



# THE UNIVERSITY *of* EDINBURGH

This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.

**Verbal problem-solving, executive functioning and language  
development in autism spectrum disorders.**

Ben Alderson-Day

PhD Psychology

The University of Edinburgh

2012



**Declaration**

I hereby declare that the following thesis is my own original work and has not been submitted in any previous application for a higher degree.

Signed ..... Date.....



## Abstract

Autism spectrum disorders (ASDs) are primarily defined by problems with social interaction and communication, but they are also associated with a complex cognitive profile. One area of difficulty for children and adults with ASD is problem-solving, or the process of identifying a solution to a puzzle or question where the answer is hidden.

This can be seen on the Twenty Questions Task (TQT), a commonly-used measure of verbal problem-solving and executive functioning. Children with autism are consistently less efficient than typically-developing children in their questioning on the task: for instance, rather than ask a general, category-based question (e.g. “Is it a living thing?”) they may ask about single items (“Is it the dog?”) or very restricted groupings (“Is it something you wear on your feet?”). This has previously been interpreted as an example of a concept formation deficit in autism, deriving from underlying difficulties with complex and integrative information processing. However, success in problem-solving relies on a number of cognitive and linguistic processes that may be impaired in ASD. This thesis attempts to identify which of these may better explain autistic problem-solving performance, using the TQT as a specific example.

The first experiment presented here examines the role of executive functioning difficulties in this profile. The performance of 22 children with ASD and 21 age- and IQ-matched typically-developing (TD) children was compared on a version of the TQT adapted to assess planning skills prior to problem-solving and selective attention during the task. Compared to controls, ASD participants were less efficient in their planning of questions, although not all ASD participants had difficulty constructing a plan. No specific effects of selective attention were evident.

The second and third experiments explore the importance of atypical language development to this profile, using the example of deafness. Experiment 2 compares the performance of deaf ( $n = 9$ ) and hearing ( $n = 27$ ) adults on the TQT, replicating prior evidence of less efficient problem-solving in deaf graduate students. Experiment 3 contrasts TQT performance in 13 deaf schoolchildren with the ASD and TD data acquired in experiment 1. Like ASD children, deaf children were less efficient in their questioning than TD participants, even when controlling for cognitive ability differences.

Both autism and deafness are associated with delays in early language development, whereas Asperger Syndrome (AS) is not. To test whether language delay explains autistic problem-solving difficulties, experiment 4 compares TQT performance in 15 children with autism, 15 AS children and 15 age- and IQ- matched typically-developing controls. Participants with autism asked less efficient questions than both AS and TD participants, between whom no differences were observed. This suggests that the problem-solving profile in autism may be better explained as a consequence of atypical language development, rather than other aspects of information processing or executive dysfunction.

### **Note**

The research presented here builds on the findings of a previous piece of work that was conducted by the author in the completion of an M.Sc. by Research. The results for the previous project are reported in the following paper:

Alderson-Day, B., & McGonigle-Chalmers, M. (2011). Is it a bird? Is it a plane? Category use in problem-solving in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 41 (5), 555-565. [DOI: 10.1007/s10803-010-1077-9](https://doi.org/10.1007/s10803-010-1077-9).

The original dissertation is available at <http://www.era.lib.ed.ac.uk/handle/1842/3635>.

In addition, chapters 1 and 2 of this thesis report on work that has also been published in the following paper:

Alderson-Day, B. (2011). Verbal problem-solving in autism spectrum disorders: A problem of plan construction? *Autism Research*, 4(6), 401-411. [DOI: 10.1002/aur.222](https://doi.org/10.1002/aur.222).

## Acknowledgements

The completion of this thesis has relied upon the help and generosity of a number of different people. First, the teachers: I would like to thank Allan Scott at Kaimes, Sian Rees at Fulford, Ann Bradbury at St John's, Clare Boyd at Donaldson's, Leslie Peacock at George Watson's, and Ellie Goodson and Clare Britton at Elmfield for their great support in finding participants and providing me with the opportunity to see their students. Big thanks also go to the Lothian Autistic Society and Parents Together West Lothian for their help with contacting families and publicising the research.

Second, I would like to thank the academics and other professionals who supported the research in a various ways. In York, thanks go to Vicky Crawley for providing BSL interpreting, and Barry Wright for making that possible, while always contributing useful questions and observations on autism and deafness. In Edinburgh, thanks are due to Marc Marschark and Rob McIntosh for their enthusiasm and willingness to discuss different aspects of the research on a number of occasions.

Third, friends and family. In the department, thanks go to Antimo, Elaine, Jennifer and Steve for their enthusiasm and willingness to not discuss the research in any way and just go to the pub. In Bristol, Amy and Matt deserve special thanks; without their input and support it is quite possible that none of the deaf research would ever have happened. In Birmingham, thank you Chris for helping me find awkward papers and updating me on the cricket score, while in Portobello, thanks to all the men of Fatneck: you know who you are. Thanks also go my parents for their constant interest and encouragement over the past three years.

Finally, I must thank three people who made this research possible. Thank you to my supervisors, Maggie McGonigle and Bob Logie; you have been a brilliant and inspiring team, always asking the right questions, reining me in when necessary but ultimately giving me the freedom to explore the issues that I found most fascinating.

And, last but not least, thank you to my wife Helen. Whenever I complained, you just told me to get on with it. Good work.





## Contents

<b>Introduction</b>	<b>11</b>
<b>Chapter 1.</b> Executive explanations of strategy inefficiency in children with autism.	<b>29</b>
<b>Chapter 2.</b> Testing the contributions of planning and selective attention to Twenty Questions performance.	<b>47</b>
<b>Chapter 3.</b> Comparing cognitive skills in deafness and autism.	<b>79</b>
<b>Chapter 4.</b> Problem-solving in deaf and hearing adults.	<b>99</b>
<b>Chapter 5.</b> Problem-solving in deaf children.	<b>121</b>
<b>Chapter 6.</b> Similarities and differences between Asperger Syndrome and autism.	<b>149</b>
<b>Chapter 7.</b> Problem-solving in Asperger Syndrome and autism.	<b>165</b>
<b>Chapter 8.</b> General Discussion.	<b>201</b>
<b>Appendices</b>	
Appendix 1. Participant data for deaf adults and children.	<b>218</b>
Appendix 2. Pilot analysis of the Semantic Decision Task (SDT).	<b>219</b>
Appendix 3. Pilot comparisons of problem-solving and free recall.	<b>221</b>
<b>References</b>	<b>223</b>



## **Introduction**

Autism is a developmental disorder that is characterised by difficulties in social functioning, problems with communication, and the presence of restricted interests and repetitive behaviours. A diagnosis of autism (sometimes referred to as childhood autism or autistic disorder) requires evidence of each of these characteristics, and is typically associated with delays in language and intellectual development in the first three years of life (APA, 1994; WHO, 1993). The term “autism spectrum disorder” (ASD) is a collective label that is used to refer to autism and other conditions on the autistic spectrum, such as Asperger Syndrome (AS) and Pervasive Developmental Disorder – Not Otherwise Specified (PDD-NOS; also known as atypical autism). These diagnoses are similar to autism in terms of their primary difficulties with social interaction, but differ in other respects: AS is associated with relatively intact development of language skills, while children with PDD-NOS may show autistic behaviours in only two out of three of the areas outlined above (APA, 1994).

The first clinical descriptions of autism were made by Leo Kanner in 1943. Kanner (1943) described a series of children with an “inability to relate themselves in the ordinary way to people and situations from the beginning of life” (p. 242). They struggled with co-operative and imaginative play, while their overall behaviour was “monotonously repetitious” (p. 245) and “governed rigidly and consistently by the powerful desire for aloneness and sameness” (Kanner, 1943, p. 249). In early years, their language skills were mostly limited to echolalia and often included inappropriate reversal of pronouns such as “you” and “me”. By the age of 5-6 years, though, “...language becomes more communicative, at first in the sense of a question-and-answer exercise, and then in the sense of greater spontaneity of sentence formation” (p. 249). Generally, their intellectual abilities were described as being low, but in many cases this situation appeared to improve with age, and did not preclude the existence of specific areas of skill, such as rote memory (p. 243).

In fact, underlying the characteristic behaviours of autism is a highly complex cognitive profile that encompasses a wide range of strengths and weaknesses (Frith, 2003; Mayes & Calhoun, 2003; Pellicano, Maybery, Durkin, & Maley, 2006). Most well-known are problems with Theory-of-Mind (ToM), or the ability to represent and understand the mental states of others (Baron-Cohen, 1989; Baron-Cohen, Leslie, & Frith, 1985; Heavey, Phillips, Baron-Cohen, & Rutter, 2000; Jolliffe & Baron-Cohen, 1999; Kaland et al., 2002). The classic test of this ability is the false-belief paradigm, in which participants are witness to the

displacement of an object unbeknownst to a toy character. When participants are asked to indicate where the character will search for the object, four-year children are typically able to point to where the character would think the object is. That is, they judge that the character will retain a false belief and search for the item in the wrong location. Younger children, in contrast, point to its current location, suggesting a failure to take into account the knowledge state of another agent (Wimmer & Perner, 1983).

Baron-Cohen, Leslie and Frith (1985) conducted the first study that applied this paradigm to a group of children with autism. In contrast to four-year old, typically-developing children, and older children with a different developmental disability (Down Syndrome), the large majority of autistic children (80%) failed to pass the false belief task, despite having greater chronological age (10-15) and mental age (8-10). For the authors, this suggested a specific deficit of Theory-of-Mind in children with autism. Since then, a range of studies with ASD groups have documented persistent difficulties with theory-of-mind on false-belief tasks and other measures (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Castelli, Frith, Happé, & Frith, 2002; Colle, Baron-Cohen, & Hill, 2007; Happé, 1995; Kaland, et al., 2002; Senju, Southgate, White, & Frith, 2009). Problems with other cognitive skills relating to social functioning have also been reported, such as joint attention (Mundy, Sigman, & Kasari, 1990), gaze following (Charman et al., 1997), and recognising emotional expressions (Weeks & Hobson, 1987).

Alongside difficulties with social cognition, individuals with autism show evidence of atypical cognitive processes in other, non-social domains. As already referred to, Kanner (1943) noted prodigious rote memory skills in the children that he saw. Participants with ASD have also been reported to excel on perception-based puzzle tasks where specific parts must be identified from a whole, such as on the Embedded Figures Test (Shah & Frith, 1983) and Block Design task (Shah & Frith, 1993). In contrast, they sometimes struggle on tasks where information must be processed globally or holistically, such as in the use of context to disambiguate the meaning of a sentence (Frith & Snowling, 1983). Such a pattern of skills has been referred to as demonstrating “weak central coherence”, or a difficulty with drawing together “diverse information to construct higher level meaning” (Frith & Happé, 1994, p. 121).

Finally, problems in a number of higher-order cognitive processes have been associated with autism. These have typically been grouped under the umbrella term of executive functions (EFs), or cognitive skills where task-related information must be maintained “online” and other information inhibited in order to achieve a goal (Pennington & Ozonoff, 1996).

Individuals with autism have been observed to have difficulty with EF skills such as planning (Ozonoff, Pennington, & Rogers, 1991), working memory (Bennetto, Pennington, & Rogers, 1996), cognitive flexibility (C. Hughes, Russell, & Robbins, 1994), inhibition (Rinehart, Bradshaw, Tonge, Brereton, & Bellgrove, 2002) and selective attention (Burack, 1994; see Hill, 2004; and Kenworthy, Yerys, Anthony, & Wallace, 2008 for reviews of EF skills in ASD).

### **Problem-solving in autism spectrum disorders**

Another complex cognitive skill, not typically included under the EF umbrella and not studied as extensively in autism research, is problem-solving. Problem-solving is a wide ranging term that has been used to refer to a variety of different tasks and cognitive processes<sup>1</sup>. Here, it is defined as the process of identifying a solution to a puzzle or question where the answer is not immediately apparent. Following its historical use within experimental research, problem-solving is also typically understood to require the generation of strategies and the completion of a series of moves, or achievement of certain sub-goals, in order to solve the problem (Newell & Simon, 1972).

Meant in this sense, a number of studies have reported apparent difficulties in the problem-solving of individuals with ASDs (Channon, Charman, Heap, Crawford, & Rios, 2001; Goddard, Howlin, Dritschel, & Patel, 2007; Minshew, Meyer, & Goldstein, 2002; Minshew, Siegel, Goldstein, & Weldy, 1994). When a scenario is presented in which participants have to come up with their own solution to a problem, or try out a variety of strategies that might help them work towards such a solution, individuals with ASDs tend to struggle.

For example, in a study with adults with Asperger Syndrome, Channon et al. (2001) asked participants to generate strategies to deal with a scenario where a neighbour has a noisy dog. Typical strategies that may be expected are ones that involve negotiating with the neighbour, or, failing that, contacting the council or police. In contrast to typically-developing participants, individuals with AS came up with solutions that were often disproportionate (“Move house”) or socially inappropriate (“Poison the dog”) (Channon et al., 2001). Here, their self-generated answers failed to solve the problem in an effective way. Moreover, this

---

<sup>1</sup> For an example of this just within autism research, compare its use in Rumsey (1985) to refer to performance on the Wisconsin Card Sorting Test (Heaton, 1993) with its use in Soulieres et al. (2009), who class matrix reasoning as a form of problem-solving. In a minimal sense of the solution being hidden, these may be considered examples of non-verbal problem-solving, but they arguably lack the step-based structure of more classic problem-solving tasks.

is likely to extend to many other situations in the everyday lives of people with ASD: if coming up with the right solution is difficult even for hypothetical scenarios, then the ad-hoc problem-solving of the day-to-day could be very challenging indeed.

The kind of scenarios presented by Channon and colleagues clearly involve a considerable social component, meaning that their results could just reflect a lack of social understanding in their participants. If that were to be the case, then problem-solving *per se* would not seem to be the problem; the definitional impairments in social interaction and function that are inherent in autism would appear to be the likely source of such difficulties. But problem-solving in itself is actually an important process to look at, because it also relies on other, non-social processes that may function abnormally in people with ASDs.

First, problem-solving requires the top-down imposition of a strategy in order to be successful. Here, “top-down” refers to the agent’s application of their own knowledge, or own organisation of information, onto task stimuli. This is in contrast to a “bottom-up” process where an agent’s response is guided simply by information inherent in the task. For example, a category matching task, where participants identify examples of a particular category (e.g. “red” or “square”), would be a bottom-up process. In contrast, a free sorting task, where participants choose the basis on which they sort exemplars, would be a top-down task.

A number of studies on autism have observed apparently intact bottom-up skills, but difficulties or atypicalities in the application of top-down processes. Taking the above example, various studies have reported intact performance in ASD individuals in basic categorisation, i.e. bottom-up, tasks (Geurts, Verté, Oosterlaan, Roeyers, & Sergeant, 2004; Tager-Flusberg, 1985a, 1985b; Ungerer & Sigman, 1987; Verté, Geurts, Roeyers, Oosterlaan, & Sergeant, 2005). However, impaired or atypical performance has been reported on tasks where more abstract or complex forms of categorisation are required (Gastgeb, Strauss, & Minshew, 2006; Shulman, Yirmiya, & Greenbaum, 1995) or where participants are free to sort according to their own rule (Kovattana & Kraemer, 1974; Ropar & Peebles, 2007) – both of which require a more top-down application of former knowledge and strategy.

An analogous discrepancy is seen in research on memory with autism. In cued recall, the stimulus to remember words is supplied bottom-up, whereas in free recall, the participant must generate their own top-down strategy to support performance. Many studies have reported intact levels of cued recall for word lists in autistic participants, alongside reduced

performance on free recall tasks, or a failure to strategically draw on semantic groupings to support recall (Cheung, Chan, Sze, Leung, & To, 2010; Gaigg, Gardiner, & Bowler, 2008; Hermelin & O'Connor, 1970; Minshew & Goldstein, 1993). Problems with generating and using effective strategies in individuals with ASD have also been reported on other cognitive tasks, such as multi-tasking (Mackinlay, Charman, & Karmiloff-Smith, 2006), and anecdotally by parents and teachers: in the classroom, children with autism often have difficulty with applying knowledge spontaneously to new problems, such as answering exam questions.

A second significant characteristic of problem-solving is that it commonly involves an “open-field” or open-ended structure; that is, participants are free to respond to a task in a range of ways. This is often associated with tasks where a top-down response is required, but the two characteristics pose slightly different kinds of challenges. In a top-down task, individuals must generate their own strategy, or apply their own knowledge to a task, but the particular constraints of that task might mean that only one or a small number of strategies are actually appropriate to use – in other words, a strategy is required, but the potential set of responses is actually fairly constrained. In contrast, an open-field task will often require a top-down strategy, but there may be multiple potential routes to success: participants can choose to solve the problem in whichever way they see fit.

When participants with autism attempt open-field tasks, they tend to perform worse than they would for tasks with a more directed set of responses (White, Burgess, & Hill, 2009). On the one hand, this could reflect a difficulty with handling the less directed format of such a task (White, In Press). On the other hand, it could reflect atypical styles or preferences in how individuals with ASD approach such tasks, given the freedom to choose their own strategies (Happé, 1999). A relatively recent trend in research on autism has been to recognise the possibility of an autistic “cognitive style” (Happé & Frith, 2006), where local details and features tend to be concentrated on at the expense of global or holistic forms (see Baron-Cohen, 2006, for another example of a style-based account). Much of the evidence for this idea is drawn from tests of visual perception (e.g. Shah & Frith, 1983), but such a style is proposed to extend to other, higher cognitive domains, where individuals with ASD may be approaching tasks in a different way to typically-developing individuals. Thus, in the case of open-field tasks, what may be being measured is an autistic cognitive style at work; a difference from the norm, rather than a difficulty or deficit.

Thus, difficulties with problem-solving in autism, though often seen in social contexts, are not necessarily “purely” social. The cognitive requirements of solving a problem –



specifically, the application of top-down strategies in open-field contexts - are suggested by a range of studies to be abnormal in people with ASD; possibly in the form of an impairment or deficit, possibly as a bias or tendency in cognitive processing. Whether deficit or style, though, it is still not clear why this happens in autism. And this is an important thing to explain, because – as the anecdotal reports attest to – not being able to apply knowledge or come up with problem-solving strategies can have a clear impact upon everyday skills for people with ASDs.

A few theories have previously been put forward: Minshew, Meyer & Goldstein (2002) have proposed that this reflects a specific deficit in the spontaneous formation of integrative, organising strategies. In the context of memory research, Bowler and colleagues (Bowler, Gardiner, & Berthollier, 2004; Bowler, Matthews, & Gardiner, 1997) suggested that people with ASD rely on external “task supports” to deploy effective cognitive strategies more than typically-developing individuals. Finally, a recent paper by White (In Press) has proposed that participants with ASD may struggle with the implicit, unsaid demands of such tasks.

What these ideas do not do, however, is amount to any sort of consensus as to what may be driving problem-solving difficulties in autism. What follows is an attempt to clarify and explain some of the cognitive factors that may underlie these kinds of difficulties. This was done by focussing in depth on one particular problem-solving measure, the Twenty Questions Task (Mosher & Hornsby, 1966) and exploring how the problem-solving of individuals with autism on this particular task may be best explained. As will be argued in the following section, this is a task with a rich history in experimental psychology, and the potential to speak to a number of the issues concerning higher cognitive skills in autism.

In focussing on Twenty Questions, this inevitably constrains the following work to certain theories and ideas at the expense of others. Based on their previous use of the TQT, the research presented here specifically engages with the theories of Minshew and colleagues (e.g. Minshew et al., 1994; Minshew et al., 2002). In contrast, the “task support” hypothesis of Bowler and colleagues (1997; 2004) is largely not discussed here, because of its primary concern with memory, rather than problem-solving. This idea is briefly returned to in the General Discussion, and Appendix 3 contains some preliminary data on memory and the TQT. The ideas of White (In Press) concerning open-ended tasks are not directly tested because they were published very late in the process of writing.

## **Twenty Questions as a measure of problem-solving**

In the game *Twenty Questions*, one player thinks of an everyday item and another has to ask questions to establish its identity. The questions can only be answered yes or no, and only a limited number of questions are permitted. Typically the best strategy is to ask questions that eliminate a number of alternatives, while establishing the sets or categories that the item belongs to. For instance, a player might ask “Is it a living thing?” or “Is it man-made?” followed by questions that further constrain the search: “Is it an animal?”, “Is it a mammal?”, “Is it a pet?” and so on. Once the search is narrowed sufficiently, a guess can be made: “Is it a dog?”, “Is it a cat?”.

While it may be best known as a fairly repetitive game played on long, boring car journeys, *Twenty Questions* and similar games have a long history as experimental tasks within psychological research. The *Twenty Questions Task* (TQT) and its analogues have been used to examine cognitive skills in typically-developing children and adults (Denney, 1972, 1974; Denney, Denney, & Ziobrowski, 1973; Drumm, Jackson, & Magley, 1995; Drumm & Jackson, 1996; Herwig, 1982; Laughlin, Moss, & Miller, 1969; Mosher & Hornsby, 1966; Siegler, 1977; Taylor & Faust, 1952; Thornton, 1982; Van Horn & Bartz, 1968), children and young people with learning disabilities (Barton, 1988; Borys, 1979; Copeland & Weissbrod, 1983; Simmonds, 1990; Tant & Douglas, 1982), adults with chronic alcoholism (Laine & Butters, 1982; Saarnio, 1993), and adults with various types of brain injury (Baldo, Delis, Wilkins, & Shimamura, 2004; F. C. Goldstein & Levin, 1991; Klouda & Cooper, 1990; Levin et al., 1993; Levin et al., 1997; Marshall, Harvey, Freed, & Phillips, 1996; Marshall et al., 2004; Marshall, Karow, Morelli, Iden, & Dixon, 2003a, 2003b; Marshall, McGurk, Karow, & Kairy, 2007; Upton & Thompson, 1999; Vilkki, 1988).

Although an apparently simple game, the TQT taps a range of skills important to understanding typical and atypical cognitive processes. To solve the TQT, participants need to gather more information about the target, and its first main use in research was as a test of information-seeking in typically-developing children (Laughlin, et al., 1969; Mosher & Hornsby, 1966; Van Horn & Bartz, 1968). To recognise that more information is required, and to select an appropriate strategy that identifies the target quickly and efficiently, is to engage executive functioning skills, such as planning and selective attention. It is in this role, as an executive task, that the TQT has been deployed in research with neuropsychological populations (Baldo, et al., 2004; Marshall, et al., 2003b).

Primarily, though, it is a measure of verbal problem-solving ability, in that the solution – the identity of the target item - is not immediately apparent from the task materials (Siegler, 1977; Taylor & Faust, 1952). Because the solution is hidden, participants must generate a top-down strategy, via questioning, that allows them to move towards a state where they can guess at the identity of the target. On the TQT, the strategy is in the questions that participants select, some of which may be more effective than others in eliminating possible options. The “moves” towards the goal state are represented by the systematic elimination of alternatives. And though the range of possible answers is typically limited to a set number of options (usually 24-48 items), participants are free to ask whichever questions they consider useful to identify the target. In this sense, participants are presented with an open set of possible responses – allowing for assessment of not just whether participants can solve the problem, but *how* they go about doing so.

Aside from the cognitive skills that it measures, a further advantage of a task like the TQT is its game format. On the one hand, this makes it very accessible for participants from a range of population groups, especially children. On the other hand, its open-field structure is thought to be more ecologically valid than many standard experimental tasks. For example, when trying to work out how best to search for a lost item round the house, there is no single right way to do it – you could search through each room systematically, or you could go first to the room you thought you last had it in, or ask someone, *etc.* In contrast, a number of laboratory-based experimental tasks only allow for a set range of responses. Some commonly used tasks - particularly in executive functioning research - have been questioned for their relevance to real-world challenges for people with autism (Kenworthy, Yerys, Anthony, & Wallace, 2008). Using a task such as the TQT allows for measurement of cognitive skills, but in a format that gets closer to everyday cognition.

### **Autism and twenty questions**

In the most commonly used version of the TQT, participants are presented with an array of 42 black and white line drawings of everyday objects and items (see figure 1). The experimenter then chooses one of the items, and gives the participant instructions similar to the following:

“Now I am going to think of one of these objects. Your task is to figure out which object I am thinking of by asking questions. You can ask any question you like, but I can only answer by saying “Yes” or “No”. The idea of the game is that you figure out the object I am thinking of with as few questions as possible” (Minshew et al., 1994, p.34).

Typically, participants have a maximum number of questions that they can ask (often 10 or 20) and performance on the task has been measured in terms of overall success (the number of problems solved), efficiency (the number of items eliminated per question, or the number of questions used per problem) and general questioning strategy. Strategy is usually divided into grouping questions, where more than one item is referred to in the set, and guess questions, where only one item is eliminated (Mosher & Hornsby, 1966)<sup>2</sup>.

A study by Minshew, Siegel, Goldstein and Weldy (1994) was the first to apply the TQT to a group of participants with autism. Performance in an ASD group of adolescents and adults (Age<sub>Range</sub> = 11-41) was compared to performance in a sample of typically-developing participants matched for age, gender and IQ. The two groups attempted four trials of a Twenty Questions Task. Compared to typically-developing controls, participants with autism completed significantly fewer successful trials of the TQT. In addition, controls and autistic participants significantly differed in their strategy use. Whereas control participants used questions that eliminated groups of items for approximately 50% of each game, participants with autism only used such questions for 25% of their queries. Instead, participants with autism used guess questions as much as 50% of the time (Minshew et al., 1994).

For example, instead of questions about general categories, such as “Is it an animal?”, participants with autism asked questions that only eliminated one item at a time, either by directly naming the item (“Is it a spoon?”) or referring to it in an elliptical way (“Is it something you eat soup with?”). This meant that participants with ASD were much less efficient in how they eliminated items to solve the problem. Similar findings were observed by Minshew and colleagues in follow-up studies involving adults (Minshew, et al., 2002) and children (D. L. Williams, Goldstein, & Minshew, 2006a) with ASDs.

As outlined above, autism is associated with delays in language and intellectual development in early life (APA, 1994). In some cases, spoken language may only develop after six or seven years, if it all (Bennett et al., 2008; Luyster, Kadlec, Carter, & Tager-Flusberg, 2008). As such, differential performance on a task like Twenty Questions may not be thought to be surprising: a verbal and cognitive task would clearly pose a number of challenges for a group with linguistic and intellectual delays.

---

<sup>2</sup> These are sometimes referred to as “constraint-seeking” and “hypothesis-scanning” or “hypothesis-testing” questions, based on Mosher & Hornsby’s (1966) original terminology. Also reported in some studies are “pseudo-constraint-seeking” questions, where participants ask a question phrased in a way that sounds like it would pick out a set of multiple items, but in actual fact only removes one item in the set. For example, “Do you use it in the rain?” could refer to multiple items, but in the standard set would only eliminate one item, the umbrella. For clarity, here only *grouping* questions and *guess* questions will be used.

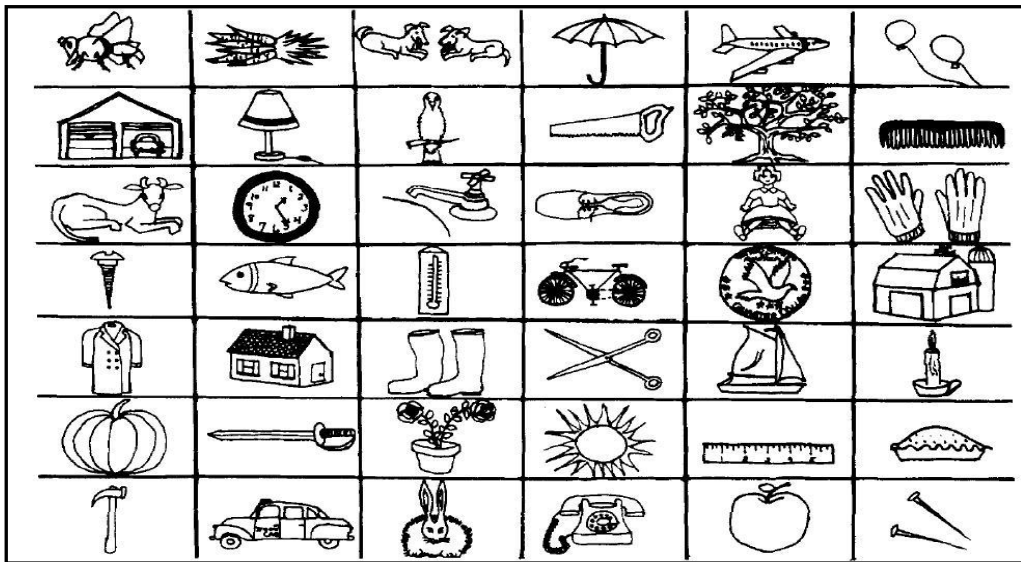


Figure 1: The Twenty Questions Task array (From Mosher & Hornsby, 1966).

But the above research was conducted with individuals with “high-functioning autism” (HFA), who fall within the normal range (full-scale IQ>70) for general cognitive ability and generally have at least functional levels of speech. In addition, IQ abilities, which often differ between autism and typically-developing groups of participants, were matched in these studies. This suggests that the atypical performance of participants with autism on the task is not attributable to deficits in the general abilities required to engage in such a game (such as vocabulary, or knowledge of semantic categories). Furthermore, the fact that differences in questioning have been observed in autistic children and adults (Minshew et al., 2002) suggests that is not simply a feature of general delay that will disappear with age. It appears instead that participants with autism have the general skills required to play the game, and yet still do not always perform in the same as way as typically-developing counterparts. Here, then, would appear to be a prime example of individuals with ASD failing to effectively use a top-down strategy to solve a problem. But how best to explain it?

As referred to above, participants with autism have often done well on category matching and recognition tasks, despite seemingly not using categories to effectively guide their search on the TQT. Based on performance on the TQT and similar tasks, Minshew et al. (2002) proposed the existence of a dissociation between intact *concept identification* and impaired *concept formation* in autism. That is, they suggested that participants could essentially recognise category groupings that they knew (“concept identification”), when prompted by a direct question, or by task stimuli that serve to cue the relevant category, but they struggled with tasks that demand “concept formation”, or the unprompted combination of different items together in the form of a category. For Minshew and colleagues, the problem-solving performance of participants with ASD reflected an underlying “inability to spontaneously form schemata or paradigms that organize information” (Minshew et al., 2002; p. 333).

The notion of a distinction between concept formation and concept identification predates the work of Minshew and colleagues – they attribute it to Bourne (1966) – but they were the first to explicitly apply such an idea to cognitive skills in autism. In its referral to using organising strategies “spontaneously”, their definition of a concept formation deficit also resembles proposals made by other researchers that children with autism have a problem with generativity, or the ability to produce novel responses without cues or prompts (Dichter, Lam, Turner-Brown, Holtzclaw, & Bodfish, 2009; Turner, 1997).

On its own, pointing out such a dissociation would not necessarily have much explanatory power: as already discussed, problems with generating strategies, and differences in performance on tasks that do and do not cue a specific response, are reasonably common

observations in autism research (White et al., 2009). But Minshew and colleagues' explanation sits within a wider theory about information processing in autism. They suggest that people with autism have a domain-general problem with complex, integrative forms of information processing, underpinned by disruptions to networks of distal brain regions (Minshew, Goldstein, & Siegel, 1997). The proposed mechanism for this is the presence of reduced communication in cortical networks in autism, making it harder for the integration of individual pieces of information, required for forming conceptual strategies, to occur (for further explanation of this theory, see Minshew & Goldstein, 1993a; Minshew & Goldstein, 1998; Minshew, Goldstein, & Siegel, 1996, 1997; Minshew, Sweeney, & Luna, 2002).

At the time that Minshew and colleagues were first proposing their ideas, relatively few studies had reported on connectivity and network-based neural abnormalities in autism. Such a hypothesis stood in contrast to other cognitive deficits in autism, which were generally assumed to result from focal brain abnormalities: for example, problems with Theory-of-Mind were proposed to result from atypical development of the amygdala, a structure strongly associated with emotional processing and social cognition (e.g. Baron-Cohen et al., 2000). However, a number of subsequent studies have gone on to offer support for Minshew's ideas, with evidence of reduced connectivity between a wide range of brain regions in ASD being reported, both structurally (Barnea-Goraly et al., 2004; C. Cheung et al., 2009; Courchesne & Pierce, 2005; Kana et al., 2006) and functionally (Castelli, et al., 2002; Just, Cherkassky, Keller, Kana, & Minshew, 2007; Just, Cherkassky, Keller, & Minshew, 2004; Kleinhans et al., 2008; Villalobos, Mizuno, Dahl, Kemmotsu, & Müller, 2005; Welchew et al., 2005)<sup>3</sup>.

Thus, the idea of a "concept formation deficit" appears to describe and explain why individuals with autism struggle to perform effectively on the TQT. By drawing on what is known about categorisation skills in other studies on autism, and contrasting with autistic problem-solving performance, Minshew and colleagues provide a theory for a specific difficulty with the use of organising strategies in ASD. Moreover, evidence from neuroimaging studies on autism appears to provide an explanatory basis for why this occurs: a failure of top-down strategy, due to more general problems with the effective integration of information across cortical networks in autism.

---

<sup>3</sup> For reviews on connectivity in autism, see Belmonte et al. (2004), Geschwind & Levitt (2007), Hughes (2007), Rippon et al. (2007), and, more recently, Müller et al. (In press), Vissers, Cohen, & Geurts (In press).

This is one interpretation of what is happening on the TQT – but is it a correct one? As already explained, problem-solving is a complex process. This is especially the case for the Twenty Questions Task, which relies on a number of executive and linguistic skills to be performed effectively – as attested to by the wide range of ways in which the task has been used in previous research (Laine & Butters, 1982; Marshall, et al., 2003b; Mosher & Hornsby, 1966). The performance of participants with ASD on the TQT could reflect a specific problem with concept formation, but it could equally have arisen from a number of other demands that the task makes on a participant. The following section describes a Master’s thesis and subsequent paper by the author (Alderson-Day & McGonigle-Chalmers, 2011) that attempted to address some of these concerns.

### **“Is it a bird? Is it a plane?”: establishing the factors that affect TQT performance in autism**

While a potentially rich task, one problem with the TQT is that it arguably taps a number of cognitive processes that are thought to be impaired or develop atypically in autism. First, in asking participants to identify a target being “thought of” by the experimenter, the standard version of the task could be interpreted as requiring a certain amount of theory-of-mind ability. By placing the answer in the mind of the experimenter, this set-up may make it hard for some individuals with ASD to represent what the answer may be, or they may simply lack any motivation to find out what it is.

Second, as reflected by its use with other neuropsychological populations, the TQT is a measure that relies on good executive skills, and failure on the task is often taken to reflect specific kinds of executive dysfunction (Delis, Kaplan, & Kramer, 2001). For instance, by having to ask a range of questions to reach the target, participants need a certain amount of cognitive flexibility: persisting with exactly the same questions, irrespective of the answer, is a strategy that is unlikely to identify the target quickly. In addition, for the standard TQT participants do not remove items as they proceed through the task; that is, the task array remains unchanged throughout each trial. This entails a demand on other executive functioning skills, especially working memory (which questions have been asked?), but also selective attention (which items do I still need to attend to?) and inhibition (which items do I need to ignore?). As explained above, studies on executive functioning in autism have reported problems in each of these areas, and all of them could conceivably be contributing to problem-solving difficulties on the TQT.



Finally, an effective approach to the TQT may require good “concept formation” skills, but not all concepts are necessarily equal. The TQT presents a mixture of everyday items that belong to relatively common semantic categories. Some organising strategies on the TQT may be based on purely perceptual categories (“Is it red?”), while others may rely on a much more sophisticated level of conceptual knowledge, such as knowledge for more abstract categories (“Is it man-made?”). On free sorting tasks, individuals with ASD have been reported to prefer concrete over abstract sorting principles (Ropar & Peebles, 2007): this may reflect a “top-down” failure of some sort, or it could indicate an adverse sensitivity to more abstract forms of grouping. Similarly, on the TQT, a problem with asking effective questions may be caused by a problem with flexibly using certain semantic categories, rather than a general problem with using organising strategies.

Alderson-Day & McGonigle-Chalmers (2011) designed a study to test these explanations, by applying a modified TQT to a group of 14 young people with ASD and a group of age- and full scale IQ-matched typically-developing controls. In order to avoid the potential confound of theory-of-mind, the hidden item was selected by the experimenter using a “random generator” animation on a laptop computer, rather than asking participants to identify the item that the experimenter had in mind. Then, when participants identified the target, it was also revealed on the computer screen. This was done in an attempt to place the answer not in the mind of the experimenter, but as a matter of fact stored by the computer (akin to an answer on a quiz).

The second set of modifications varied the executive demands of the TQT. To assess cognitive flexibility, the contents of the set were changed after three trials, such that the internal structure of categories that were available in the set had differed. This was done in order to prompt a change in questioning, as questions that had been useful previously would not necessarily be as effective following the switch. Also varied was whether participants could or could not eliminate items from the set after each question, by using a hinged-frame *Guess Who?* board. Participants attempted six trials where they could eliminate items from the set (by knocking them down), followed by two trials where this was prohibited. Any decrement in performance was taken to indicate that participants struggled with the additional executive demands – demands that would usually be in place for all trials on a standard TQT.

The final modification made was to vary the task stimuli. To contrast with the use of everyday semantic categories in the standard task, a set of robot characters were included in a separate TQT condition. The robots were designed to vary only in perceptual

characteristics, in order to assess whether problems with strategy use on the task were actually more specific to the handling of semantic-conceptual categories in the standard TQT. The content of questions was also recorded, to assess for any preferences in participants' strategies.

What was observed partially supported Minsheu and colleagues' earlier findings. The study observed less grouping and more guessing in the ASD group than the TD group, even with the removal of theory-of-mind elements from the task. This tendency a) was observed irrespective of the changes in set structure (designed to test flexibility of questioning), b) was present whether item elimination was or was not permitted, and c) was evident for both standard stimuli and the perceptually-categorised Robot stimuli.

