Joseph et al. (2005b) looked at the extent to which working memory was verbally mediated in ASD. Twenty four children with ASD and 24 comparison participants (closely matched for age, VIQ, and PIQ) completed two versions of a self-ordered pointing task, counterbalanced across participants. In a standard self-ordered pointing task, participants are presented with a series of arrays, each array of which contains the same items but in a different spatial arrangement on each trial. The aim is to point to a different item on each trial, which requires the storage and processing of visually-presented information. The crucial experimental manipulation in Joseph et al.’s study was that in a verbal condition, items were easily nameable, single-syllable objects (e.g., cake, shirt), whereas in a non-verbal condition items were difficult-to-name abstract shapes. Joseph et al. predicted that participants with ASD would show impaired self-ordered pointing in the verbal condition only, because unlike comparison participants they would not benefit from the spontaneous use of inner speech to internally label and rehearse the easily-nameable items in the condition. Conversely, Joseph et al. predicted equivalent performance across groups in the non-verbal condition, because items were non-nameable and thus both groups would be forced to encode, store, and manipulate purely visual representations of items, meaning that comparison participants could not benefit from a natural tendency to use inner speech to support task performance. As predicted, Joseph et al. observed group x condition interactions with respect to the dependent variables of memory span and the number of error-free trials completed, respectively. These interactions reflected that participants with ASD a) performed significantly less well than comparison participants in the verbal condition of the task only, and b) showed a flat profile of performance across conditions, unlike comparison participants who performed better in the verbal condition than in the nonverbal condition.

The results of Joseph et al. (2005b) appear out of keeping with the results reported by Williams et al. (2008, 2012) and Russell et al. (1996). In both studies by Williams et al. it was clear that individuals with ASD were capable of spontaneously recoding visually-presented information and rehearsing it. Why, then, did individuals with ASD in Joseph et al.'s study not do the same and benefit from verbal mediation in the verbal condition of the task? When considering this it is important to bear in mind that Joseph et al.'s study is rigorous in every detail and avoids the potential confounds that we believe are likely to have influenced other studies of inner speech use in ASD. One possible reason for the discrepancy between studies is that Joseph et al. focussed on working memory (i.e., the short-term storage *and* concurrent processing of information in mind), whereas the remainder of the studies explored short-term storage only. Perhaps individuals with ASD spontaneously recode visual information into a verbal form and then rehearse that information with a view to maintaining it in mind, but do not use inner speech for the purpose of updating and *processing* the verbal information that is stored. In the short-term memory tasks employed by Williams et al. and Russell et al., participants were provided with a sequence of items. In contrast, for the self-ordered pointing task, participants need to generate a sequence of points based on a comparison of a stored representation of choices that have already been made with the choice that is to be made on each trial. Such generativity and online processing/updating of working memory can be considered executive in nature. As such, it may be the executive component of working memory, rather than the storage component, for which inner speech is underutilised among people with ASD. This is the interpretation favoured by Joseph et al. While this is plausible (and see Williams et al., 2012, for a related discussion of the distinction between monologic and dialogic inner speech use), it is not in keeping with the findings from other studies of executive functioning discussed above. There are strikingly few unambiguous findings (other than those reported by Joseph et al.) that inner speech is *not* utilised to mediate executive functioning tasks. In our view, it would be important to replicate and extend the findings of Joseph et al. to resolve this discrepancy. In particular, it would be informative to establish whether individuals with ASD whose self-ordered pointing ability is unaffected by the ease with which studied items are label-able (suggesting that they do not employ inner speech to mediate the task) nonetheless show a phonological similarity effect in their serial recall. If this were the case, it would support Joseph et al.'s hypothesis that it is the executive demands of the self-ordered pointing task that are impacted by a diminished tendency or diminished/atypical quality of inner speech use in ASD.

Summary

The majority of studies in this area appear to indicate developmentally appropriate verbal mediation of short-term memory in ASD. Two studies (involving closely matched ASD and comparison participants) have found that adults and children with ASD show typical phonological similarity effect in the serial recall of pictorial stimuli, indicating that short-term memory is mediated verbally (Williams et al., 2008; 2012). Further supporting this conclusion, the Williams et al. (2012) study showed that serial recall performance was significantly detrimentally affected by articulatory suppression (although this study failed to include an appropriate control secondary task condition). A further study by Whitehouse et al. (2007) suggested that individuals with ASD showed a diminished word length effect in serial recall. This contrasts with not only the findings of Williams et al. (2008; 2012), who found an undiminished *phonological* similarity effect, but also with an early study by Russell et al. (1996) who observed an enhanced *word* length effect in children with ASD. A final study by Joseph et al. (2005) was the only study to explore working, rather than short-term, memory. The well-designed Joseph et al. study showed that individuals with ASD found the storage and processing of unnameable stimuli and easy-to-name stimuli equally difficult (unlike comparison participants who showed an advantage for nameable stimuli). Understandably, Joseph et al. interpreted this as showing that individuals with ASD, unlike comparison participants, do not spontaneously use inner speech when stimuli are nameable. This might suggest working memory is not verbally mediated in ASD, although independent replication of the finding would be welcome given the contradictory evidence from studies of short-term memory.

Motor control

Gidley Larson and Suchy (2014) investigated whether motor control is verbally mediated in ASD. Twenty one adolescents with ASD and 22 neurotypical comparison participants performed the “Push-Turn-TapTap” task - a test of motor sequence learning from the Behavioural Dyscontrol Scale-Electronic Version (Kraybill & Suchy, 2011). In this task, participants are instructed to perform a repeated sequence of movements across 21 trials. The sequence involved pushing a joystick forward, then turn it clockwise, then tapping a button twice. The task was completed under three different conditions. First, participants completed 21 trials in silence. Following this, they completed 21 trials under conditions that involved saying a task-*congruent* instruction before each step in the sequence (i.e., “push”, “turn”, “tap-tap”) on each trial. Finally, they completed 21 trials under conditions that involved saying a task-incongruent instruction before each step in the sequence (i.e., “tap-tap” before *pushing* the joystick, “push” before *turning* the joystick, and “turn” before *tapping* the button). In short, the rationale for this design was that if inner speech is *not* used to support control of motor learning, then relative to completion in silence, task performance should not be negatively affected by task-incongruent verbalisations, but it should be facilitated by task-congruent instruction (because participants are being forced to utilise language to guide their motor control; cf. Russell-Smith et al., 2014; Russell et al., 1999). Four dependent variables were considered

(a) Sequencing speed, which refers to the average speed of completion of the entire sequence of three moves across all 21 trials
(b) Sequencing accuracy, which refers to the total number of sequencing errors made across the 21 trials
(c) Motor speed, which refers to the average amount of time taken to complete one element of the sequence only, namely the tap-tap function (i.e., the latency between the first and second tap on each trial)
(d) Motor accuracy, which refers to the number of errors made on the tap-tap element of the sequence (e.g., triple tap, rather than double tap).