However, the performance of ASD and control participants also varied in other ways. For example, although ASD participants asked fewer grouping questions throughout the task, they were more likely to require extra questions to solve the problem when elimination was prohibited. This was taken to indicate a greater difficulty with searching the set when the executive demands of the task increased. In addition, the variation of perceptual vs conceptual task stimuli highlighted group differences in performance: ASD participants were generally more efficient and more successful on the task when dealing with purely perceptual stimuli than when handling the standard stimuli, whereas controls performed comparably on both forms of the task. ASD participants were also less likely than control participants to ask questions that referred to abstract categories (such as "Is it a mammal?").

A final key finding concerned the efficiency of questions asked by the ASD group. Participants with ASD not only asked fewer grouping questions on the task, they also eliminated fewer items per question (a measure referred to as "question quality"; Alderson-Day & McGonigle-Chalmers, 2011). That is, each question that they asked tended to be less efficient and too restricted to subdivide the set effectively.

In principle, this could have been produced by participants with ASD taking more guesses, as a guess by definition will only eliminate one item at a time. But differences in question quality were observed even when only grouping questions were analysed. That is, participants with ASD, even when they were using grouping strategies, were still eliminating too few items - they were casting their net in a way that was too restricted to be fully effective. For example, rather than ask a very general question first ("Is it a living thing?") they may ask about a more restricted category ("Is it footwear?").

This prompts a subtle but important distinction in how best to describe ASD performance on the task. Minshew and colleagues proposal of a concept formation deficit refers to a failure to spontaneously use organising strategies (Minshew et al., 2002). The analysis of grouping and guessing rates from the TQT supported this idea, as participants with ASD were taking guesses, rather than organising the items together in the form of category-based questions. But the analysis of question quality shows that even when they adopted a level of organisation, even when they were grouping, their focus was too restricted. That is, rather than the deficit being a categorical one, of either using or not using an organising strategy, it appeared to be a continuous one – a tendency towards smaller groupings when a grouping strategy is being deployed.

Such a tendency was reminiscent of the aforementioned autistic “cognitive style”. Participants with ASD were focussing on the local (“Do you use it to brush your teeth”) rather than the global (“Is it a tool?”); on the particular (“Is it a dog?”) rather than the general (“Is it a living thing?”). This did not appear to be a purely perceptual bias – autistic participants, for example, were no more likely to ask questions about perceptual characteristics, or individual features, such as wheels – but a conceptual bias: a tendency towards the local in terms of how semantic categories were being utilised.

### **Asking the right questions**

The data from Alderson-Day & McGonigle-Chalmers (2011) pointed to a range of ways in which the performance of ASD children on the TQT may differ from typically-developing children. Some of the results were compatible with the ideas of Minshew and colleagues, while others suggested the presence of additional factors that may impinge upon problem-solving in this particular context. Participants with ASD were affected by the change in executive demands of the task; they appeared to be sensitive to the level of conceptual material in the task stimuli, and when grouping, they asked about fewer items than typically-developing participants.

These data could be taken to suggest that there are multiple cognitive difficulties affecting problem-solving performance in autism. Or, they could be used to argue for the absence of a concept formation deficit and the presence of an autistic cognitive style at work; a tendency towards the local in ASD problem-solving skills. But to endorse either of these positions, just based on the data from Minshew et al. (1994; 2002) and Alderson-Day & McGonigle-Chalmers (2011), would be premature, and would leave many questions about ASD

performance unanswered. In particular, three points can be made that highlight the need for further examination of this issue.

First, the experiment conducted in Alderson-Day & McGonigle-Chalmers (2011) realistically only represented the starting point of a long process. Given the complexity of the TQT, many more cognitive factors remained that could explain why participants with ASD fail to ask effective questions, both in terms of what knowledge they bring to the task and how they are affected by the demands of the TQT itself. For instance, if participants with ASD apparently could use top-down strategies, but nevertheless chose ones that happened to be less effective, then one needs to understand how they think about questions in advance – in essence, how they plan their questions, and what they understand to constitute a “good question”. Also to be explained was why participants with ASD were affected by not being able to eliminate items from the testing set: it could be that a failure to allocate attention to relevant stimuli was affecting ASD performance, throughout the task.

Both of these possibilities have the support of previous data suggesting problems with planning and attention skills in autism (Christ, Holt, White, & Green, 2007; Ozonoff, Pennington, et al., 1991). These are, in essence, further executive explanations of problem-solving in autism. What they represent are examples of candidate explanations, drawn from previous autism research, that need to be tested before any particular account of problem-solving in ASD can be endorsed. The research presented in the first part of this thesis (Chapters 1 & 2) examines these explanations.

A second point concerns the explanatory power of a cognitive style. If the problem-solving of autistic individuals was taken to represent a style at work, rather than a deficit in concept formation, then still lacking from this explanation is an account of where such a style may come from. To say that this is evidence of a cognitive style is simply to redescribe the performance in terms of a wider tendency in autistic cognitive processes, but it does not provide a developmental story about why people with autism would be biased towards solving problems in a particular way. Any account that attempts to explain the pattern of problem-solving skills in autism, needs also to make some attempt at saying why a bias or tendency would develop – settling for style alone does not do this. The second and third parts of this thesis (Chapters 3-7) represent an attempt to provide such an account.

Finally, whether understood as a deficit or style, these ideas only explain autistic problem-solving with reference to theory and data from autism research. This, though, might ignore important ideas from elsewhere. Given the variety of studies that have used the TQT to study

cognitive processes, in both typical and atypical populations, there is a huge amount of knowledge available that could be used to inform understanding of problem-solving in autism. By using the examples of other atypical groups with similar difficulties on the TQT, new explanations and hypotheses can be formed and tested – explanations that may be more powerful than those that are only drawn from autism research.

If these kinds of comparisons are not made, then there is a danger that an “autism-only” story emerges. That is, by only considering explanations of problem-solving from autism research, a bespoke explanation is constructed that fits the available data for autism in a “just-so” way. If the problem-solving difficulties seen in autism are indeed autism-specific, then this may be valid. But if they are *not* autism-specific, and if the same problems occur in other population groups, then an explanation is likely needed that draws on more general characteristics of cognitive development. Throughout the following chapters, the examples of other population groups are drawn on in an attempt to achieve this.

A series of studies was planned to further clarify and explain problem-solving difficulties in people with autism spectrum disorders. The general question addressed was why people with ASD deploy inefficient strategies in their problem-solving. The specific question at hand was why this occurs on the Twenty Questions task; why ASD participants do not ask questions that effectively narrow down possibilities. This was done primarily via a task-analysis of the executive demands of the TQT, and secondly with reference to the problem-solving performance of other population groups.

## **Chapter 1: Executive explanations of strategy inefficiency in children with autism**<sup>4</sup>

Being too specific, focussing on the the local and ignoring the global are familiar and well-worn characteristics in research on autism. But an inefficiency in asking effective questions must be interpreted in the first instance as just that: an inefficiency. And if participants can attempt the task in other respects, and yet perform inefficiently, then a number of factors may be driving this kind of performance. Evidence from studies on typical development, brain injury, and executive functioning skills in autism combine to highlight two areas that ASD participants may also struggle with on the task: planning questions, and selectively allocating attention.

### **Planning in Twenty Questions**

If a participant asks less efficient questions on the Twenty Questions Task, this may reflect a problem with how they prepare their questions. Selecting the right question to ask is a matter of assessing which categories will effectively subdivide the set. In most cases, this will be a category that refers to a sizeable proportion of items while still leaving a considerable out-group, thus guaranteeing an elimination of multiple items whether the answer to the question is yes or no. A simple heuristic to achieve this is to use general questions in the initial approach to the task, but selecting consistently effective questions involves being able to think ahead about the consequences of different questions; in essence, being able to *plan* questions effectively. Thus, formulating good questions involves the weighing up of multiple options, adapting a strategy based on the set items remaining, and being able to think ahead through multiple game states.

There is evidence from Mosher & Hornsby's (1966) original verbal inquiry research that the ability to recognise effective questions may actually precede a tendency to consistently deploy such questions. Participants in the study played an open-set questioning game where a scenario was described, such as a car crash, and the player had to ask questions to ascertain the cause of the event. Six, eight and 11 year-old participants were then questioned after the task about their approach to the game. Participants were given a choice between a general

---

<sup>4</sup> Sections of this chapter and chapter that follows are published in Alderson-Day (2011) and a poster presentation at the International Meeting for Autism Research (IMFAR) 2011 in San Diego, California. ("Asking the right questions: Planning differences during verbal problem-solving in children with autism spectrum disorders". B. Alderson-Day, IMFAR, May 12<sup>th</sup>-14<sup>th</sup>, 2011).

question (e.g. “Was there anything wrong with the man?”) and a specific question (“Did he have a heart attack?”) and asked which they would choose if they started again and “wanted to get an answer in as few questions as possible” (p. 99). While the majority of six year-olds would select the more specific question (equivalent to asking a guess question on the TQT), eight and 11 year-olds would consistently select the more general question. However, when attempting the task, eight year-olds would actually ask specific questions instead of more general ones. To the authors, this suggested that they “appeared able to *recognise* a better strategy in the verbal game, but they seemed less able to mount the strategy on their own initiative.” (p. 99, italics theirs).

There is little evidence regarding the extent to which children with autism may recognise the right question on the TQT, as such measures were not included in Minshew and colleagues’ previous uses of the task (Minshew et al., 1994; Williams et al., 2006). Evidence of intact sorting and matching abilities in categorisation tasks might be considered indirect support for the suggestion that ASD participants would be able to recognise efficient categories. If they did choose the right questions in a similar forced-choice format to Mosher & Hornsby’s, this would seem to suggest that the ability to discriminate more and less effective questions is unimpaired in ASD and that their inefficiency is genuinely to be found in the deployment of questions. Alternatively, if they did not consistently identify effective questions, this would suggest that their difficulty is more one of understanding what constitutes a “good” question.

Another aspect of planning effective questions is being able to string together a sequence of categories that progressively narrow down available options. Asking one question that eliminates half the set is effective, but any advantage that this provides is largely lost if the question is followed up by a series of direct guesses. Truly effective strategies contain a series of questions that dissect the set each time; as in the sequence of “living/animal/pet/four legs”.

From around the age of 11, children report strategies that resemble this approach (Drumm, et al., 1995; Drumm & Jackson, 1996). The choice of such a strategy appears to be mostly “on-line”, that is, it occurs during the task itself, each question at a time. Very few players will work out a series of questions fully in advance, because the path of questioning will differ depending on the answers that the experimenter provides. Often the first, and maybe second, questions can be planned, but after that the set must usually be inspected again to see what questions will fit best.

However, there are reasons to believe that planning to narrow down possibilities is still important. Firstly, when people who have a traumatic brain injury (TBI) have attempted the TQT they have been reported to struggle with deploying narrowing strategies (Marshall, et al., 2003b). People who have had a TBI can show a range of executive functioning impairments, especially concerning inhibition and the ability to not act on a prepotent response. As such, direct guessing might be expected to contribute to narrowing failures in this group (as has been reported in other participants with cortical damage; Klouda & Cooper, 1990; Upton & Thompson, 1999). However, in some cases a much more subtle profile is evident. In Marshall et al. (2003b) participants with a TBI often asked effective grouping questions, but would then follow up those questions with restricted grouping, or “pseudo-constraint” questions, i.e. questions that only referred to one item, but in a roundabout way rather than a guess. This was interpreted by the study authors as a problem of planning and strategy shifting – participants were choosing a good initial question, but then failing to recognise the importance of maintaining a narrowing plan, and struggling with shifting their focus to a different category that could narrow down possibilities again in an efficient way (Marshall, et al., 2003b).

A second piece of evidence is that planning out questions prior to a task appears to improve performance, at least for typically-developing young people. Siegler (1977), in an experiment utilising a numbers and letters analogue of Twenty Questions, observed improved question efficiency in a group of 13-14 year olds following the instruction to plan ahead the sequence of questions they wanted to ask. This was the case even when participants departed from the plan itself, suggesting that it was the the adoption of a planning strategy and the act of pre-planning that allowed participants to improve in their questioning, rather than actually possessing a plan and sticking to it during the task. Siegler (1977) argued that pre-planning had this effect because it prompted participants to conduct a deeper search of the “problem space” created by the task.

If problems with planning a strategy can impair TQT performance, and the experience of making a plan facilitates effective problem-solving, then a failure to ask effective questions may not only reflect a difficulty in recognising which questions are best to use, but a problem with planning a *series* of questions to interrogate the TQT. If this is true, it raises a candidate explanation of difficulty for ASD individuals, who in many executive tasks display impairments in planning (Hill, 2004). The following sections summarise the evidence for this.



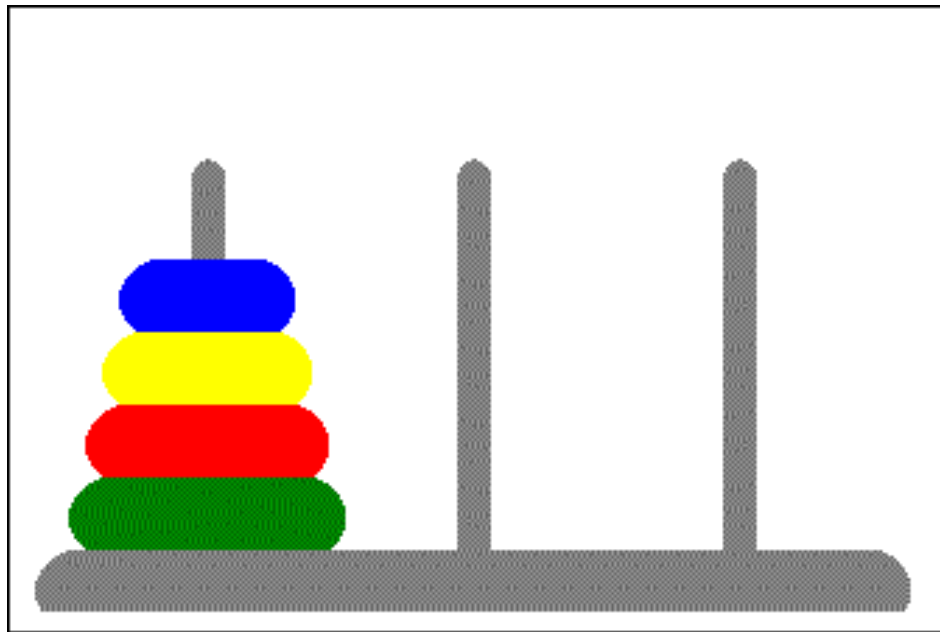
## Measuring planning

As a core process in executive functioning, there is an extensive background of research on planning abilities in individuals with autism (Bramham et al., 2009; Geurts, et al., 2004; Goldberg et al., 2005; C. Hughes, et al., 1994; Ozonoff et al., 2004; Ozonoff, Pennington, et al., 1991). The large majority of studies have examined planning using basic non-verbal problem-solving tasks, where a series of moves or game states must be thought through in order to reach the solution quickly and efficiently.

Much of this work has either used the classic “Tower of Hanoi” (ToH) problem, or close analogues of this task. In the Tower of Hanoi, participants must move a series of rings between three poles to arrange them in a new sequence, while limiting the number of moves used. In the simplest version, three rings of different sizes must be moved from the first pole to the third pole, although more complex versions are often used (See figure 2 for an example of a four-ring version of the task). Only one ring can be moved at a time, and a large ring can never be placed on top of a small ring. The most well-known use of the Tower of Hanoi task in experimental psychology is the work of Newell and Simon (1972), who used the puzzle as the basis of their formalised approach to problem-solving in healthy adults.

The use of Tower tasks as specific measures of planning, particularly in neuropsychological populations, is typically attributed to Shallice (1982). Shallice sought a task that could be used to highlight deficits in the executive control abilities of people with anterior lesions to the brain. Based on the work of Simon and colleagues (Anzai & Simon, 1979) and others working in the field of artificial intelligence (e.g. Sussman, 1975), Shallice argued that Tower tasks involve a specific requirement to plan ahead because they demand the decomposition of the puzzle into a series of subgoals. Failure to plan ahead by thinking through such subgoals may lead to moves being attempted in the wrong order, or needing to be reversed in the course of the puzzle (Shallice, 1982).

To test this idea, Shallice (1982) developed an analogue of the Tower of Hanoi, the “Tower of London” (ToL) and applied it to a group of 61 patients with localised unilateral lesions to the brain. The ToL differed from a Hanoi puzzle in requiring the movement of three coloured balls of the same size across three poles of differing size (a design chosen to allow greater variation of difficulty across trials). The task therefore differed in terms of its logical



**Figure 2.** A four-ring form of the classic Tower of Hanoi Task. Participants must move all of the rings, one at a time, to a third peg, without placing a larger ring on top of a small ring.

structure, but the cognitive demands of the task were thought to be largely similar to the ToH. The results supported the existence of specific task difficulties in certain patients: in contrast to participants with lesions to either posterior or right anterior regions of the brain, participants with left anterior lesions were much less successful than controls in solving 3-, 4- and 5-move puzzles on the ToL. As this was seen alongside intact performance on other neuropsychological tests, such as Block Design, the results were taken to support the existence of specific planning impairments in this particular patient group (Shallice, 1982).

Since then Tower of London tasks have been used to study planning processes for a variety of conditions, including schizophrenia (Rasser et al., 2005), attention deficit hyperactivity disorder (Riccio, Wolfe, Romine, Davis, & Sullivan, 2004), dyslexia (Reiter, Tucha, & Lange, 2005), and Tourette Syndrome (Lavoie, Thibault, Stip, & O'Connor, 2007). Tower tasks have also been used in research with healthy adults to highlight specific subcomponents of planning and the main cognitive resources that effective planning appears to rely upon (Davies, 2003; Gilhooly, Wynn, Phillips, Logie, & Della Sala, 2002; Phillips, 1999; Phillips, Wynn, McPherson, & Gilhooly, 2001).

One main development coming out of this work has been the making of a distinction between pre-planning and on-line planning on Tower tasks (Davies, 2005; Rattermann, Spector, Grafman, Levin, & Harward, 2001; Ward & Allport, 1997). On the one hand, when participants are instructed to plan out their moves in advance on the ToL, the amount of time they initially spend planning has been observed to increase in line with the complexity of the puzzle (Ward & Allport, 1997). On the other hand, the amount of time spent planning does not always translate into better performance in solving the puzzle (Phillips, et al., 2001). It has also been suggested that pre-planning may only have a beneficial effect for puzzles of easy to moderate difficulty (Davies, 2003), possibly because the visuo-spatial working memory demands become too great on more complex problems. In addition, groups of older adults, who sometimes show impoverished or reduced levels of pre-planning, still perform as well as younger adults during the “move” stage of the task itself (Gilhooly, Phillips, Wynn, Logie, & Della Sala, 1999). Thus, though Tower puzzles demand a certain amount of planning ahead, the ability to pre-plan is not necessarily as important as once thought; instead, much of the planning skill may consist in being able to think one or two moves ahead during the task<sup>5</sup>.

---

<sup>5</sup> Such a process has also sometimes been referred to as “opportunistic” planning (Hayes-Roth & Hayes-Roth, 1979)

## Planning in autism

The first study to specifically report on planning processes in individuals with autism was conducted by Ozonoff, Pennington and Rogers (1991). As part of a study comparing theory-of-mind and executive functioning deficits in ASD, Ozonoff and colleagues tested 23 high functioning children with autism ( $Age_M = 12.05$ ) and 20 typically-developing controls matched for age and IQ on a 3-disc Tower of Hanoi task. Compared to controls, participants with autism scored significantly lower for their “planning efficiency”, a composite measure based on how many trials they required to solve each problem (lower efficiency scores reflected the use of more trials to solve the problem). For the authors, this was interpreted as reflecting an example of core executive dysfunction in autism, possibly with a basis in damage to prefrontal regions of the brain (Ozonoff, Pennington, et al., 1991).

The findings of Ozonoff et al. (1991) have largely been supported in subsequent studies that have applied Tower tasks to groups of participants with autism. ASD participants have been reported to require more moves to solve problems (C. Hughes, et al., 1994; Ozonoff, et al., 2004; Robinson, Goddard, Dritschel, Wisley, & Howlin, 2009), complete fewer puzzles overall (Lopez, Lincoln, Ozonoff, & Lai, 2005; Pellicano, 2007), and take more time to complete solutions than typically-developing controls (Geurts, et al., 2004; Verté, et al., 2005). In comparison to some other areas of executive functioning, the evidence of planning difficulties drawn from such studies has been considered to be among the strongest, in the conclusions of reviews by Pennington & Ozonoff (1996) and Hill (2004).

More recently the strength of such conclusions has come under some scrutiny, as seen in a recent review of EF abilities conducted by Kenworthy and colleagues (Kenworthy, et al., 2008). The first question relates to the consistency of evidence between desk-based and computer-based versions of Tower tasks. Studies of autism in the 1990s and early 2000s typically used desk-based versions of the ToH and ToL, while a number of studies after this used the Stockings of Cambridge (SoC), a computerised version of the ToL included in the Cambridge Neuropsychological Test Automated Battery (CANTAB; Robbins et al., 1994). In contrast to previous findings, studies using the SoC have often reported either less marked group differences in planning, or no differences at all (Corbett, Constantine, Hendren, Roche, & Ozonoff, 2009; Goldberg, et al., 2005; Happé, Booth, Charlton, & Hughes, 2006).

For example, Corbett et al. (2009) compared a group of 18 children with ASD ( $Age_M = 9.44$ ), a groups of children with attention deficit hyperactivity disorder (ADHD) and a group of typically-developing children on a battery of executive tasks that included the SoC.

Planning performance was measured in terms of the number of problems solved with the minimum moves possible (“Min Moves”), the amount of time participants spent thinking before attempting the task (“Initial Thinking”) and the average amount of time spent completing each move (“Subsequent Thinking”). In contrast to a number of group differences for other EF variables (such as response inhibition and working memory) no group differences at all were observed on the three indexes of planning performance (Corbett, et al., 2009).

One reason that Kenworthy et al. (2008) suggest for this disparity in findings is the removal of social components from computerised tasks. Whereas the original ToH would require a participant to understand instructions and feedback from a human experimenter, the use of a computer largely eliminates this interaction, making it easier for an individual with autism to attempt.

Another criticism of Tower tasks raised by Kenworthy and colleagues is that they actually conflate a number of executive demands, as in the case of other widely used executive tasks (Ozonoff, 1995). The initial approach to the task may require planning moves in advance and on-line monitoring of possible moves, but this also depends on inhibition of unwanted moves and execution of intended moves. In this respect, deficient performance on one or more of the overall task outcomes does not necessarily indicate a specific problem with planning, but may reflect a separate executive difficulty, or even one that emerges with the interaction of multiple task demands (Kenworthy, et al., 2008).

These concerns are important to consider when establishing the degree of planning difficulties in autism, and it is certainly true that the evidence for a “planning deficit” has been less consistent since earlier reviews. It is not the case, though, that such concerns are unanswerable.

First – as Kenworthy et al. note - evidence of specific planning difficulties is certainly much less strong when SoC findings are considered, but normal performance by ASD individuals has not always been observed on the computerised task either. Studies by Landa & Goldberg (2005), Sinzig et al. (2008) and Ozonoff et al. (2004) have reported planning impairments in ASD participants across the age and ability range. In the latter case, a relatively large group of ASD participants ( $n = 79$ ) were observed to solve fewer puzzles in the minimum amount of moves, used more moves on average, and take longer than controls to think about their moves during the task (Ozonoff, et al., 2004). Thus, specific problems with Tower puzzles do not appear to be confined to tasks that require interaction with an experimenter.

Second, although most of the evidence for planning difficulties in autism may be drawn from Tower tasks, research using other tasks has also pointed to specific problems with planning in individuals with ASD (Bramham, et al., 2009; Mackinlay, et al., 2006; Rajendran et al., 2010). For example, Bramham et al. (2009) tested a large sample of ASD ( $n = 45$ ), ADHD ( $n = 53$ ), and neurotypical adults ( $n = 31$ ) on an executive battery that included the Zoo Map task from the Behavioural Assessment of Dysexecutive Syndrome (BADS: Wilson et al., 1996). The Zoo Map task is a measure of spatial planning that requires participants to plan a route round a series of markers while following certain rules. In contrast to both ADHD and control participants, ASD participants were significantly slower in their creation and execution of a plan to visit a series of fixed points on a map. The study authors interpreted this as an indicator of problems with initiating a strategic plan in the ASD group (Bramham et al., 2009).

Similar problems with planning have also been observed in studies on multitasking in autism. Rajendran et al. (2010) used a computer-based measure of multitasking to assess EF skills in a group of 18 adolescents with ASD and 18 typically-developing controls, who were matched for age and verbal IQ but differed in non-verbal IQ ( $ASD < TD$ ). The paradigm – which was based on the Virtual Errands Task (McGeorge et al., 2001) – required participants to complete a series of jobs around a virtual reality office environment within an eight-minute period. The jobs were presented as a written list to participants that could be referred to throughout the task, although the order of the list did not present the jobs in their optimal order: efficient performance required the formation of a newly ordered plan.

Overall, participants with ASD scored lower than controls in the number of tasks they successfully completed. In addition, the efficiency of their routes around the environment was reduced compared to controls, suggesting a failure to form an effective plan of which tasks to complete in what order. Participants with ASD were also less likely to depart from the initial order of the list of tasks, even when it made sense to do so, which the study authors described as a possible “inability to engage in on-line planning...forcing them to fall back on the strategy of completing the errands in list order” (Rajendran, et al., 2010, p. 1453). Discrepancies in non-verbal IQ may have been expected to underlie some of these group differences, but planning difficulties were still observed in the ASD group even when NVIQ scores were controlled for using analysis of covariance (Rajendran, et al., 2010).

In an earlier study on multitasking, Mackinlay, Charman and Karmiloff-Smith (2006) adapted a paradigm designed for adults (the Greenwich Multitask Test, Burgess, 2000) to study goal-directed skills in 14 children with ASD and 16 typically-developing children

matched for age and IQ. The paradigm consisted of three colouring and sorting tasks that needed to be completed as much as possible within a limited time (three minutes) and in accordance with certain rules, such as attempting all the games at least once within the allotted time, and only attempting tasks one at a time. Before beginning, participants were asked to produce a plan of how they were going to attempt the tasks, which was given a score based on its complexity and optimality. Compared to control participants, children with autism achieved significantly lower planning scores. ASD participants also scored lower for the following of their plans when attempting the task, although this difference disappeared when the problems with initial planning were taken into account. Thus, participants with ASD appeared to specifically struggle with effective pre-planning in this case (Mackinlay, et al., 2006).

Any generalisation about planning skills made across these different measures must be done so tentatively, as the above studies clearly differ from classic Tower tasks in what participants are being asked to do<sup>6</sup>. In the case of measures like the Zoo Map task, the requirement is one of efficient spatial planning round a fixed route, rather than manipulation of moves “in the mind’s eye”. Planning skills across these tasks do not necessarily always correspond: for example, in a study of young children with autism ( $Age_M = 5.5$  years), Pellicano (2007) observed impaired performance on a ToL task, but intact performance on the Mazes task (a measure of spatial planning).

Such tasks may also differ in the level of pre-planning and on-line planning that they require. As already outlined, success on Tower tasks appears to be less related to initial planning and more to on-line or concurrent planning (Phillips, et al., 2001). The kinds of difficulties reported by Rajendran et al. (2010) may point to on-line planning difficulties, but the paradigm used by Mackinlay et al. (2006) highlighted problems primarily in the pre-planning of what tasks to do when.

Nevertheless, studies such as these are informative exactly because they relate specific planning difficulties to specific task measures. Rather than talk of a general “planning deficit”, measured by overall task performance, they point to specific problems, both in initiating effective plans *and* executing plans on-line. Evidence of pre-planning difficulties is particularly significant, because this begins to answer the concern raised by Kenworthy et al. about planning tasks measuring multiple executive processes. A problem implementing a plan mid-task might plausibly be caused by other related executive difficulties, but

---

<sup>6</sup> See also Scholnick, Friedman & Wallner-Allan (1997) for an analysis of how planning demands differ between Tower-based tasks and other measures.

differences in the initial time and depth devoted to preparation *prior* to the task arguably supports a specific planning problem. Interestingly, there is also evidence for this within the ToH literature: in Verté et al. (2005), ASD participants were observed to increase their planning time less than controls in relation to increases in task difficulty, suggesting a possible failure to use the planning stage effectively in attempting the task.

In summary, there is arguably evidence for specific difficulties with planning among ASD individuals, despite recent concerns about the consistency and clarity of evidence drawn from planning tasks. Previous reports of impairments on Tower of Hanoi–style measures have not always extended to more recent, computerised versions of the task, but evidence from other tests does seem to suggest difficulties in the use of planning to support problem-solving. Furthermore, the use of specific task measures has highlighted specific problems with planning efficiently, both in terms of pre-planning and executing plans “on-line”. For a task like Twenty Questions, where multiple moves must be weighed up, selected or rejected, this could have a considerable impact, both upon how questions are evaluated and how planning is effectively utilised in approaching a task.

### **Selective attention during Twenty Questions**

Difficulties with planning may affect how participants with autism think ahead about the effectiveness of their questions. But what about other executive demands that affect performance specifically during the task? When ASD participants attempted the Twenty Questions task in Alderson-Day & McGonigle-Chalmers (2011) their inefficiency in questioning was in general lower than controls. However, group differences were also evident in the number of questions used when participants could not physically eliminate items from the set, suggesting a sensitivity in the ASD participants to the increase in executive demands.

The task condition where elimination was prohibited was designed to test participants’ search abilities when they were required to track their own questions and the experimenter’s responses. Without physical elimination, a participant would be forced to recall which questions they had already asked, and what the answers to those questions were, in order to establish which items remained at each stage of their search. This requirement was hypothesised to be especially problematic for ASD participants, given previous evidence of difficulties with verbal working memory (Bennetto, et al., 1996).

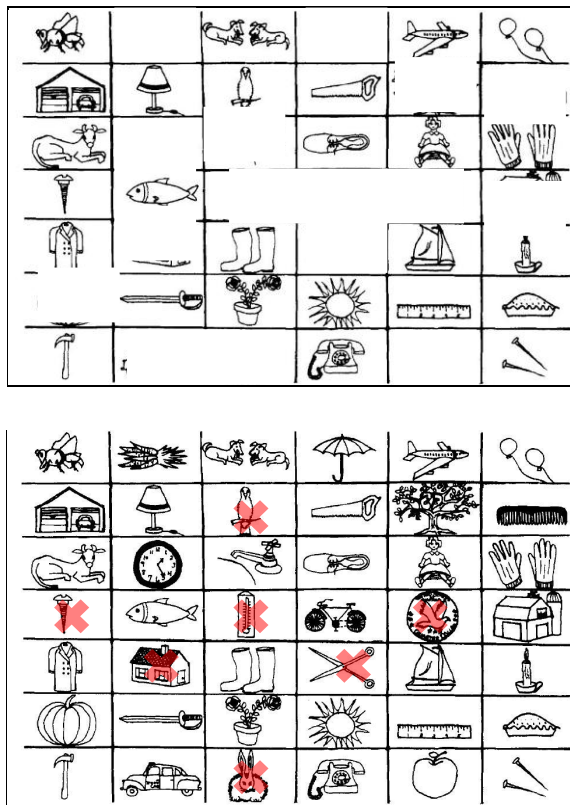


However, the prohibition of physical elimination also introduces another aspect of executive demand into the task procedure. Not only do participants have to remember their questions to mentally update the set; they also have to ignore all the items in the visual array that have already been eliminated, but are still present. Physical elimination removes items from the visual array at every stage, so when such elimination is prohibited, the redundant items remain in set throughout the search. This is significant, because it requires participants to ignore the eliminated items at each stage of their questioning, and selectively divert attention towards items which are still active in the search. Figure 3 displays an illustration of this.

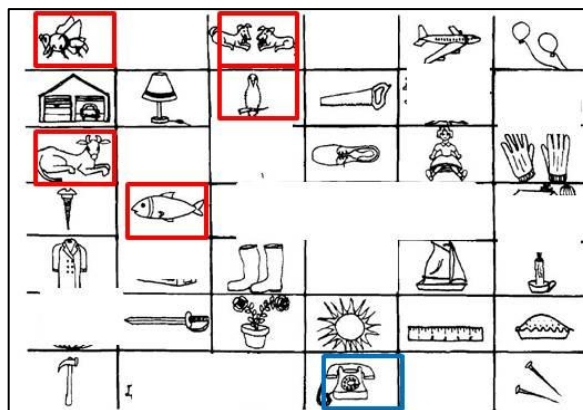
Not being able to physically update the set, then, not only draws on working memory resources, but also the ability to *selectively attend* to relevant items in the set, and ignore irrelevant ones. The executive demand is multi-faceted, and if any of those components were impaired in a participant, this would be expected to lead to poor selection of questions during the task.

Furthermore, if there was a problem with selective attention in the face of perceptual interference, this could impact subtly on overall game performance, i.e. even when physical elimination was allowed. An inability to filter out redundant information might be most strongly expressed when inactive items are constantly part of the perceptual array, but it may also affect the extent to which a player can choose an appropriate question even when all the items are still active.

For example, in the array shown in figure 4, a reasonable question to ask would be “Is it an animal?”, which would be guaranteed to eliminate five of the 27 remaining items. If physical elimination was not allowed, choosing that question could be tricky as the array would also contain many other redundant items that are no longer relevant to the search. But equally, choosing the right question might be disrupted if attention was captured by minority items present in the set: failing to selectively attend to the most global category (animals) might mean that the participant focuses on the appliances, or vehicles, and asks “Is it a telephone” or “Can it fly?” instead. Ignoring these items could be tricky because they offer a potentially quicker route to success on the task if the questioner is lucky. Thus, a salient response is available in the set and must be ignored in order to select questions which are the most likely to be efficient.



**Figure 3. An example of interference when items cannot be removed from the Twenty Questions array.** When redundant items can be removed (top), distractors are no longer present. When this is prohibited (bottom), participants must ignore previously eliminated items.



**Figure 4. Interference from minority items when elimination is permitted.** Even when items can be removed, a problem with ignoring distractors could lead participants to ask a more specific question (blue) over a more general question (red).

In this sense, the inefficiency ASD participants demonstrate in their questioning could be understood as a failure to selectively attend to relevant categories within the array and selectively inhibit redundant or minority categories. In extreme cases (where all items are present in the array), this would be expressed as a marked inefficiency in questioning, but even where redundant items were removed, this could lead to the inappropriate selection of questions mid-game.

### **Measuring selective attention (and inhibition)**

As with planning, a large number of studies have examined attention abilities in people with autism (see Ames & Fletcher-Watson, 2010, for a recent review). Because the process of selectively attending to a stimulus necessarily involves ignoring other stimuli, this area has often been examined in tandem with research on response inhibition abilities, i.e. the ability to deploy and withhold a specific response, even when there is a strong impulse to do so. One consequence of this, though, is that sometimes paradigms end up mixing attentional and inhibitory demands, or using one term (inhibition) to describe areas that are, in principle, separable.

A range of paradigms have been used in this area of research. Typical measures of selective attention that have been used in autism studies include flanker tasks (e.g. Christ, Kester, Bodner, & Miles, 2011), spatial cueing tasks (Wainwright-Sharp & Bryson, 1993) and anti-saccade tasks (Minschew, Luna, & Sweeney, 1999). Most flanker tasks have followed the methodology developed by Eriksen & Eriksen (1974), where participants must respond to the presentation of specific letters (e.g. H or K) while other letters are presented that “flank” the central stimulus. In some cases this will be the same letter, or another letter that is nevertheless associated with the same task response. In other cases, the flanking letters will be associated with an opposite response, creating a conflict in the participants’ response set. When this happens, participants are typically slower to respond to the central stimulus (Eriksen & Eriksen, 1974).

Rather than focussing attention on a central stimulus, spatial cueing tasks and anti-saccade tasks are designed to manipulate the allocation of attention in a visual array. In the case of spatial cueing, arrows, eyes or other markers are often used to guide participants’ attention away from a central fixation point towards specific areas or stimuli, with responses being measured either via task performance (usually in terms of reaction times) or sometimes via eye-movements directly. In anti-saccade tasks, participants must suppress particular eye-

movements when specific stimuli appear, and instead perform the opposite movement. For example, if a light appears to the left of the screen, participants are required to look only at the right hand side (Minshew, et al., 1999).

Tasks that have been used in autism research to look more specifically at inhibition abilities, without involving the allocation of visual attention, have included Stroop tasks (e.g. Lopez, et al., 2005), Go-No Go tasks (Nydén, Gillberg, Hjelmquist, & Heiman, 1999) and measures of negative priming (Brian, Tipper, Weaver, & Bryson, 2003). The classic Stroop test consists of a series of colour words that are printed in incongruous colours (such as “RED” printed in blue ink). Participants must report on the colour of the words while inhibiting the automatic and potent response to read the words themselves (Comalli Jr, Wapner, & Werner, 1962). On Go-No Go tasks, participants are entrained in providing a response to a series of stimuli, but must then withhold that response entirely when a specific stimulus appears (Drewe, 1975). In negative priming paradigms, the information from a previously encountered stimulus must be inhibited in order to respond appropriately to a target. For example, participants may be asked to indicate whether certain strings of letters are the same or different, when the immediately preceding stimuli had prompted the opposite response (Tipper, 1985).

### **Selective attention in autism**

A first look at attention studies on autism may give rise to the conclusion that there is little consistency in findings to date. ASD children and adults have been reported to successfully ignore distractors and potent responses in a number of studies (Adams & Jarrold, 2009; Ames & Jarrold, 2007; Brian, et al., 2003; Griffith, Pennington, Wehner, & Rogers, 1999; Iarocci & Burack, 2004; Ozonoff & Strayer, 1997; Schmitz et al., 2006), while other studies have reported almost exactly opposite findings (Burack, 1994; Christ, et al., 2007; Ciesielski, Courchesne, & Elmasian, 1990; Geurts, Luman, & Van Meel, 2008; Geurts, et al., 2004; Henderson et al., 2006; Luna, Doll, Hegedus, Minshew, & Sweeney, 2007; Minshew, et al., 1999; Ozonoff, Strayer, McMahon, & Filloux, 1994; Pellicano, 2007; Rinehart, Bradshaw, Moss, Brereton, & Tonge, 2008; Verté, et al., 2005). However, as two recent papers on this topic argue (Adams & Jarrold, 2012; Christ, et al., 2011), a lot of this inconsistency can be made sense of once differences in paradigms are taken into account.

First, although impaired performance has been reported in some studies on children with autism (Adams & Jarrold, 2009; Chan et al., 2011; Lopez, et al., 2005; Pellicano, 2007) ASD

participants show largely intact performance on tasks where attention is not modulated, but a certain task response must be inhibited. For example, Goldberg and colleagues (2005) compared inhibition performance on a Stroop task in 17 children with autism, 21 ADHD children and 32 typically-developing controls, all aged 8-12. No group differences at all were observed, despite the ASD group being of generally lower ability (in terms of full-scale, verbal and non-verbal IQ) than both of the comparator groups (Goldberg, et al., 2005). Similarly intact performance for ASD participants on the Stroop Task has been reported in a number of other studies (Bryson, 1983; Christ, et al., 2007; Johnston, Madden, Bramham, & Russell, 2011; Ozonoff & Jensen, 1999; Schmitz, et al., 2006).

One problem with the Stroop test, as argued by Adams and Jarrold (2009), is that the interference it creates depends on participants having a certain level of reading ability. That is, if participants do not automatically read the words in front of them, then focussing on the colour of the words rather than the text may not require any inhibitory skill. Many children with autism show delays or have ongoing problems with reading, particularly for reading comprehension (Nation, Clarke, Wright, & Williams, 2006). Thus, apparently typical performance on this task by participants with ASD may mask underlying problems with inhibition.

However, the presence of intact inhibition skills on other tasks argues against this interpretation. In studies using Go-NoGo tasks and similar measures, unimpaired performance in participants with autism has been observed in studies by Griffith et al. (1999), Happé et al. (2006), Ozonoff and Strayer (1997), Schmitz et al. (2006) and, most recently, by Adams and Jarrold themselves (Adams & Jarrold, 2012). Similarly, intact inhibition skills in ASD groups have been reported in a studies using measures of negative priming (Brian, et al., 2003; Christ, et al., 2011; Ozonoff & Strayer, 1997).

In addition, the findings of two recent studies suggest that difficulties with inhibition, where they are seen, may actually be confined to certain subsets of ASD participants. In studies by Sinzig et al. (2008) and Bühler et al. (2011), impaired performance on a Go-No-Go task was observed in subgroups of ASD participants who also showed features of ADHD, but not in ASD participants without such characteristics. In contrast to autism, ADHD is thought to be strongly associated with inhibition impairments (Scheres et al., 2004). Thus it may be that, for at least some of the above cases, examples of problems with inhibition that have been observed are actually driven by the performance of individual ASD participants who qualify for an alternative diagnosis.

Contrasting with these examples of intact inhibition in autism, tasks where attention must be selectively allocated between competing stimuli have revealed more consistent differences between ASD and control participants. On flanker-style tasks, impaired performance in ASD children and adults has been observed regularly (Adams & Jarrold, 2012; Burack, 1994; Christ, et al., 2007; Christ, et al., 2011; Geurts, Luman, & Van Meel, 2008).

For example, Christ et al. (2007) tested 18 ASD children, 23 siblings of ASD children and 25 typically-developing children on a flanker task that used simple shapes rather than letters as the central and distractor stimuli. Compared to the two control groups (who did not differ in performance), participants with ASD were significantly slower to respond when incongruent flankers were present, suggesting a specific problem with ignoring distractors in the visual array (Christ, et al., 2007). Supporting these findings, results from anti-saccade tasks and other measures of selective attention have indicated impaired performance in ASD participants compared to neurotypical controls (Ciesielski, et al., 1990; Goldberg et al., 2002; Luna, et al., 2007; Minshew, et al., 1999; Mosconi et al., 2009). In contrast to the mixed pattern of findings for inhibition, very few studies have reported intact ASD performance on flanker-style tasks (Iarocci & Burack, 2004; Keehn, Lincoln, Müller, & Townsend, 2010).

The evidence seems stronger, then, for a more specific problem in ASD with selective attention and the basic visual inhibitory processes that this requires, alongside intact skills for response inhibition in other contexts, such as providing a specific verbal or motor response. One caveat to this conclusion, that needs to be acknowledged before continuing, comes from a recent study by Remington, Sweetenham, Campbell and Coleman (2009).

In an experiment with ASD adults, Remington et al. (2009) reported that ASD participants attempting a distractor task required greater load in the perceptual set than control participants before their attention focussed on specific stimuli. That is, control participants would begin selectively attending to specific stimuli once a certain amount of information was present in the perceptual array, but for ASD participants this only happened when a greater amount of information was present. As this was observed alongside intact levels of speed and accuracy, this was interpreted as indicating “enhanced perceptual capacity” in the ASD group (Remington et al., 2009).

If this is correct, then it may be that the findings of previous studies need to be reassessed in terms of group differences in perceptual load; it could be that ASD participants are more distracted by flankers, for instance, because they “take more in” in the first place (see Adams & Jarrold, 2012, for this kind of argument). Thus far there has not been enough follow-up

research to establish the validity of this claim, barring the exception of one very recent replication by the same group (Remington, Swettenham, & Lavie, 2012), but it represents an important counter-explanation for atypicalities in the attentional performance of people with autism.

Nevertheless, it seems fair to say that if people with ASD have difficulties with attention and inhibition processes, they are most likely to be evident in the selective allocation of attention – inhibition *per se* would not appear to be a central issue. Returning to the Twenty Questions task, inhibiting certain questions may therefore not be expected to be a problem for ASD participants, but an unchanging task array may pose a selective attentional demand which could significantly affect performance. If this is the case, it needs to be measured and isolated (where possible) from other executive demands. The following chapter presents an experiment that sought to test this idea, alongside assessment of the contributions of planning difficulties to TQT performance in autism.

## **Chapter 2: Testing the contributions of planning and selective attention to Twenty Questions performance**

An experiment was devised to examine whether difficulties in planning and selective attention could explain the problem-solving performance of children with ASD on the Twenty Questions Task. Two new measures were devised to investigate the role of planning in Twenty Questions performance: a *question discrimination* task and a *plan construction* task. The tasks were primarily designed to see whether problems in selecting effective questions were evident at what might be termed the “planning stage”, that is, in participants’ approach prior to playing the task. Evidence of difficulties in question discrimination and plan construction would suggest a difference in the way questions are thought through by participants with autism, independent of the immediate attentional demands of the game itself. To examine difficulties with selective attention, the within-task conditions of the TQT were varied to moderate the amount of information that participants were required to ignore.

### **Assessing planning on the TQT**

The aim of the *question discrimination* task was to measure whether ASD participants can recognise effective questions outside of the game. Participants were given 10 picture arrays showing example sets, and were asked in each case which of two questions was the “best question to ask first”<sup>7</sup>. In each case, one of the questions was clearly more effective in eliminating a sizeable proportion of the array. In terms of planning, this assessed the extent to which participants could think through the potential consequences of two different questions. Thus, while the task asked participants to think about potential questions in advance – a form of pre-planning- it was also relevant to the on-line forms of planning that participants engage in during the TQT, when thinking through possible options for each new

---

<sup>7</sup> The request to choose the “best” question was defined, according to conventional understanding of the game, as the question that was guaranteed to eliminate the most items. Questions which eliminate close to 50% of the set are guaranteed to substantially subdivide the set. In contrast, “high-risk” questions, i.e. those that pick out only a small subset of the available items, may significantly narrow the search if answered yes, but will minimally affect the overall set if answered no. If participants attempted the question discrimination task thinking that the latter were better questions, this would result in a poor score on the task. However, this would also reflect a misunderstanding about the quality of questions for the actual Twenty Questions game. That is, if ASD participants were answering on the basis of considering high-risk questions to be better, then this reflects something important about their approach to the game.



question. Primarily, though, the task represented a check on understanding; assessing whether participants understood what constitutes a “good” question.

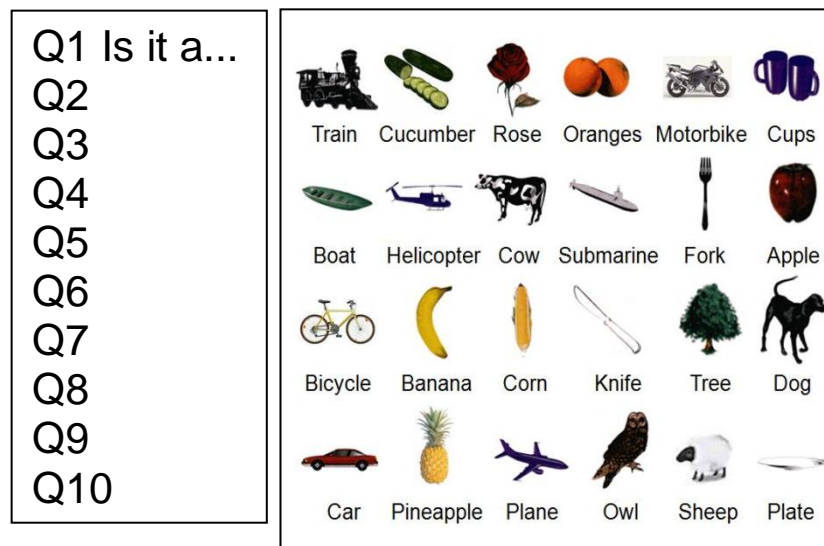
The second task measured *plan construction*. Participants were presented with a selection of possible questions, printed in black on plain white cards, and asked to select five questions that they thought might be useful to use when playing Twenty Questions. They were then asked to arrange those questions into an order, according to which question “they think they would ask first, which one next, *and so on*”. In asking for a whole sequence of questions to be planned out, this measure was a more specific assessment of pre-planning abilities.

The key question that both of these tasks were designed to answer was whether there was any evidence of a planning difficulty in participants with autism. Based on evidence from previous studies on planning in autism, it predicted that ASD participants would struggle with the plan construction task in particular, due to the level of pre-planning that it requires.

A secondary question, related to the first, was whether performing either of these tasks prior to attempting the TQT improved performance for ASD participants. To measure this, a “baseline” measure of Twenty Questions performance was established, and then participants were tested on the TQT for a series of trials after each planning task. If ASD participants were actually having a problem with planning, then the specific instructions to plan might be expected to have a disproportionately beneficial influence on their subsequent task performance. Examining the effects of specific planning activities could also shed light on ways in which ASD performance might be facilitated.

### **Assessing selective attention on the TQT**

As noted above, the prohibition of physical elimination on the Twenty Questions creates a complex executive demand. To keep track of where one is in the task (which items have been eliminated, which are active, *etc*), previous questions must be remembered, and the set array must be selectively attended to. In order to separate these demands, the TQT procedure in the present study was adapted by introducing a “supported” questioning round and comparing it to a “baseline” round (where item elimination was permitted) and an “unsupported” round (where elimination was prohibited). In the supported round, item elimination was prohibited, so participants were required to manage an unchanging perceptual array (see figure 5). However, participants’ questioning was supported by



**Figure 5. The “supported” condition.** Participants’ questions are recorded and presented alongside the set during each trial, although physical elimination of items is prohibited.

recording each question they asked and displaying the record of questions alongside the game board during the trial.

In this way, the dual demand of remembering questions and managing the perceptual array becomes a single demand. If the difficulty with an unchanging array that ASD participants previously displayed related to verbal working memory differences, relatively intact performance would be expected in the supported condition. In contrast, if their difficulty was with ignoring the irrelevant information in the perceptual array, then this should still be evident even when their questions are recorded for them<sup>8</sup>. Based on the evidence for specific problems with selective attention, it was predicted that ASD participants would demonstrate difficulty on the Twenty Questions Task across both supported and unsupported conditions.

## **Experiment 1: Planning and Selective Attention on Twenty Questions**

### **Method**

#### *Participants*

A mixed clinical group of 22 ASD participants (21 male, 1 female) were recruited from families in the local community and a specialist school for children with autism spectrum disorders. All had a diagnosis of either autism ( $n = 21$ ) or Asperger syndrome ( $n = 1$ ) in accordance with ICD-10 research diagnostic criteria (World Health Organisation, 1993). All

---

<sup>8</sup> A potential confound here is in task difficulty and the role of multiple task demands. As the condition label indicates, the supported condition does not just work to minimise verbal working memory demands; it provides more help to participants in simplifying the task. In the unsupported condition, a dual demand is in place, and so relative decreases in performance on this part of the task could reflect general drop-offs with task difficulty and/or specific difficulties with the dual demand. Nevertheless, arguably information can be drawn from the relative level of performance in the supported condition. If ASD performance in the supported condition is closer to their baseline performance, this would suggest that their problems relate more to remembering questions than handling the perceptual array. If, instead, their performance is more similar to the unsupported condition, this suggests verbal working memory is not primarily at fault and that their difficulties lie in selectively attending to parts of the task array.

A full demonstration of what is causing difficulty in this area would ideally include a condition where no inhibition or selective attention is required – as in the baseline condition – but verbal working memory was still put under pressure. If participants were impaired in this condition, but not the supported condition, it would show that their difficulties with the unsupported condition are more likely to relate to working memory deficits rather than attentional deficits. However, if item elimination is permitted, the set itself provides a cue to which questions have already been asked, and it is not obvious how the task could be adapted to selectively measure verbal working memory without interrupting the game itself.

participants received their original diagnoses via contact with local clinical services, who use the Autism Diagnostic Interview – Revised (ADI-R; Lord, Rutter & Le Couteur, 1994) and Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000), among other diagnostic measures. Nine participants had also had their diagnosis confirmed within the previous two years by a trained university researcher using the ADI-R.

Exclusion criteria for the clinical group included any presence of conditions known to be comorbid and similar in symptomatology to ASD (such as Fragile X Syndrome or Rett’s Syndrome) or other neurological conditions which may be expected to affect cognitive functioning (e.g. Tourette Syndrome). In addition, children with specific and severe impairments in language processing separable from ASD symptoms (such as verbal dyspraxia, or a hearing difficulty) were not recruited. Although specific screening tools for dyslexia and ADHD were not included, no children included in the study were reported to have such difficulties by parents or teachers.

Twenty-one children (19 male, 2 female) were recruited from local mainstream schools for an age-matched typically-developing (TD) control group. Participants were matched to within 12 months of their clinical counterpart.

#### *Recruitment and settings*

Participants were recruited in two stages. Initially families were contacted in an information letter distributed by the local school. The letter invited families to take part in a new study on problem-solving in board games. On return of a reply slip and consent form, arrangements were made to test consented participants during school time. Following this a second wave of participants were recruited by information letters sent directly to families from the local ASD community who attended other schools. Families that responded to the letter were contacted by phone to further explain the study and provide the opportunity to ask questions, after which a testing session was arranged. The majority of testing sessions for second-wave participants took place at a laboratory at the university, although a minority were tested in participants’ homes (in cases where this was more convenient for families). In total 21 participants were tested in a school setting (12 ASD, 9 TD); the remaining participants were tested at the university or at home. Before testing, parents of participants were asked to provide full written consent for study participation and participants were reminded that they could withdraw at any time. All study procedures were approved by a university ethics committee prior to testing.

### *Cognitive profile*

Full-scale, verbal and non-verbal IQ profiles were acquired for participants in testing sessions at local schools and university settings. IQ scores were estimated using the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999). The Verbal IQ subscale of the WASI consists of two subtests, Vocabulary and Similarities, while the Non-Verbal or Performance IQ subscale consists of Block Design and Matrix Reasoning. Full-scale and Verbal IQ scores for all participants were based on WASI performance. Full WASI data was not available for all participants; for 10/21 TD participants the Similarities subtest was not run due to time constraints put in place by the participating school that they were recruited from; VIQ estimates in these participants were derived from their Vocabulary performance only.

For four ASD participants non-verbal IQ estimates were already available from a previous study using the Wechsler Intelligence Scale for Children – Third Edition (WISC-III; Wechsler, 1991). The non-verbal component of the WISC-III consists of five subtests: Picture Completion, Coding, Picture Arrangement, Block Design, Object Assembly, and Symbol Search. Non-verbal IQ scores for all remaining participants were calculated based on performance on the WASI non-verbal subscale. As non-verbal IQ estimates on the WISC-III and WASI have been reported to correlate highly in both clinical (Scott, Austin, & Reid, 2007) and non-clinical samples (Wechsler, 1999) the scores were treated as functionally equivalent. Clinical and control participants were matched to within 10 full-scale IQ points.

Table 1 shows the mean age and cognitive ability scores for each group. T-tests indicated no significant differences between the groups (age:  $t = 0.665$ ,  $p = .510$ ; FSIQ:  $t = -0.512$ ,  $p = .611$ ; VIQ:  $t = -0.112$ ,  $p = .911$ ; NVIQ:  $t = 0.639$ ,  $p = .526$ ; all  $df = 41$ , all Cohen's  $d < 0.02$ ). As a further check on the matching of the groups, the raw scores for IQ subtests were converted to age-equivalent scores and averaged to produce estimates of verbal and non-verbal mental age (VMA and NVMA respectively)<sup>9</sup>. The mean VMA for ASD ( $M (SD) = 14.11 (6.40)$ ) and TD ( $M (SD) = 15.03 (7.44)$ ) participants did not significantly differ ( $t = -0.433$ ,  $df = 41$ ,  $p = .667$ , n.s.). Similarly, no difference was observed in mean NVMA for ASD ( $M (SD) = 12.66 (4.60)$ ) and TD ( $M (SD) = 11.75 (3.29)$ ) participants ( $t = 0.745$ ,  $df = 41$ ,  $p = .460$ , n.s.).

---

<sup>9</sup> For a description of this procedure see Whitaker (2008).

**Table 1. Age and IQ scores for autism spectrum disorder (ASD) and typically-developing (TD) participants**

	ASD		TD		<i>Sig.</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Age (y:m)	13:8	1:5	13:4	1:5	<i>n.s.</i>
Full-scale IQ	94.82	21.49	98.10	20.45	<i>n.s.</i>
Verbal IQ	95.55	22.23	96.29	20.92	<i>n.s.</i>
Non-verbal IQ	94.23	19.50	98.29	22.08	<i>n.s.</i>

A subset of participants (10 ASD; 9 TD) had taken part in a study using the TQT that was run 18 months previously. These participants were marked for later analysis to assess any possible advantages of practice, although previous research on the TQT has documented minimal learning effects (Marshall & Karow, 2008; Marshall et al., 2003a).

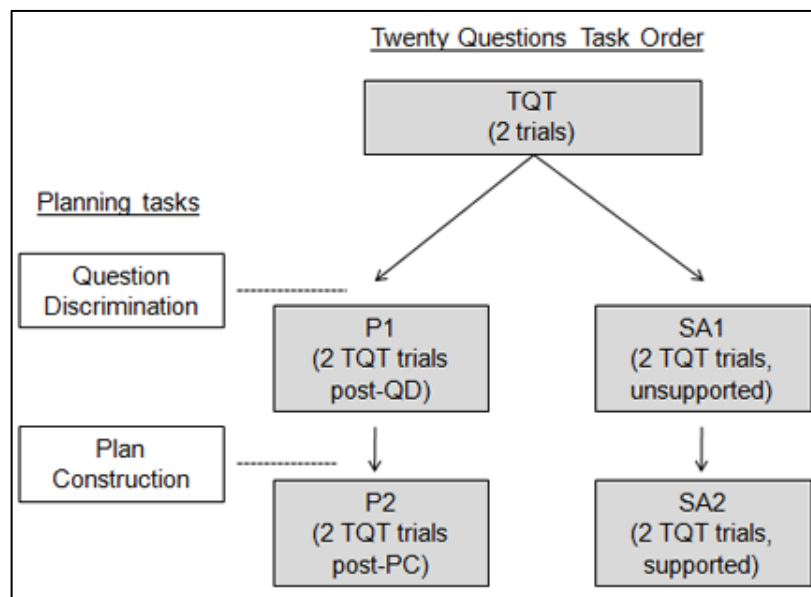
### *Design*

For practical reasons, each participant could only be tested in a single session. This influenced the design of the study, as the measurement of planning and attention skills in relation to the TQT could potentially change as a result of the order that participants attempted the tasks in. For example, measuring planning also tends to provide participants with examples and ideas for new questions, potentially obscuring underlying group differences.

To address this, a mixed 2x3 design was used to compare participant groups (ASD/TD) across three TQT variations: *baseline*, *planning* and *selective attention* (see figure 6). In the *baseline* TQT, participants completed two trials of a standard Twenty Questions task. This established any existing differences in TQT performance between the groups. In the *planning* variation, participants' planning skills were assessed by attempting two tasks, question discrimination (QD) and plan construction (PC). Following each task, participants completed two more standard TQT trials (2 post-QD, 2 post-PC), and this performance was compared to the baseline group scores.

The *selective attention* variation consisted of two TQT trials where elimination was prohibited (referred to as the *unsupported* condition) and two similar trials where participants' questions were recorded during search (the *supported* condition). Task performance on these variations was compared to the baseline version, where elimination was allowed.

In total participants attempted 10 TQT trials (2 baseline, 4 planning, 4 selective attention). Following the baseline condition, the order of variations (planning/selective attention) and their subcomponents (question discrimination/plan construction/unsupported/supported) was counterbalanced across participants in both groups. Potential order effects were assessed using repeated measures analyses of variance.



**Figure 6. Task and condition order for experiment 1.** TQT = Twenty Questions Task; QD = Question Discrimination; PC = Plan Construction; P1, P2 = Planning conditions 1 & 2; SA1, SA2 = Selective Attention conditions 1 & 2.



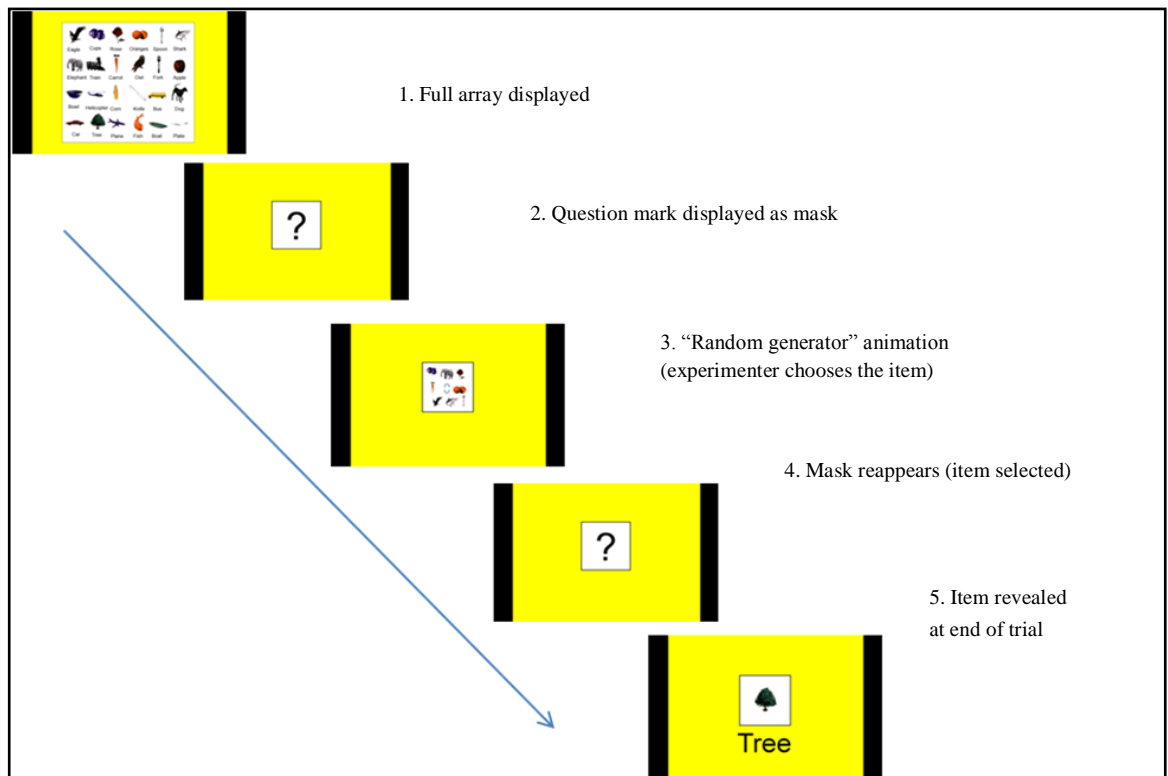
## *Materials & Procedure*

### *- The Twenty Questions Task (TQT)*

The Twenty Questions set used consisted of 24 everyday items that can be organised into roughly equal conventional categories such as animals, plants, vehicles, and household objects (Delis et al., 2001; Laine & Butters, 1982; Mosher & Hornsby, 1966). Previous studies that have used the Twenty Questions procedure have typically tested participants on three to four trials, whereas in the present study a testing session consisted of 10 trials. This created the potential for practice effects occurring and the specific possibility of participants identifying maximal questions which could be used on every trial. To avoid this, the set contents were varied for each condition. Three different sets were devised and rotated across the three testing conditions. For example, participant A was presented with sets 1, 2 and 3, participant B attempted sets 2, 3 and 1, and participant C completed sets 3, 1 and 2.

The sets were devised in a way that would retain a roughly conventional structure, as in the standard task. To maintain a minimal representation of conventional groupings in each set, eight “core” items were selected first, based on their use in previous versions of the Twenty Questions procedure (e.g. Delis et al., 2001; Mosher & Hornsby, 1966). The eight core items were DOG, COW, CAR, PLANE, KNIFE, FORK, APPLE and ORANGE. In addition to the core items, 16 items were then randomly selected from a list of 46 everyday items. The additional items in the list of 46 were collected based on their common frequency and, in the majority of cases, because they had appeared in at least one version of the Twenty Questions procedure in previous research. The make-up of this list was equally apportioned to represent living and non-living items, and, within those categories, sub-groupings of animals, plants, vehicles and household items. However, the random selection of items from the list would not guarantee their equal representation in the three 24-item testing sets, providing each set with a subtly unique category structure. Table 2 lists the items used in each set in addition to the “core” of those items listed above.

Three 30cm x 30cm plastic game boards were used for presentation of each set. Each board contains 24 hinged frames which can be used to physically eliminate redundant set items during search. Each item was displayed in colour printed on white card, with its name printed at the bottom of the card in black ink. Coloured plastic tokens were used to keep count of the number of questions used by participants and the number of trials they had successfully completed. Participants were given 10 question counters at the start of each round, denoting the maximum number of questions they could ask. As with the set contents,



**Figure 7. Procedure of visual feedback on game trials.** The “random generator” animation was deployed alongside the Twenty Questions Task, to avoid participants having to guess what the experimenter was thinking of in order to find the answer.

**Table 2. Additional items used in each testing set.** Sixteen items were added to the eight “core” items (dog, cow, car, plane, knife, fork, apple and orange).

Set 1	Set 2	Set 3
ELEPHANT	OWL	FISH
ROSE	CORN	TREE
CUPS	PLATE	OVEN
LORRY	HELICOPTER	LORRY
POTATO	MOTORBIKE	TRAIN
TANK	GRAPES	BUS
HORSE	CUCUMBER	HORSE
SHOES	PINEAPPLE	CAT
CARROT	BICYCLE	ONION
FRIDGE	SHEEP	GRAPES
GRAPES	SUBMARINE	COMPUTER
OVEN	BANANA	SHARK
BOWL	TRAIN	GLOVES
PLATE	ROSE	BOWL
TV	CUPS	TROUSERS
HAMMER	BOAT	PIG

the identity of target items for each trial was determined randomly prior to testing. Within each set, the identity of the target items was always the same. (Set 1: CAR, POTATO, LORRY, ELEPHANT; Set 2: SHEEP, FORK, PLATE, ORANGES; Set 3: BUS, TROUSERS, GLOVES, OVEN).

Prior to the first trial, the following instructions were given to participants:

“Today we are going to play a game which involves looking for objects. As you can see, all of these cards have a different object on them. I am going to use the computer to choose one of them, and your job is to try and find out which one it is. You can do this by asking me questions about them, but I can only answer yes or no to your questions. Every time I answer a question, you can push down all the ones you don’t need on the board, like this. The aim of the game is to find the object in as few questions as possible. Are you ready to start? [If appropriate] Would you like a practice go?”<sup>10</sup>

Participant questions were recorded using *Audacity*, an Open Source sound editing software programme (<http://audacity.sourceforge.net/>). The audio-visual feedback used was the same as in Alderson-Day & McGonigle-Chalmers (2011). A 17” Dell Inspiron Laptop running Microsoft PowerPoint was used to provide the visual and auditory feedback. Figure 7 displays the procedure of visual feedback presented to participants during each trial. This included a “random generator” animation, which was run prior to each trial when the experimenter was choosing the target item, and a short fanfare, applause or game-show organ audio clip following successful trials. Unsuccessful trials were followed by the revelation of the item, but without sound.

#### - *Planning task 1: Question Discrimination (QD)*

The question discrimination task consisted of a forced-choice format in which participants were presented with two possible questions (e.g. Is it an animal? Is it food?) and an example array from the Twenty Questions game. Participants were asked to identify which of the two questions would be better to ask first. Each example array was designed so that one question

---

<sup>10</sup> For participants who appeared to have difficulty in understanding the task instructions, a practice round was offered. The practice trial consisted of only eight items. If the target was identified by the participant in no more than eight questions then the full task was begun. If participants still did not appear to understand how to play at this point (if, for example, they had not asked any questions or only asked irrelevant questions) then the experimenter would take on the role of participant and model an average questioning strategy (i.e. asking questions of medium efficiency). Participants who required this step were denoted in later analysis and their responses examined for any apparent modelling effects (Denney, 1975).

could clearly eliminate more items from the array, irrespective of the answer. Figure 8 displays an example array and the two question options that would be presented with it. The question discrimination task consisted of 10 forced-choice trials, which are listed in table 3.

Each trial was presented on a white A4 sheet, containing the example array at the top and the two question options printed below. The trials were organised in increasing difficulty, in two respects. Firstly, trials 1-5 included arrays of only 12 items, i.e. half the size of arrays used in the Twenty Questions task. Trials 6-10 included full-size arrays. Secondly, the difference in scope of each question was designed to be large in early trials, but much closer in later trials. For example, trial 1, as displayed in figure 8, compared the question “Is it an animal?” (referring to six items and being guaranteed to at least eliminate half of the set) with the question “Is it a fruit?” (referring to only two items and potentially eliminating only two of the 12 items available). In contrast, trial 9 presented a choice between a question which potentially only eliminated 5/24 items (“Does it have wheels?”) and a question which could only be guaranteed to eliminate 3/24 items (“Is it cutlery?”). Question 10 was of specific difficulty as it presented a full category overlap, where one question actually picked out a subset of another (“Is it a living thing?” vs “Is it an animal?”).













The specific instructions given to participants were as follows:

*“Now, before we play the next game we’re going to look at some examples of different situations from the game. For each example I’m going to show you two different questions that you might ask to find the item. All I want you to do is to tell me which one you think would be best to ask **first**.”*

A point was granted for each correct question selection. On completion of the discrimination task, participants were tested on two trials of the Twenty Questions task (following the procedure described above).

#### *- Planning task 2: Plan Construction (PC)*

The *plan construction* task was loosely based on a measure deployed by Siegler (1977) where participants were encouraged to plan their questions prior to attempting a TQT-style task. Participants were presented with an array of 32 common questions and were asked to select five of them to use when playing Twenty Questions. (The TQT board was on display, showing the distribution of items to be searched). Participants were then asked to put these questions in order, creating a rough plan that could be used to search the set. There was no

 Spoon	 Cow	 Oranges	 Boat
 Apple	 Dog	 Pig	 Computer
 Car	 Fish	 Cat	 Elephant
IS IT AN ANIMAL?		IS IT A FRUIT?	

**Figure 8. Example trial from question discrimination task.** Participants were asked to pick which of the two questions would be best to ask first.

**Table 3. Trials in the question discrimination task.** Questions were designed to increase in difficulty.

1. Animal or Fruit?	6. Living thing or Transport?
2. Food or Cutlery?	7. Kitchen or Zoo?
3. Wheels or Fly?	8. Fruit or Manmade?
4. Living thing or Fly?	9. Wheels or Cutlery?
5. Swim or Wheels?	10. Living thing or Animal?

time limit for question selection. Questions were presented in size 72 font on white 5cm x 10cm cards. The following specific instructions were given to participants:

“Before we play the next game we’re going to look at some different questions. Most people ask about five questions when they are playing the game. Here are some questions you might use to look for items in this set (testing set indicated). I want you to pick five questions that you think might be useful in the next game. Once you have got five questions, put them in order (a bit like a plan), so choose which one you would ask first, which one next (and so on).”

Once five questions had been chosen, the experimenter read them out loud back to the participant.

Following completion of the planning task participants attempted two more trials of Twenty Questions. Participants were allowed to keep their prepared questions in front of them as a guide (as in Siegler, 1977) although they were informed that they “didn’t have to use those questions and could ask completely different ones”.

#### - *Selective Attention*

The effect of selective attention abilities was assessed by varying the level of perceptual interference facing participants and the level of verbal support given to them during search. These variations were named SA1, the *unsupported* condition, and SA2, the *supported* condition.

SA1. The unsupported version of the task was like the standard Twenty Questions game in all but one respect. During search, participants were not permitted to physically eliminate items on the game board. Two trials were run on the Twenty Questions game using this format.

SA2. During the supported version participants were again not permitted to physically eliminate items during their search. In this way, the level of perceptual interference present in the set was the same as in the previous (unsupported condition). However, participants were also given the additional support of a record of their questions being kept by the experimenter. Every time a question was asked, the experimenter would note the question and place it next to the game board. Questions were written on blank white cards by the experimenter during the trial. Participants completed two trials of Twenty Questions on this version of the task.

## *Scoring*

### *- Twenty Questions Task*

Performance on the TQT was scored in terms of question efficiency and question types. Efficiency on the TQT was primarily assessed via *question quality* (Alderson-Day & McGonigle-Chalmers, 2011). Question quality (QQ) is an efficiency measure indicating the proportion of items guaranteed to be eliminated by any question. It is defined as the minimum number of items eliminated by a question, divided by the total number of items in a set. For example, the question “Is it a fruit?”, when asked about a set containing 3 fruits and 7 non-fruits, would receive a score of 0.33 (3 items at least will be eliminated/ 10 items available in set = 0.33). As no question can be guaranteed to eliminate more than 50% of the available items, a maximum score for any question is 0.5. A minimum score would be 0, if the question used referred to none of the items left in the set. However, a more common low score response would be a question referring to a single item in the whole set. (For example, “Is it the cat?”, in a set of 24 items, could only guarantee the elimination of 1 item. The QQ for this question would be  $1/24 = 0.04$ ).

Measurement of performance using QQ is preferable to other measures because it is sufficiently variable across individuals but also robust across task trials. Other measures - such as the numbers of trials completed or questions used per trial - can often be skewed by lucky guesses and ceiling or floor effects, making them less ideal as outcome measures. However, the number of questions per trial (QT) that a participant uses can sometimes be sensitive to variations in memory and attention demands (see Alderson-Day & McGonigle-Chalmers, 2011). To allow for this, QT was also recorded to provide a secondary measure of efficiency. Other TQT variables typically measured are the relative frequencies of grouping questions and guesses used. To allow for comparisons with previous TQT studies (Minshew, et al., 2002; Minshew, et al., 1994), these question types were also noted.

### *- Planning tasks*

Performance on the question discrimination task was assessed as a score out of 10, reflecting the number of good questions correctly identified per participant. Plan construction was assessed by assigning an “expected” QQ score to each question selected, based on the average number of items they eliminated in the testing sets. For example, the question “Is it a vehicle?” would refer to four items from set 1, eight from set 2 and five from set 3. Across

the three sets, the mean number of items that this question could be expected to eliminate is therefore 5.66 out of 24, producing an expected QQ score of  $5.66/24 = 0.24$ . For a 5-question plan, this created a sequence of five expected QQ scores for each participant, indicating how many items they would usually eliminate with their first question, next question and so on. Plans where the expected QQ progressively reduced with each question indicated a narrowing strategy, with more general questions being asked to begin with and more specific questions coming later in the plan.

### *Analysis*

All analyses were conducted using SPSS v16.0 statistical software. Preliminary analysis included normality testing for all outcome variables and assumptions testing appropriate for analyses of variance and covariance. Mean scores across different set and task combinations were assessed for order effects using a MANOVA containing the following dependent variables: question quality, percentage of grouping questions, and percentage of guess questions.

TQT performance outcomes were assessed using analyses of covariance, with age and VIQ included as covariates, although guessing rates were also analysed non-parametrically because of skew. Age and VIQ were included as covariates because they had been previously observed to significantly contribute to TQT performance (Alderson-Day & McGonigle-Chalmers, 2011). Alongside group matching, the use of these variables as covariates provided an additional measure of their influence, and provided information on the extent to which they affected each group equally.

The question discrimination component of the planning tasks was also assessed with an ANCOVA. Due to non-normal data, plan construction scores were assessed using Two-Sample Kolmogorov-Smirnov tests (between-groups) and Friedman's ANOVA (within-groups).  $\eta_p^2$  values indicate partial eta squared effect sizes. Unless otherwise noted, pairwise comparisons are presented with uncorrected  $p$ -values.

### **Results**

No significant main effects were observed for any of the variations in task or set order (all  $p > .05$ ; all  $\eta_p^2 < .01$ ) so scores across different presentation orders were combined.



Comparison of mean QQ scores for participants who had ( $n = 19$ ) and had not ( $n = 24$ ) attempted the TQT before also showed no significant differences,  $F(1, 39) = 0.456$ ,  $p = .504$ ,  $\eta^2_p = .012$ .

### *Twenty Questions Task (TQT)*

ASD participants recorded significantly lower QQ scores than TD participants on the TQT at baseline,  $F(1, 38) = 7.246$ ,  $p = .01$  (see table 4). This result was supported by the second efficiency measure, questions-per-trial (QT). ASD participants took significantly more questions to solve their baseline trials than controls ( $ASD_{M(SD)} = 6.16 (2.07)$ ,  $TD_M = 5.29 (0.90)$ ,  $F(1, 39) = 4.314$ ,  $p = .044$ ). The main effect of group was observed despite controlling for significant effects of age,  $F(1, 38) = 7.615$ ,  $p = .001$ , and verbal IQ,  $F(1, 38) = 8.571$ ,  $p = .001$ , suggesting a significant difference in the questioning efficiency of each group<sup>11</sup>. These analyses were also run using VMA as a covariate, which produced very similar results (e.g. question quality: VMA  $F(1, 39) = 8.205$ ,  $p = .007$ ; group  $F(1, 39) = 6.287$ ,  $p = .001$ ).