The authors reported on p.2153 that "across all analyses, these values were used as dependent variables, Condition [Silence/Task-congruent/Task-incongruent] was used as the within subjects factor, and Group [ASD/comparison] was used as the between-subjects factor". This would have been a parsimonious strategy, and should have yielded a specific Group x Condition interaction if the authors' predictions were borne out by the data (assuming measurement was valid and reliable; Strauss, 2001). However, rather than completing a 2 x 3 ANOVA with respect to each dependent variable (Sequencing speed, sequencing accuracy, motor speed, and motor accuracy), the authors instead completed two separate 2 (Group) x 2 (Condition) ANOVAs for each dependent variable. Specifically, for each dependent variable, the authors conducted one ANOVA comparing group performance in the silent vs. task-*incongruent* condition. The authors subsequently conducted a second ANOVA for each dependent variable comparing group performance in the silent condition vs. task-*congruent* condition. Hence, eight ANOVAs were conducted on this data in all. In each ANOVA, a Group x Condition interaction was predicted. We have raised concerns about such a piecemeal approach to data analysis in relation to the study by Holland and Low (2010) above, so we will not repeat these here although they apply equally (and perhaps more so, given the large number of analyses performed). We do note, however, that Gidley Larson and Suchy (2014) do not consistently break down significant interaction effects, sometimes reporting only within-participants contrasts and sometimes only between-participants contrasts, but never both. This makes it difficult to interpret effects clearly, but we have attempted to do this on the basis of the information provided.

Results from these ANOVAs are summarised in Table 3 and described as concisely as possible below, starting with the result that was most in keeping with the authors' predictions:

In sum, only three out of the eight predicted interaction effects emerged (and not all of these three were straightforward to interpret): With respect to i) *sequencing speed*, a significant Group x Condition interaction effect emerged when comparing group performance in silent vs. task-*congruent* conditions. This reflected slower sequence speed among participants with ASD than comparison participants in the silent condition, but not the task-congruent condition (within-participants contrasts were not conducted). Within-subjects analysis showed that participants with ASD benefitted significantly more than comparison participants from task-congruent verbalisations. Most importantly, *between*-subjects analysis showed that the sequencing speed was significantly slower among ASD compared to control participants in the silent condition only. This is exactly what would be expected if individuals with ASD became quicker at sequencing as a result of benefitting from overt self-talk. This is a potentially very important finding (and contradicts the finding of Russell-Smith et al., 2014, who could not elicit an increase in task performance among participants with ASD by encouraging overt self-talk). However, it needs to be interpreted in light of the results from the seven other key analyses conducted in the study. Unfortunately, none of these other results was unambiguously in keeping with the predictions made by the authors:

(ii) with respect to *sequencing accuracy* (rather than speed) on the task, neither of the predicted interaction effects emerged (all *p*s for interaction effects $\geq .23$) . That is, both groups benefitted equally from overt task-relevant self-instruction and, conversely, were equally detrimentally affected by overt task-irrelevant self-instruction. Moreover, overall sequencing accuracy was *superior* among participants with ASD than among comparison participants, with the superiority reaching statistical significance in the ANOVA comparing group performance in the silent vs. task-incongruent conditions.

(iii) with respect to *motor accuracy*, neither of the predicted interaction effects emerged (results in the paper are reported as "NS"). That is, both groups benefitted equally from overt task-relevant self-instruction and, conversely, were equally detrimentally affected by overt task-irrelevant self-instruction.

(iv) with respect to *motor speed*, the predicted interaction effect did not emerge when group performance was compared in the silent vs. task-congruent conditions (result in the paper are reported as "NS") . That is, with respect to motor speed, both groups benefitted equally from overt task-relevant self-instruction.

(v) following from point (iv), with respect to *motor speed*, there was a significant Group x Condition interaction when group performance was compared in the silent vs. task-incongruent conditions. Within-subjects analysis showed that participants with ASD were significantly less negatively affected by overt task-incongruent self-talk than comparison participants were. This may suggest that participants with ASD were not employing inner speech to mediate motor speed. However, even if this was the case, it is not clear that the under-utilisation of inner speech had a detrimental performance on motor speed in general. That is, it is not clear that motor speed was significantly slower among participants with ASD than among comparison participants in any of the secondary task conditions, but surprisingly between-subjects analyses were not performed so it is not possible to know for sure.

Summary

Although there was one potentially positive result from Gidley Larson and Suchy (2014), all of the other results indicated either that participants with ASD did employ inner speech to mediate the task (contrary to predictions) or that if they did not employ inner speech typically it did not detrimentally affect their performance. Like many of the other studies of inner speech use in ASD, therefore, there is a gap between theoretically-based predictions and actual empirical results. The precise reasons for this gap are unclear and more work is needed to elucidate this issue. But, in our view, future work should be informed by an acknowledgement of this gap - as we discuss below.

Overall summary and future directions

While we believe there are good theoretical reasons to believe that thinking is not verbally mediated among people with ASD to the same extent as it is among neurotypical individuals – and that this contributes to executive dysfunction and diminished behavioural flexibility in this disorder – the quality of the empirical evidence available to support such inclinations is not generally high, in our view. To our knowledge, 13 experiments have directly measured the extent to which self-talk is used by people with ASD to mediate (various domains of) cognition. Of these 13 experiments, there are clear trends in the results reported. For example, six out of seven studies that employ articulatory suppression as a means of assessing inner speech use reported that such suppression does not negatively influence task performance in ASD. In each of these six studies, this lack of an articulatory suppression effect was interpreted as showing that task performance was not verbally mediated among participants with ASD. This is impressive consistency across studies and could be taken as indication of the robustness of the findings. However, most of the studies of articulatory suppression in

ASD have potential methodological confounds (some very serious confounds, arguably), which could have compromised the validity of the results. It is certainly possible to take the view that, given the apparent consistency of results across studies, the overall *conclusion* that inner speech is impaired in ASD is reliable even if the *results* from some individual studies are not valid. We urge caution in that case though, because the study of inner speech use in ASD is only emerging and a small number of invalid results can quickly bias the field when it contains relatively few studies. Moreover, the well-known "file drawer" problem in psychology has an especially large influence on emerging fields of study and it is possible that failures to replicate findings of diminished inner speech use in ASD go un-submitted to journals or unpublished if they are submitted. We hope that these points will be borne in mind by authors and used to justify publication of any future studies that find inner speech usage to be typical in ASD.