In the analysis of question types, ASD participants asked a significantly lower percentage of grouping questions ( $ASD_{M(SD)} = 43.75\% (30.10)$ ,  $TD_M = 68.77\% (11.44)$ ,  $F(1, 38) = 14.369$ ,  $p = .001$ ) and used guesses more ( $ASD_{MEDIAN} = 20.19\%$ ,  $TD_{MEDIAN} = 15.38\%$ ,  $Z = 1.334$ ,  $df = 42$ ,  $p = .026$ ). However, these differences did not fully explain the observed group differences in efficiency: even when they were using grouping questions, ASD participants asked questions that eliminated fewer items ( $ASD_M = 0.22 (0.14)$ ,  $TD_M = 0.29 (0.6)$ ,  $F(1, 38) = 5.406$ ,  $p = .026$ ).

### *Planning*

#### *- Question discrimination*

Performance on the question discrimination task was high in both groups, as indicated in table 4. No significant differences were observed between the two groups,  $F(1, 39) = 0.213$ ,

---

<sup>11</sup> A marginally significant group-by-VIQ interaction effect was also observed,  $F(1, 38) = 4.236$ ,  $p = .046$ , indicating that VIQ may have influenced group scores unevenly. As this violated one of the assumptions of analysis of covariance (homogeneity of regression slopes), the analysis was rerun with VIQ removed, but a significant group effect remained,  $F(1, 40) = 7.922$ ,  $p = .008$ . Uneven influences of verbal IQ were also observed in Alderson-Day & McGonigle-Chalmers (2011).

**Table 4. Group performance on the Twenty Questions Task, Question Discrimination (QD) and Plan Construction (PC) Tasks. QQ = question quality.**

Twenty Questions (QQ)		ASD		TD		Sig.
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Baseline		0.24	0.14	0.32	0.04	0.01*
Planning	1 (Following QD)	0.24	0.12	0.37	0.05	<.001***
	2 (Following PC)	0.25	0.11	0.36	0.06	<.001***
Attention	1 (Unsupported)	0.23	0.11	0.32	0.06	<.001***
	2 (Supported)	0.25	0.12	0.32	0.06	0.003**
Planning		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>Sig.</i>
Question Discrimination (Score/10)		7.27	1.61	7.43	1.78	0.647
Plan Construction (Mean QQ over 5 questions)		0.19	0.07	0.25	0.05	0.006**

\*  $p < .05$       \*\*  $p < .01$       \*\*\*  $p < .001$

$p = .647$ ,  $\eta^2_p = .005$ , suggesting that both groups could identify effective questions.

Covariate contributions of VIQ,  $F(1, 39) = 4.982$ ,  $p = .031$ , and age,  $F(1, 39) = 3.057$ ,  $p = .088$ , were also evident on this outcome, with better question discrimination associated with higher VIQ ( $r = .300$ ) and age ( $r = .213$ ).

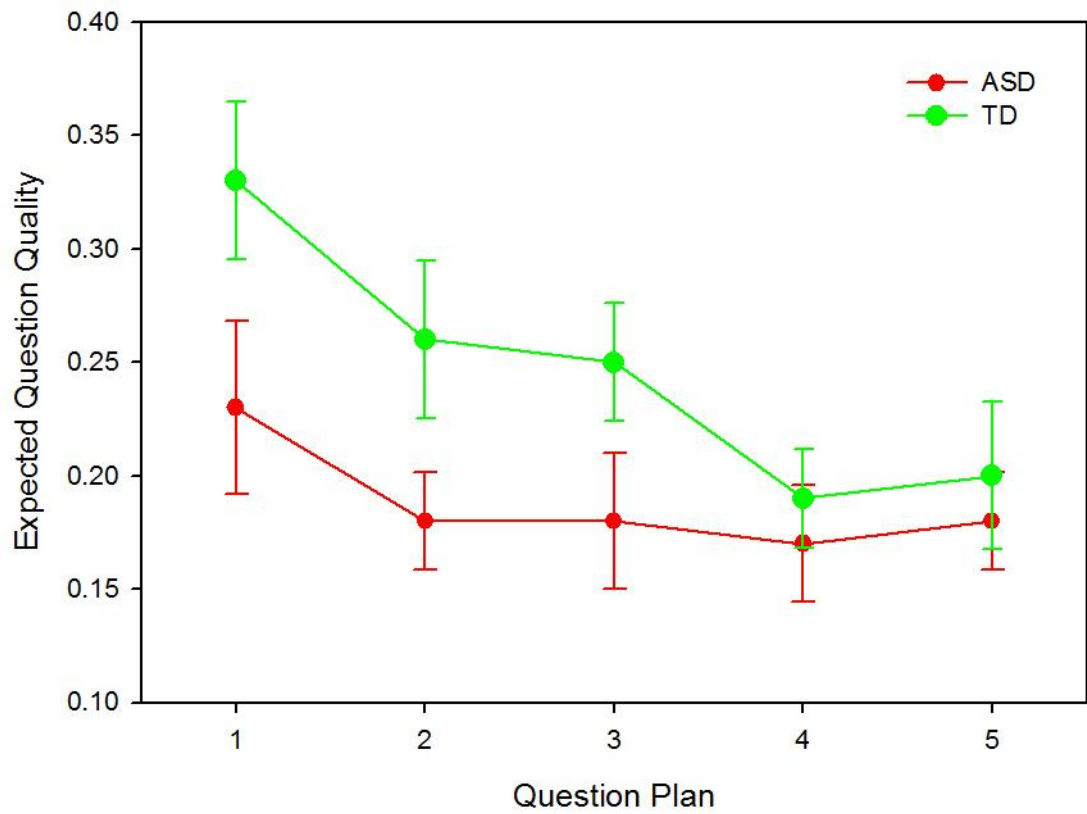
#### *- Plan construction*

Participants' plans were analysed using separate Friedman's ANOVAs for each group. Whereas TD participants selected questions with significantly different average QQ scores,  $X^2 = 11.520$ ,  $df = 4$ ,  $N = 21$ ,  $p = .021$ , the questions selected by ASD participants did not differ in question quality,  $X^2 = 1.630$ ,  $df = 4$ ,  $N = 22$ ,  $p = .803$ . Figure 9 shows the difference in plan construction, with TD participants progressively narrowing the scope of their questions and ASD participants asking a more even set of questions. Two-Sample Kolmogorov-Smirnov tests indicated that the two groups differed in their average efficiency across the whole plan ( $ASD_{MEDIAN} = 0.19$ ,  $TD_{MEDIAN} = 0.24$ ,  $Z = 1.611$ ,  $p = .006$ ; means are presented in table 4). This difference was most evident on the first three questions selected (all  $Z > 1.121$ , all  $p < .072$ ), indicating that TD participants were selecting more general questions than ASD participants for the start of their plans.

#### *- Relations of planning to TQT performance*

No significant correlations were observed between scores on the question discrimination task and subsequent QQ performance on two trials of the TQT, for each group or both groups combined ( $N = 43$ ,  $r = .141$ ,  $p = .367$ ). On the plan construction task, questions scores for questions 1 ( $r = .415$ ,  $p = .006$ ), 2 ( $r = .399$ ,  $p = .008$ ) and 3 ( $r = .537$ ,  $p = .001$ , all  $N = 43$ ) significantly correlated with QQ scores on the TQT, indicating that better question selection during planning was associated with proficiency during Twenty Questions. This was also reflected when correlations were calculated separately for each group (questions 1-3: ASD range,  $r = .322-.619$ ; TD range,  $r = .241-.440$ ).

Although question selection during planning and TQT performance was related, the act of planning did not subsequently improve question quality scores for either group (all  $F < 1.158$ , all  $p > .3$ , all  $\eta^2_p < .03$ ). However, ASD participants were more likely to use grouping questions after the planning activities, as indicated by a significant group-by-



**Figure 9. Efficiency of plans constructed by autism spectrum disorder (ASD) and typically-developing (TD) participants. Plans were scored in terms of expected question quality.**

condition interaction effect for this variable ( $ASD_{baseline} = 43.75$ ;  $ASD_p = 47.90$ ,  $F(2, 76) = 3.210$ ,  $p = .046$ ).

### *Selective Attention*

ASD participants recorded lower QQ scores than TD participants during the selective attention variations of the TQT,  $F(1, 38) = 5.511$ ,  $p = .024$ , but no group\*condition effects were observed on this outcome when scores were compared across conditions. This suggested that the group difference in efficiency seen on the baseline trials did not significantly change with the variations in attentional load (all  $F < 1.650$ , all  $p > .3$ , all  $\eta^2_p < .05$ ). Main effects of VIQ,  $F(1, 39) = 17.382$ ,  $p < .001$ , and age,  $F(1, 39) = 13.513$ ,  $p = .014$ , were also observed across the selective attention conditions.

A trend was observed for the interaction effect of group\* condition in the number of questions used per trial (QT),  $F(2, 78) = 2.787$ ,  $p = .068$ ,  $\eta^2_p = .067$ . ASD participants tended to require slightly more questions during the unsupported condition in particular ( $ASD_{change} = +0.95$ ,  $TD_{change} = -0.04$ ). Performance on the supported condition, contrastingly, was very similar to performance on the standard TQT for ASD participants (Baseline QQ = 0.24, Supported QQ = 0.25; Baseline QT = 6.16, Supported QT = 6.07). This suggests that, where such a difficulty was evident in ASD participants, it was only observable when questions were not being recorded, rather than when irrelevant items were present in the array.

### *Post-hoc analysis: reading abilities and VIQ profile*

Both the question discrimination and plan construction tasks required participants to read a number of written questions. To check that any group differences were not simply reflecting reading abilities, participants were subsequently tested on their reading of the 32 question words in a follow-up session. Data from all but six participants (3 ASD, 3 TD) were available in this way. No significant differences in reading scores were apparent ( $ASD_{MEDIAN} = 32$ ;  $TD_{MEDIAN} = 32$ ;  $Z = 0.480$ ,  $df = 36$ ,  $p = .366$ ). Furthermore, reanalysis of the planning data using only those participants who scored 32/32 on the reading test (14 ASD/16 TD) indicated that the differences in planning were still evident between groups: TD participants narrowed their questions from general to more specific ( $QQ1-5_{MEANS} = 0.36$ ,

0.29, 0.27, 0.17, 0.20;  $X_2 = 14.214$ ,  $df = 4$ ,  $N = 16$ ,  $p = .007$ ), whereas ASD participants did not (QQ1-5<sub>MEANS</sub> = 0.29, 0.19, 0.20, 0.19, 0.22;  $X_2 = 2.672$ ,  $df = 4$ ,  $N = 14$ ,  $p = .614$ ).

The functioning range for the ASD sample in the study was also relatively large (VIQ<sub>range</sub> = 65-141) and during testing it became apparent that some ASD participants were much more able to recognise and select good questions on the planning tasks than others. To investigate this further, both groups were divided into two: those above and below the mean VIQ score (95.91). This produced a lower-VIQ ASD group ( $n = 12$ ) and a higher-VIQ ASD group ( $n = 10$ ) and corresponding lower-VIQ ( $n = 11$ ) and higher-VIQ ( $n = 10$ ) TD groups.

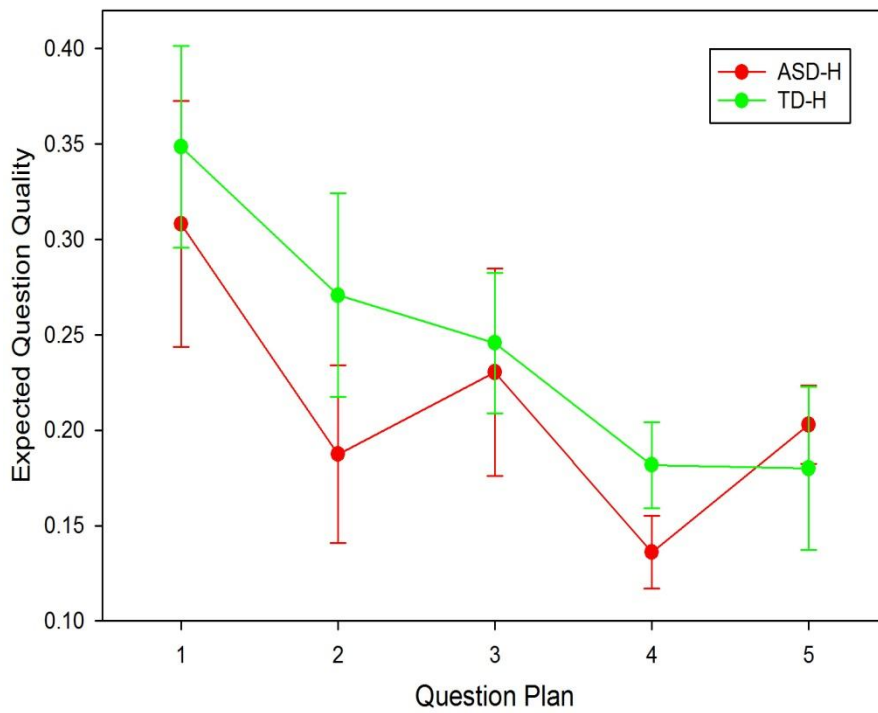
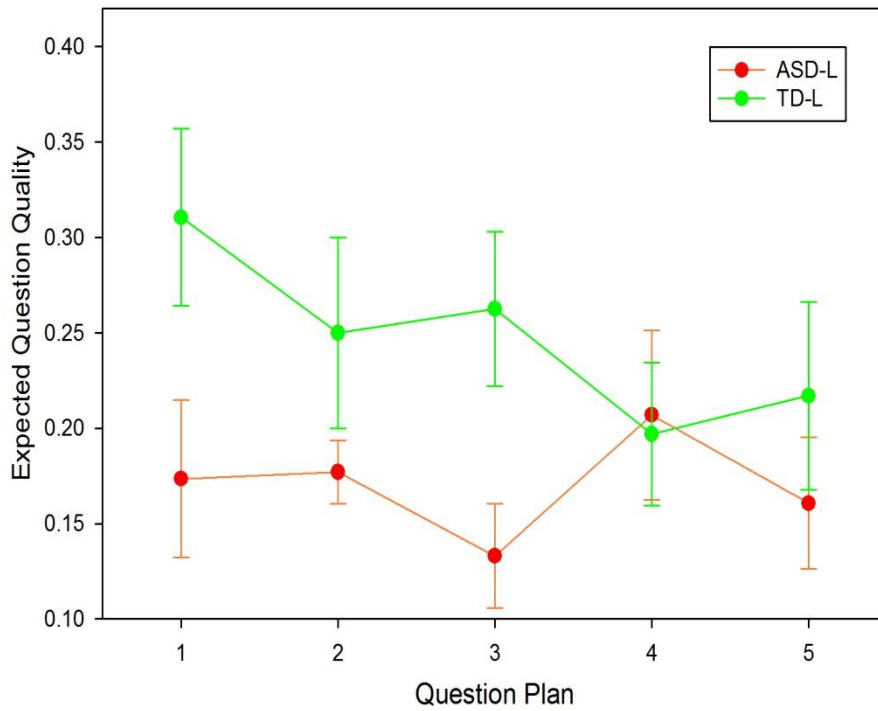
On question discrimination, a 4-way univariate analysis of variance comparing the four groups still indicated no significant difference in accuracy ( $F(1, 38) = 1.220$ ,  $p = .316$ ,  $\eta^2_p = .088$ , n.s.); suggesting that even lower-VIQ ASD participants were able to identify good questions in a forced choice format.

More noticeable differences were evident, though, on the plan construction task. As figure 10 shows, the plans of lower-VIQ ASD participants had relatively low efficiency (as indicated by their average question quality) and generally did not move from asking more general to more specific questions as the plan progressed (ASD-L Mean change from Q1 to Q5 = 0.01). In contrast, higher-VIQ ASD participants produced plans that did appear to show evidence of narrowing, albeit with slightly less efficient questions than their higher-VIQ TD counterparts (ASD-H Mean change from Q1 to Q5 = 0.11).

A mixed ANOVA comparing groups x questions (1-5) produced a significant group effect ( $F(3, 39) = 3.913$ ,  $p = .016$ ,  $\eta^2_p = .231$ ), highlighting the overall difference in efficiency between the four groups. Subsequent pairwise comparisons indicated that the lower-VIQ ASD group recorded significantly lower scores than both TD groups (L-ASD < L-TD,  $p = .028$ ; L-ASD < H-TD,  $p = .042$ ) but not the higher-VIQ ASD group (L-ASD = H-ASD,  $p = .682$ , n.s.). This suggests that, although the two ASD groups could be separated in their tendency to construct a narrowing plan, they do not significantly differ in their problems with efficiency<sup>12</sup>.

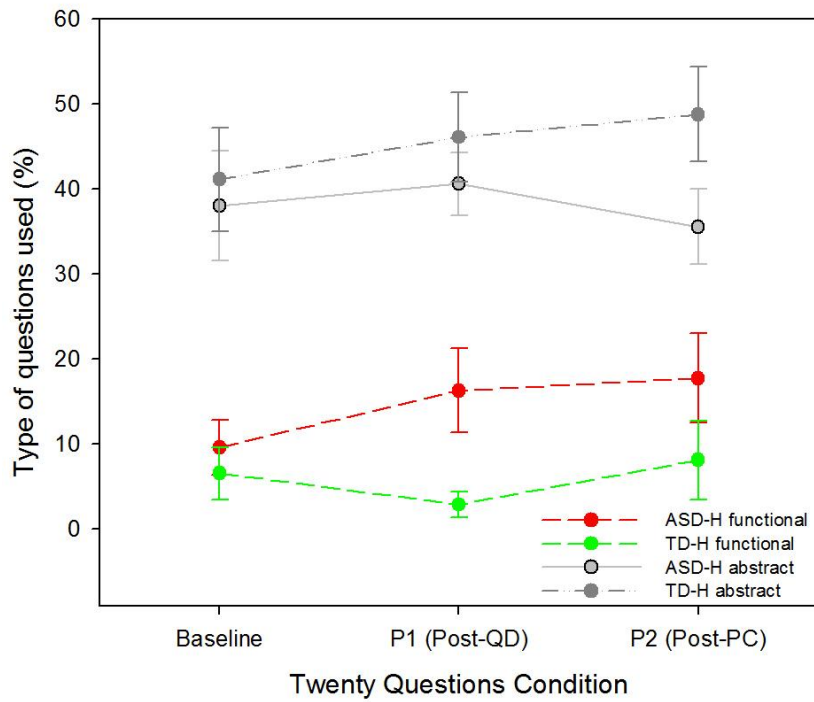
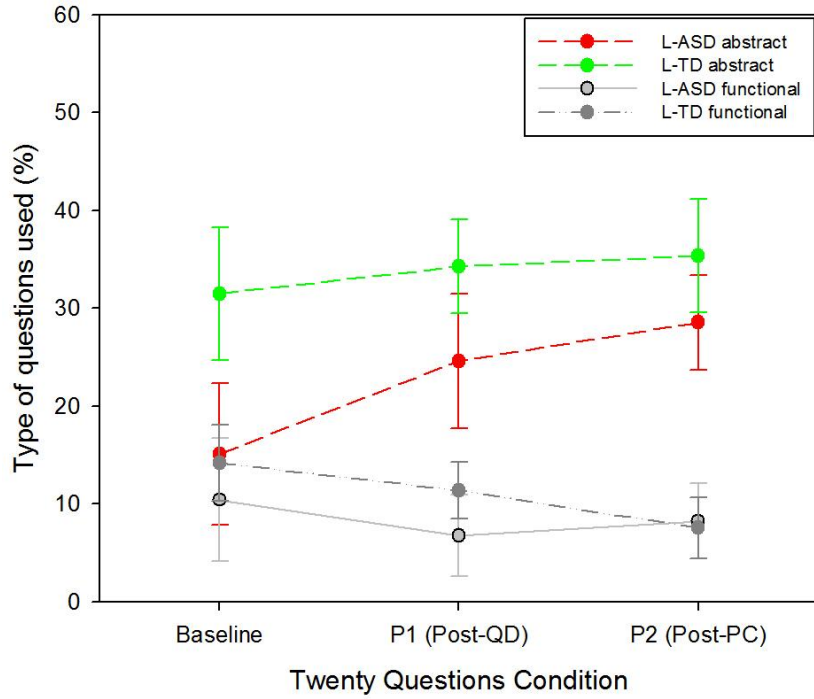
---

<sup>12</sup> As a post-hoc set of analyses the comparison of these subgroups suffers to a certain extent from a lack of power, having two consequences. First, although a Friedman's ANOVA was used in the initial analysis for this variable, the reduction of power precluded the running of individual Friedman analyses within each group. As such, a parametric analysis was run here to provide indicators of potential subgroup differences, but it should be interpreted with caution. Second, the pattern of pairwise comparisons suggests that higher VIQ ASD participants are still less efficient than their counterparts, but this is only evident in the comparison of each group to the lower-VIQ ASD group



**Figure 10. Efficiency of plans constructed in lower VIQ (top) and higher VIQ (bottom) groups.**

(which is being treated effectively as a baseline) – a proper demonstration of this effect would also show pairwise H-ASD vs H-TD differences, but this is not evident with the power available.



**Figure 11. Changes in use of abstract and functional questions following planning in lower VIQ (top) higher-VIQ (bottom) participants. (Grey lines show no significant group differences).**



The suggestion of important differences between higher and lower VIQ groups was supported by a further subgroup analysis of how participants responded to the planning activities when subsequently attempting the TQT. As reported above, few significant changes in performance were evident on an overall group level, barring some improvement in the use of grouping questions in the ASD group. Analysis of the *content* of participants' questions revealed further changes, but only in specific ASD subgroups. On the Twenty Questions Task the content of questions can be coded into queries that refer to *abstract* category names (e.g. "animal", "vegetable", "vehicle"), those that refer to *function* (e.g. "is it used to cook things?") and those that refer to individual *perceptual* or *conceptual* features (e.g. "Does it have wheels?" or "Does it use electricity?") (Alderson-Day & McGonigle-Chalmers, 2011).

After planning, ASD participants in the lower group asked more abstract questions ( $F(2, 34) = 4.613, p = .017$ ); while ASD participants with higher VIQ asked more questions about function ( $F(1.545, 30) = 4.769, p = .048$ ) as indicated by significant group\*round interaction effects on these outcomes (see figures 11a and 11b for the changes in low and high groups respectively).

## **Discussion**

The data supported the hypothesis that ASD participants have difficulty planning their questions during verbal problem-solving. Although participants in the ASD group showed no difficulty in identifying effective questions via the question discrimination task, their question selection during planning was less efficient than question selection in the typically-developing group. Furthermore, this correlated with the questions they used during the task itself, especially for the initial questions used in their plan.

A trend was also observed relating to the attentional demands of the task. The direction of the trend suggested that ASD participants were likely to increase the number of questions that they used when items could not be physically removed from the set. Although this effect was not significant, it was consistent with the observation made by Alderson-Day & McGonigle-Chalmers (2011) that ASD participants require more questions when elimination is prohibited. There was, however, no evidence for this sort of change if participants were provided with a written record of their questions during the task. As the item array did not change across these conditions, this suggests that ASD participants did not use more questions because of a problem with selectively attending to relevant information in the

array. Rather, any difficulty with these particular task conditions appeared to reflect problems with tracking one's own questions without additional support. Such an interpretation should be made with caution, given the marginal nature of this difference, but if correct it would implicate difficulties with verbal working memory, not selective attention. Before discussing how these findings fit with the existing literature on executive functioning in autism, a few points must be considered concerning the methodology of the study.

### *Caveats and Limitations*

First, the participant age range was relatively wide (9:3 – 16:0) and covers a period in which problem-solving abilities may change considerably (Drumm & Jackson, 1996; Yurgelun-Todd, 2007). By age 11, most typically-developing children would be expected to use a variety of abstract categories to effectively interrogate a Twenty Questions set (Drumm, Jackson, & Magley, 1995; Mosher & Hornsby, 1966). The main effects of age observed in the present study point towards its influence on both the TQT and related planning processes. However, the specific developmental trajectory of problem-solving skills in autism likely requires further examination. The results of some studies have pointed to impairments in both concept identification *and* formation in younger children with autism (Solomon, Bauminger, & Rogers, 2010) but only more specific problems with concept use in older children and adults (Minshew et al., 2002), suggesting that some components of problem-solving ability may “catch-up” in ASD individuals while others remain problematic. It may be that the difficulties seen on the TQT represent a delay in the development of specific abstract skills, or, alternatively, a divergence as the skills of neurotypical children advance. To better understand how and when specific problem-solving abilities develop in children with ASD, tighter age ranges and comparisons between different age groups would be required.

A second limitation concerns the reading demands of the planning tasks used. In both of the planning tasks, participants were required to read a number of different questions. As participants were not specifically matched on reading ability or screened for specific difficulties with reading, problems in planning could instead reflect poorer reading skills in the ASD group. There are good reasons to think this is not the case; the main reason being that the majority of category words used in the plan construction task were also used on the question discrimination task. If reading was a problem on one planning task, group differences should also have been evident on the other. Nevertheless, future research could

utilise a more extensive standardised reading measure (such as the WTAR; Glutting & Wilkinson, 2005) to clarify the contribution of reading ability to question selection.

A final point to consider concerns the wider executive profile of the participants who took part in the study. As already discussed, executive difficulties in ASD individuals can sometimes overlap with those seen in other developmental disorders, particularly ADHD (Goldberg, et al., 2005; Sinzig, et al., 2008; Yerys et al., 2009). If any of the participants here had additional ADHD-like difficulties, such as a problem with inhibition, then that could also have affected their overall performance on the TQT. As the present study did not characterise ASD participants in terms of their possible ADHD profile, this cannot be ruled out, although two points can be made that begin to answer such a concern. First, very few data exists on the TQT profile for ADHD children, but studies on children with hyperactivity problems (Tant & Douglas, 1982) and adults with attention and inhibition difficulties (Upton & Thompson, 1999) suggest that guessing rates associated with this profile tend to be much higher than those observed in the present study. Second, as noted in the participant information, no concerns about possible ADHD were raised by parents or teachers in the initial recruitment of participants. Thus, though it remains a possibility, it seems unlikely that these results can be explained by underlying, ADHD-like difficulties in the ASD group.

#### *Executive dysfunction and problem-solving in ASD*

Despite these caveats, the above findings are broadly consistent with existing understanding of executive skills in autism. Although selective attention problems did not appear to influence TQT performance, the data here implicated the presence of possible verbal working memory difficulties in the ability of ASD participants to track their own questions and avoiding repetition. Anecdotally, it is the impression of the author that individuals of all ages and participant groups often use explicit verbalising strategies when elimination is prohibited. For example, they may say the following when attending to the set:

“OK, so it *is* a living thing, and it *is* an animal, but it *is not* a pet...is it the elephant?”

Although not always observed (Lopez, et al., 2005; D. L. Williams, Goldstein, & Minshew, 2005), problems with verbal working memory and difficulties using verbal mediation to support working memory have been reported in a number of studies with ASD individuals (Bennetto, et al., 1996; Joseph, Steele, Meyer, & Tager-Flusberg, 2005; Minshew & Goldstein, 2001; D. L. Williams, Goldstein, & Minshew, 2006b). However, difficulties with

working memory during search would not directly explain why ASD participants were less effective in their use of questions even when elimination on the TQT *was* allowed. It seems more likely that this reflects an additional difficulty related to specific task conditions, rather than a factor that underlies their overall problem-solving performance. The presence of differences in planning processes, in contrast, may shed further light on the issue.

### *Problems with planning*

On the question discrimination task, ASD participants did not significantly differ from typically-developing participants, suggesting that they were able to choose between more and less effective questions and could plan through the immediate consequences for each pair of options. On the plan construction task, ASD participants generally selected plans that would be less effective than in searching the TQT than the plans of control participants. The presence of group differences when constructing plans would seem to indicate difficulties during the selection and sequencing of questions. This is consistent with previous research on planning skills in children and adults with autism (Bramham, et al., 2009; Ozonoff, et al., 2004; Verté, et al., 2005) and similar to problem-solving difficulties seen in people with frontal lobe injuries (Klouda & Cooper, 1990; Upton & Thompson, 1999).

When the ASD group was analysed as a whole, the plans that they constructed tended to be less efficient in referring to fewer items. The difficulty with selecting general questions, particularly for initial questions on the plan construction task, implicates problems with utilising hierarchical sequences, as seen in the TQT performance of people with traumatic brain injury (Marshall, et al., 2003b). The construction of a series of questions that effectively narrow possibilities requires good understanding of categories and their relations to one another: LIVING>ANIMAL>PET works as an effective narrowing sequence because the latter two categories are logical subsets of the former, in virtue of their semantic properties.

In the group of lower-VIQ ASD participants, narrowing sequences of this kind were generally not observed. The questions that they selected were both less efficient overall (reflected roughly by the intercept of their planning graph in figure 10) and failed to move from general questions at the start of the plan to more specific questions at the end of the plan (as reflected by the gradient of their graph). This was suggestive of a difficulty in pre-planning a series of questions that effectively searched the set, by narrowing down possibilities step by step. Narrowing, however, was not necessarily absent for all ASD

participants: those with higher VIQ were able to select plans that consisted of more general questions first and more specific questions later. In their case, thinking through a hierarchical series of questions in advance did not seem to be a problem. This suggests that difficulties with pre-planning may have been specific to the less able ASD participants.

Where the ASD participants showed a common difficulty was in the general efficiency of their plans. Although higher-VIQ participants showed evidence of narrowing, the questions that they selected were still lower in efficiency than would be expected for their IQ level, and generally only matched the lower-VIQ group of typically-developing participants. That is, the “gradient” of their planning was reasonable, but the “intercept” was still too low. Thus, although more able ASD participants could move from general to specific questions, their questions were still not general enough. This suggests a consistent problem, across the whole ASD group, with selecting effective questions in their planning.

The presence of intact question discrimination in both low and high ability ASD participants would seem to suggest a working understanding of the individual categories being used by each question. The fact that they are not then used effectively suggests a specific problem with their deployment. On the one hand, such a difficulty would seem consistent with Minshew et al.’s (2002) concept formation deficit. Defined as an “inability to spontaneously form schemata or paradigms that organize information” (p.333), difficulties with selecting questions could be taken to reflect a problem with re-organising information from the set into categories that will be effective when used in the game.

On the other hand, Minshew et al.’s definition of this deficit refers to a difficulty with having to “spontaneously form schemata”, as if the problem is actually one of generating strategies in an unprompted context. This, though, does not seem to apply in the case of the plan construction task. When planning, participants did not have to produce questions; only choose them from a range of options. They were not being asked to spontaneously deploy some form of strategy; the strategies, in the form of questions, were made available to them, and yet they still chose options that were too restricted and would not be efficient when applied to a TQT set. Arguably, the generative demand placed on ASD participants in this context was relatively low. In this sense, their failure to choose effective questions would appear to reflect a problem of concept *selection*, not formation – all the legwork of “formation” had already been done for them.

### *Chasing concepts – an autism-specific problem?*

The question that remains is how accessible semantic concepts - and possibly verbal strategies in general - are to ASD individuals. It may be that the concepts used in Twenty Questions are present and intact within the autistic mental lexicon, but the strength of their associations to a) their exemplars and b) each other is somehow weakened or irregularly organised. Both could lead to inefficient category deployment in a context of strategic use (such as problem-solving) yet intact performance when categories or their exemplars are directly cued (as in a concept identification task).

However, such problems are not necessarily exclusive to autism. It is noteworthy that another group, deaf individuals, show a very similar profile to ASD participants on Twenty Questions (Marschark & Everhart, 1999) and other cognitive tasks that utilise semantic relations (Marschark, 2006; Ottem, 1980). Within deaf research this profile has been interpreted as a possible consequence of abnormal language development and experience, rather than executive dysfunction or other types of information processing (Marschark, Convertino, McEvoy, & Masteller, 2004). In the case of autism, the barriers to language are clearly different, but it may be that similarly atypical experiences and atypical development of language skills lead to problems with semantic accessibility and verbal strategy use in later life. Direct comparison of problem-solving and categorisation skills in deaf and ASD participants would further elucidate this possibility.

---

In summary, experiment 1 tested two candidate deficits, drawn from executive functioning research, that could explain inefficient problem-solving by individuals with autism: planning and selective attention. As in previous studies, participants with ASD asked questions that eliminated fewer items than controls on the TQT. Although a selective attention hypothesis was not supported, group differences in planning skills were observed, with ASD participants generally constructing less efficient plans than controls. This observation, along with the differences in TQT performance, combined to suggest that participants with ASD had difficulty not with *forming* strategic questions, but *selecting* the right semantic categories on which to search.



### **Chapter 3: Comparing cognitive skills in deafness and autism.**

The population of people who are deaf or hard of hearing is highly heterogeneous, varying hugely in terms of cause of hearing loss, educational experience, preferred mode of communication and cultural identity (Ladd, 2003; Mitchell, 2005)<sup>13</sup>. Nevertheless, the performance of deaf individuals on cognitive tasks can arguably be used to inform the understanding of cognitive strengths and weaknesses in autism. The respective aetiologies and outcomes of deafness and autism differ considerably, but both involve atypical experience and development of language. Accordingly, there is a considerable overlap in the kinds of tasks deployed and theories tested in experimental research with people from each group. But because this has mostly been done by different researchers, in different fields, the explanations that have been given and the conclusions that have been drawn have differed and diverged in interesting ways.

Contrasting autism and deafness on the TQT provides a specific example of this. Marschark & Everhart (1999) used a version of the TQT to explore problem-solving skills over two experiments in children and adults who are deaf. In the first experiment, a cross-sectional sample of 36 children aged 7-8, 10-11 and 13-14 attempted Mosher & Hornsby's original 42-item task over three trials (Mosher & Hornsby, 1966). In the second experiment, 14 graduate students from the National Technical Institute for the Deaf in Rochester, NY attempted the same task. In both experiments, deaf students were compared to samples of age- and education-matched hearing participants.

All the children in the first experiment were recruited from a residential school for the deaf. Every child had either congenital or early-onset hearing loss and all but five of the children had severe to profound hearing loss in both ears; that is, at least 70dB hearing loss<sup>14</sup>. All the children used sign language as their preferred mode of communication and all had hearing parents, as is the case for the large majority of children who are deaf (Mitchell & Karchmer, 2004; NDCS, 2011). As in Mosher & Hornsby (1966), Marschark & Everhart examined

---

<sup>13</sup> With reference to the latter, identification with a specific culture and community associated with deafness is typically denoted with a capital *d* (as in "Deaf culture" or "Deaf individuals"). Here, unless otherwise noted, the term "deaf" with a small *d* will be used to refer generally to people with a hearing loss who have taken part in the above studies; no claim is made concerning their specific cultural identity.

<sup>14</sup> The remaining five children had only moderate hearing loss (55dB loss) but were not observed to perform any differently to other children in the experiment (Marschark & Everhart, 1999).



rates of grouping questions and guesses on the task, alongside overall rates of trial completion, the number of questions used, and the number of items eliminated per question.

Compared to hearing children, children from the deaf group completed significantly fewer trials than hearing children and required more questions on average to find the target. Deaf children were also more likely to make guesses and used very few grouping questions in their attempts at the task: only six out of 36 deaf participants used grouping questions at all, compared to 19 of the hearing students. While younger participants in both groups were more likely to guess than older participants, deaf participants in the middle and older age groups were also less efficient in their questioning than their hearing counterparts (Marschark & Everhart, 1999).

This suggested consistent group differences in problem-solving in deaf and hearing adolescents, but Marschark and Everhart expected that such differences would largely disappear in a group of young adults who were deaf. However, the performance of deaf graduates in the second experiment did not show this. Compared to hearing adults, deaf adults completed as many trials and used as many questions, but they were still less likely to ask grouping questions. Furthermore, the average efficiency of the questions that they asked was significantly lower for deaf participants. This suggests that differences between deaf and hearing participants in problem-solving performance may persist into adulthood, even in very able deaf individuals (Marschark & Everhart, 1999).

This profile on Twenty Questions – less grouping, more guessing, but also less efficiency when grouping - is very similar to that seen for participants with autism in the studies by Minshew et al. (1994), Minshew et al. (2002), Alderson-Day & McGonigle-Chalmers (2011) and the research reported above. But the interpretation of this pattern of performance, developed in the context of deaf research, differs notably from the explanations put forward in autism research.

### **Interpreting deaf performance on Twenty Questions**

Marschark & Everhart (1999) tested and discussed a range of possible explanations for why participants who are deaf performed differently from hearing participants on Twenty Questions. Two of these, the role of impulsivity and the influence of prior game experience, were directly tested within their study design.

For a number of years, researchers have suggested that a problem with impulsivity is evident in some deaf individuals (e.g. Altshuler, Deming, Vollenweider, Rainer, & Tendler, 1976; see R. I. Harris, 1978, for an early review; and Hauser, Lukomski, & Hillman, 2008, for a more recent discussion of executive skills in deafness). Children who are impulsive tend to be less efficient in their approach to the Twenty Questions Task, as documented in a number of early studies (Ault, 1973; Denney, 1973; McKinney, 1975; Zelniker, Renan, Sorer, & Shavit, 1977). In particular, problems with impulsivity may lead to increased guessing, as a lucky guess might provide a very quick solution to the problem: if Marschark & Everhart's deaf participants were indeed more impulsive than hearing participants, then this may have led them to guess at a quick answer, rather than forming a more strategic question. To test this explanation, Marschark & Everhart (1999) assessed their younger participants on the Porteous Mazes test, a measure of impulsivity-reflectivity (Porteus, 1950). Although significant age-related effects were observed, no significant differences between deaf and hearing participants were apparent, suggesting that deaf participants were no more impulsive than their hearing counterparts (Marschark & Everhart, 1999).

A second explanation was that participants in the deaf group may have been less familiar with the game of Twenty Questions itself, leading to the adoption of different and less effective strategies to search for targets. To test this, Marschark & Everhart asked all the participants if they had ever played the game before. While there were more hearing participants overall who had played the game (11 deaf & 20 hearing across both experiments combined), the more important finding was that this appeared to make a specific difference to the strategy adopted by deaf children. Whereas deaf children who had prior experience of the game asked a majority of grouping questions (51%), inexperienced deaf children did not (20%). In contrast, both experienced and inexperienced hearing children asked a large proportion of grouping questions (55% & 46% respectively).