Setting aside any potential methodological difficulties with studies or with any hypothetical unpublished failures to replicate, there exists an anomaly in the results of existing research that has not been fully addressed to date. Assuming that inner speech use is diminished in ASD, there is very little evidence that such a diminution has a substantial negative effect on cognition. For example, in all but two of the studies that report diminished inner speech use in ASD, no between-group differences in baseline cognitive task performance were observed. That is, in the baseline condition of each study –performance relative to which the effect of the experimental manipulation (e.g., articulatory suppression) was judged– participants with ASD did not show impaired performance. Rather, between-group differences in these studies tended to emerge from the *superior* performance of individuals with ASD, relative to comparison participants, in the key experimental conditions of the experiments (e.g., in articulatory suppression conditions). Thus, even if participants with ASD in these studies were not employing inner speech to mediate the baseline (i.e., silent) condition of the experimental tasks, there was little evidence that this was harming their performance. Note, also, that in the only study of private speech use in ASD, Winsler et al. (2007) found that executive functioning task performance was diminished in their sample of ASD participants, despite the fact that self-talk was not diminished. Thus, some studies observe diminished self-talk in ASD alongside unimpaired cognitive task performance, whereas other studies observe diminished cognitive task performance alongside unimpaired self-talk. This is problematic for theories that propose a) inner speech is necessary for (or at least contributes to) efficient executive functioning and behavioural flexibility, and b) that diminished inner speech use explains (or at least contributes to) executive dysfunction and behavioural inflexibility in ASD. There are several possible explanations for this seeming anomaly, four of which follow.

First, it might be that inner speech use *is* diminished in ASD and that this diminution contributes to everyday, real-world difficulties with executive functioning and self-regulation, but that the laboratory-based measures of executive functioning used in these studies were not sensitive enough to detect impairments. While this is possible, it is purely speculative as things stand. It would be interesting to explore in future studies whether, for example, the extent to which articulatory suppression affects executive task performance is related to self-reported difficulties with executive functioning in everyday life among people with ASD (e.g., using the Behavior Rating Inventory of Executive Function, which has revealed real-world executive functioning impairments in both children [Granader et al., 2014] and adults [Wallace et al., 2016] with ASD). Perhaps more important still would be to explore the effects of both articulatory suppression and overt labelling on ecologically-valid tasks, such as subtests (e.g., Key Search or Zoo Map) from the Behavioural Assessment of Dysexecutive Syndrome that tap executive functioning across various subdomains (e.g., Planning and Flexibility).

Second, it might be that inner speech use *is* diminished in ASD, but that individuals with these impairments have available to them alternative strategies to successfully negotiate (laboratory and real-world) executive tasks. For example, several studies suggest that, at least in some respects, individuals with ASD have enhanced visual processing skills (e.g., Mottron et al., 2006). These skills could be employed as a successful means of mediating executive tasks among individuals with ASD, whereas neurotypical individuals rely heavily on inner speech use for success. This possibility is supported by the finding of Williams et al. (2012, Experiment 2) that successful planning was associated with performance on an indirect measure of visuo-spatial processing among individuals with ASD, but not among TD participants. Williams et al.'s data were purely correlational, however. What is needed to test this second possibility is a study that requires participants to complete a secondary task that is equivalent to articulatory suppression in terms of general processing requirements, but which a) taps visuo-spatial resources, rather than inner speech use, and b) does not tap the key resources required for performance on the primary (e.g., executive) task. If people with ASD really do employ visual resources to mediate executive functioning tasks, then such a visuo-spatial secondary task should negatively affect performance among individuals with ASD significantly more than it should negatively affect performance among neurotypical individuals.

Third, it may be that the differences in findings across studies are the result of sampling differences. ASD is a highly heterogeneous disorder (at all levels of explanation/description). Perhaps reduced inner speech use does contribute to behavioural and cognitive impairments in some individuals, or sub-group(s) of individuals, with ASD, but not others. Future research would benefit from assessing the relation between inner speech use and participant characteristics (cf. Lidstone et al., 2009; Williams & Jarrold, 2010).

Fourth, it may be that inner speech use is typical in ASD per se, at least in some cognitive/behavioural domains. This possibility should be taken seriously, in our view (as should possibility 2; even though both contradict the general theory of inner speech development that we espouse).

Finally, it is important to consider the clinical/practical relevance of the findings from this critical review. A surface reading of the literature on inner speech in ASD might lead clinicians and educators to prioritize verbal labelling and encouragement of self-talk to facilitate everything from solving mathematics problems at school/home to aiding in dealing with difficult transitions that people with ASD might experience in their daily lives. Therefore, it is crucial that future work clarifies what appears on further scrutiny to be a more clouded picture of the (a)typicality of inner speech use and function in ASD. It is also possible, if not probable, that some interventionists already employ verbal labelling and other techniques in attempts to improve language function (if not explicit attempts to target inner speech itself). For example, a recently developed intervention that targets executive function deficits in ASD, “Unstuck and On Target!”, makes critical use of verbal labelling and verbal scripts throughout its curriculum (Cannon, Kenworthy, Alexander, Werner, & Anthony, 2011; Kenworthy et al., 2014). However, to our knowledge, no studies to date have focused specifically on the utilization of these strategies and their effectiveness in groups of individuals with ASD. We noted above that several studies of verbal thinking in ASD have explored the effect on cognitive task performance of enforcing overt (private) speech. The logic driving these study designs, as expressed more or less explicitly by the study authors, is that, if people with ASD do not spontaneously utilize inner speech to mediate a cognitive task and if this contributes to their usual limitations on those tasks, then enforcing overt speech will improve cognitive task performance (perhaps to the level of neurotypical individuals who do spontaneously mediate the task verbally). In other words, artificially enforcing verbal mediation of a task will eliminate a usual deficit on that task that is caused by a failure to utilize verbal mediation spontaneously. When discussing each of these studies above, it was important to draw attention to the fact that almost all studies *failed* to support the authors’ hypothesis. Specifically, apart from the study by Gidley Larson and Suchy (2014), no study has found that enforcing overt speech significantly improved the cognitive task performance of participants with ASD (although the baseline performance of participants was not generally impaired anyway). It is always important to make clear when theory-driven hypotheses are not supported, because it often implies that the theory is not entirely correct. However, another possibility in this particular case is that the predictions stemming from the theory, rather than the theory itself, are invalid. The logic of attempting to enforce verbal mediation assumes that the effect will be the same as if the normal social-cognitive/developmental routes to verbal mediation had been followed. Yet, it may well be that such explicit instruction fails to replicate the effect on cognition that neurotypical individuals develop and harness via implicit/long-term learning (Diaz & Berk, 1995). From a neo-Vygotskian perspective, verbal thinking develops out of complex social interactions via an implicit internalisation process over many years. It may be too much to expect instructional training, especially in short-term experimental studies, to achieve the desired impact of increasing the influence of language on thought among people with ASD (although, note that it *does* appear to achieve the desired impact in other disorders; Perry et al., 2001). Having said this, as stressed multiple times in this paper, the evidence that people with ASD do not employ inner speech is not watertight and the evidence that a failure to employ inner speech results directly in cognitive deficits is even less evident. Thus, based on evidence to date, it is not clear that intervention efforts *should* necessarily be targeted at the enhancement of verbal thinking in ASD. It seems clear that carefully controlled investigation is required in this area to interrogate these questions more clearly. One possible future direction is to systematically target inner speech for intervention in ASD compared to one or more control groups (e.g., individuals with ASD vs. typically developing controls, or children with ASD vs. children with specific language impairment) to examine the malleability of these abilities and whether they in turn have a cascading influence on executive function and other problem-solving and real-world skills.

Relatedly, the influence of common co-morbidities in ASD, such as ADHD and/or specific language impairment, on inner speech usage has not been examined. Given that prior studies demonstrate atypical utilization of private speech in children with ADHD and other behaviour problems (Berk & Potts, 1991; Winsler et al., 2000) and possible delays in inner speech development in specific language impairment (Lidstone et al., 2012), this is clearly an important question for future research to tackle. There has been indirect examination of co-morbid intellectual disability and specific language impairment via (verbal) IQ subgrouping and/or matching on metrics of mental age, but it has not been systematic. Particular ASD co-morbidities/subgroups might be driving the inner speech impairments documented in prior studies or alternatively could be clouding the picture leading to failed replication of study findings in some cases. Similarly, more studies that employ additional clinical control groups could be helpful in determining the specificity of these findings to ASD. Different patterns of intact and impaired inner speech function in various clinical groups could eventually drive differential approaches to intervention.

Finally, future research in this area should address heterogeneity and developmental processes in ASD. Evidence regarding group-level differences (or similarities) in levels or patterns of performance on cognitive tasks is important, but it may hide underlying heterogeneity in cognitive and behavioural profiles of people with ASD. All of the studies reviewed in this paper employed ASD participants who were diagnosed according to pre-DSM-5 criteria, but future studies will employ the more recent DSM-5 guidelines, which may have an effect on results (albeit only a subtle effect, most likely; Volkmar & Reichow, 2013). Regardless of diagnostic criteria, *if* ASD is characterised by diminished verbal thinking, it is very unlikely that *all* people with ASD fail to utilize inner speech. It will be important to establish what drives individual differences in this capacity among people with ASD, as well as to investigate its developmental trajectory.

Research on inner speech use in ASD has the potential to inform general theories of the development and function of verbal thinking, as well as to have a transformative effect on clinical practice. It is precisely because of this potential importance that we believe a detailed critical analysis of results is necessary and timely. The aim of the review is to take stock of, and draw accurate conclusions from, existing empirical findings, as well as to help develop the directions for future research that will resolve the outstanding issues in this field. At first glance, research on inner speech in ASD to date has tended to support Vygotskian notions of the nature, development, and function of self-talk, as well as point out potentially useful treatment targets (e.g., verbal labelling, self-talk, etc.) for people with ASD. However, upon closer examination of the evidence to date with a critical eye, methodological shortcomings and anomalous findings temper these expectations. Only through careful and systematic future research can the outstanding theoretical and clinical questions posed here (as well as many others) be answered more definitively as we seek to understand the role that self-talk plays in mediating cognitive and behavioural functioning in ASD.

Footnotes

1. Winsler et al. also investigated private speech use among the same sample when completing a general problem-solving task (the "Building Sticks Task") and observed a very similar pattern of results to those from the WCST study.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Arlington, VA: American Psychiatric Publishing.

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

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

Baldo J.V., Dronkers, N.F., Wilkins, D., Ludy, C., Raskin, P., & Kim, J. (2005). Is problem solving dependent on language? *Brain and Language, 92,* 240-250.

Bayliss, E.M., Jarrold, C., Baddeley, A., & Gunn, D. (2005). The relationship between short-term memory and working memory: Complex span made simple? *Memory*, *13*, 414-421.

Berk, L.E., & Potts, M.K. (1991). Development and functional significance of private speech among attention-deficit hyperactivity disordered and normal boys. *Journal of Abnormal Child Psychology, 19,* 357-377.

Cannon, L., Kenworthy, L., Alexander, K.C., Werner, M.A., & Anthony, L.G. (2011). *Unstuck and on target!: An executive function curriculum to improve flexibility for children with autism spectrum disorders*. Baltimore, MD: Paul H. Brookes.

Corsi, P.M. (1972). Human memory and the medial temporal region of the brain. Unpublished doctoral dissertation, Abstracts International, 34(02), 891B (University microfilms No. AA105-77717).

Davidson, D. (1973). ‘Radical Interpretation’, *Dialectica*, 27: 314-328.

Emerson, M.J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language, 48,* 148-168.

Fernyhough, C. (2008). Getting Vygotskian about theory of mind: Mediation, dialogue, and the development of social understanding. *Developmental Review*, *28*, 225–262.

Fleiss, J.L., & Tanur, J. M. (1973). The analysis of covariance in psychopathology. In M. Hammer., K. Salzinger, & S. Sutton (Eds.), *Psychopathology: Contributions from the social, behavioural, and biological sciences* (pp. 509-527). New York: Wiley.

Fodor, J.A. (1975). The language of thought: A philosophical study of cognitive psychology. New York: Crowell.

Ford, S., & Silber, K.P. (1994). Working memory in children: A developmental approach to the phonological coding of pictorial material. *British Journal of Developmental Psychology, 12,* 165-175.

Gelman, A., & Stern, H. (2006). The difference between “Significant” and “Not Significant” is not itself statistically significant. *The American Statistician, 60(4),* 328-331.

Gidley Larson, J.C., & Suchy, Y. (2014). Does language guide behaviour in children with autism? *Journal of Autism and Developmental Disorders, 44(9),* 2147-2161.

Granader, Y., Wallace, G. L., Hardy, K. K., Yerys, B. E., Lawson, R. A., Rosenthal, M., Wills, M. C., Dixon, E., Pandey, J., Penna, R., Schultz, R. T., & Kenworthy, L. (2014). Characterizing the factor structure of parent reported executive function in autism spectrum disorders: the impact of cognitive inflexibility. *Journal of Autism and Developmental Disorders, 44*, 3056-3062.

Grandin, T. (1995). *Thinking in Pictures*. Doubleday: New York.

Hale, S., Myerson, J., Rhee, S.H., Weiss, C.S., & Abrams, R.A. (1996). Selective interference with the maintenance of location in working memory. *Neuropsychology, 10,* 228-240.

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

Holland, L., & Low, J. (2010). Do children with autism use inner speech and visuospatial resources for the service of executive control? Evidence from suppression in dual tasks. *British Journal of Developmental Psychology, 28,* 369-391.

Hurlburt, R., Happé, F., & Frith, U. (1994). Sampling the form of inner experience in three adults with Asperger syndrome. *Psychological Medicine, 24,* 385-396.

Joseph, R.M., McGrath, L.M., & Tager-Flusberg, H. (2005a). Executive dysfunction and its relation to language ability in verbal school-age children with autism. *Developmental Neuropsychology, 27,* 361-378.

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

Kenworthy, L., Anthony, L.G., Naiman, D.Q., Cannon, L., Wills, M.C., Werner, M.A., Alexander, K., Strang, J., Bal, E., Sokoloff, J.L., & Wallace, G.L. (2014). Randomized controlled effectiveness trial of executive function intervention for children on the autism spectrum. *Journal of Child Psychology & Psychiatry, 55,* 374-383.

Kraybill, M.L., & Suchy, Y. (2011). Executive functioning, motor programming, and functional independence: Accounting for variance, people, and time. *The Clinical Neuropsychologist, 25*, 210-223.