For Marschark & Everhart this result suggested that the experiences of deaf children play an important role in their problem-solving performance. In a specific sense, they may have had less opportunity to play Twenty Questions or other guessing-based games. In a more general sense, this may reflect a broader difference in linguistic experience, both in home and at school:

“At a more general level, lack of experience with games of this sort are consistent with the fact that most deaf children have less linguistic interaction with their hearing parents and less social interaction with peers than either deaf children with deaf parents or hearing children with hearing parents. This lack of access to effective language is likely to have a broader impact on social and cognitive development,

through restrictions of diversity in meaningful linguistic, social and cognitive interactions..... It may turn out that deaf children really do have difficulties in problem-solving when tasks require particular kinds of strategies (e.g. hierarchical organisation). Alternatively, it may be that younger deaf children have not yet acquired the problem-solving strategies exhibited by hearing peers only because those peers had a ‘head start’ through effective early communication in the home.” (Marschark & Everhart, 1999, pp. 78-79).

This interpretation is very general, and – as the authors note – requires further specification and testing. Before doing that though, it is worth contrasting this kind of explanation with those applied to problem-solving in autism. For Minshew and colleagues, difficulties on the TQT and similar tasks were evidence of a “inability to spontaneously form schemata or paradigms that organize information” (Minshew et al., 2002; p333), arising out of a general problem with complex cognitive processing. Based on previous applications of the TQT, the experiments in Alderson-Day & McGonigle-Chalmers (2011) and in the previous chapter assessed hypotheses about the specific executive processes that contribute to the autistic profile on this task.

In both cases, the emphasis is on cognitive processes; difficulty or atypical performance on the task is interpreted as an indicator of cognitive deficit or bias on the task. And this is done with good reason: as with many other cognitive experiments with autistic participants, the individuals seen represent the higher-functioning end of the autistic spectrum, with good functional language and the general abilities to engage sufficiently with an experimental task. With good matching and adequate control for confounding variables, any remaining group differences are taken to reflect specific information about cognitive skills in autism.

But if the participants in Marschark & Everhart’s study did not perform as efficiently as controls due to the effect of their atypical linguistic experiences, then a similar explanation may apply for the problem-solving skills of people with autism: what may be being observed is not a cognitive deficit or impairment per se, but a reflection of their language development. The cardinal features of autism spectrum disorders are difficulties with social interaction and communication – difficulties that will clearly shape the linguistic experience of people on the spectrum, even for very able ASD individuals. In particular, those individuals on the spectrum who show considerable delays in their language development may be expected to be affected, given their relatively fewer opportunities to interact linguistically with others in early years. If a deaf child in a hearing family misses out on such opportunities, then the autistic child who did not speak until their sixth year will have missed out too. Such a child may appear to “catch up” in other respects, in terms of structural

language skills or academic performance (Boucher, 2012), but still show evidence of subtle difficulties in higher cognitive processes that depend on language.

### **Cognitive overlaps in deafness and high-functioning in autism**

In this sense, the problem-solving profiles seen in autism and deafness could represent an overlap arising out of similarly atypical linguistic development. This kind of overlap is not without precedent: despite the differences between the two groups, there are a number of similarities in the cognitive profile of children who are deaf and children with autism (Emmorey, Klima, & Hickok, 1998; Ottem, 1980; Peterson & Siegal, 1999).

The possibility of similarities between these two groups was in fact explored in some of the earliest cognitive research on autism. As part of a programme of research on memory, language and reasoning skills, Hermelin and O'Connor often used groups of ASD and deaf children to examine how atypical sensory processing affects cognitive development in similar ways across different populations (Hermelin & O'Connor, 1970; O'Connor & Hermelin, 1978). For example, in one study (O'Connor & Hermelin, 1973) they compared groups of ASD children, deaf children, and hearing, typically-developing children on a short-term memory task, where three-digit sequences had to be recalled by participants. The digits were presented visually one at a time, but their spatial position did not correspond to their temporal order (for example, the first digit appeared on the right, the second on the left, and the third in the middle). O'Connor and Hermelin observed that their hearing-typical group tended to recall items in their temporal order, whereas deaf and autistic participants were more likely to recall them in their spatial order, from left to right. This was interpreted as an example of deaf and autistic children relying more on spatial representations to support their recall, in contrast to hearing-typical participants, who used temporally-ordered, language-based strategies (O'Connor & Hermelin, 1973).

Since then few studies have directly compared groups of ASD and deaf participants, but separate studies have highlighted common areas of strength and weakness. As in studies on autism (Bennetto, et al., 1996; Landa & Goldberg, 2005; D. L. Williams, et al., 2006b), research has documented problems with working memory (WM) in people who are deaf (Krakow & Hanson, 1985; Lichtenstein, 1985). Similarly, groups of deaf and ASD participants have shown advantages in visuo-spatial skills in comparison to hearing counterparts (Emmorey, et al., 1998; Emmorey & Kosslyn, 1996; Neville & Lawson, 1987; Todman & Cowdy, 1993; M. Wilson, Bettger, Nicolae, & Klima, 1997). Theory-of-mind

skills, at one time proposed to be *the* cognitive deficit that explains autism, are often delayed in many children who are deaf (Courtin & Melot, 1998; Peterson & Siegal, 1995, 1999). Most pertinently for Twenty Questions, there are a number of similarities in the categorisation performance of children who are deaf and children with autism (Ormel et al., 2010; Shulman, et al., 1995).

When examined more closely, some of these similarities actually mask subtly different profiles. For example, while working memory difficulties have been reported in both groups, the specific WM components affected in deafness and autism would appear to differ. In deafness, participants tend to perform worse than controls for serial recall of language-based stimuli (speech or sign) and those tasks requiring use of the phonological loop (Conrad, 1972, cited in Luftig, 1983; M. Harris & Moreno, 2004; Logan, Maybery, & Fletcher, 1996; M. Wilson, et al., 1997). In contrast, verbal working memory difficulties have been reported in studies on autism (e.g. Bennetto, et al., 1996), but the most consistent WM impairments documented have tended to be drawn from studies on spatial working memory (Landa & Goldberg, 2005; D. L. Williams, et al., 2006b).

In other aspects of visuo-spatial skill (such as mental rotation), strengths are apparent in both groups but the cause of such skills appears to differ. In deafness, such skills seem to be closely related to signing proficiency and experience of using signed languages (Emmorey, et al., 1998), rather than some form of compensation for auditory deprivation or a “lack” of spoken language (Marschark & Hauser, 2008; Mayberry, 2002). In autism the source of such skills is arguably still unclear. They could reflect some form of innate ability, cognitive style or a compensatory response to abnormal language development (Baron-Cohen, 2002; Happé & Frith, 2006; Mottron, Dawson, Soulieres, Hubert, & Burack, 2006). What is more certain, though, is that such skills are probably not the product of using sign language, as signed forms of communication are not used with anywhere near the same prevalence and extent by people with autism, as compared to the deaf community.

### **Theory of mind research with deaf children**

A more significant overlap in skills has been observed in research on theory-of-mind. As discussed in the introductory chapter, the difficulties that children with autism show in Theory-of-Mind (ToM) performance are well documented (e.g. Baron-Cohen, et al., 1985). In deaf research, studies by Candida Peterson and colleagues (Peterson, 2004, 2009; Peterson

& Siegal, 1995, 1999, 2000; Peterson & Wellman, 2009; Peterson, Wellman, & Liu, 2005) have provided consistent evidence of delays in ToM development.

Specifically, signing deaf children of hearing parents have been observed to perform less well than hearing children on the false-belief measure used by Baron-Cohen, Leslie and Frith, the Sally-Anne task (Peterson & Siegal, 1995). Deaf children have also been reported to struggle on other false-belief tasks, such as the Smarties test (Perner, Frith, Leslie, & Leekam, 1989), in which participants are shown that a tube of sweets has unexpected contents (e.g. pencils), and are asked to indicate what another person will think is in the tube (Peterson & Siegal, 1999).

In Peterson and Siegal's initial research, oral deaf children – who are assumed to have more opportunity to communicate with hearing parents – also performed much better than signing children (Peterson & Siegal, 1999). However, subsequent data suggests that they too exhibit delays in performance compared to hearing counterparts, albeit to a lesser degree (de Villiers, Hosler, Miller, Whalen, & Wong, 1997, cited by Peterson & Siegal, 2000; Peterson, 2004). The only group who have consistently shown equal abilities in ToM performance are deaf children who are native signers, either by having deaf parents who sign or another significant relative in the home who is fluent in signing (Courtin, 2000; Jackson, 2001; Peterson, 2004; Peterson & Siegal, 1999). For Peterson and colleagues, this strongly suggests that deafness *per se* is not the cause of ToM difficulties. Rather, having limited opportunities to interact linguistically with others and specifically engage in talk about mental states, leads to later delays in socio-cognitive skills (Garfield, Peterson, & Perry, 2001; for evidence for this relation see Moeller & Schick, 2006, for example).

Although a continuing matter of debate, a prevailing assumption within autism research is that ToM difficulties reflect a modular deficit, which is in some way central and innately specified in autism. Research on ToM development in deaf children, on the other hand, has prompted attempts to explain both the autistic and deaf profiles as a consequence of their language environments (e.g. Garfield, et al., 2001), alongside other theories that have emphasised a contributory role for language in the acquisition of ToM skills (de Villiers, 2005; Tager Flusberg & Joseph, 2005). A similar, older idea within autism research is that those individuals on the spectrum with more typical language histories, such as children with Asperger Syndrome, can “hack” standard theory-of-mind tasks by drawing on their relatively good structural language skills (see Happé, 1995, for a discussion of this issue).

Subsequent research has not always supported the idea of a common, language-mediated overlap in ToM abilities in deafness and autism. For example, a recent paper by Peterson's own research group (Peterson, Wellman, & Slaughter, 2012) compared performance on six tasks of Theory-of-Mind ability in deaf children, hearing children, children with autism and children with Asperger Syndrome (AS) aged 3-12. If early language interactions serve to scaffold the development of Theory-of Mind abilities, then children with Asperger Syndrome, who have more typical language histories than children with autism, should perform much more like typically-developing children on ToM tasks. However, in Peterson et al. (2012) children in the AS group were no better than deaf or autistic participants in their performance on the ToM tasks, and significantly worse than typically-developing, hearing children. In addition, the six tasks were designed to be passed in a certain order, depending on their difficulty and the developmental stage of the child. The order in which AS and autistic participants were able to complete the tasks differed from that of deaf and hearing children, suggesting a qualitative difference in the theory-of-mind development of ASD participants in general (Peterson, et al., 2012).

Thus in the case of theory-of-mind, though deafness and autism may appear to share certain aspects of performance, it is unlikely that the cause of their ToM difficulties is the same. And this might not be considered surprising – after all, problems with social interaction are by definition central to autism, but in the case of deafness only secondary to problems with hearing loss.

Where the example of ToM research is useful, though, is in highlighting a mechanism that could in theory affect the experience and development of both deaf and autistic children in similar ways; namely, reduced opportunity to engage in parent-child communication in early years. This is useful for understanding problem-solving because of its potential impact upon category learning and semantic development. If parents and children engage in less talk about mental states, then the link with later understanding of mental states would seem straightforward, at least in deaf children (although see Jackson, 2001, for a discussion of complicating factors). But if parents and children have less opportunity *in general* to engage in these sorts of interactions, then typical processes of semantic category learning and concept acquisition may also be affected. If this is the case, then similarities in the categorisation skills of children who are deaf and children with autism should be apparent.

## **Categorisation research in deaf children**

As referred to in prior chapters, children with autism have been shown to perform well on perceptual or basic categorisation tasks (Tager-Flusberg, 1985a; Ungerer & Sigman, 1987), appear to struggle on tasks that require manipulation of more complex or abstract categories (Shulman, et al., 1995), and they may not spontaneously use categories to organise information in the same way as typically-developing children (Ropar & Peebles, 2007). The following section reviews what corresponding findings there are from research with deaf individuals.

A common assumption made by many deaf researchers in the early to mid- 20<sup>th</sup> century was that all conceptual and abstract thought must be impaired in deaf individuals, as a necessary consequence of their auditory deprivation (Marschark, 2003). For example, Oléron (1953) begins his paper with the following claim:

“Studies by psychologists, as well as observations by educators, indicate that deaf people are inferior to those with normal hearing, particularly in the domain of abstract mental activities.” (Oléron, 1953, p. 304)

Along the same lines, Youniss & Furth (1966), in their summary of work by Oléron and others, refer to deaf participants being “dominated by obvious perceptual cues” (p.75), while Blanton and Nunnally (1964) considered the possibility that their results can be explained by “the greater concreteness of the deaf” (p. 401) in a discussion of categorisation skills.

It is likely that early psychological theories and experiments on deafness underestimated both the spoken language demands of the task deployed and the conceptual abilities of deaf participants: subsequent research that deployed primarily non-verbal measures of categorisation did not support the idea of a general conceptual impairment. For example, in a series of card-sorting experiments, Furth (1964) reported equivalent levels of performance in age- and IQ-matched deaf and hearing adults when categorising according to different colour and shape combinations. Similarly, Kates et al. (1961) observed no differences between a group of deaf adolescents and two groups of hearing high-school students in successful categorisation performance on the Goldstein-Scheerer Object Sorting Test (K. Goldstein & Scheerer, 1941). While there were group differences in the extent to which deaf participants could provide verbal labels for different category groupings, these differences disappeared once academic achievement was controlled for (Kates, Kates, Michael, & Walsh, 1961; see also Rosenstein, 1960). For Furth - a researcher known for claiming that deaf cognition fundamentally lacked language (Furth, 1966) - findings such as these indicated that concept



processing abilities were intact in people who are deaf, and clearly separable from any difficulties with language (see Vernon, 1967, for a similar conclusion)<sup>15</sup>.

For more semantic forms of categorisation the division between conceptual ability and language skill is a lot less clear. Basic semantic categories such as “DOG” or “TREE” may be based primarily on perceptual similarities, but understanding of their extension and reference will inherently require linguistic knowledge (e.g. what counts as a dog, as opposed to a wolf?). Superordinate groupings, where the perceptual similarities across exemplars may be few and far between, likely rely heavily on language skill, both in initial acquisition and use (Horton & Markman, 1980). Accordingly, the profile of semantic categorisation skills in deafness is much less straightforward than that seen for perceptual categorisation.

A number of studies have investigated semantic categorisation skills in deaf participants, usually utilising picture-based paradigms in order to avoid the confounding effects of language ability (D'Hondt & Leybaert, 2003; Farjardo, Arfé, Benedetti, & Altoé, 2008; Friedman, 1985; Furth & Milgram, 1965; Ormel, et al., 2010; Pettifor, 1968; Silverman, 1967; Yi et al., 2011). Pettifor (1968) compared deaf and hearing school-children (age range 5-14) on a card sorting task where pictures could be classified according to their visual characteristics (e.g. size) or semantic associations (child and adult). While hearing children performed better overall, difficulties in sorting accuracy in the deaf group were most prominent when semantic categorisation was required, as shown by a significant group-by-categorisation interaction effect. For Pettifor, this reflected the dependence on language skill of more conceptual forms of categorisation (Pettifor, 1968).

Relatively more intact categorisation abilities have been reported in studies with deaf participants conducted by Friedman (1985), Furth & Milgram (1965), Silverman (1968) and D'Hondt & Leybaert (2003). On a sorting paradigm contrasting perceptual, basic and superordinate categories, Friedman (1985) reported equal performance for perceptual and

---

<sup>15</sup> Another idea that has been proposed based on early categorisation research in deafness is that categorising according to a conjunctive or two-part rule may be impaired. Ottem (1980), in a review of largely French and German studies, highlighted that deaf participants have tended to struggle on tasks that require sorting according to two principles simultaneously (e.g. shape *and* colour). Early studies by Höfler (1927, 1934), Oléron (1951) and Vincent (1957) reported intact sorting in deaf participants for single sorting principles, but reduced performance for simultaneous sorting, or when sorting rules must be switched at different points during the task (Ottem, 1980). Even in Furth's (1964) study, where deaf and hearing participants performed seven different categorisation tasks at equal levels, Ottem (1980) notes that an eighth task involving simultaneous sorting produced significant group differences. This idea has been taken up by Marschark (2006) with the suggestion that relational processing in general may be atypical in deafness. It also has a counterpart in research on autism: Shulman et al. (1995) report on specific problems with sorting according to two criteria simultaneously in a group of children with autism.

basic sorting rules in a sample of deaf and hearing schoolchildren. For sorting according to superordinate criteria, deaf participants were less successful than hearing participants on their initial attempts at sorting, but improved to a typical level of performance following a single, scaffolded trial where the experimenter illustrated the sorting rule (Friedman, 1985). Similarly intact identification of basic semantic stimuli were also reported in a study by Furth & Milgram (1965), while Silverman (1967) reported a greater tendency towards thematic over taxonomic forms of sorting in deaf schoolchildren compared to hearing children, but equivalent levels of accuracy for superordinate associations. On a word-based semantic decision task, where participants were asked to judge whether different basic exemplars were part of the same superordinate category, deaf adolescents have also been reported to be as fast and as accurate as controls (D'Hondt & Leybaert, 2003).

In contrast to earlier findings, three recent studies have reported less successful categorisation performance in deaf participants (Farjardo, et al., 2008; Ormel, et al., 2010; Yi, et al., 2011). Ormel et al. (2010) compared categorisation skills in a sample of deaf and hearing children (mean age = 9.5 years). Across two experiments that compared basic and superordinate categorisation skills – the former in both words and pictures – deaf children were less accurate and slower to categorise semantic stimuli than hearing controls. As in Pettifor (1968) and Friedman (1985), differences in deaf and hearing performance were most prominent for superordinate stimuli (although as these stimuli were only presented verbally, this may have confounded participants' difficulties in response).

Farjardo et al. (2008) compared deaf and hearing high-school students (mean age 17.9 years) on a forced-choice semantic decision task where participants had to judge whether an item was an example of a superordinate category. Items were either presented as pictures or words, and categories were either similar (fruit and vegetable) or dissimilar (fruit or animal) in their perceptual qualities. While deaf and hearing participants did not differ in terms of accuracy, hearing participants were significantly faster overall and showed more of a benefit in accuracy for word-based categorisations.

Finally, Yi et al. (2011) compared groups of deaf and hearing teenagers (mean age = 13.0) on a forced choice categorisation task where items could either be grouped according to a theme (e.g. chalk and a blackboard) or a taxonomic relation (e.g. chicken and ostrich). Items were presented either as words or pictures. As in Farjardo et al. (2008), accuracy rates were high and did not significantly differ between deaf and hearing participants. However, reaction times were longer for deaf participants than hearing participants, irrespective of stimulus type (Yi et al., 2011).

Differences in semantic skills have also been evident when participants need to utilise categories to support strategic cognitive processes, such as memory. In research on recall processes in deaf people, normal encoding of semantic categories (Hoemann, Andrews, & DeRosa, 1974) and typical use of semantic clustering strategies (Liben, 1979, 1985) has been documented. However, the efficiency and success of such strategies has not always matched that of hearing participants (Liben, 1985). For example, Koh, Vernon and Bailey (1971) tested young (13-14) and older (18-20) deaf adolescents on their free recall of words from the Peabody Picture Vocabulary Test (L. Dunn, 1965). Overall, recall rates in the deaf participants were lower than hearing participants of the same age. In addition, a range of measures of subjective organisation (the amount of semantic grouping within each recall string) indicated that deaf adolescents utilised less grouping than their hearing counterparts (Koh, et al., 1971).

Methodologically the above studies vary in a number of ways, making it hard to draw clear conclusions. Most notably, relatively little is known about the language histories or current language skills of the deaf participants included in these studies. In contrast to the ToM research literature, none of the above categorisation studies prior to 1985 report any information on the hearing status of parents or the language used at home (Furth, 1964; Kates et al., 1961; Koh et al., 1971; Liben, 1979; Pettifor, 1968; Silverman, 1967). Given the large numbers of deaf people who come from hearing families (Mitchell & Karchmer, 2004), it is probably fair to assume that this was also the case for most of the participants in the above studies. Indeed, in studies that have collected this information, they either report “no use of signing in [participants’] homes” (Friedman, 1985, p70; D’Hondt & Leybaert, 2003) or indicate that only a small minority of their participants came from deaf families (Ormel et al., 2010; Yi et al., 2011). Nevertheless, no study to date has systematically compared groups of deaf participants with different language backgrounds in the same way that this has been done from ToM research, so the possible effects of linguistic experience remain unclear.

Information on the preferred or primary method of communication for deaf participants in the above studies is more readily available. Reflecting the educational practises of the time, participants in early studies were either reported to be attending oral-based schools (Kates et al., 1961; Rosenstein, 1960) or were provided with their task instructions orally (Pettifor, 1968). In some cases references were made to use of sign, gesture or pantomime where it was needed (Furth, 1964; Silverman, 1967) but the implication given is that deaf participants were expected to understand and attempt the task via Spoken English. Subsequent studies mention use of sign (e.g. Koh et al., 1971) and tailoring of spoken, signed or combined

methods of instruction depending on the skills of each participant (Liben, 1979), but only recent studies have characterised their participants in any detail, either as users of cued speech (D'Hondt & Leybaert, 2003), sign-language users (Farjardo, et al., 2008), or bilingual users of speech and sign (Ormel, et al., 2010; Yi, et al., 2011).

In addition to language differences, the above studies vary considerably in the level of matching between deaf and hearing participants. Almost all have deployed some form of age matching and most have attempted some form of non-verbal IQ matching (Friedman, 1985; Furth, 1964; Ormel, et al., 2010; Pettifor, 1968; Yi, et al., 2011), but only one study has reported matching for both VIQ and NVIQ abilities on standardised measures (Kates, et al., 1961). Other studies, where it is reported, have assumed participants to all be within the normal IQ range via their grade level and inclusion in mainstream classes (Liben, 1979; Silverman, 1967), or have deployed other matching measures, such as reading ability (D'Hondt & Leybaert, 2003).

The lack of verbal IQ matching reflects the difficulties with measuring such abilities fairly in deaf individuals, given the level of spoken language that typical VIQ measures both a) assume for their administration, and b) require for successful completion (Braden, 1994). This means that it is not at all clear that deaf and hearing participants in the above studies had comparable levels of general, verbal-cognitive skills (and in fact this would seem unlikely) – although it should be noted that more recent studies have also measured levels of signing ability, for example, in order to gauge the contribution of such skills (Ormel, et al., 2010). The lack of consistent NVIQ matching is more puzzling, as a number of measures can be used with appropriate administration (Hill-Briggs, Dial, Morere, & Joyce, 2007). It could be that in some studies this reflects differences in research practice: whereas in autism research close age- and IQ-matching is assumed to be the norm, this does not appear to be the case for studies on deafness in general.

Despite these caveats, a number of parallels with categorisation skills in autism are arguably apparent. As in the autism literature, perceptual and basic categorisation skills appear intact, but more complex and semantically abstract forms of categorisation appear problematic for some deaf participants. In addition, deaf and autistic participants show atypicalities in how effectively they can use categories to support other cognitive processes, such as memory. Without exception, the above studies have interpreted the deaf categorisation profile as a product of their linguistic (or, rather non-linguistic) development, whereas in autism categorisation differences are often taken to reflect fundamental differences in information processing (e.g. Gastgeb, et al., 2006). But if both groups experience delays in their language

development and reduced opportunity to engage communicatively with others, then their initial learning and subsequent use of categories seems likely be affected in similar or overlapping ways.

### **Two hypotheses concerning problem-solving in autism and deafness**

Returning to Twenty Questions, it seems there is at least a *prima facie* argument to be made that problem-solving which relies on the flexible use of semantic categories may be impaired in both deafness and autism as a result of atypical language development.

A “strong” form of this argument would be to propose a specific causal chain, linking

- 1) commonly delayed language development in autism and deafness,
- 2) atypical acquisition and organisation of semantic categories in both groups, and
- 3) inefficiencies in verbal problem-solving tasks such as the TQT.

That is, difficulties in verbal problem-solving in autism and deafness are proposed be a long-term consequence of early language development, because of the impact that language delays have on semantic organisation. Abnormal early language experiences (1) lead to atypical semantic organisation (2), making it harder for semantic categories to be used efficiently when they are required on verbal problem-solving tasks (3). The cause and developmental pathway of this profile is proposed to not only be similar, but the *same* across both groups. Figure 12 provides an illustration of this argument.

A “weak” form of the argument is to propose the following: if problems with semantic categorisation and problem-solving can occur in deafness, apparently as a product of their language development, then an *analogous* process could also occur in autism. In contrast to the “strong” argument outlined above, the following causal story is proposed:

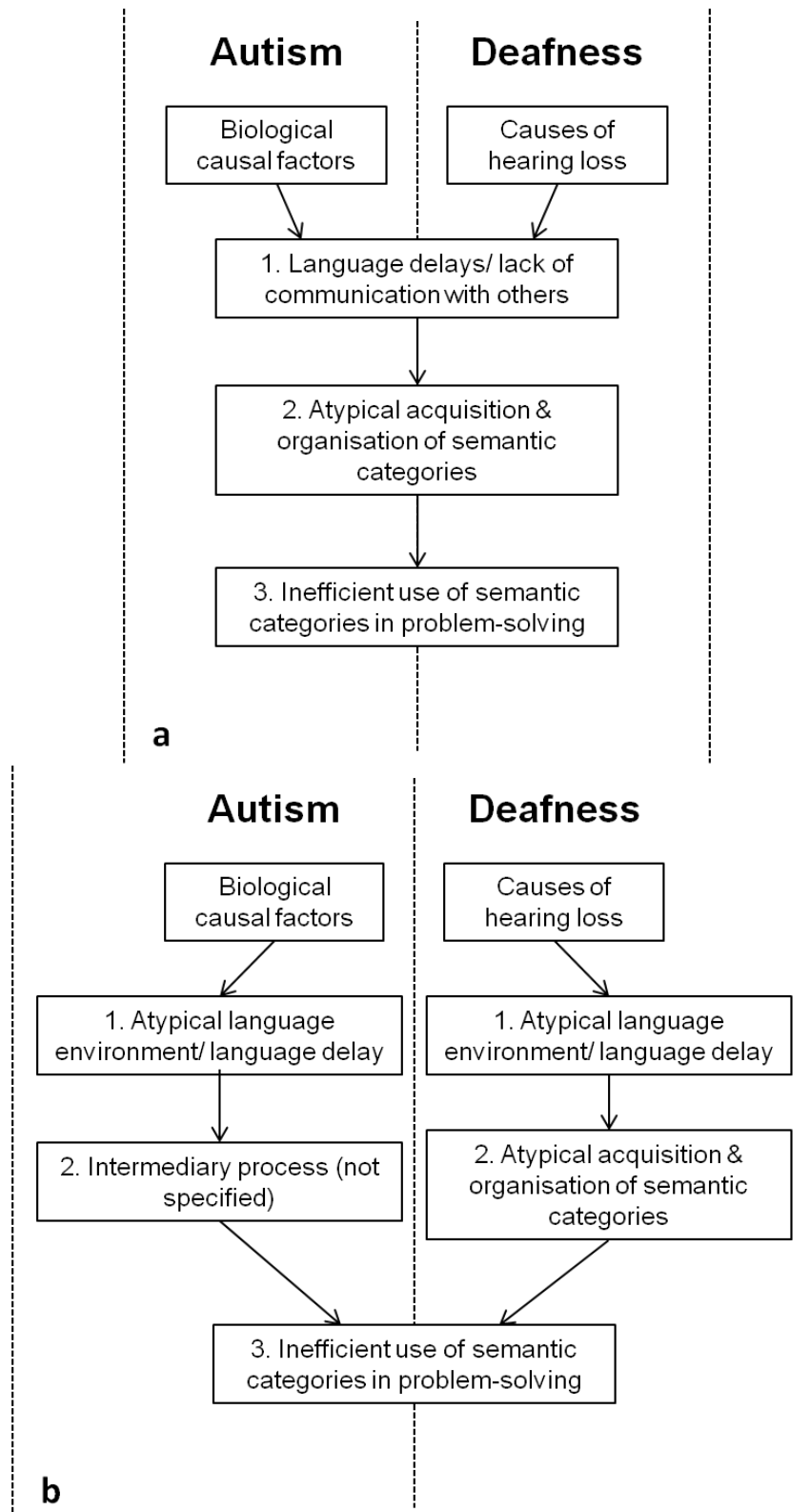
- 1) commonly delayed language development in autism and deafness, *leading to*
- 2) atypical semantic organisation in deafness; an unspecified process in autism, *resulting in*
- 3) inefficiencies in verbal problem-solving tasks such as the TQT.

For both groups, the original source of their problem-solving difficulties is in their early language development (1). But the route by which this happens is not necessarily the same: in deafness it could follow the path of atypical semantic organisation (2), while in autism this could be some other form of intermediary process, or combination of processes associated with their wider psychopathology.

For example, a lack of early language in autism could lead to problems with how language-related regions of the brain communicate with areas linked to other cognitive processes. The development of structural connections between different brain regions is thought to be abnormal in autism - Courchesne and Pierce (2005), for example, refer to the frontal cortex “talking to itself” (p. 225) – and it could be that a failure to use typical language areas and networks by ASD children leads to atypical connectivity patterns at later ages. On a task like Twenty Questions, executive skills have to be applied in concert with semantic knowledge, likely requiring effective communication between frontal (executive) and temporal (semantic) regions in the brain. Disruption to such a process could lead to inefficient performance when using semantic strategies to solve a problem.

This represents just one possible alternative to the “semantic route” specified in the strong argument, and there may be other good candidates for the link between language development and problem-solving in autism. The important point about the weak argument is that it is not committed to a specific intermediary process: it simply proposes that problem-solving atypicalities in this group are ultimately a product of early linguistic development, as opposed to an explanation that only refers to cognitive skills or information processing.

To argue for either of these hypotheses, however, would require a firm empirical basis, and a confidence that Marschark and Everhart (1999) were correct in interpreting their data as they did. To date, no other researchers have attempted to replicate Marschark & Everhart’s findings on the TQT or similar measures: their result could be a one-off. It could be that the differences observed between deaf and hearing participants in their study were produced by other cognitive discrepancies not measured by their paradigm. Or, alternatively, such differences may be an artefact of the communication method with which participants



**Figure 12. Causal models for the strong and weak arguments.** The strong argument (a) differs from the weak argument (b) in specifying a common causal route for problem-solving difficulties in autism and deafness.

attempted the task. To establish whether either (or neither) of the strong or weak arguments can be endorsed, these alternative explanations must first be eliminated.

### **Alternative explanations**

A first point to make about Marschark & Everhart's (1999) study is that the deaf and hearing participants were not explicitly matched on general cognitive ability. In experiment 1, the sample of children who were deaf were reported to be "within the range of normal intelligence according to school authorities" (p69), and the sample of hearing children was recruited from local public schools. In experiment 2, both deaf and hearing participants were recruited from Rochester Institute of Technology and assumed to possibly have "exceptional" problem-solving skills based on their acceptance to courses at the college. Nevertheless, in neither experiment did deaf or hearing participants attempt a standardised test of intelligence, such as one of the Wechsler batteries (e.g. Wechsler, 1999). As such, any major group differences in problem-solving performance may reflect underlying group differences in general cognitive ability.

Generalised cognitive differences between deaf and hearing populations are a matter of considerable debate. A full discussion of all the findings and issues in relation to this area is beyond the scope of this project (for a review, see Braden, 1994), but some brief, general points can be made.

First, there is a great lack of standardised cognitive tasks that are suitable for administration with deaf individuals (Haug & Mann, 2008). Most non-verbal or performance IQ subtests of cognitive batteries yield normal range scores for deaf individuals (Vernon, 2005) and can be used as long as they are administered with a suitable interpreter (Hill-Briggs, et al., 2007), but there are no appropriate verbal IQ measures that currently exist. This is because verbal IQ measures primarily measure cognitive skills via knowledge of spoken language, which for many deaf individuals may not be their preferred mode of communication. For example, in the WASI (Wechsler, 1999), the Vocabulary subtest requests definitions of words and the Similarities subtest requires category names for pairs of words: both tests require knowledge of Spoken English. In some cases, translation into a signed language such as British Sign Language (BSL) or American Sign Language (ASL) may be possible, but for some words a signed counterpart will not exist, or the iconicity of the sign for the word will illustrate its meaning in some way. For example, the sign for "mountain" resembles the sloped, triangle shape of a mountain (Brien & Brennan, 1992). If verbal IQ as a concept is understood as a



measure of language-based cognitive abilities, rather than knowledge of a specific language, this suggests verbal IQ *per se* is not being measured in these cases. Or, alternatively, it is being measured, but only in a piecemeal and indirect fashion.

Of course, for a sizeable proportion of deaf individuals spoken language may be actually their preferred mode of communication; many people who are deaf are taught orally, learning to lip-read and attending schools that encourage speech. However, the use of spoken language VIQ measures even in this population is still problematic, because acquisition of speech, lexical knowledge and speech comprehension will often show considerable age delays (see Geers, 2006; Mayberry, 2002; Quigley & Paul, 1984, for reviews). This means that VIQ measures again are not neutral measures of language-based cognitive abilities, reflecting a person's "general level", but specific measures of difficulty in the test language. Thus, when VIQ measures have been deployed, samples of deaf participants have shown notably uneven VIQ-NVIQ profiles, with verbal skills lagging behind non-verbal skills (Braden, 1992, 1994). For problem-solving, this means that poorer performance on a language-heavy task may only be the result of poorer language skills, not a specific difficulty or divergence in cognitive skills.

A second factor that could potentially influence problem-solving performance concerns the actual modality of response that Marschark & Everhart's (1999) participants used. As noted above, the children and adults in the study all used ASL as their preferred mode of communication. This could have affected the kind of questions they deployed, as the linguistics of signed languages differs considerably from spoken languages (Stokoe, 1960/2005; Sutton-Spence & Woll, 1999). One difference relevant to Twenty Questions is that some superordinate categories are not represented in signed languages (a point noted in categorisation research on deafness; Wolff, 1985). Depending on the language, a signing dictionary may contain a few thousand signs (e.g. Brien & Brennan, 1992), while dictionaries of spoken language contain hundreds of thousands (e.g. Soanes, Stevenson, Pearsall, & Hanks, 2005). Many abstract words are not represented at all in sign language, including superordinate terms: for example, there are no specific signs for reptile or weapon in British Sign Language (Brien & Brennan, 1992). For games like Twenty Questions, grouping questions often rely on superordinates to collect more basic items together – if a specific superordinate sign does not exist, that grouping may not be as available for participants who sign.

A related - but more contentious – suggestion is that sign language is less abstract than spoken language *in general*. If it were the case that there was less abstractness in signed

languages, then this could affect how likely participants would be to use more abstract or general category terms. Whether this is the case is a matter of longstanding debate, and one which cannot be fully addressed here (compare, for example, the views in Taub, 2001, and Greene, 1975). It is also important to note that, while certain terms may not have a corresponding sign in BSL or ASL, there are also concepts in sign that are not easy to express in English. It would be a considerable mistake to assume that one language is necessarily less rich or less flexible than another. Despite this, the possibility remains that the way in which sign is commonly used to refer and represent may be different from spoken language in subtle ways, emphasising the tangible over the vague, or exemplars over categories, for example. This, in turn, may impact on how concepts are used in a setting like Twenty Questions – above and beyond the more concrete question of whether specific terms have an equivalent in each language<sup>16</sup>.

In fact, there is evidence - albeit tentative – to suggest that communication method could be an important factor to consider in Twenty Questions performance. In the only other study to apply a TQT-type paradigm to a deaf sample, Remine, Care, and Brown (2008) assessed the problem-solving performance of a group of deaf adolescents ( $Age_M = 14.02$ ) whose first language was Spoken English, and who attempted the task using speech. As the study was focussed on identifying linguistic and executive predictors of problem-solving in deafness specifically, no comparison group of hearing students was tested. Nevertheless, when compared to hearing norms from the DKEFS testing battery (Delis, et al., 2001), Remine and colleagues observed that all their deaf participants performed within the normal range of scores for their age group. Furthermore, their raw scores would seem to indicate greater use

---

<sup>16</sup> A contrasting example of the role language differences might play is provided by the work of Courtin (1997, 2000) on native deaf signers using French Sign Language (LSF). Courtin (1997) argues that some aspects of LSF may actually facilitate more advanced abstract categorisation skills in deaf children, based on the iconic structure of certain basic-level signs and their exemplars. For example, in the case of different flowers or trees, Courtin notes that the FSL sign for “flower” or “tree” is always signed before specifying the particular variety. This kind of general sign provides information about the nature of the category that the specific exemplar (e.g. an oak, or a larch) belongs to:

“The generic sign TREE, for example, sometimes refers to the prototypical element of the category... But it also refers to the intensional properties of the category in that it encompasses some of its characteristic properties (e.g., the trunk and branches) that cannot be taken away from the sign, the category, without “destroying” it. For these two reasons, categorization may be easier for a deaf child who is a native FSL signer than it would be for a hearing one.” (Courtin, 1997, p. 162).

of grouping questions than that seen in Marschark & Everhart's (1999) first experiment (Remine et al., 2008).

Remine et al.'s (2008) findings are hard to interpret without comparing performance with a control group, but at the very least they add support to the idea that the language deaf participants use may be important to their problem-solving. This, combined with the issues about category signs discussed above, raises the possibility that problem-solving on the TQT may require signing participants to use strategies (i.e. semantic category terms) that are not a usual part of the everyday language use. A reduced lexicon of such terms in signed languages would not seem to be as much of a problem: while some category terms may not exist in signed languages, the typical grouping terms on a TQT such as ANIMAL, TRANSPORT, HOUSE, FRUIT do tend to have corresponding signs. But if there is a tendency in signed languages to use superordinate terms less in general, then this may lead to less efficient use of such terms on a task like the TQT.

In summary, there are two explanations, one concerning cognitive skills, the other language differences, which may account for Marschark and Everhart's findings. If they can be tested and rejected, then deaf performance on Twenty Questions would not be directly explained by general differences in cognitive ability or language use. This would then leave open the possibility that similarities in problem-solving and categorisation in deafness and autism arise out of a similar developmental path, for early language delay and later semantic organisation.