Kunda, M., & Goel, A.K. (2011). Thinking in Pictures as a cognitive account of autism. *Journal of Autism and Developmental Disorders, 41(9)*, 1157-1177.

Lidstone, J.S.M., Fernyhough, C., Meins, E., & Whitehouse, A.J.O. (2009). Brief report: Inner speech impairment in children with autism is associated with greater nonverbal than verbal skills. *Journal of Autism and Developmental Disorders, 39*, 1222-1225.

Lidstone, J.S., Meins, E., & Fernyhough, C. (2012). Verbal mediation of cognition in children with specific language impairment. *Development and Psychopathology, 24(2),* 651-660.

Liss, M., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., Waterhouse, L., & Rapin, I. (2001). Executive functioning in high-functioning children with autism. *Journal of Child Psychology and Psychiatry, 42,* 261-270

Mackinlay, R., Charman, T., & Karmiloff-Smith, A. (2006). High functioning children with autism spectrum disorder: A novel test of multitasking. *Brain and Cognition, 61*, 14-24.

Mervis, C.B., & Klein-Tasman, B.P. (2004). Methodological issues in group-matching designs: α levels for control variable comparisons and measurement characteristics of control and target variables. *Journal of Autism and Developmental Disorders, 34,* 7-17.

Miller, G.A., & Chapman, J.P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology, 110,* 40-48.

Miyake, A., Emerson, M.J., Padilla, F., & Ahn, J. (2004). Inner speech as a retrieval aid for task goals: The effects of cue type and articulatory suppression. *Acta Psychologia, 115,* 123-142.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex ‘‘frontal lobe’’ tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49-100.

Mottron, L., Dawson, M., Soulieres, I., Hubert, B. & Burack, J. (2006) Enhanced perceptual functioning in autism: An update, and eight principles of autistic perception. *Journal of Autism and Developmental Disorders, 36,* 27-43.

Pennington, B.F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychological Psychiatry, 37(1),* 51-87.

Perry, W., Potterat, E.G., & Braff, D.L. (2001). Self-monitoring enhances Wisconsin Card Sorting Test performance in patients with schizophrenia: Performance is improved by simply asking patients to verbalize their sorting strategy. *Journal of the International Neuropsychological Society, 7,* 344-352.

Russell, J., Jarrold, C., & Henry, L. (1996). Working memory in children with autism and with 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(2),* 103-112.

Russell-Smith, S.N., Comerford, B.J.E., Maybery, M.T., & Whitehouse, A.J.O. (2014). Brief Report: Further evidence for a link between inner speech limitations and executive function in high-functioning children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44,* 1236-1243.

Strauss, M.E. (2001). Demonstrating specific cognitive deficits: A psychometric perspective. *Journal of Abnormal Psychology, 110*(1), 6-14.

Van Eylen, L., Boets, B., Steyaert, J., Evers, K., Wagemans, J., & Noens, I. (2011). Cognitive flexibility in autism spectrum disorder: explaining the inconsistencies? *Research in Autism Spectrum Disorders, 5,* 1390-1401.

Volkmar, F. R. & Reichow, B. (2013). Autism in DSM-5: Progress and challenges. *Molecular Autism, 4*:13 DOI: 10.1186/2040-2392-4-13

Vygotsky, L.S. (1981). The genesis of higher mental functions. In J.V. Wertsch (Ed.), *The concept of activity in Soviet psychology* (pp. 189-240). Armonk, NY: Sharpe.

Wallace, G. L., Kenworthy, L., Pugliese, C. E., Popal, H. S., White, E. I., Brodsky, E., & Martin, A. (2016). Real-world executive functions in adults with autism spectrum disorder: Profiles of impairment and associations with adaptive functioning and co-morbid anxiety and depression. *Journal of Autism and Developmental Disorders, 46,* 1071-1083.

Wallace, G.L., Silvers, J.A., Martin, A., & Kenworthy, L.E. (2009). Brief report: Further evidence for inner speech deficits in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39,* 1375-1379.

Whitehouse, A.J.O., Maybery, M.T., & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry, 47,* 857-865.

Whorf, B.L. (1956). *Language, Thought, and Reality* New York: John Wiley & Sons, and The Technology Press of M.I.T.

Williams, D. M., Bowler, D. M., & Jarrold, C. (2012). Inner speech is used to mediate short-term memory, but not planning, among intellectually high-functioning adults with autism spectrum disorder. *Development and Psychopathology, 24,* 225-239.

Williams, D., Happe´, F. & Jarrold, C. (2008). Intact inner speech use in autism spectrum disorder: Evidence from a short-term memory task. *Journal of Child Psychology and Psychiatry, 49,* 51-58.

Williams, D., & Jarrold, C. (2013). Assessing planning and set-shifting abilities in autism: Are experimenter-administered and computerised versions of tasks equivalent? *Autism Research, 6,* 461-467.

Williams, D., & Jarrold, C. (2010). Predicting inner speech use in autism spectrum disorder: The roles of verbal ability and cognitive profile. *Journal of Autism and Developmental Disorders, 40,* 907-913.

Winsler, A., Diaz, R.M., Atencio, D.J., McCarthy, E.M., & Chabay, L.A. (2000). Verbal self-regulation over time in preschool children at risk for attention and behavior problems. *Journal of Child Psychology & Psychiatry*, *41*, 875-886.

Winsler, A., Manfra, L., & Diaz, R.M. (2007). “Should I let them talk?”: Private speech and task performance among preschool children with and without behavior problems. *Early Childhood Research Quarterly*, *22*, 215-231.

Table 1: Techniques for measuring verbal thinking

Technique	Procedure	Rationale
Articulatory suppression (dual task design)	Participants repeat a simple word or phrase while completing the primary task of interest.	Primary task performance completed under articulatory suppression is compared to primary task performance under alternative secondary conditions (e.g., silence; foot tapping). If a person uses inner speech to mediate the primary task under normal circumstances, then articulatory suppression should *diminish* their primary task performance significantly more than alternative secondary task conditions disrupt performance, because articulatory suppression selectively disrupts inner speech.
Overt labelling (dual task design)	Participants articulate task instructions or verbal strategies relevant to primary task completion.	Primary task performance completed under enforced self-talk is usually compared to primary task performance under silent conditions. If a person does *not* use inner speech to mediate the primary task under normal circumstances (i.e., silence), then enforced overt self-talk should *enhance* primary task performance, because it enforces the kind of verbal mediation that is absent in spontaneous situations.
Similarity effects in memory	During short-term/working memory task (usually during the study phase), the properties of (pictorial) items are manipulated to influence the extent to which verbal mediation is possible or probable	If short-term/working memory is verbally mediated, then pictures of objects that have long, phonologically similar, or difficult-to-articulate, names should be less reliably recalled than pictures of objects that have short, phonologically dissimilar, or easy-to-articulate, names. If the phonological properties of pictorial stimuli affect recall, then this must show that the images have been recoded from a visual into a verbal representation. However, if memory is *not* verbally mediated, then the phonological properties of items should not impact on recall.
Analysis of private speech	Record participants while they complete a task. Analyse	If task performance is verbally mediated, then "task-relevant" private speech (i.e., overt speech

	recordings for the number of overt utterances made and code these according to whether they were a) fully overt or partially internalised; b) relevant to the task, c) associated with task performance	that is *about* the task one is completing) should be used, particularly during moments of difficulty on the task.

Table 2: Studies examining verbal thinking in autism spectrum disorder.

Study	Age/Sex/IQ	Baseline performance diminished in ASD?	Measure of verbal mediation	Cognitive task	Design	Diminished verbal mediation in ASD *reported*?	Possible confounds
Planning							
Wallace et al. (2009)	ASD (n=28): Age: 15.74 (2.10); 93% male; FSIQ: 110.25 (16.84) TD (n=25): Age: 16.36 (1.83); 96% male; FSIQ: 113.84 (10.02)	Yes	Effect of articulatory suppression	Tower of London	Group (ASD/NT) × condition (articulatory suppression/silence). DV = No. of extra moves (above the minimum number required) taken to complete each trial	**No**, according to omnibus test; Group (ASD/NT) × condition (articulatory suppression/silence) interaction: $F(1,51) = 1.08$, $p = .30$. **Yes**, according to post hoc tests; *Within-participants t tests*: NT performance significantly worse in articulatory suppression condition than in silent condition: $t(24) = 2.34$, $p = .03$. ASD performance non-significantly affected by condition: $t(27) = 1.13$, $p = .27$. *Between-participants t tests*: ASD participants showed significantly poorer planning in the silent condition, $t(51) = 2.28$, $p = .03$, but not the articulatory suppression condition, $t(51) = 0.01$, $p = .99$	Lack of “visuo-spatial” control condition
Williams et al. (2012, Exp. 2)	ASD (n=17): Age: 42.13 (14.14); VIQ: 112.82 (11.84);	No	Effect of articulatory suppression	Tower of London	Group (ASD/NT) × condition (articulatory suppression/silence).	**Yes**. Significant Group (ASD/NT) × Condition (articulatory suppression/silence) interaction, $F(1, 29) = 5.69$, $p = .02$. *Within-participants t tests*: NT performance significantly worse in articulatory suppression condition than in silent condition: t (15)	Lack of “visuo-spatial” control condition

	PIQ: 112.88 (15.33) FSIQ: 114.00 (13.39) TD (n=17): Age: 39.43 (12.51); VIQ: 117.59 (13.13); PIQ: 112.59 (11.05); FSIQ: 16.71 (13.32)				DV = No. of moves taken to complete each trial	= 3.46, p = .003. ASD performance non-significantly affected by condition: *t* (14) = 0.20, p = .85. *Between-participants t tests*: ASD and NT participants show equivalent planning in the silent condition, *t* (29) = 0.72, p = .48, but ASD participants show better planning than NT participants in the articulatory suppression condition, *t*(29) = 2.15, p = .04.	
Holland & Low (2010, Exp. 3)	ASD (n=13): Age: 10.75 (2.33) VMA: 11.42 (5.25) TD (n=13): Age: 9.33 (1.58) VMA: 11.25 (4.00)	No	Effect of articulatory suppression & visuo-spatial suppression	Tower of London	Group (ASD/NT) × Condition (articulatory suppression/visuo-spatial suppression/silence). DV = *Time* taken to complete each trial	**Yes**. Significant Group (ASD/NT) × Condition (articulatory suppression/visuo-spatial suppression/silence) on completion time $F(2, 23) = 9.70$, $p < .01$. *Within-participants contrasts* (Bonferroni adjusted): NT performance significantly slower in articulatory and visuo-spatial conditions than in the silent condition, "all *p*s < .01". ASD performance significantly slower in visuo-spatial condition than in the silent or articulatory suppression conditions, "all *p*s < .01", but no significant difference between silent and articulatory suppression conditions, "*p* > .05".	Use of a task single trial in all conditions; groups not matched for age or VIQ; Inappropriate visuo spatial control task; piecemeal data analysis

						Between-participants t tests: ASD and NT participants show equivalent planning speed in the silent condition, *t* (24) = 1.30, p > .05, and the visuo-spatial suppression condition, *t* (24) = 0.66, p > .05. However, ASD participants show *faster* planning than NT participants in the articulatory suppression condition, *t*(24) = 4.56, p < .001.	
Task switching							
Whitehouse et al (2006, Exp.3)	ASD (n=23): Age: 10.91 (1.75); 100% male VMA: 9.42 (2.83) Non-verbal ability *raw* score: 38.1 (6.7) TD (n=23): Age: 8.33 (0.83); 100% male VMA: 9.17 (1.42) Non-verbal ability raw score: 35.5 (4.7)*	No	Effect of articulatory suppression	Arithmetic task: Two columns of digits. In the *switching task*, rows of digits need to be added and substracted in alternating fashion; in non-switching condition,	Group (ASD/NT) × Task (Switching/non-switching) × Condition (articulatory suppression/silence) DV = Time taken to complete each set of lists	**Yes** (but *not* clearly in a reanalysis of the data by Lidstone et al.). Significant Group (ASD/NT) × Task (Switching/non-switching) × Condition (articulatory suppression/silence) interaction in an ANCOVA (controlling for non-verbal ability), *F*(1, 43) = 7.30, *p* < .02. *Within-participants contrasts*: NT performance in both switching (adding, then subtracting lists of numbers) *and* non-switching (e.g., only adding lists of numbers) significantly worse in articulatory suppression condition than in silent condition: t(22) = 5.66, p < .001 (switching task), t(22) = 4.23, p < .001 (non-switching task). ASD performance non-significantly affected by condition: t(22) = .76, p = .45 (switching task), t(22) = 1.94, p = .06 (non-switching task). *Between-participants t tests*: ASD and NT participants show equivalent performance in the switching task in the silent	Groups not matched for age, VIQ, or PIQ; reanalysis of data by Lidstone et al. suggested verbal mediation was developmentally appropriate in ASD (see Williams & Jarrold, 2010

				rows of digits need to be added together (or subtracted) in blocked fashion (hence no witching between mental operations required)		condition, F(1,43) = 0.23, p = .63, but ASD participants show *better* performance than NT participants in the switching task in the articulatory suppression condition, F(1,43) = 3.88, p = .05.	
Holland & Low (2010, Exps 1 and 2)	ASD (n=13): Age: 10.75 (2.33) VMA: 11.42 (5.25) TD (n=13): Age: 9.33 (1.58) VMA: 11.25 (4.00)	No	Effects of articulatory suppression & visuo-spatial suppression	Same as Whitehouse et al., 2006, Exp. 1.	Exp 1: Group (ASD/NT) × Task (Switching/non-switching) × Condition (articulatory suppression/silence) DV = Time taken to complete each set of lists	**Yes**. In Exp. 1, significant Group (ASD/NT) × Condition (articulatory suppression/silence) interaction in an ANOVA, *F*(1, 24) = 4.30, *p* < .05. But note: the three way interaction involving Task (switching/non-switching) was apparently non-significant, although inferential statistics were not reported by Holland and Low. *Within-participants contrasts*: NT performance across switching and non-switching tasks *collapsed* significantly worse in articulatory suppression condition than in silent condition: F(1,12) = 21.20, p < .01 ASD performance across switching and non-	Groups not matched for age or VIQ; Inappropriat visuo-spatial control task piecemeal data analysis (data from exps 1 and 2 should have been combined for analysis)