#### **Chapter 4: Problem-solving in deaf and hearing adults<sup>17</sup>**

The second experiment aimed to replicate Marschark & Everhart's (1999) findings for deaf performance on the TQT, while testing some of the alternative explanations for why deaf and hearing problem-solving may differ. For feasibility reasons, and to rule out other confounding factors (such as developmental delays), this was done in the first instance by comparing TQT performance in a pilot group consisting of nine deaf adults and 27 hearing adults.

The first explanation to test was that general cognitive ability differences may have driven the group differences observed. As discussed in the previous chapter, there are a number of methodological difficulties with assessing IQ accurately and reliably in deaf populations – but that does not mean that it is impossible to establish some level of general ability. This is definitely possible with an appropriate NVIQ task, such as the Matrix Reasoning task from the WASI (Wechsler, 1999). In matrix reasoning, participants must complete a series of pattern-based sequences. Success depends on recognising the logic behind such sequences, but this is not thought to rely on language-based skills. Provided that task instructions are provided in a format accessible for deaf participants or with the assistance of a suitable interpreter, such tasks can provide fair assessments of non-verbal reasoning skill (Braden, 1992).

The measurement of verbal abilities is more problematic: at the time of writing, no standardised indices of sign language vocabulary knowledge or ability were available, although some were in development (Mann, personal communication). In the absence of such a measure, the Similarities task of the WASI was deployed as a rough index of Spoken English category knowledge, rather than general VIQ. On the Similarities task, the experimenter reads out a pair of words (such as COW and BEAR) and the participant must respond with a word or description that indicates in what way they are alike (e.g. ANIMALS). The test was chosen instead of the Vocabulary subtest of the WASI because of the greater expressive demands of the latter. For vocabulary, participants must define the meaning of different terms and sometimes describe a context in which that word is used. The Similarities test, in contrast, can largely be answered with single-word responses such as animal, clothes, etc. (Wechsler, 1999).

---

<sup>17</sup> Findings from this chapter and the one that follows were presented as a poster at the International Meeting for Autism Research 2012 in Toronto, Canada. (“Verbal problem-solving in deafness and autism spectrum disorders”. B. Alderson-Day, IMFAR, May 17<sup>th</sup>-19<sup>th</sup>, 2012).

Based on the assumption that Marschark & Everhart (1999) were correct in considering their groups of participants to be equal in general abilities, it was hypothesised that deaf participants would ask less efficient questions on the TQT even when controlling for the influence of NVIQ level (on the Matrix Reasoning task) and Spoken English knowledge (on the Similarities task).

The second explanation to test was whether differences in problem-solving performance were attributable to differences in category use in signed languages. A way to examine this is to measure the use of basic and superordinate category terms in ordinary descriptions of scenes, in a similar manner to neuropsychological measures of descriptive abilities, such as the Boston Naming Test (Kaplan, Goodglass, & Weintraub, 1976). If superordinate category terms are used less in signed languages in general, then this may be expected to be evident in how users of sign language go about describing such scenes, in contrast to users of Spoken English.

To assess this, a novel picture description task was designed. The task consisted of a range of illustrated common scenes that could be either described in terms of the superordinate groupings present, or in terms of more basic-level categories. For example, a farmyard scene may contain *animals, plants, vegetables* and *vehicles*. Alternatively, the same scene could be described as containing *horses, cows, trees, flowers, carrots* and so on. On the task, participants were asked to either a) generally describe what was happening in each scene, or b) identify the main differences between two very similar scenes containing subtle differences. Both a) general descriptions and b) specific differences were used, in order to provide two potential ways of eliciting the use of category terms. Based on the differences in superordinate use and frequency in signed languages, it was hypothesised that deaf participants may utilise superordinate terms less than hearing participants in their descriptions of scenes. In addition, it was predicted that the use of such terms on the descriptions task would be significantly related to their use in questions on the TQT.

## **Experiment 2: Cognitive and linguistic differences in deaf and hearing problem-solving**

### **Method**

#### *Participants*

A group of nine adults who are deaf (DA henceforth) were recruited from the Bristol area to take part in the study (6M, 3F). All participants had a hearing loss of some sort since the age

of 5 or younger. Eight of the nine participants were moderately to profoundly deaf; one participant referred to herself as “severely hard-of-hearing”<sup>18</sup>. The causes of hearing loss varied considerably across the group and were in some cases unknown (see Appendix 1 for a full breakdown). All were users of British Sign Language, although three participants stated that they were happy using BSL or speech. The DA group were compared with a larger group of hearing adult (HA) participants recruited from undergraduates and postgraduates at the University of Edinburgh ( $n = 27$ ; 14 M/13 F). Age and education data for the two groups are shown in table 5. No significant differences were observed between the groups for age or years of education.

### *Recruitment and settings*

All participants in the DA group were recruited via contact with a local community centre and with the assistance of a British Sign Language interpreter with links to the Deaf community in Bristol. Potential participants were provided with study information prior to consent and had the study explained by the interpreter. Following this, written consent was sought prior to participation. All testing with DA participants took place at the community centre. For HA participants, consenting and testing took place in a quiet room at the Psychology Department in Edinburgh. (All study procedures were approved by the Psychology Ethics committee at the University of Edinburgh).

### *Cognitive profile*

Table 5 displays group statistics on the cognitive profile of DA and HA participants based on the Similarities and Matrix Reasoning subtests of the WASI (Wechsler, 1999). The groups significantly differed in their T-scores for Similarities (HA>DA) and a trend was observable in the same direction for matrix reasoning T-scores (HA>DA). Age-equivalent (i.e. mental age) scores were not an appropriate check on matching in this instance, due to the age of participants, so raw subtest scores were used instead<sup>19</sup>. As for the T-scores, the groups

---

<sup>18</sup> This participant’s data were marked for later analysis and excluded from the remaining group data where they notably differed.

<sup>19</sup> Age-equivalent scores in the WASI are generally only available for child and adolescent samples (Wechsler, 1999). Scores over a certain cut-off are given an “adult level” age, but this does not continue to be adjusted for older participants. For example, a raw score of 35 or over is given an age-equivalent score of 27.5 years. In a high-scoring, adult sample, scores over 35 will be common, and

significantly differed in Similarities raw score. ( $DA_{M(SD)} = 33.89 (4.43)$ ,  $HA_{M(SD)} = 38.44 (3.86)$ ;  $t = -2.988$ ,  $df = 34$ ,  $p = .005$ ). The groups also differed in Matrix Reasoning raw score, albeit to a lesser degree ( $DA_{M(SD)} = 26.44 (1.88)$ ,  $HA_{M(SD)} = 29.30 (3.43)$ ;  $t = -2.365$ ,  $df = 34$ ,  $p = .024$ ).

### *Design*

A between-groups design was used to compare deaf and hearing participants on the Twenty Questions Task and a short battery of control tasks.

### *Materials & Procedure*

#### *- The Twenty Questions Task (TQT)*

The materials and procedure deployed in this experiment were broadly similar to the baseline condition in experiment 1: participants viewed an array of everyday items and were permitted 10 questions per trial to identify a selected target. Key ways in which the procedure differed were as follows.

First, DA participants were permitted to attempt the task using their preferred mode of communication, with an interpreter present to translate task instructions and participant questions. Second, to avoid ceiling effects in testing an adult sample, the task array for all participants was doubled from the 24-item set used in the preceding experiment, to a 48-item set. (This also made the set size more similar to the 42-item set used by Marschark & Everhart, 1999). Finally, all participants completed three TQT trials; two allowing elimination followed by one where it was prohibited. (A non-elimination condition was included as an indicator of changes in the profile as a result of executive load).

In addition, participants were asked to rate their level of familiarity with Twenty Questions, on a scale from 1 (Very Familiar) to 5 (Very Unfamiliar) before attempting the task.

---

result in most participants receiving the same age-equivalent score (27.5). For this reason, raw scores are preferable to age-equivalents in older groups.

### *- Picture Descriptions*

The picture descriptions task was devised specifically for the present experiment. To test basic and superordinate label use, a set of five images were selected that showed various everyday objects and scenes in cartoon form. Scenes showing a banquet, the inside of a greenhouse, a farmyard, a busy street and a desert were chosen from a children's picture book (Brocklehurst, Doherty, & Milbourne, 2002). An example scene is displayed in figure 13.

Two methods were used to elicit verbal descriptions of the scenes. First, participants were asked to "give a brief description of what's happening in the scene" and their responses were scored for use of basic and superordinate terms. Descriptions were scored for all the participants' utterances or signs up until they stopped describing the scene (usually no longer than one or two sentences). Second, each scene was displayed next to a counterpart scene which was identical with the exception of specific details (See figure 14). In the twin scene some collections of items (e.g. all the tools, or all the animals) were moved or changed in orientation, and participants were asked to describe three main differences between the scenes. This "spot the difference" paradigm was used in addition to the general descriptions to create contexts where it made sense to describe certain pictures in basic terms and others in superordinate terms<sup>20</sup>.

### *- Plan Construction*

To allow for comparisons with the data from experiment 1, the plan construction task was also deployed after the three TQT trials. As participants were no longer planning a set of questions to use on a subsequent TQT trial, participants were asked to construct a plan based on the premise of what they would do if they were about to play the game again. Participants were also allowed to view the test array from the previous TQT.

---

<sup>20</sup> Pictures for the task were piloted with a sample of psychology class undergraduates ( $n=47$ ) to establish sufficient elicitation of basic and superordinate terms.



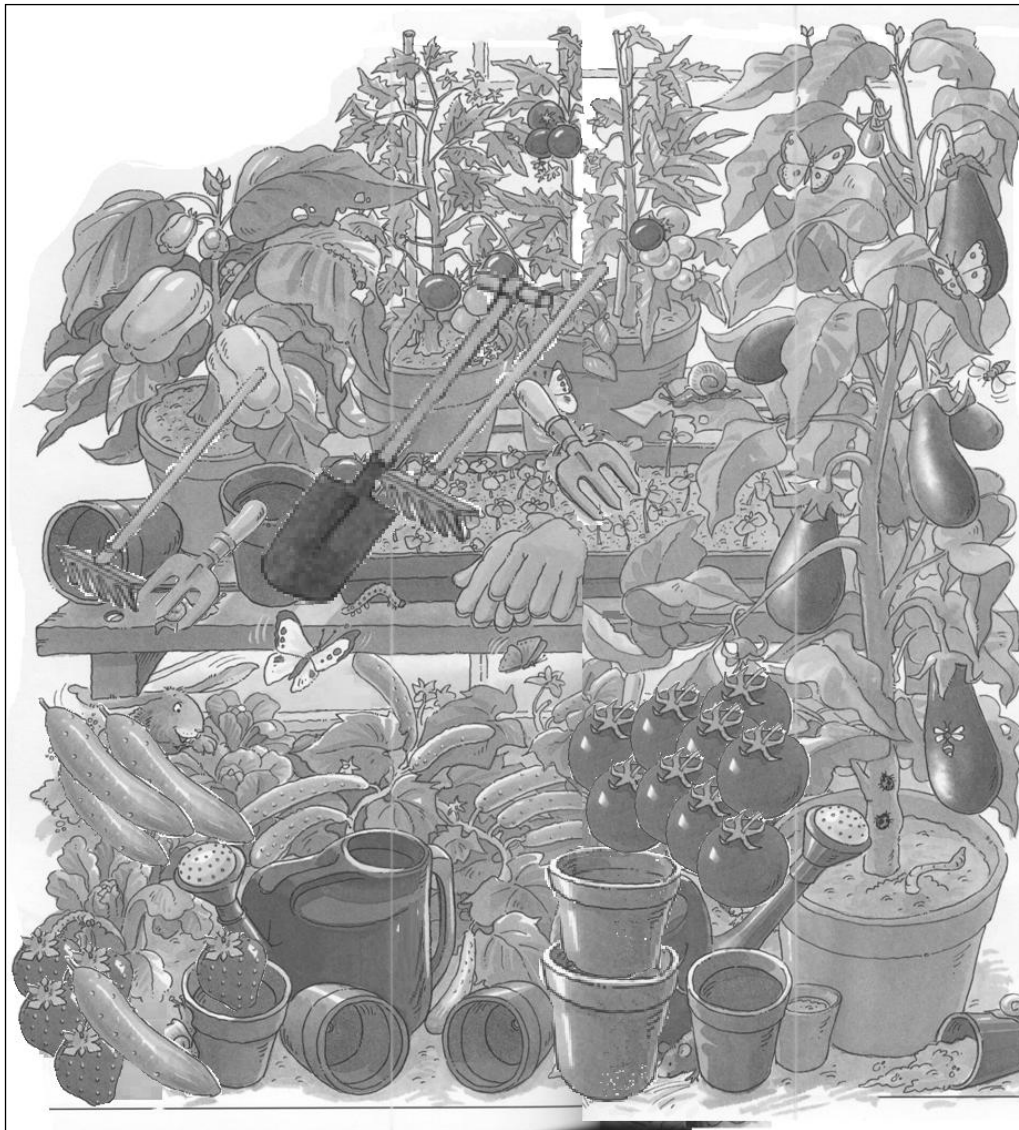


Figure 13. Example stimulus for the general description on the picture descriptions task.

Table 5. Age and cognitive characteristics for deaf and hearing adults

	Deaf adults (DA)		Hearing adults (HA)		Sig.
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Age	26.00	2.18	25.26	4.33	0.628
VIQ: Similarities (T-score)	47.67	7.73	55.44	6.51	0.005**
NVIQ: Matrix Reasoning (T-score)	52.67	3.97	58.00	8.05	0.066
Years of Education	16.63	4.21	17.59	2.50	0.552

\*\*  $p < .01$

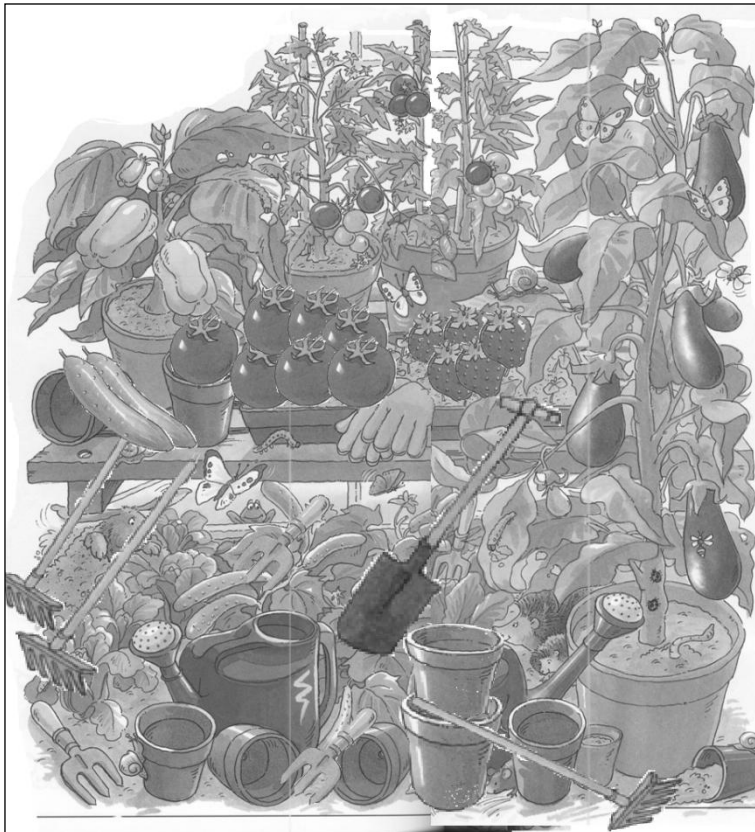
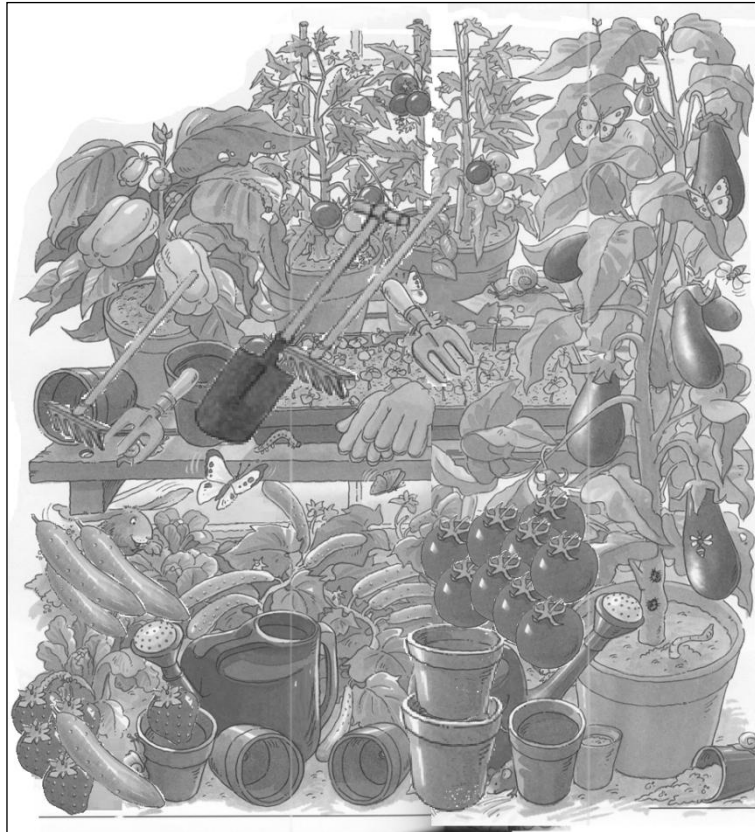


Figure 14. Example stimuli for picture differences on the picture descriptions task.

Also scored was the number of questions used per trial (QT) and the types of questions used (grouping and guessing). No hypotheses were made as to the content of participants' questions, so this was not analysed. Analysis of variance, analysis of covariance and unpaired t-tests were used to compare performance in the two groups.

#### *- Picture Descriptions*

Picture descriptions were scored as frequency counts of basic and superordinate category terms used within each condition (general description vs picture differences). Counts for basic and superordinate terms were compared in each condition for each group using a multivariate analysis of variance, while regression analysis was used to assess the relationship of each of these variables to TQT performance.

#### *- Plan Construction*

The scoring for plan construction also followed experiment 1, with expected QQ scores being assigned to each question selected in each plan. As the TQT array was larger for the adult study, expected QQ scores were assigned that reflected the 48-item array. (As QQ reflects a proportion of information eliminated this still provided scores on the same scale as the previous study). Friedman's ANOVAs were then run in each group to assess narrowing of plans.

## **Results**

### *Twenty Questions Task*

As would be expected for an adult sample, participants in both groups performed the task with relative ease, with no participant failing to identify the target within 10 questions on any trial. An unpaired t-test indicated that the two groups had comparable levels of game experience ( $DA_{M(SD)} = 2.22 (0.97)$ ,  $HA_{M(SD)} = 2.67 (1.54)$ ;  $t = -1.01$ ,  $p = .323$ , *n.s.*). Table 6 shows the mean group scores for TQT performance. An analysis of variance indicated that the mean differences between the groups were observable for question quality ( $F(1,34) = 4.437$ ,  $p = .0436$ ,  $\eta_p^2 = .115$ ) with participants in the HA group asking slightly more efficient questions on average.

**Table 6. Task scores for deaf and hearing adults on Twenty Questions , Picture Descriptions and Plan Construction tasks.**

	Deaf adults (DA)		Hearing adults (HA)		Sig.
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
<b>TQT</b>					
QQ	0.33	0.04	0.37	0.04	.044*
QT	5.89	0.87	6.22	0.91	.341
Grouping (%)	75.35	4.55	75.79	9.63	.895
Guessing (%)	18.96	7.68	14.67	12.34	.336
<b>PD</b>					
General: Basic	8.33	4.98	10.11	4.76	.336
General: Superordinate	4.56	1.24	5.26	2.03	.344
Difference: Basic	11.89	2.37	11.56	2.78	.252
Difference Superordinate	2.22	0.97	2.85	1.51	.749
<b>PC</b>					
QQ1	0.37	0.13	0.44	0.11	.092
QQ2	0.24	0.14	0.22	0.11	.655
QQ3	0.24	0.16	0.19	0.12	.307
QQ4	0.22	0.13	0.14	0.09	.045*
QQ5	0.19	0.15	0.15	0.09	.308

\* significant at  $p < .05$

TQT = Twenty Questions Task, PD = Picture Descriptions, PC = Plan Construction, QQ = Question Quality, QT = Questions Taken.

Two methods were deployed to take into account the contribution of cognitive ability scores to question quality scores. First, analyses of covariance were deployed using a) the T-scores from the Matrix Reasoning and Similarities tests and b) the raw scores from those same tests as covariates. Second, subsets of deaf and hearing participants were matched for WASI scores on each measure (to within 5 points), and then compared using t-tests.

With the inclusion of reasoning scores as a covariate in an ANCOVA model, a trend towards a main effect of group was observed ( $F(1,33) = 3.126, p = .086, \eta_p^2 = .087$ ), suggesting that deaf and hearing participants may have differed in their efficiency even when NVIQ was taken into account. Almost exactly the same trend was evident when raw scores for matrix reasoning were included instead (group main effect:  $F(1,33) = 3.137, p = .086, \eta_p^2 = .087$ ). With the inclusion of Similarities T-scores, this trend was no longer evident ( $F(1,33) = 2.191, p = .148, \eta_p^2 = .062$ ), and this was also the case when raw scores were used ( $F(1,33) = 2.179, p = .149, \eta_p^2 = .062$ )<sup>21</sup>. In each of the models, the observed covariate effects were never as strong as the group contrast: for matrix reasoning T-scores the partial eta squared effect size was 0.015 ( $F(1,33) = 0.505, p = .482$ ), while for similarities it was 0.022 ( $F(1,33) = 0.740, p = .396$ ). This implied that neither covariate was as effective a predictor of question quality as group membership (and may simply have been reducing power with their inclusion in the model). Nevertheless, group differences in mean QQ could not be clearly demonstrated once all of the relevant IQ covariates were taken into account, especially for VIQ<sup>22</sup>.

Comparisons using matched subsets added to this picture. A t-test comparing the 9 DA participants with 19 HA participants who were matched for NVIQ ( $DA_{M(SD)} = 52.67 (3.97)$ ,  $HA_{M(SD)} = 54.11(5.99)$ , *n.s.*) showed a significant mean difference in question quality ( $DA_{M(SD)} = 0.334 (0.042)$ ,  $HA_{M(SD)} = 0.369 (0.040)$ ,  $t(26) = -2.075, p = .048, d = -0.83$ )<sup>23</sup>. A sample of 11 HA participants matched to the DA group for VIQ ( $DA_{M(SD)} = 47.67 (7.72)$ ,  $HA_{M(SD)} = 49.55(4.41)$ , *n.s.*) also scored higher for question efficiency, at a level approaching significance ( $DA_{M(SD)} = 0.334 (0.042)$ ,  $HA_{M(SD)} = 0.371 (0.036)$ ,  $t(18) = -2.052, p = .055, d = -$

<sup>21</sup> As per the assumptions underlying analysis of covariance, group\*covariate interaction effects were also tested to check for homogeneity of regression slopes. For both covariates no significant interactions with group were observed.

<sup>22</sup> An ANCOVA model including both NVIQ and VIQ covariates was also run to assess for their combined contribution, although for parsimony this is not reported here. With both covariates included, the observed effect size for Group was similar to that seen for the Similarities only model ( $\eta_p^2 = 0.62$ ) and again greater than the contribution of either of the covariates, or their combination.

<sup>23</sup> The groups were also comparable for matrix reasoning raw score: ( $DA_{M(SD)} = 26.44(1.88)$ ,  $HA_{M(SD)} = 27.68(2.60)$ , *n.s.*).

0.93)<sup>24</sup>. Thus, group differences in question quality were apparent when matching subgroups of participants (in contrast to using ANCOVA), but such differences were still not consistently separable from the influence of verbal IQ.

No group differences were evident on the number of questions used per trial ( $p = .341$ ,  $\eta_p^2 = .027$ , *n.s.*), suggesting that deaf and hearing participants largely used the same number of questions to reach the target on each trial. As table 6 indicates, the percentages of grouping and guess questions used by each group were also very similar: both DA and HA participants utilised grouping for about 75% of their questions, and no significant differences were observed (all  $p > .300$ , all  $\eta_p^2 < .03$ ).

#### - Within-group differences

Changes across the three TQT trials in each group were assessed using a mixed 2x3 ANOVA (group x trial). In particular, the difference between the first two trials, where elimination was allowed, and the third trial, where it was prohibited, were of interest. For question quality a significant group\* trial interaction effect was evident ( $F(1,34) = 4.312$ ,  $p = .045$ ,  $\eta_p^2 = .113$ ), although as figure 15 shows, this reflected a widening difference in QQ between the groups across the three trials. There was no evidence of a specific problem with the third, non-elimination trial in deaf participants, as QQ in both groups dropped for this trial<sup>25</sup>. No significant group\*trial interaction effects were observed for the number of questions used per trial, or the respective rates of grouping or guessing (all  $p > .700$ , all  $\eta_p^2 < .005$ ). This strongly suggested that deaf and hearing participants were largely unaffected by not being able to eliminate items from the TQT set.

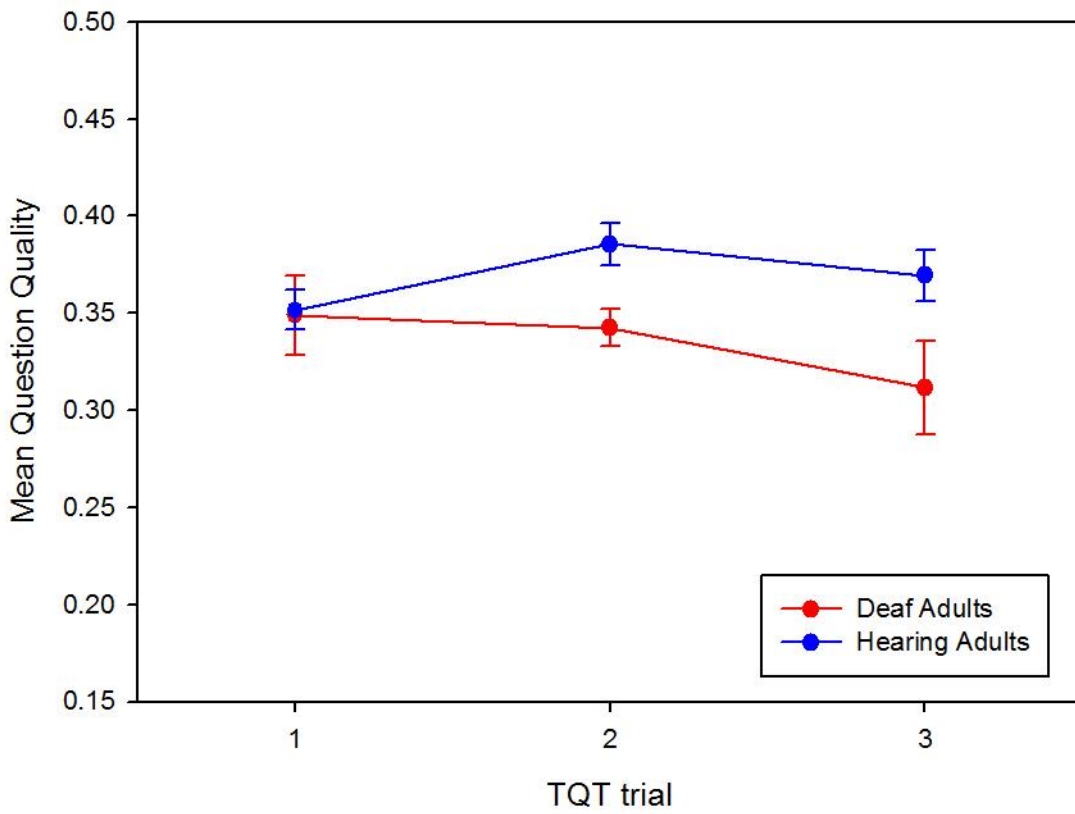
#### Picture Descriptions

A multivariate mixed analysis of variance was used to compare frequencies of basic and superordinate category use in both groups across both forms of the picture descriptions task (general descriptions, and describing differences). No significant differences were observed

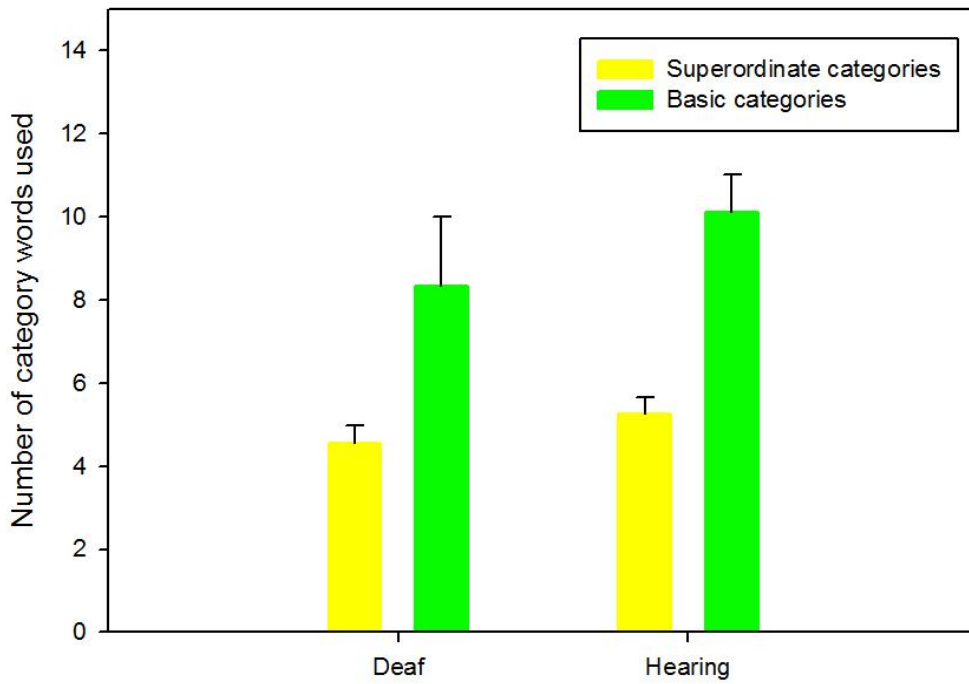
---

<sup>24</sup> No significant difference was evident in raw scores between these groups ( $DA_{M(SD)} = 33.89(4.43)$ ,  $HA_{M(SD)} = 35.09(2.43)$ , *n.s.*). T-tests comparing age, years of education and game experience were also run for each pair of subsets, but no significant differences were evident.

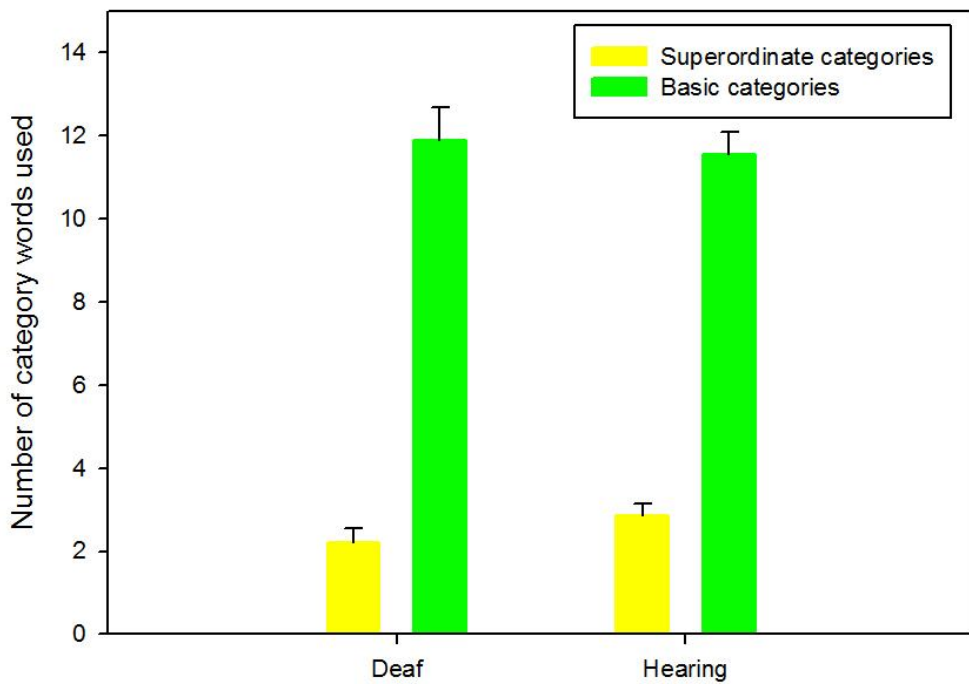
<sup>25</sup> As with the analysis of mean QQ, these analyses were also run with VIQ and NVIQ covariates, producing similar results to the overall mean effect. Importantly neither covariate interacted significantly with specific trials, suggesting that they had a uniform effect on performance.



**Figure 15. Change in question quality across the Twenty Questions Task (TQT) in deaf and hearing adults.**



**Figure 16.** Use of basic and superordinate category terms in general descriptions made by deaf and hearing adults.



**Figure 17.** Use of basic and superordinate category terms in picture differences described by deaf and hearing adults.



between the groups (all  $p > .250$ , all  $\eta_p^2 < .04$ ; see table 6). As figures 16 and 17 show, the general descriptions elicited more superordinate terms ( $F(1, 34.000) = 33.381, p = .0001, \eta_p^2 = .495$ ) whereas picture differences prompted increased use of basic categories ( $F(1, 34.000) = 6.285, p = .017, \eta_p^2 = .156$ ). This suggested that there was no evidence of reduced superordinate use in the deaf participants, nor was this use irregularly modulated by task demands: DA and HA participants used similar rates of basic and superordinate categories, across two different conditions.

To assess for the relation of picture description performance to TQT performance, the counts for basic and superordinate category use in each condition were entered into a backwards regression model with QQ as the dependent variable (using all participants combined). No significant model was returned, indicating that none of the picture description scores significantly predicted TQT performance either individually or in combination ( $N = 36$ , all stan. betas  $< .270$ , all  $p > .149$ ). The lack of group differences in picture descriptions and lack of any predictive relations between picture description scores and TQT scores suggested that the use of basic and superordinate category terms could not explain group differences in questioning efficiency.

### *Plan Construction*

On the planning task both groups showed evidence of selecting a hierarchical sequence of questions (see figure 18). The running of Friedman's ANOVAs within each group indicated significant differences between questions for both DA ( $X^2 = 9.216, df = 4, p = .049$ ) and HA ( $X^2 = 53.529, df = 4, p < .001$ ) participants. Comparisons between groups indicated no significant difference in the mean efficiency of participant plans ( $DA_{M(SD)} = 0.253 (0.067)$ ,  $HA_{M(SD)} = 0.230 (0.060)$ ;  $F(1, 34) = 1.017, p = .320, \eta_p^2 = .029, n.s.$ )<sup>26</sup>, nor for most of the individual questions in each plan.

However, differences approaching significance for questions 1 (HA>DA) and 4 (DA>HA) suggested some differences in the mean gradient of participants' plans, with hearing participants starting with more general questions but narrowing more than deaf

---

<sup>26</sup> Participants' question scores for individual questions in plans were non-normally distributed (as in experiment 1), necessitating the use of non-parametric tests. The mean question quality of a plan, being an average *across* questions selected, was normally distributed however. For this outcome and the following post-hoc analysis of plan gradient it was possible to use parametric comparisons.

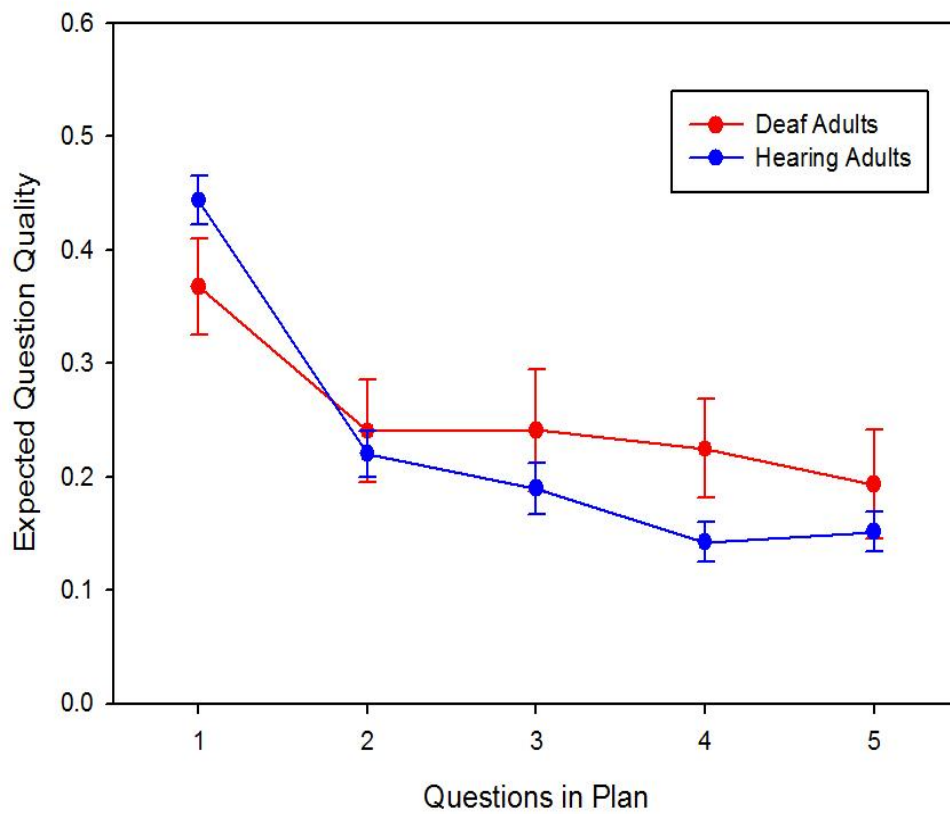


Figure 18. Mean efficiency of plans constructed by deaf and hearing adults.

participants<sup>27</sup>. A post-hoc comparison of plan gradient (simply defined as the difference between question 1 and question 5) confirmed this impression: HA participants were quicker to move from general to specific questions in their plans than DA participants ( $DA_{M(SD)} = 0.174 (0.182)$ ,  $HA_{M(SD)} = 0.292 (0.118)$ ;  $F(1,34) = 5.085$ ,  $p = .031$ ,  $\eta_p^2 = .130$ ), an effect that was evident even when controlling for WASI performance ( $F(1,32) = 5.140$ ,  $p = .030$ ,  $\eta_p^2 = .138$ )<sup>28</sup>. Furthermore, this measure correlated strongly with mean question quality on the TQT itself ( $r = .468$ ,  $n = 36$ ,  $p = .004$ ), indicating that greater plan gradient was associated with more efficient question use during problem-solving<sup>29</sup>.