					Exp 2: Group (ASD/NT) × Task (Switching/non-switching) × Condition (visuo-spatial suppression/silence) DV = Time taken to complete each set of lists	switching tasks collapsed was non-significantly affected by condition: F(1,12) = 1.88, p > .05. *Between-participants contrasts*: Not reported. Figure 1 in Holland and Low (p.376) appears to indicate no between-group differences in performance in the silent condition, but superior (faster) performance among ASD participants than NT participants in the articulatory suppression condition. In Exp. 2, *non*-significant Group (ASD/NT) × Condition (Visuo-spatial suppression/silence) interaction in an ANOVA, F(1,24) = .01, p > .05. Again, the three way interaction involving Task (switching/non-switching) was not reported by Holland and Low. *Within-participants contrasts*: Both ASD and NT participants performed less well (slower) in the visuo-spatial suppression condition than in the silent condition, all ts > 4.51, all ps < .01. *Between-participants contrasts*: Not reported. Figure 1 in Holland and Low (p.376) appears to indicate no between-group differences in performance in either the silent or visuo-spatial suppression condition.
Cognitive flexibility						
Russell-Smith et	ASD (n=17): Age: 11.93 (1.90);	No	Effects of articulatory	Wisconsin Card	Group (ASD/NT) × Verbal ability	**Yes**, as measured by effect of articulatory suppression; **No**, as measured by effect of overt labelling:

al. (2014)	82% male; VIQ: 101.4 (14.8) PIQ: 101.6 (14.6) TD (n=18): Age: 10.69 (2.25); 89% male; VIQ: 109.6 (14.7) PIQ: 102.9 (11.7)	suppression & overt self-talk	Sorting Test	(High/Low) × Order of conditions (4 levels) × Condition (Silent/mouthing/articulatory suppression/talk-aloud) DVs = No. of perseverative errors; No. of categories completed; No. of correct responses Covariate = age	Significant Group × Condition interaction in an ANOVA for two of the three DVs - perseverative errors, $F(3, 81) = 4.97$, $p = .01$, and responses correct, $F(3, 81) = 3.68$, $p = .02$. The statistics for the Group × Condition interaction effect associated with the other DV (no. of categories completed) were not reported by Russell-Smith et al. *Within-participant contrasts*: Among NT participants, performance was poorer in the articulatory suppression condition than in the silent condition, in terms of both no. of perseverative errors, $F(1,14) = 6.59$, $p = .02$, and no. of responses correct $F(1,14) = 4.88$, $p = .04$. In contrast, performance in the talk-aloud condition was significantly better than in the silent condition, in terms of no. of perseverative errors, $F(1,14) = 4.50$, $p = .05$. Also, NT participants had more correct responses in the talk aloud condition than in the mouthing condition, $F(1,14) = 5.16$, $p = .04$. Among ASD participants, the effect of condition was non-significant for both perseverative errors and correct responses, "$F(3,39) < 1$". *Between-participants contrasts*: Not reported. Figure 2 in Russell-Smith et al. (p.1239) appears to indicate no between-group differences in performance in any of the conditions with respect to no. of responses correct or no. of categories completed (based on overlapping error bars). With respect to no. of perseverative errors, participants with ASD appear to make significantly more perseverative errors than NT participants in the

						silent and talk-aloud conditions, but this is not certain because inferential statistics are not reported.	
Winsler et al. (2007)	ASD (n=33): Age: 11.0 (2.3); 97% male TD (n=28): Age: 10.3 (3.2); 68% male	Yes	Private speech use (total amount and proportion task-relevant)	Wisconsin Card Sorting Test	Main effect of Group (ASD/ADHD/Typical) on DVs = Total private speech/minute; Task relevant private speech/minute	**No**, private speech use was as frequent and frequently task relevant as in typically developing children	Groups not matched for verbal ability, non-verba ability, or gender
Short-term and working memory							
Whitehouse et al (2006, Exp.2)	ASD (n=23): Age: 10.91 (1.75); 100% male VMA: 9.42 (2.83) Non-verbal ability *raw* score: 38.1 (6.7)	No	Word length effect	Serial recall of pictorial stimuli	Group (ASD/NT) × Condition (Long items/short items) × Secondary task condition (Silence/labelling)	**Yes**, as reflected by a diminished word length effect. Significant Group × Condition × Secondary task interaction in an ANCOVA: $F(1, 43) = 4.14$, $p < .05$. Whitehouse et al. then explored Group × Condition interactions in each secondary task condition separately. The interaction effects were significant both in the silent secondary task condition, $F(1,43) = 12.66$, $p < .01$, and overt labelling condition, $F(1,43) =$	Groups not matched for age or VIQ

	TD (n=23): Age: 8.33 (0.83); 100% male VMA: 9.17 (1.42) Non-verbal ability *raw* score: 35.5 (4.7)*				DV = no of items recalled correctly. Covariate = non-verbal ability	33.77, p < .001. ASD participants showed a diminished word length effect in *both* secondary task conditions, which was unexpected by Whitehouse et al.	
Russell et al. (1996)	ASD (n=33) Age: 12.38 (2.95) VMA = 6.27 (1.19) MLD (n=33) Age: 10.82 (1.94) VMA = 6.24 (1.18) TD (n=33): Age: 6.28 (1.19)	No	Word length effect	Serial recall of auditory stimuli	Group (ASD/NT/MLD) × Condition (Long items/short items) × Recall mode (Verbal/non-verbal) DV = memory span	**No**, normal word length effect. Significant Group × Condition × Recall mode interaction: $F(2, 96) = 3.69$, $p = .03$. Post hoc contrasts: Significant Group × Condition interaction for nonverbal recall, $F(2, 90) = 3.12$, $p = .05$, but not for verbal recall, $F(2, 90) = 1.44$, $p = .24$. Figure 1 and Table 2 in Russell et al (p.676) show that ASD participants had a non-significantly *larger* word length effect than NT participants and a significantly *larger* word length effect than MLD participants in the nonverbal recall condition. Between-group differences in the size of the word length effect were non-significant in the verbal recall condition. Within-participants analyses showed that participants with ASD were unique in showing a significantly larger word length effect in the non-verbal than verbal recall, $F(1, 90) = 5.82$, $p = .02$.	Groups not matched for age or VIQ
Williams	ASD (n=25):	No	Phonological	Serial	Group (ASD/NT)	**No**, developmentally normal phonological similarity effect, and	-

et al. (2008)	Age:12.25 (0.08); 88% male; VIQ: 77.16 (15.25); PIQ: 76.84 (20.27); FSIQ: 74.84 (15.99) TD (n=20): Age: 12.08 (2.75); 80% male; VIQ: 73.20 (13.84); PIQ: 74.39 (22.53); FSIQ: 71.56 (18.57)		similarity effect & visual similarity effect	recall of pictorial stimuli	× Condition (Phonologically similar stimuli/visually similar stimuli/control stimuli) × verbal mental age (< 7 years/>7 years) DV = no of correctly recalled trials	visual similarity effect. No significant Group × Verbal mental age interaction, F(1, 41) = .07, p = .80, or Group × Condition × Verbal mental age interaction , F(2, 82) = 0.52, p = .60. There was a significant Condition × Verbal mental age interaction, F (2, 82) = 7.54, p = .001. *Post hoc contrasts* revealed that, among participants with verbal mental ages over seven years, control stimuli were recalled significantly better than phonologically similar stimuli (indicating a phonological similarity effect), F(1, 31) = 15.74, p < .001, but not than visually similar stimuli (indicating no visual similarity effect), F(1, 31) < .01, p > .99. Among participants with verbal mental ages under seven years the opposite pattern was present with control stimuli better recalled than visually similar stimuli (indicating a visual similarity effect), F(1, 12) = 4.59, p = .05, but not than phonologically similar stimuli (indicating no phonological similarity effect), F(1, 12) = .18, p = .68.	
Williams et al. (2012, Exp. 1)	ASD (n=17): Age: 42.13 (14.14); VIQ: 112.82 (11.84); PIQ: 112.88 (15.33) FSIQ: 114.00 (13.39)	No	Phonological similarity effect and effect of articulatory suppression	Serial recall of pictorial stimuli	Group (ASD/NT) × Condition (Phonologically similar stimuli/control stimuli) × Secondary task	**No**, normal phonological similarity effect and normal articulatory suppression effect. No significant main effect of group, F (1, 32) = 0.87, p = .36; no significant Group × Secondary task interaction, F(1, 32) = 0.71, p = .41; no significant Group × Condition interaction F (1, 32) = 0.58, p = .45; no significant Group × Condition × Secondary task interaction, F(1, 32) = 0.04, p = .85.	Lack of “visuo-spatial” control condition

	TD (n=17): Age: 39.43 (12.51); VIQ: 117.59 (13.13); PIQ: 112.59 (11.05); FSIQ: 116.71 (13.32)				(Articulatory suppression/silence)		
Joseph et al. (2005b)	ASD (n=24): Age: 8.92 (2.33); 87.5% male; VIQ: 94 (19); PIQ: 99 (20); FSIQ: 96 (18) TD (n=24): Age: 8.92 (2.17); 79% male; VIQ: 89 (12); PIQ: 94 (14); FSIQ: 92 (13)	Yes (in the verbal condition, as predicted)	Memory for nameable vs. non-nameable stimuli	Self-ordered pointing	Group (ASD/NT) × Condition (Namable images/non-namable images) DV = no. of errors	**Yes**, ASD deficits in the namable condition only. Significant Group × Condition interaction, F(1,46) = 10.8, p < .001. *Within-participants contrasts*: Among comparison children, performance was better in the namable than non-namable condition, t(23) = 4.7, p < .001, whereas there was no difference in performance across conditions among ASD participants, t(23) = 0.2, n.s. *Between-participants contrasts*: Participants with ASD performed significantly less well than comparison children in the nameable condition, *t*(46) = 2.4, *p* < .02, but not the non-namable condition, t(46) = 1.0, n.s.	-
Motor control							

Gidley Larson & Suchy (2014)	ASD (n=21): Age: 14.3 (1.8); 100% male; FSIQ: 103.5 (14.7); TD (n=22): Age: 13.4 (1.5); 100% male; FSIQ: 104.7 (11.1);	No (except for one of the four variables, namely sequencing *speed*; but sequencing *accuracy* was superior in ASD, suggesting a trade-off between speed and accuracy)	Effect of enforcing task-relevant vs task-irrelevant self-talk	Group (ASD/NT) × Condition (Silence/Task-congruent/Task-incongruent)* DVs = Sequencing speed, sequencing accuracy, motor speed, and motor accuracy	**Ambiguous**: In 5 of 8 analyses, the predicted Group x Condition interactions did not emerge and in the three that did emerge the cause was not always clear. *Note: The authors departed from this design in the data analysis by conducting <u>eight</u> 2 (Group) × 2 (Condition) ANOVAS, rather than <u>four</u> 2 (Group) × 3 (Condition) ANOVAs.	Piecemeal data analysis; Four ANOVAs should have been completed, on for each DV. Instead, eight were conducted, because condition effects were analysed individually

Notes: NS = non-significant; DV = dependent variable; VMA = verbal mental age; VIQ = verbal IQ; PIQ = performance IQ; FSIQ = full-scale IQ; NT = neurotypical; *Baseline characteristics (age, VMA, non-verbal raw score) for participants in Whitehouse et al. (2006, exps 2 and 3 are based on those provided in their Table 1 (p.858). Seven additional children (four ASD; three NT) were added for experiments 2 and 3, but means and SDs were not reported for this extended sample.

Table 3: Summary of Gidley Larson & Suchy (2014)

Conditions compared	Dependent variable	Group × condition interaction effect	Group contrasts
silent vs task-*incongruent*	Sequencing speed	$F(1, 41) = 8.46$; $p = .006$	*Within-participants*: NT group: poorer in incongruent than silent condition, $p = .002$; ASD group: no effect of condition, $p = .53$ *Between-participants*: Not reported, but no main effect of Group in the ANOVA, $p = .18$
	Sequencing accuracy	$F(1, 41) = 1.46$; $p = .23$	None reported (because interaction non-significant), but significant main effect of Group in ANOVA, reflecting *superior* accuracy in ASD than comparison participants, $p < .05$
	Motor speed	$F(1, 41) = 5.55$, $p = .02$	*Within-participants*: NT group: poorer in incongruent than silent condition, $p = .001$; ASD group: no effect of condition, $p = .85$. *Between-participants*: Not reported, but no main effect of Group in the ANOVA, $p = .81$
	Motor accuracy	NS	"no significant main effects or interaction (all p values >.10" (p.2154)
silent vs task-*congruent*	Sequencing speed	$F(1, 41) = 12.81$; $p = .001$,	*Within-participants*: Not reported *Between-participants*: ASD slower than NT in the silent condition, $p = .01$, but not in the task-congruent condition, $p = .18$
	Sequencing accuracy	$F(1, 41) = .454$; $p = .504$,	None reported (because interaction non-significant), but no main effect of Group in ANOVA $p = .272$.
	Motor speed	NS	"no significant main effects or interactions (all p values >.10)" (p.2155)
	Motor accuracy	NS	"no significant main effects or interactions (all p values >.10)" (p.2155)