## Discussion

In experiment 2, deaf adult participants appeared to ask less efficient questions than hearing adult participants on the Twenty Questions Task. There were no differences between the groups in age, years of education or game experience, suggesting that they did not play a significant role in explaining group differences. In contrast, there were differences in verbal and non-verbal IQ scores between the two groups, whether measured as T-scores or raw subtest scores. Using two different methods to take into account the influence of these factors, trends were apparent to suggest that deaf adults asked less significant questions than hearing counterparts with similar levels of verbal and non-verbal ability. Nevertheless, group differences in this kind of problem-solving could not be clearly separated from effects of category knowledge (Similarities) and non-verbal reasoning (Matrix Reasoning) in this sample.

Notwithstanding these caveats, the pattern of results seen here across different task measures is in line with previous findings. The presence of potential group differences in questioning efficiency, but not other other measures (such as guessing rates) is consistent with the results of the second experiment conducted by Marschark & Everhart (1999), that is, their follow-up experiment with deaf college students. As in that study, the deaf adults tested here were not more likely to require more questions on the task, but the amount of information they eliminated with each question was generally smaller than hearing participants. The present data add to this picture by suggesting that this is not *just* the product of gross

---

<sup>27</sup> As table 6 shows, the p-value for question 4 was in fact 0.045, but this is only approaching significance once the alpha value is corrected for multiple comparisons.

<sup>28</sup> This difference was also evident when raw scores for Similarities and Matrix Reasoning were used as covariates.

<sup>29</sup> These correlations were also run for each group separately, yielding consistent results for both.

differences in general cognitive ability, but it is important to note that they do not demonstrate a completely independent effect of hearing status.

There were some dissimilarities between these data and Marschark & Everhart's that need to be noted. It appears that the participants in this study performed at a higher level on the task than the deaf college students. No differences in grouping were evident here, with both deaf and hearing participants using such questions about 75% of the time. In contrast, the deaf adults in Marschark & Everhart's study used grouping questions only 50% of the time, which was significantly lower than the questioning rate observed in the hearing group. Why exactly this is the case is unclear without further information: as college students it is quite possible that Marschark's participants were on average younger than those tested here ( $Age_M = 26$ ), although no specific age data were provided in the original paper. Another possibility is that the participants in the present study were more familiar with the game; of the nine participants tested here, only one described themselves as in any way unfamiliar with the game, whereas in Marschark & Everhart (1999) only five of the 14 deaf participants tested had previous game experience (compared to 12/14 hearing students). Finally, Marschark's study appeared to follow Mosher & Hornsby's (1966) paradigm in not allowing elimination of items during search, whereas two of the three trials deployed here did allow removal of eliminated items. As such, it may be that this form of the TQT was easier for participants, although it is not clear why this would make them use grouping questions more.

Before discussing some of the potential explanations of this profile, it needs to be acknowledged also that these data are drawn from a small sample of deaf adults. Rather than suggest that these data are truly representative of the wider population of deaf individuals, it should be treated as it was intended – as a pilot sample, where some preliminary concerns about Marschark & Everhart's findings could be explored.

#### *Basic and superordinate category use in picture descriptions*

The picture description task was designed to elicit use of basic and superordinate category terms, and compare their use in deaf and hearing participants. Some differences were evident between conditions on the task: when participants were asked to provide brief descriptions they were more likely to use superordinate terms, and when asked to describe differences they used more basic terms. It is possible that the picture differences paradigm encourages a focus on local details, rather than more general groupings (being essentially a game of "Spot the Difference"). For example, on the former condition a typical response was "It's a

greenhouse of some sort, there's some fruit, and some tools, the tools are on the table". In contrast, picture differences prompted descriptions like "The tomatoes are on the floor, the rakes are on the table, and the strawberries have moved".

Importantly, though, no group differences were evident on any outcome or condition of this task. It is possible that the task, being custom designed for the present study, was not sufficiently powerful to elicit differences in superordinate use between deaf and hearing participants and it may be that more obscure category terms (e.g. mammal) would be required to do so. But all of the category terms used on the picture description task, while common and familiar, were terms that would be required on the TQT. As such there is at least an ecological validity in focussing on the use of those terms specifically in picture descriptions. Furthermore, these descriptions did not predict performance on the TQT task to a significant degree. This fact, combined with the lack of group differences in basic and superordinate use, suggests that differences in general category use cannot explain the group differences observed on Twenty Questions.

#### *How similar are these data to the ASD Twenty Questions profile?*

Direct comparison of these data with those reported for ASD participants in the previous experiment and in prior studies is tricky because of differences in the age of participants tested and the use of a larger task array for deaf adult participants. As with the comparison with Marschark & Everhart's participants, the deaf participants in this study appeared to use fewer questions overall and more grouping than ASD participants, and this could be due to age differences. Nevertheless there are some key similarities and differences that can be observed.

The obvious similarity with the ASD profile seen previously (Alderson-Day, 2011; Alderson-Day & McGonigle-Chalmers, 2011) is the suggestion of reduced question quality in deaf participants compared to hearing participants. The presence of this, despite otherwise competent performance, is similar to the most able ASD participants described in the previous chapter, suggesting a common or analogous difficulty with selecting effective category questions in both groups.

Differences between the groups were evident in the within-task variations of the TQT, and in the responses to plan construction. Within the task, the deaf participants differed from ASD participants in their responses to elimination: whereas ASD participants specifically required

more questions to reach the target when elimination was prohibited, deaf adult participants showed no specific drop in their performance (on any outcome) in relation to this change.

On the planning task, deaf participants produced plans that narrowed items hierarchically in a way that was not typical of the whole ASD group. The most able ASD participants could produce “narrowing” plans, but the efficiency of their plans was still lower than the most able typically-developing controls. Deaf participants did not create less efficient plans in general, but the gradient of their plans was significantly different from controls. As a post-hoc measure this should be interpreted with caution, but its relation to question quality rates during Twenty Questions arguably lends it a certain amount of validity: participants who narrowed their plans to a greater degree also performed more efficiently on the problem-solving task.

To a certain extent, this pattern of similarities and differences between deaf and ASD participants is not surprising. If differences with questioning efficiency are indicative of differences in concept and category processing, then this is exactly what ASD and deaf participants would be proposed to possibly share. In contrast, being able to handle an unchanging TQT array effectively, or plan a series of narrowing questions, is an executive capability (as argued above). Given that executive dysfunction is strongly associated with ASD, but not with deafness, it would be surprising if the two groups *did* resemble each other in this way as well.

However, what is important to establish is whether this profile is similar across deaf and autistic participants of a similar age, on the same task. The apparent presence of this profile in deaf adults highlights a similarity, and serves to rule out some alternative hypotheses about problem-solving differences, but it does not show that the development of such skills is the same in both deaf and ASD individuals. It is also worth reiterating that the data here do not show that such differences can be clearly separated from IQ factors in the case of deafness. Testing this question again in a sample of deaf children would serve to clarify this issue further.

In addition, it is also important to consider the next question: if deaf and ASD participants were to exhibit a similar profile on the TQT, specifically in the efficiency of the questions they select, the why does this occur? That is, what is it about the processing of semantic categories that makes them harder to utilise during problem-solving?

### *Semantic organisation in deaf individuals*

Marschark & Everhart's (1999) original interpretation of deaf problem-solving performance was that it reflected how the early language experiences of deaf children may differ from those of hearing children. Specifically, they suggested that games like Twenty Questions provide "informal educational experiences" that involve learning about important social and cognitive rules; experiences that the deaf children of hearing parents will have less access to than hearing children with hearing parents (Marschark & Everhart, 1999, p. 79). In essence, deaf children are likely to have less experience of such language games, and this leaves them at a disadvantage when attempting such tasks at a later age.

Following the study by Marschark & Everhart (1999), research conducted by McEvoy, Marschark & Nelson (1999) and Marschark et al. (2004) added to this hypothesis. In McEvoy et al. (1999), 136 deaf undergraduate students and a corresponding group of hearing college students were assessed on a written semantic association task designed to examine the "mental lexicon" of deaf and hearing individuals. Participants were presented with lists of common words (e.g. AIRPLANE, CLOCK, PARROT), and next to each word asked to write "the first word that came to mind that was related to the printed word" (p. 314). Participants' responses in each group were then assessed for their level of frequency and agreement, within and between deaf and hearing groups.

The main finding of the study was that, contrary to expectation, the semantic associations of deaf and hearing participants were largely similar: 59 of the 79 of stimulus words used produced high agreement ( $r = >.7$ ) between the two groups. However, there were also subtle differences evident between the groups. First, deaf participants were more variable in their responses within group: whereas hearing participants produced 13 unique association words on average for each stimulus, deaf participants produced 15 (a statistically significant difference). Second, some individual words (e.g. BARN) produced very little agreement between groups, when there was no *prima facie* reason to believe that that should be the case. Finally, deaf students were also significantly more likely to leave some responses blank, and offer no association word. McEvoy and colleagues argued that these subtle differences between the groups indicate higher variability in deaf lexical knowledge and lower accessibility of certain semantic concepts<sup>30</sup>.

---

<sup>30</sup> For earlier evidence of semantic association differences in deaf individuals see also Green & Shepherd (1975).

The suggestion of subtle differences in the lexicons of deaf and hearing students was supported in a follow-up study by Marschark et al. (2004). In this study, another large sample of deaf college students attempted a semantic association task, but this time the stimulus words were deliberately varied based on their category level. Twenty words provided a superordinate level cue (e.g. ANIMAL) where the most common response was a basic level word (e.g. DOG) and twenty words provided the reverse, i.e. a basic category word that would be followed by a superordinate (most common responses were based on the USF set of norms; Nelson, McEvoy, & Schreiber, 1998). The responses from deaf students were then compared with existing norm data from a sample of hearing students.

As in McEvoy et al. (1999), the responses of deaf participants largely agreed with those seen in the hearing norms, but the level of agreement among the sample of deaf participants was significantly lower than in the hearing sample; that is, deaf participants were more heterogeneous in their responses. In addition, the contrast between superordinate and basic level cues revealed further anomalies. The strength of associations between basic and superordinate categories did not vary between deaf and hearing students, but the direction of the association modulated responses in deaf individuals specifically. Deaf students were more likely to respond to a basic exemplar with a superordinate category name than they were to respond to a superordinate with a basic exemplar, suggesting an asymmetry in their category relations that was not evident in the data from hearing students. Among the deaf students, those with the best reading abilities were also those whose responses resembled hearing students' the closest (Marschark, et al., 2004).

The presence of such an asymmetry, alongside relatively intact and similar lexicons to hearing students, could then provide a basis for the apparent problems with concept selection seen in deaf problem-solving. To further investigate this notion, in a second experiment Marschark et al. (2004) tested a subset of their deaf sample ( $n = 18$ ) on a verbal analogies task where participants had to complete a series of category-based analogies, e.g. "mouse: animal; poison ivy: ?" where a suitable answer would be "plant". The relations that the analogies relied on were varied to test directions of association; superordinate (animal > cat), co-ordinate (helicopter > plane) and subordinate (dog – Labrador) analogies were included, alongside analogies testing rhyming, part-whole, predication and sound-related associations.



Compared to hearing participants ( $n = 21$ ), deaf participants made significantly more errors overall, and again demonstrated differences in the association of directions:

“hearing students solved significantly more analogies in which they had to provide superordinate terms (mouse : animal :: poison ivy : *plant*) than ones in which they had to provide subordinate terms (book : dictionary :: tool : *screwdriver*)...whereas deaf students solved significantly more subordinate than superordinate analogies”  
(Marschark, et al., 2004, p. 57)

Thus, whether making semantic associations or using them in reasoning, deaf participants showed asymmetries in their semantic relations that were not the same as those seen in hearing students. The importance of the *direction* of asymmetry is not clear, as on the one hand deaf participants were more likely to respond to basic exemplars with superordinates than the reverse (experiment 1), but then in analogous reasoning struggled to produce superordinate terms (experiment 2). Nevertheless, such asymmetries imply genuine issues with semantic organisation rather than the strength of relations *per se*. If a relation was weak between a basic level term and its corresponding superordinate, then producing either would arguably be a problem. But if going in one direction of association is difficult, depending on the task (superordinate to basic in experiment 1, basic to superordinate in experiment 2), this suggests a qualitative difference in the way in which basic and superordinate concepts are linked; that is, a difference in semantic organisation.

It could be that these differences underlie how accessible concepts are when they are required in problem-solving; how easy it is to select one category over another when searching the set. If so, it should be the case that levels of asymmetry are related to TQT performance; that is, more asymmetry – insofar as it can be taken as an indicator of semantic organisation – should be associated with greater inefficiency in problem-solving. If this association can be demonstrated, then this would support the strong form of argument outlined above: similarities in deaf and autistic problem-solving arise out of atypical early language development, via atypicalities in semantic organisation.

## **Chapter 5: Problem-solving in deaf children**

The aim of the following experiment was two-fold. First, TQT problem-solving in a sample of deaf children was investigated, to assess whether they show the same problems with question quality as seen in a) deaf adults and b) children with autism spectrum disorder. Second, the relation between semantic organisation and TQT performance was investigated, to establish whether measures of semantic association could explain possible inefficiencies in problem-solving for deaf participants.

To attempt the latter analysis, a semantic decision task was devised. In McEvoy et al. (1999) and Marschark et al. (2004) the method of measuring semantic associations was based on written responses from participants and level of agreement across individuals. For example, a response of DOG would receive a score based on how many individuals in that group responded with DOG, and how that compares to the number of overall responses.

There are two problems with this sort of procedure. The first is practical: whereas the college students who took part in the above studies would be expected to have writing skills that are sufficient to attempt the task, it may be trickier for younger deaf children, who often still have problems with their written language as a consequence of language delay (Antia, Reed, & Kreimeyer, 2005). As such, a paradigm that did not require written language responses is preferable for studying semantic associations in this group.

The second issue is that this method arguably does not provide a direct metric of an individual's semantic associations. In scoring participants' responses based on their frequency across a group, the measures deployed by Marschark and colleagues only really reflect group agreement, or how typical an individual is of his or her group. What they do not do is provide a measure of how strong any given semantic relation is for any given individual.

For example, if 10 out of 12 participants in a group produced DOG in response to ANIMAL, still unknown is the strength of relation between category (ANIMAL) and exemplar (DOG) for each of those participants. Within the 10 who provided that response, the strength of their ANIMAL>DOG associations may still vary considerably. Providing a sensitive measure of such relations, within each participant, is important because it allows for more accurate assessment of how semantic decision performance relates to performance on other tasks.

One way of achieving this is to measure reaction times. A number of classic models of semantic memory, such as Collins and Loftus' (1975) spreading activation model, use reaction times on semantic tasks to indicate the strength and nature of relations between categories and exemplars. In such models, conceptual knowledge is typically organised as a series of nodes that are linked to one another based on their conceptual relations. Activation of a specific concept node, such as CAR, will raise the activation level of other nodes that are closely linked to it, either via subordinate relations (different types of car), co-ordinate relations to similar exemplars of the same category (BUS or CART) or superordinate relations (VEHICLE or TRANSPORT). The strength of links between nodes is determined by the frequency and use of the particular relation, such that the most common associations will have the strongest links in the network (such as CAR and VEHICLE) and prompt the fastest reaction times. The activation of a basic concepts will "prime" their common superordinates, and vice versa (Collins & Loftus, 1975).

The data from Marschark and colleagues' research pointed to asymmetries in written word associations for deaf students when prompted by basic or superordinate categories – asymmetries that were not seen in norms based on hearing populations (Marschark et al., 2004). For reaction times, asymmetries are in fact often seen in hearing participants for the naming of pictures, with superordinate terms prompting longer responses than basic terms (Jolicoeur, Gluck, & Kosslyn, 1984; Tanaka & Taylor, 1991). This is thought to reflect the perceptual and conceptual primacy of basic category terms – showing a picture of a dog will prompt the basic category of DOG before it prompts the more general superordinate category of ANIMAL.

On semantic decision tasks, however, the mutual priming of basic and superordinate terms can produce fairly equal reaction times, provided that there is a sufficiently frequent level of association. For example, Loftus (1973) compared reaction times on a categorisation task where participants were primed with either a superordinate category word or a basic exemplar. The categories and exemplars used were varied in terms of how strongly they evoked one another: some were associated with a high-frequency (e.g. tree-oak), some with a low frequency (e.g. cloth-Orlon) while others only offered a frequent association in one direction. For instance, in the pair "butterfly-insect", the exemplar prompted the category with high-frequency (butterfly>insect), but the category term evoked the exemplar with a low frequency (insect>butterfly). Loftus observed that reaction times to category and exemplar primes varied according to frequency of association: for both kinds of primes the fastest responses were for high-frequency pairs of words. Following this, reaction times were

fastest if the prime (category or exemplar) prompted a high-frequency association; that is, butterfly prompted a fast response to insect, but not vice versa. This suggested that the strength of association, rather than the category level of the prime, largely determined the speed of response (Loftus, 1973)

It should be possible, then, to use common or high-frequency pairs of basic and superordinate terms to examine the nature and direction of semantic relations, via measurement of differences in reaction times. To attempt this, a reaction time paradigm was devised following a method used by Gaffrey et al. (2007) to study semantic abilities in ASD individuals. The task displayed basic and superordinate category words consecutively, and participants were asked to indicate which “go together”. In one condition, participants were required to respond to items that went with a superordinate category term, e.g. “Does it go with ANIMAL?” followed by “DOG” (yes), “CAR” (no), and so on. In the other condition, the direction of the association was reversed, such that superordinate terms followed a basic category, e.g., “Does it go with DOG?” followed by “ANIMAL” (yes), “FRUIT” (no), “TRANSPORT” (no), “PET” (yes), and so on. Reaction times for accurate responses in each of the conditions were then compared, controlling for word length and word frequency.

Because high frequency pairings were used, it was assumed that typical reaction times for either direction of association should be reasonably equal<sup>31</sup>. If the directions of semantic relations are asymmetrical in deaf participants, then it would be expected that reaction times in the two conditions would significantly differ (although without a direct contrast to a hearing control group, this would not demonstrate *atypicality* per se). More important than such a difference, though, is the relation that any such difference would have with TQT performance. If semantic organisation underpins problem-solving inefficiencies, then the amount of asymmetry between different directions of association should correlate with task measures on the TQT.

A further point to note is that the word association data from Marschark and colleagues did not clearly predict a direction of asymmetry. As explained above, the deaf students in Marschark et al. (2004) were more likely to respond to a basic exemplar with a superordinate category than the reverse case. That is, they may readily supply ANIMAL when prompted with DOG, but not respond as readily with DOG when prompted with ANIMAL. If this is true for individual category relations, measured with reaction times, then

---

<sup>31</sup> This was supported in a pilot of the task with a small group of hearing adults ( $n = 5$ ). No significant differences in reaction times for basic-to-superordinate and superordinate-to-basic semantic decisions were observed ( $p = .503$ ). Full details of this analysis can be found in Appendix 2.

responses to the former condition, “DOES IT GO WITH (superordinate)?”, should be slower than responses in the latter condition, “DOES IT GO WITH (basic)?”.

However, the responses on the verbal analogies task deployed by Marschark et al. (2004) suggested that deaf participants also struggled to provide superordinate responses when prompted with a basic category, e.g. “hammer > tool; dog >.....?”. Furthermore, in Twenty Questions the player is faced with basic exemplars, not superordinates – the challenge is to select the appropriate superordinate to use in a question. As such, the opposite direction of asymmetry, moving from basic to superordinate, may be expected to be more important in reasoning and problem-solving.

For these reasons, it was hypothesised that deaf participants would show *an* asymmetry in their reaction times to the two conditions, and that this would predict efficiency performance on the TQT. The specific direction of the asymmetry was not hypothesised, however, as there was no strong case to argue that either direction was more significant than the other. Instead, the presence of any asymmetry was taken as an indicator of the nature of semantic organisation in this group.

### **Experiment 3: Twenty Questions and Semantic Decision performance in children who are deaf**

#### **Method**

##### *Participants*

For this experiment 16 deaf children aged 11-16 ( $Age_M = 13:8$ ) were initially recruited from specialist schools and language units for deaf schoolchildren based in North Yorkshire ( $n = 4$ ), Edinburgh and the Lothians ( $n = 6$ ) and Bristol ( $n = 6$ ). All children (11 M: 5 F) recruited were moderately to profoundly deaf and had been since before the age of 5 years. As in the wider population, the causes of deafness and presence of additional support needs varied considerably in this group, particularly in relation to language skills (see Appendix 1 for a breakdown of individual profiles). To allow for valid comparisons with the ASD profile, no deaf children with a diagnosis of autism or severe learning disability were included. To maximise the number of possible participants, the presence of other additional difficulties was not in the first instance used to exclude participants. However, in the course

of testing three participants were excluded as they were unable to complete the Similarities/VIQ test, because of its language demands. This resulted in a group of 13 participants being used in the study (10M:3F; Age<sub>M</sub> = 14:5, range = 13-16). Those with other additional diagnoses in the resulting group were marked for supplementary analysis, just in case their task profile differed.

Table 7 displays age and IQ information for the group of deaf children (DC). Compared to the participants in experiment 1, this group ended up being significantly older than hearing, typically-developing participants, but not ASD participants (DC>HC,  $t(32) = 2.158, p = .039$ , all other age comparisons n.s.).

Although not significant, differences approaching significance were also apparent in NVIQ and VIQ scores for deaf and hearing participants, with the DC group generally matching the lower half of the TD and ASD groups (DC<HC, NVIQ:  $t(32) = 1.874, p = .070$ ; VIQ:  $t(32) = 1.896, p = .067$ )<sup>32</sup>. No pairwise differences were evident in verbal or non-verbal mental age (DC: VMA<sub>M(SD)</sub> = 12.04(4.72), NVMA<sub>M(SD)</sub> = 10.68 (2.52); HC: VMA<sub>M(SD)</sub> = 15.03 (7.44), NVMA<sub>M(SD)</sub> = 11.75 (3.29); ASD: VMA<sub>M(SD)</sub> = 14.11 (6.40), NVMA<sub>M(SD)</sub> = 12.66 (4.60); all  $p > .10$ ).

A variety of communication methods were preferred by participants: seven used speech, while six used some form of sign (either BSL or Signed Supported English). This mixture of preferred communication methods (as opposed to the mostly BSL group of adults participants in experiment 2) is common among deaf children and young people. When asked, all of the participants said that they were familiar with the game, either through playing the game before or seeing someone else play it.

### *Design*

A mixed design was used, consisting of

a) a between-groups analysis comparing problem-solving performance in deaf children (DC) and existing data for hearing, typically-developing children (HC) and autism spectrum disorder children (ASD) acquired in experiment 1, and

---

<sup>32</sup> To allow with comparison with the data from experiment 1, the WASI scores for deaf participants were converted into their VIQ/NVIQ equivalents. As in the previous experiment, however, these are better considered as indicators of Spoken English knowledge rather than reliable indicators of general VIQ.

**Table 7. Age and IQ scores for deaf children (experiment 3), compared to autism spectrum disorder and hearing, typically-developing children (experiment 1, shown in grey).**

	DC ( <i>n</i> =13)		HC ( <i>n</i> =21)		ASD ( <i>n</i> =22)		<i>Sig.</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Age	14.40	1.07	13.40	1.44	13.69	1.45	DC>HC*
VIQ estimate	84.00	13.03	96.29	20.92	95.55	22.23	n.s.
NVIQ estimate	84.69	17.71	98.29	22.08	94.23	19.50	n.s.

\*  $p < .05$ . DC= deaf children, HC = hearing children, ASD = autism spectrum disorder children.

b) a within-groups, repeated measures analysis of deaf children's performance on the semantic decision task.

### *Materials and Procedure*

#### *- Twenty Questions Task, Question Discrimination & Plan Construction*

As in the above studies, all participants attempted a battery of tasks centring on the TQT. Participants began by completing three trials of the TQT on a 24-item board, followed by the question discrimination (QD) and plan construction (PC) tasks from experiment 1. The question discrimination task was used here (but not in experiment 2) because it was important to check the level of understanding that this younger group of participants had.

Because both of the planning tasks require vocabulary knowledge and reading ability for a range of category words, the 32 question words were displayed prior to starting the task and participants were asked to a) read them to a teaching assistant/ interpreter and b) indicate if they were unsure as to the meaning of any words. If there were words that had not been seen before or were not understood, they were explained by the experimenter.

In addition to the above tasks, participants completed the Matrix Reasoning and Similarities subtests of the WASI, and a new measure of semantic decision.

#### *- Semantic Decision Task*

The new task was a reaction time-based measure divided into three conditions:

- 1) word recognition
- 2) superordinate>basic decisions, and
- 3) basic>superordinate decisions.

The basic format of the task consisted of a series of words presented on a computer screen, to which participants were required to respond YES or NO (as indicated via a button press on a laptop keyboard). In the first condition, this simply consisted of responding YES when specific words were shown (such as LOOT) and NO if a different word was observed. In conditions 2 and 3, participants responded with YES or NO depending on whether the

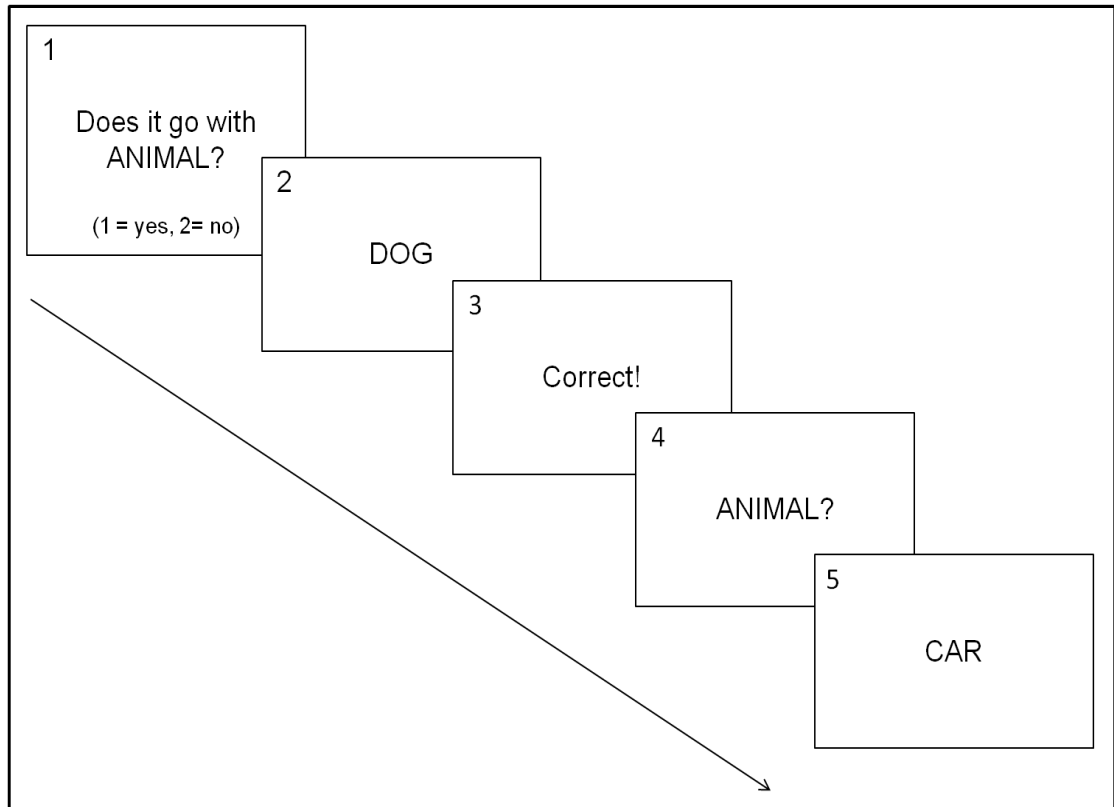


presented word was associated with a designated category. Each condition consisted of 30 trials divided into three blocks of 10. The order of conditions 2 and 3 were counterbalanced across participants. All stimuli were presented in E-Prime (Schneider, Eschman, & Zuccolotto, 2002).

*Condition 1.* The aim of the word recognition condition was to establish a baseline for response times to words of differing length and complexity. Participants were required to recognise words of four, six and eight letters across three consecutive trial blocks. In the first trial block the target was LOOT, with other 4-letter words with similar letters being used as non-targets (LOTS, LOST, etc). This was followed by similar trial blocks with CAMERA (6 letters) and BATHROOM (8 letters). Feedback on participants' responses was provided on each trial with either a screen displaying either "Correct!" or "Incorrect". Performance was measured via accuracy of responses and reaction times for each word presentation. The first set of trials (LOOT) was designated as a practice round, to ensure that participants were acquainted with the controls.

*Condition 2.* The second condition examined superordinate>basic category associations. Instead of indicating the presence of words, participants were asked "which words go together" for a superordinate term (TRANSPORT, ANIMAL, TOOL) followed by basic exemplars (e.g. CAR, DOG, HAMMER; see figure 19). In the first trial block, participants were asked "Does it go with TRANSPORT?" and then shown ten words containing exemplars and non-exemplars in a random order (e.g. CAR, DOG, BUS, TRAIN, TREE, SOFA, PLANE, TRUCK, FORK, ORANGE). Each time, participants were required to indicate YES or NO to indicate category membership. This was followed by the feedback screen (correct or incorrect), a reminder screen displaying the name of the category again "TRANSPORT?", and then the next exemplar. Trial blocks followed with the superordinates ANIMAL and TOOL.

*Condition 3.* In the final condition the order of associations was reversed, so that participants would first be primed with a basic-level exemplar and required to respond to superordinate categories. For example, in the first trial block they were asked "Does it go with APPLE?" and then shown a series of words including superordinates such as FOOD, FRUIT, PLANT, ANIMAL, MAMMAL, PET, CUTLERY, TOOL, and so on. This was followed by similar trial blocks for DOG and FORK.



**Figure 19. Example of presentation order for condition 2 of the Semantic Decision Task (superordinate>basic).** The first screen (1) displayed the category for the trial block. This was followed by the first exemplar (2), to which participants responded “yes” or “no”, and then a feedback screen (3). A reminder screen then appeared with the name of the category (4), followed by the next trial (5).

### *Scoring & Analysis*

Measurement of performance on the TQT, QD and PC tasks was identical to the deployment of these tasks in the studies presented earlier. Efficiency (question quality, questions used per trial) and question types (grouping, guessing) were measured on the TQT; total correct answers were noted for the QD task, and the expected question quality scores for the five-question plans were recorded for the plan construction task. Univariate ANCOVAs were used to compare the three groups, with the inclusion of age and IQ estimates as covariates.

Because comparisons between ASD and TD participants had already been conducted (in experiment 1), only those pairwise contrasts involving deaf participants were focussed on. As such, for the main TQT outcomes (especially question quality), the pairwise contrasts presented here are not corrected for multiple comparisons. However, for tasks involving multiple data points - such as the plan construction task - correction via techniques such as the Bonferroni method was still deemed necessary.

For the semantic decision task, though accuracy scores were recorded, the primary outcome measure across all three conditions was reaction time for accurate responses. Reaction times for inaccurate responses were not used as they may simply reflect a lack of knowledge about a category term, rather than a weaker association between a category and an exemplar. To address outliers, responses in the bottom and top 5% of distributions were trimmed for each individual within each condition. Then, parametric statistics (univariate analyses of variance) were run comparing the reaction times in the three conditions after a log<sub>10</sub> transform (conducted to normalise the response distribution). For ease of interpretation, raw scores for reaction time are presented in the results section.

Comparing inter-task correlations (SDT – TQT performance) on a group level, a standard practice would be to use medians as a measure of central tendency for each individual's RTs. However, as medians can be unreliable indicators of reaction time if the number of trials differs across conditions (Miller, 1988), trimmed means were used as an indicator of each individual's average reaction time. The differences between participants' mean reaction times for conditions 2 and 3 were then used as an index of any asymmetry in their category associations, and compared to the main TQT outcomes.

## Results

### *Twenty Questions Task*

All participants managed to attempt three trials of the Twenty Questions Task and all successfully completed at least one trial (i.e. identified the target in ten questions or less). Within the DC group, no significant changes were observed for any of the main task outcomes across the three trials (all  $p > .250$ , all  $\eta^2_p < .1$ ), suggesting that performance was relatively stable in the group. Therefore, analysis proceeded by just examining the mean TQT scores for deaf participants, averaged across the three trials.

Table 8 shows the mean values for question quality (QQ), questions used per trial (QT), grouping and guessing for the DC group, as compared to the data from experiment 1. Like ASD participants, the questions that DC participants asked were on average lower in question quality, less likely to be grouping questions and more likely to be guesses than for hearing participants. As in the previous experiment, covariate analysis and subgroup matching were used to address the contribution of existing group differences.

First, a univariate analysis of covariance (containing age VIQ and NVIQ estimates as covariates) was used to compare mean question quality in the three groups<sup>33</sup>. This analysis produced a significant main effect of group ( $F(2,50) = 7.592, p = .001, \eta^2_p = .233$ ), alongside significant contributions of age ( $F(1,50) = 4.727, p = .034, \eta^2_p = .086$ ) and VIQ ( $F(1,50) = 5.559, p = .022, \eta^2_p = .100$ ) and an effect approaching significance for NVIQ ( $F(1,50) = 3.840, p = .056, \eta^2_p = .071$ ). Pairwise comparisons indicated that the DC group were significantly less efficient in their questioning than HC participants ( $p = .048$ ), but not ASD participants ( $p = .244$ , uncorrected)<sup>34</sup>.

Analysis of the mean number of questions used per trial (QT) revealed a main effect of group ( $F(2,45) = 6.708, p = .003, \eta^2_p = .212$ ). However, pairwise comparisons on this variable indicated that this mainly reflected the performance of ASD participants, who took more questions on average than DC ( $p = .005$ ) and HC participants ( $p = .003$ ). A Kruskal-Wallis test comparing mean rates of grouping questions also indicated a significant group

---

<sup>33</sup> This analysis was run based on the mean QQ scores from each experiment. This, however, represents differing numbers of trials, due to the extended nature of the TQT paradigm in experiment 1. To check the effect of this, the analysis was run with baseline QQ scores from the prior study instead: this reveals very similar results, suggesting that any apparent group differences were not just a function of the number of trials tested.

<sup>34</sup> This analysis was also run using VMA and NVMA as alternative covariates. This yielded largely similar results (group effect:  $F(2,50) = 8.724, p = .001, \eta^2_p = .259$ ) although the pairwise contrast between deaf and hearing participants was only observed to be approaching significance ( $p = .055$ ).

**Table 8. Twenty Questions and Planning Performance in deaf, ASD and HC children.**

	DC		HC		ASD		<i>Sig.</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
<b>TQT</b>							
QQ	0.26	0.10	0.33	0.05	0.24	0.11	HC>DC,ASD*
QT	5.45	1.44	5.36	0.72	6.34	1.19	HC,DC<ASD*
Grouping (mean %)	48.84	24.73	66.73	8.36	45.69	24.25	HC>DC,ASD*
Guessing (mean %)	32.71	33.93	13.54	7.64	30.55	33.75	<i>n.s.</i>
<b>QD</b>	6.69	1.55	7.43	1.78	7.27	1.61	<i>n.s.</i>
<b>PC<sup>†</sup></b>							
QQ1	0.33	0.19	0.33	0.16	0.24	0.18	HC>ASD*
QQ2	0.21	0.16	0.26	0.16	0.18	0.11	HC>DC*
QQ3	0.17	0.14	0.26	0.12	0.18	0.14	HC>DC*, HC>ASD*
QQ4	0.13	0.08	0.19	0.10	0.18	0.12	HC>DC*
QQ5	0.14	0.09	0.20	0.15	0.18	0.10	
Mean QQ	0.20	0.06	0.25	0.05	0.19	0.70	HC>ASD*

\*  $p < .05$  † (DC  $n=12$ )

DC = deaf children, HC = hearing children, ASD = autism spectrum disorder, TQT = Twenty Questions Task, QQ = question quality, QT = questions per trial, QD = question discrimination, PC = plan construction. Data and contrasts from experiment 1 in grey.

difference ( $X^2(2) = 12.438, p = .002$ ), which non-parametric pairwise comparisons indicated to be based in the higher rates of guessing used by HC participants (HC>DC:  $U = 67.50, N = 34, p = .013$ ; HC>ASD:  $U = 92.50, N = 34, p = .001$ ). In contrast, no significant group difference was evident on the Kruskal-Wallis test for rates of guessing ( $X^2(2) = 1.878, p = .391$ ).

Following this, the DC group was compared to IQ-matched subgroups of HC and ASD children, representing the bottom 13 participants for VIQ (HC-13  $VIQ_{M(SD)} = 83.00 (11.75)$ ; ASD-13  $VIQ_{M(SD)} = 80.23 (10.39)$ , all DC/HC/ASD pairwise comparisons  $p > .420$ ). Along with matching for VIQ, the groups did not significantly differ for NVIQ (HC-13  $NVIQ_{M(SD)} = 88.92 (21.83)$ ; ASD-13  $VIQ_{M(SD)} = 81.92 (14.83)$ , all pairwise comparisons  $p > .337$ )<sup>35</sup>. Differences in age were evident between deaf and hearing participants (HC-13 Age  $M(SD) = 13:7 (0:11)$ ,  $t(24) = 1.856, p = .076$ , DC>HC) albeit at a level approaching significance. The ASD subgroup did not differ from either of the other groups in terms of age (ASD-13 Age  $M(SD) = 13:9 (0:9)$ , all pairwise comparisons  $p > .110$ ).

Because of potential age influences, an ANCOVA was again deployed in the comparison of mean question quality across the three groups, including age as a covariate. However, age was observed to make a negligible contribution to this comparison ( $F(1,35) = 0.076, p = .784, \eta^2_p = .002$ ). A group main effect was observed ( $F(2,35) = 5.478, p = .009, \eta^2_p = .238$ ), which pairwise comparisons indicated to reflect ASD<HC differences ( $p = .002$ ) and a trend towards a difference between DC and HC participants ( $p = .093$ ). As in the whole-group comparison, the subgroup of HC participants ( $QQ_{M(SD)} = 0.316 (0.042)$ ) asked questions that eliminated more alternatives than ASD participants ( $QQ_{M(SD)} = 0.199 (0.107)$ ) and DC participants ( $QQ_{M(SD)} = 0.255 (0.103)$ , as above)<sup>36</sup>. Thus, whether accounting for VIQ and NVIQ influences via ANCOVA, or matching subgroups on IQ abilities, a mean difference in question quality was still evident in the comparison of deaf and hearing children.

---

<sup>35</sup> The subgroups also showed similar levels of VMA (DC  $VMA_{M(SD)} = 12.04 (4.72)$ , HC-13  $VMA_{M(SD)} = 11.92 (5.64)$ ; ASD-13  $VMA_{M(SD)} = 10.20 (1.29)$ , all pairwise comparisons  $p > .10$ ) and NVMA (DC  $NVMA_{M(SD)} = 10.68 (2.52)$ , HC-13  $NVMA_{M(SD)} = 10.43 (3.08)$ ; ASD-13  $NVMA_{M(SD)} = 10.10 (2.38)$ , all  $p > .10$ ).

<sup>36</sup> Given the lack of a significant age effect, either planned comparisons in ANOVA or simple t-tests would also have been valid in comparing the DC and HC-13 data. On an unpaired t-test, the groups significantly differ in mean QQ ( $t(24) = 1.976, p = .030$ , one-tailed).

### *Question Discrimination*

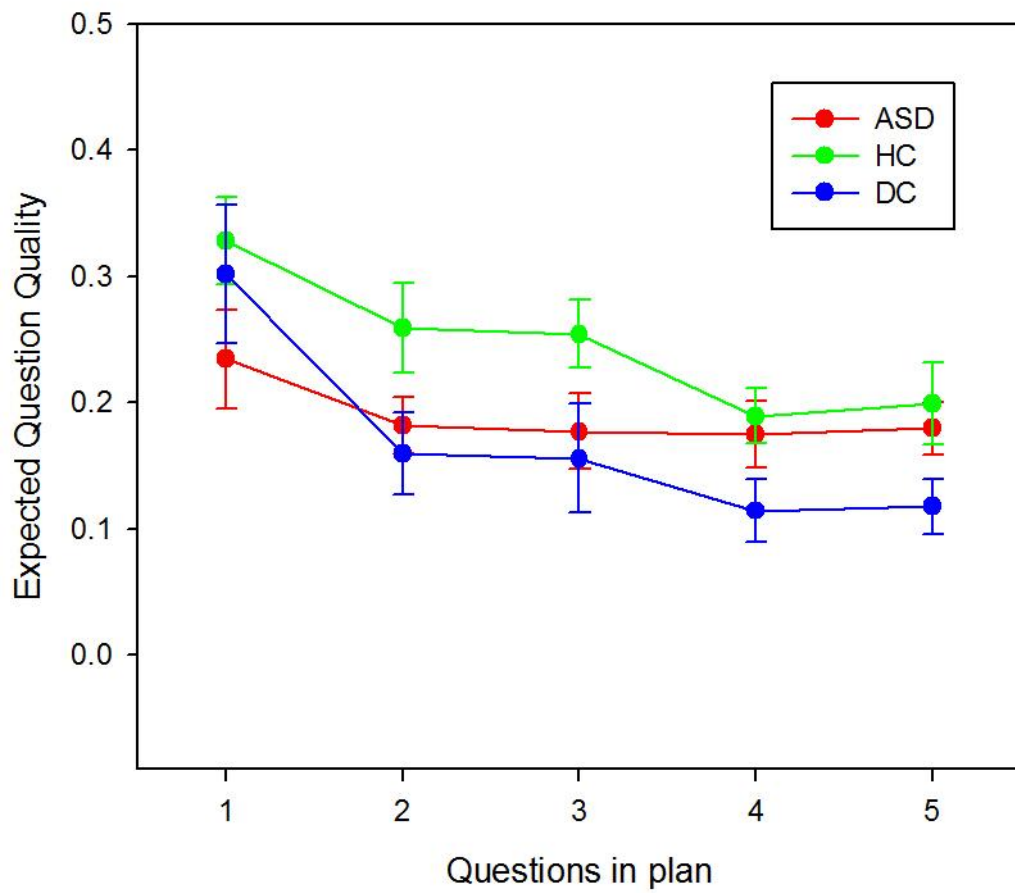
Although the mean score for question discrimination was slightly lower in the DC group (see table 8), an ANCOVA including age, VIQ and NVIQ indicated no significant group differences on the task (group effect  $F(2,50) = 0.551, p = .580, \eta^2_p = .022$ , all pairwise comparisons  $p > .300$ ). Covariate contributions, though generally greater than that of group, were also non-significant. (The highest contribution was age:  $F(1,50) = 1.676, p = .201, \eta^2_p = .032$ ). This suggested that the groups were comparable for their ability to discriminate less and more effective questions.

### *Plan Construction*

One of the DC participants could not complete the PC task because of problems with reading the question cards; this left a group of 12 deaf participants with complete data on this measure. (Plan Construction is the only task where a large number of words are presented simultaneously to participants). In experiment 1, ASD participants were less likely to select plans that narrowed down possibilities, whereas the questions selected by hearing, typically-developing participants gradually narrowed in expected question quality. Using the same method as in that analysis, DC participants did not show a significant change across their five questions (Friedman's ANOVA;  $X^2(4) = 6.996, N = 12, p = .136$ ).

However, this could have been attributable to a lack of power, as the deaf group was smaller than the two other groups. As only one question (Q3) actually failed the Shapiro-Wilk test for normality of data in the deaf group, a repeated measures ANOVA was also run to assess change in QQ across their plans. This revealed a significant linear trend ( $F(1,11) = 13.100, p = .004, \eta^2_p = .544$ ) with questions at the start of plans eliminating a greater number of items than questions towards the end of plans (see table 8). This suggested that DC participants did not in general have a difficulty with constructing a plan that narrowed possibilities.

Group differences in the efficiency of plans was assessed across the average of five questions (parametrically) and between individual questions (non-parametrically). Analysis of covariance (including age, VIQ and NVIQ) indicated a significant group main effect for mean question quality across the plan ( $F(2,49) = 5.902, p = .005, \eta^2_p = .194$ ). Pairwise comparisons showed that this group difference again reflected greater efficiency in the HC group, as compared to either DC participants ( $p = .005$ ) or ASD participants ( $p = .006$ ; see



**Figure 20. Mean efficiency of plans constructed by deaf , ASD and HC participants. DC = deaf children, ASD = autism spectrum disorder children, HC = hearing, typically-developing children.**



table 8 for group means)<sup>37</sup>. Pairwise comparisons for individual questions in the plan indicated no differences between deaf and ASD participants (all  $p > .120$ ), but greater efficiency for questions 2,3 and 4 in the plans of HC participants ( $p$ -range = .004-.047, Mann-Whitney tests used). It should be noted, however, that only one of these comparisons (question 3) would survive a Bonferroni correction for multiple comparisons.

As in experiment 2, plan gradient was also calculated, but this time no significant difference was observed between any of the groups (group main effect:  $F(2, 49) = 1.957, p = .152, \eta^2_p = .074$ )<sup>38</sup>. As figure 20 displays, this suggested that the plans of DC participants were similar to their HC counterparts in a) the presence of narrowing and b) the gradients of their plans, but dissimilar in their overall efficiency.

### *Semantic Decision Task*

The 13 DC participants who completed the TQT and IQ procedures also completed the semantic decision task. Accuracy rates were high across the three conditions. On average, participants were slightly less accurate in the third condition (basic>super, median = 25/30) than for condition 1 (word identification, median = 27/30) or condition 2 (super>basic median = 28/30), but the application of a Friedman's ANOVA to accuracy rates indicated no significant differences between the conditions ( $X^2(2) = 4.571, N = 13, p = .098, n.s.$ ). Thus, despite a slight advantage in condition 2, accuracy rates were largely comparable throughout the task.

Analysis of reaction times in the three conditions was first conducted across all participants combined, to assess for any asymmetry in the association of basic and superordinate terms, and whether this was still evident when accounting for other factors, such as condition order, word length or word frequency. Second, mean scores on each condition and the differences between them were assessed for their relation to performance on the TQT. Only correct responses were used in the RT analysis.

---

<sup>37</sup> Although not fully reported here for parsimony, analysis of this variable using the same matched subgroups as the preceding TQT analysis yielded similar results: HC-13 participants constructed more efficient plans than DC participants ( $p = .052$ ) and ASD participants ( $p = .047$ ). Similarly, the analysis was run using VMA and NVMA as alternative covariates, producing almost identical results.

<sup>38</sup> The effect size here was suggestive of an underlying group difference in the gradients of plans selected: pairwise comparisons indicated steeper gradients in deaf participants compared to ASD participants, although only at a level approaching significance ( $p = .062$ ).

As figure 21 shows, reaction times tended to be shortest for identification of words (condition 1) and longest for the identification of superordinates from a basic prompt (condition 3). A 3 (condition) x 13 (participant) multiple ANOVA was conducted on the log-transformed reaction times for each trial attempted<sup>39</sup>. (Participants were included as a separate factor to test for any differences or interactions in responses for particular children) A significant main effect of condition was evident ( $F(2,830) = 120.318, p < .001, \eta_p^2 = .225$ ): RTs were on average lowest in the first condition ( $M(SD) = 715.96\text{ms} (334.71\text{ms})$ ) followed by responses in conditions 2 ( $M(SD) = 890.14\text{ms} (400.86\text{ms})$ ) and 3 ( $M(SD) = 1041.64\text{ms} (585.16\text{ms})$ ); all pairwise comparisons  $p < .001$ ).

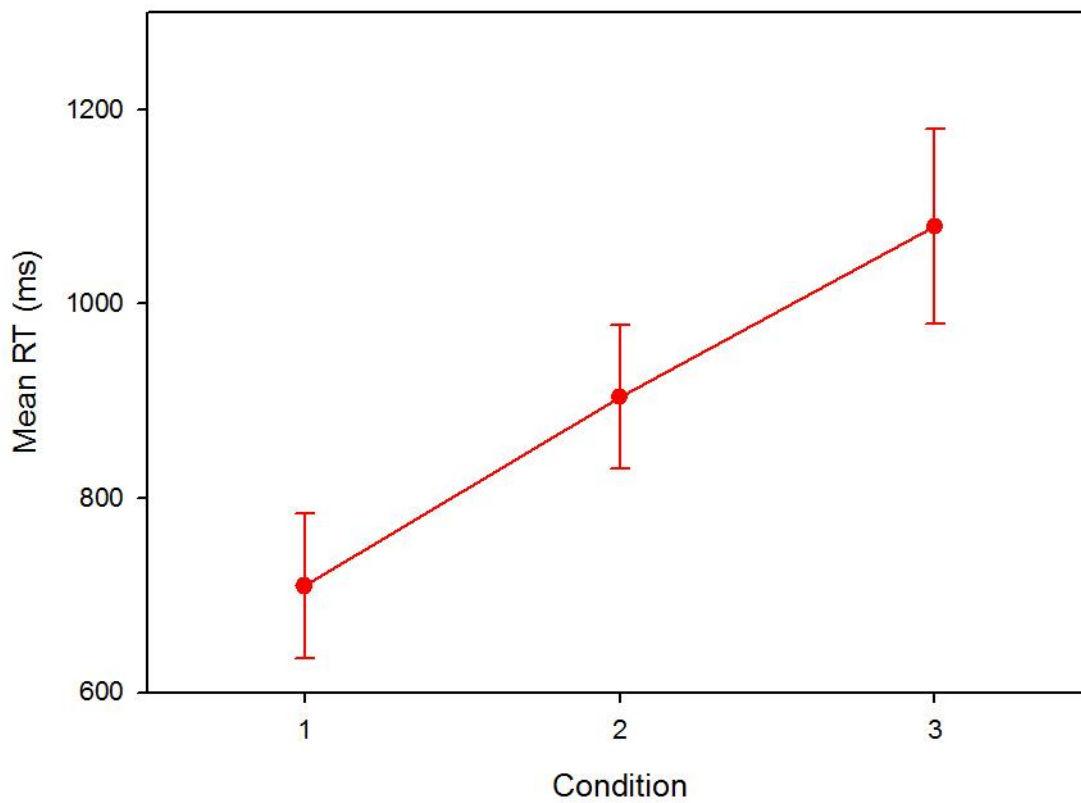
However, significant effects of participant ( $F(12,830) = 66.592, p < .001, \eta_p^2 = .491$ ) and an interaction between conditions and participants ( $F(24,830) = 5.522, p < .001, \eta_p^2 = .138$ ) were also observed, reflecting variation within the group. When the data were analysed on an individual basis, six out of 13 participants showed evidence of consistent differences between conditions 2 and 3 ( $RT_2 < RT_3$ ;  $p$ -value range = .0001-.054,  $\eta_p^2$  range = .077- .239). As these were always in the direction of condition 3 reaction times being longer (and not the reverse), this suggested a specific slowing of responses for basic-to-superordinate category decisions, albeit one that was not consistently evident across participants.

One possible confound was the order in which conditions 2 and 3 were attempted. To control for this, the order of conditions had been counterbalanced during testing, with 7/13 participants attempting condition 2 then 3, and the other six the reverse. However, inclusion of condition order as a term in an ANOVA comparing RTs for conditions 2 and 3 did not reveal a significant effect ( $F(1,635) = 1.473, p = .225, \eta_p^2 = .002$ ), suggesting that the order in which participants completed the task did not influence their reaction times.

Other potentially confounding factors were word length and word frequency. The average word length differed in the two conditions, with longer words (i.e. superordinates) appearing in condition 3 more often (No. of letters, C2:  $M(SD) = 4.80 (1.61)$ ; C3  $M(SD) = 6.17 (1.86)$ ;  $t(58) = -3.048, p = .003$ ). When word length was included as a covariate in analysis of the response times no significant contribution was observed ( $F(1, 634) = 2.702, p = .101$ ,

---

<sup>39</sup> A standard ANOVA was chosen over a repeated measures ANOVA because of the unequal numbers of trials across conditions (due to differing accuracy rates).



**Figure 21. Mean reaction times across conditions for the Semantic Decision Task.** 1 = Word Recognition, 2 = Superordinate>Basic category judgements, 3 = Basic>Superordinate Category judgements.

$eta_p^2 = .004$ ), suggesting that length did not explain any of the variance accounted for by the main effect of condition ( $F(1, 634) = 9.040, p = .003, eta_p^2 = .014$ )<sup>40</sup>. Word frequency rates did not differ between the conditions (C2:  $M (SD) = 58.33 (68.25)$ ; C3  $M (SD) = 55.50 (54.72)$ ;  $t(58) = 0.177, p = .860, n.s.$ ), but an ANCOVA was run anyway to check for any possible influence of frequency on RT rates. As with word length, no significant contribution of frequency was observed ( $F(1, 634) = 1.207, p = .272, eta_p^2 = .002$ ), suggesting that it could not explain the difference in RTs between conditions 2 and 3.

#### *- Relations to TQT performance*

To compare with performance on the Twenty Questions Task, semantic difference scores were calculated for each participant by subtracting the mean reaction time for condition 2 from condition 3. These scores were then included in individual regression models for each of the TQT performance outcomes: question quality, questions taken, use of grouping and use of guessing. Age, VIQ and NVIQ scores were also included in each of the models to control for their influence.

Semantic difference scores were not found to significantly predict question quality (stan. beta = .014,  $p = .960$ ). For the prediction of QQ a model was retained that only included NVIQ ( $R^2 = .429, F(1,11) = 8.267$ , stan. beta = .655,  $p = .015$ ), with greater non-verbal IQ being associated with greater question quality. No significant associations were observed for age or VIQ (stan. beta range = .177-.381,  $p$ -value range = .138-.539).

In contrast, an association was observed between semantic difference scores and the mean number of questions used per trial (QT; see figure 22). For predicting QT only semantic difference score was retained in the model ( $R^2 = .584, F(2,10) = 7.014$ , stan. beta = .601,  $p = .012$ ), with greater difference scores being associated with taking more questions to find the target (see figure 22). That is, participants with a greater difference between conditions 2 and 3 also had to take more questions to find the target on the TQT. NVIQ was also retained in

---

<sup>40</sup> Correlational analysis did in fact show a small but significant relation between reaction times and word lengths (as might be expected), but this was insufficient to explain the difference between conditions ( $r = .097, N = 639, p = .014$ ).

the model (stan. beta =  $-.382$ ,  $p = .094$ ), but not VIQ or age (stan. beta range =  $-.021$  to  $-.099$ ,  $p$ -value range =  $.689$ -. $949$ )<sup>41</sup>.

To explore this association further, the predictive power of semantic difference scores was compared with mean reaction times in each of the task conditions. None of the individual conditions showed a greater association with questions used on the TQT than the discrepancy between conditions (stan. beta range =  $.148$ - $.200$ ,  $p$ -value range =  $.533$ -. $545$ ). This supported the significance of *differences* in reaction time between different semantic conditions, rather than speed of response in any particular condition. That is, problem-solving was specifically associated with the level of asymmetry or inequality in responses for basic-to-superordinate and superordinate-to-basic decisions.

In models for grouping and guessing rates, only VIQ was retained as a significant predictor (grouping model:  $R^2 = .358$ ,  $F(1,11) = 6.131$ , VIQ stan. beta =  $.598$ ,  $p = .031$ ; guessing model:  $R^2 = .487$ ,  $F(1,11) = 10.443$ , VIQ stan. beta =  $-.232$ ,  $p = .008$ ). Higher VIQ was associated with increased rates of grouping and decreased rates of guessing: no other predictors were significant (all  $p$ -values  $>.120$ ).

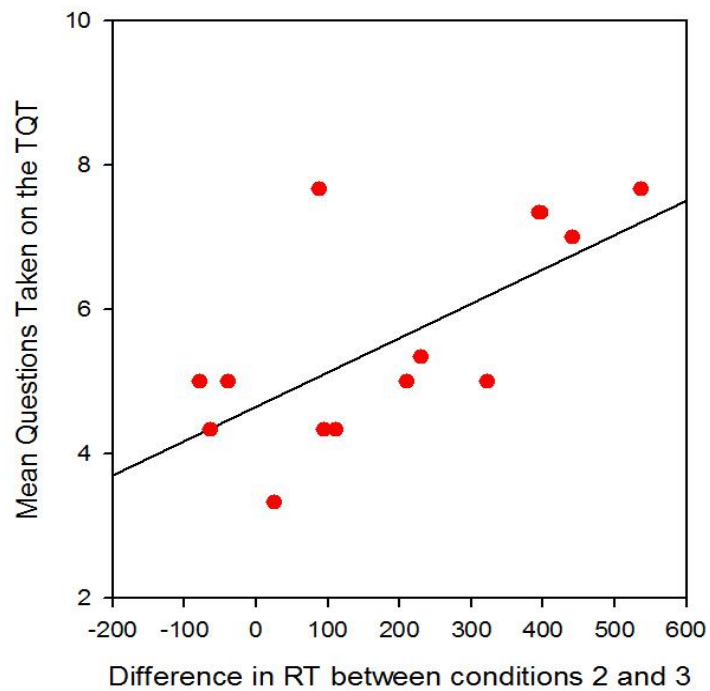
#### *Post-hoc analysis: handling heterogeneity*

In order to check for the influence of additional support needs or diagnoses that could have affected TQT performance, a “leave one out” procedure was followed where individual DC participants were excluded from the analysis one at a time. This included one participant with CHARGE syndrome, one with ADHD and one with possible hyperactivity problems who scored very low on mean QQ ( $0.05$ ). Despite the loss of power associated with reducing the sample size in this way, comparisons between DC and HC participants were still evident either significantly or at a level approaching significance (deaf mean QQ range =  $0.253$ - $0.272$ ;  $p$ -value range =  $.043$ -. $061$ ,  $d$ - range =  $0.65$ - $0.77$ ). Thus, despite the presence of such diagnoses, it did not appear to be the case that specific participants were disproportionately affecting the overall group results.

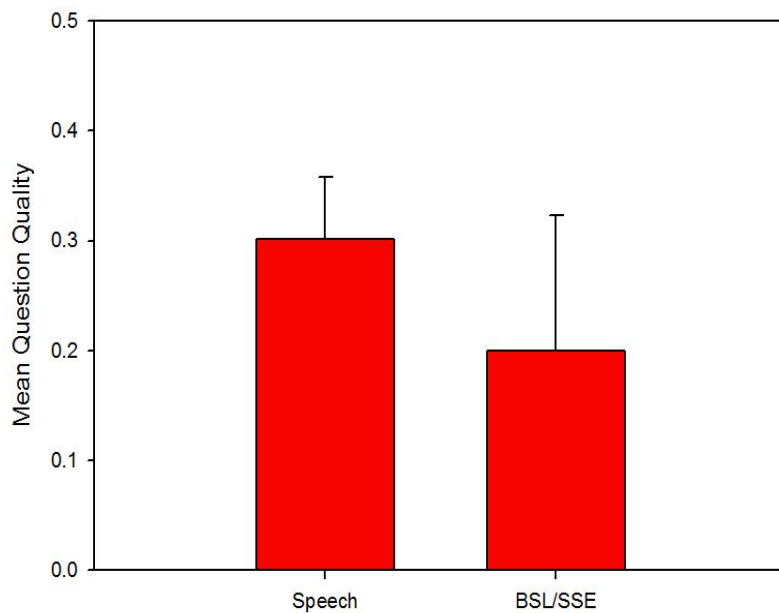
Assessment of other ways in which DC participants differed was hampered by the small numbers involved, but some descriptive analysis was possible. As figure 23 shows, DC participants who used speech appeared to score higher for mean QQ than users of BSL or

---

<sup>41</sup> Because of skew in the TQT outcome variable, this analysis was also run using a log10 transform of the number of questions taken. This, however, produced identical results, so for clarity only the model using raw QT scores is reported here.



**Figure 22.** The relationship between differences in reaction time on the Semantic Decision task and the number of question used per trial on the Twenty Questions Task (TQT). The level of asymmetry between different directions of category judgement predicted efficient question use on during problem-solving.



**Figure 23.** Mean question quality in deaf children with different preferred modes of communication. BSL = British Sign Language, SSE = Sign-Supported English.

SSE. However, this should be interpreted with caution, as these groups also differed considerably in mean IQ scores (Sims/VIQ:  $Speech_M = 92.85$ ,  $BSL_M = 73.67$ ; MR/NVIQ:  $Speech_M = 93.57$ ,  $BSL_M = 74.33$ ).

Eight of the DC participants had at least one cochlear implant, compared to five who either used hearing aids or nothing at all. Those who had an implant on average scored lower on QQ ( $Cochlear_M = 0.23$ ;  $No\ implant_M = 0.29$ ) but were also more variable in their scores ( $Cochlear_{SD} = 0.12$ ;  $No\ implant_{SD} = 0.07$ ). As such, it is not clear that there were any consistent differences in problem-solving that could be attributed to these additional factors.

## **Discussion**

The main finding of experiment 3 was that deaf children also exhibit inefficiency in their problem-solving on the TQT. Like ASD participants, deaf participants appeared to be less efficient in the questions that they ask on the Twenty Questions Task, and the types of questions that they asked were different from hearing, typically-developing participants. Specifically, deaf participants scored low on question quality (like ASD participants) and were less likely to use grouping questions. This effect, evident when accounting as much as possible for influences of age, verbal and non-verbal IQ, is the first direct demonstration of similarities in the problem-solving profile of deaf and ASD children.

In addition, exploration of semantic decision skills in deaf children highlighted a potential mechanism underlying problem-solving differences. Namely, deaf children, on average, took longer to make judgements about superordinates following a basic category cue, than when doing the reverse (i.e. judging which basic exemplars fit a superordinate cue). This asymmetry within their own category associations was related to the number of questions deaf participants used on the TQT, independent of age and IQ abilities.

### *Caveats and limitations*

Before discussing how exactly the TQT and SDT data fit together, some limitations of the present experiment must be acknowledged. First, as in experiment 2, what conclusions can be drawn from these data are limited by the size of the target group. The initial recruitment of 16 participants, over a period of 12 months, represented the combined efforts of contact with three specialist deaf schools (in Bristol, North Yorkshire and the Lothians) and one mainstream Edinburgh school with a dedicated language unit. A number of other schools (special and mainstream) and deaf charities (Deaf Action and the National Deaf Children's

Society) were also contacted, including over 20 of the secondary schools in the Edinburgh area. Among those families that were approached (once exclusion criteria were taken into account), roughly 50% of parents consented to their children participating. It is not immediately clear why recruitment was so difficult. In the case of less able children, very often the presence of significant additional needs precluded their recruitment. In contrast, it is the impression of the author that the families of more able children, including those in mainstream education, seemed in general to have less engagement and possibly less interest in deaf research (certainly compared to the families of children with autism).

A second, related limitation is the heterogeneity present in the sample of deaf children seen. Although heterogeneity is the norm for samples of deaf participants, the presence of such a mixture of children obviously makes drawing reliable, group-level conclusions much more challenging. It is important to acknowledge that the individuals seen in this group differed in their preferred methods of communication, their use of hearing devices and the presence of additional difficulties.

In terms of communication, those who used speech appeared to perform better on the TQT than those who signed. This is consistent with the contrast in performance between signing children in Marschark & Everhart (1999) and speaking children in Remine et al. (2008), but the difference observed here cannot be clearly separated from differences in IQ between participants. There were also potential differences in question quality between those with and without a cochlear implant – although contrary to what might be expected, those who had such a device appeared to perform slightly worse on the TQT than those who did not. For both of these factors (communication and device use), such differences at most provide pointers for future research. Without bigger, ability-matched groups of participants, however, genuine differences in problem-solving performance cannot be established with any certainty.

The presence of additional difficulties in some individuals in the sample introduces a potential confound in the data – namely, that particular participants with additional diagnoses may have unduly influenced the overall group result. This was a concern throughout the study, and care was taken to gather as much information as possible in order to monitor the influence of other diagnoses. Arguably, this factor did not clearly bias the sample: while the participants with ADHD-like difficulties were among the poorer performers, the participant with CHARGE syndrome was one of the best within the group. Moreover, systematic



removal of these individuals from the sample did not clearly change the overall group result. Nevertheless, a replication of the paradigm in a sample of highly able deaf children without any additional support needs is clearly needed.

A final limitation concerns the suitability of the stimuli used in the tasks. Analysis was constrained to only consider those participants who were able to attempt the majority of the tasks, and, in the case of the semantic decision task, reaction times were not used if participants could not accurately respond to certain words. Thus, any words that participants were not familiar with should have been largely filtered out of the analysis. It is still possible, though, that the language demands were quite high for an experiment with deaf children. The plan construction task in particular seemed to be challenging due to the simultaneous presentation of a large amount of text (over 30 question cards), and only 12 of the 13 participants were able to attempt this measure. With hindsight, this and some of the other tasks would have benefitted with further piloting with young deaf participants. As explained above, in this instance this was not possible (the recruitment of the experimental group itself was challenging enough), but future work would certainly benefit from a more extensive stage of piloting.

#### *Similarities in profile for deaf and ASD children*

Despite these considerations, there are some tentative conclusions that can be drawn out of the data from above sample. First, the question quality scores seen for the group of deaf children are a) consistent with the data from more able deaf adults and b) apparently similar to those recorded by ASD participants. The deaf children in this group did not appear to ask questions that efficiently eliminated information at a level that would be expected for their age and cognitive ability. This was not clearly a result of experience with the task, or lack of knowledge about effective questions: deaf children were as able to identify good questions as hearing and ASD participants. Compared to hearing participants, the deaf participants in experiment 3 were also less likely to use grouping questions, consistent with the performance of deaf participants in Marschark & Everhart (1999) and autistic participants in Minshew et al. (1994), Alderson-Day & McGonigle-Chalmers (2011) and Alderson-Day (2011).

It is worth noting too that there were also some differences between the deaf and ASD profiles on the TQT. Unlike ASD participants, deaf participants were not likely to require more questions to identify the target, despite the lower quality of their questions. In addition,

the performance of deaf children did not significantly change across the three trials they attempted, suggesting that they did not have any particular difficulty in handling an unchanging array, or having to remember their questions.

The data from the plan construction task pointed to similarities and differences between the deaf and ASD profiles. Like deaf adults, the deaf children here did not appear to have any problem with constructing a narrowing plan. This contrasted with the less able ASD participants in experiment 1, but was consistent with the more able participants in that group. However, like all of the ASD children, the mean efficiency of their plans was lower than that of hearing, typically-developing children. Thus, whether using questions (on the TQT) or planning questions, ASD and deaf children tended to select semantic categories that formed a less efficient basis for searching the set.

#### *Relations with semantic organisation*

On the semantic decision task, basic-to-superordinate category associations took longer for deaf participants to make than superordinate-to-basic category associations. Without further testing of hearing children, the presence of a such an asymmetry does not necessarily indicate *atypicality* of semantic organisation: it should be thought of as a within-subjects effect only. Put another way, there is nothing necessarily abnormal about this pattern of reaction times, however asymmetric they may appear to be: hearing children may be just as likely to react slower when making basic-to-superordinate associations compared to the opposite type of association.

Nevertheless, the regression analysis of SDT and TQT performance suggested that the more asymmetrical participants' category responses were, the more questions they required on average to solve TQT trials – a relation that was not explained by age or IQ factors. This is at least suggestive of a relationship between category organisation and problem-solving within this sample of deaf children.

In interpreting the difference between conditions on the semantic decision task, an important thing to notice is that the direction of difficulty observed here is the opposite of that reported by Marschark et al. (2004). On their measure of semantic organisation, deaf adults were observed to be less likely to respond to a superordinate category name with a basic exemplar, whereas here participants were slower to respond to a basic exemplar with a superordinate category name.

One possible explanation for this is that it reflects an actual asymmetry in category structures. There simply are many more exemplars than superordinates for any given category. That is, ANIMAL includes many cases, whereas DOG has only a few common superordinates, such as animal, pet, or mammal. It is possible to add to those, but to do so is to quickly decrease the frequency and specificity of terms used (one could add living creature, vertebrate, earth-based *etc*). Furthermore, even relatively common superordinates stand in unequal relation to one another: the superordinate of DOG is most commonly ANIMAL, or possibly PET, whereas MAMMAL is much less likely to be used. As such, each condition on the semantic decision task is asking a slightly different kind of question. To go from superordinate to basic exemplars is necessarily to ask about the many contents of a given set, to go from one to many. In contrast, to go from a basic exemplar to multiple superordinates is to take one exemplar, and ask for many possible category names where usually only one or two would frequently be used. Thus, while the frequencies of individual words may be similar in each condition, the frequency of relations between category and exemplar may differ. When asking “Does it go with ANIMAL?”, a number of items with frequent relations to that word will be available (dog, cow, horse). In the converse “Does it go with DOG?”, there may be only one or two frequent associations.

While inconsistent with their first experiment, these data are consistent with the results of the analogical reasoning task in Marschark et al.’s (2004) second experiment, where deaf participants were less likely to produce superordinate terms when primed with a basic exemplar. They also, as argued in the the hypotheses for this experiment, fit with what is being asked of participants on the TQT itself: participants are being faced with exemplars, from which they need to produce appropriate superordinates. In this sense, inefficiency in questioning may be related to semantic organisation in that atypical or infrequent associations with superordinates could be somehow weaker or less accessible for deaf participants. Without such access, the range of potential category terms that come to mind when deciding what question to ask may be limited, even when knowledge of such terms is present when specifically prompted or tested. When this is contrasted with stronger associations in the other direction, towards basic exemplars, this may create a tendency to ask about a more restricted category; to focus on the particular over the general.

For this to be correct though, deaf participants would have to demonstrate an asymmetry on the semantic decision task when hearing participants do not. Or, if hearing participants showed such an asymmetry as well, one would predict a greater level of discrepancy within

the deaf group. The presence of intertask correlations suggests that these possibilities are plausible, but without a control group the present experiment cannot demonstrate that either of these scenarios are in fact the case.

A final caveat here is that the correlation with TQT did not involve question quality. The correlation between response asymmetry on the SDT and the number of questions used per trial (QT) indicates a relation between semantic skills and efficiency, but this is not the measure of efficiency that consistently discriminates deaf and hearing participants, or ASD and typically-developing participants. As discussed above, questions-per-trial offers a rough measure of efficiency. On the one hand, it usually correlates negatively with question quality, but on the other hand it can randomly fluctuate due to lucky guesses made by participants. As such it is not clear why this would correlate with semantic performance while QQ would not. In this case, it may by chance have produced a more continuous distribution of data across the deaf group than that seen for QQ. It may be that with more data (and a better spread of data for QQ) significant relations would be seen for both efficiency measures.

At the very least, the relation observed between SDT and TQT performance raises the question as to whether a similar effect can be observed in ASD participants. Returning to the “strong” and “weak” arguments introduced in chapter 3, the relation with semantic skills observed here lends support to one part of the strong argument outlined earlier: deaf and autistic problem-solving profiles overlap because of similar effects of atypical language development, via semantic organisation. If the same relation can be observed in a group of ASD participants, the strong argument would be bolstered further.

If, alternatively, ASD participants do not show this relation, then this proposed “semantic route” to problem-solving differences may be one that is specific to deaf individuals. Such a scenario, though, would still be consistent with the weak form of the argument, if it could be shown that ASD performance is moderated by differences in language development. This would demonstrate that the comparison with deaf individuals is still a useful one; that if deaf participants attempt the TQT in a different way because of their language history, then a similar kind of process may occur in people with ASD.

A way to examine this is to compare performance in children with autism and Asperger Syndrome. Diagnostically, children with autism exhibit delays in early language

development. Children with Asperger Syndrome do not. If language development is relevant to understanding problem-solving in ASD, a point that both the strong and weak forms of the argument rely upon, then participants with Asperger Syndrome should be relatively unimpaired in their performance on the TQT.

## **Chapter 6: Similarities and differences between Asperger Syndrome and autism.**

If delays in language development have an effect on semantic organisation, and by extension verbal problem-solving, then this should be reflected within the autism spectrum itself.

While a diagnosis of autism typically includes parental report of delayed language in the first three years of life, the diagnosis for Asperger Syndrome does not (APA, 1994). Individuals with Asperger Syndrome show all the social, communicative and repetitive characteristics of autism, but may have relatively intact language skills and intellectual abilities within the normal range (Frith, 2004). Therefore, individuals with Asperger Syndrome may be expected to be more successful in their problem-solving on a task like the TQT.

Because some individuals diagnosed with autism will also go on to develop good linguistic and cognitive skills despite a history of language delay (Boucher, 2012), there has always been considerable debate as to whether or not the three-year cut off represents a real distinction in autistic symptomatology and outcome (see, for example, Ghaziuddin, 2010; Mayes & Calhoun, 2001; Schopler, 1996). A large body of research has examined potential differences in the long term cognitive and linguistic profiles of Asperger Syndrome (AS) and high-functioning autism (HFA), in an attempt to clarify their separation. This debate has culminated in the recent decision to remove Asperger Syndrome from the forthcoming DSM-V (due in 2013), in favour of treating autism, AS and other conditions on the spectrum as examples of a dimensional “autism spectrum disorder” (APA, 2012). Under the DSM-V, Asperger Syndrome is simply taken as another example of high-functioning autism, rather than a distinct disorder of its own.

However, the proposed changes for DSM-V do not entail that there are *no* cognitive differences between those with a diagnosis of either autism or Asperger Syndrome. The decision primarily concerns the validity of a clinical distinction between the two diagnoses, and in this instance, the judgement made is that a) they do not differ in terms of core autistic symptomatology, and that b) any differences in cognition or language that have been seen are not significant enough to warrant diagnostic separation. The question of whether or not autism and Asperger Syndrome differ from each other in a clinically significant way is a separate one, though, from the question asked here; namely, do differences in early language development within the autistic spectrum have consequences for specific higher cognitive skills in later life?

The following sections summarise the main cognitive and linguistic differences that have been proposed to exist between Asperger Syndrome and autism. First, the original clinical descriptions of AS are outlined. Second, research on general cognitive abilities and cognitive profile is reviewed. Finally, studies looking at language skills and executive functions are examined. The aim of doing so is to establish what differences may support a distinction in verbal problem-solving skills between autism and AS, and what differences may need to be controlled for in a comparison between the two groups.

### **Clinical descriptions: Asperger, Wing and “Autistic Psychopathy”**

The key characteristics of Asperger Syndrome are derived from the original case series of four children described by Hans Asperger (1944) in his paper “Die ‘Autistischen Psychopathen’ im Kindesalter” [Autistic Psychopathy in Childhood]. As in Kanner’s case reports, the central problem observed by Asperger was one of social interaction. Asperger (1944/1991) introduced his case reports as follows:

“The children I will present all have in common a fundamental disturbance which manifests itself in their physical appearance, expressive functions and, indeed, their whole behaviour. This disturbance results in severe and characteristic difficulties of social integration.” (p. 37)

These difficulties were not, according to Asperger, due to a lack of feeling for others, but some form of disconnect in their affective processing:

“...the children cannot be understood simply in terms of the concept “poverty of emotion”, used in a quantitative sense. Rather, what characterises these children is a qualitative difference, a disharmony in emotion and disposition.” (p. 83)

Also in accordance with Kanner’s reports were a number of typically autistic behaviours and characteristics: repetitive behaviours (p. 43), an insistence on sameness (p. 61) and irregular play or use of toys (p. 66).

Asperger’s cases largely differed from Kanner’s in their linguistic abilities. All four of Asperger’s cases showed precocious language skills: Fritz V. “quickly learned to express himself in sentences and soon talked ‘like an adult’” (p.39), while Harro L. “had an ‘unusually mature and adult manner of expressing himself’” (p.52). Even in the cases of Ernst K. and Hellmuth L., where a degree of delayed speech within the second year is referred to, both were described as speaking “like an adult” in their clinical presentation at 7

and a half and 11 years old, respectively. Good skills in reading (Harro L.) and spelling (Hellmuth K.) were also noted by Asperger.

Some, but not all, of Asperger's cases also demonstrated strong cognitive skills in other domains. Fritz V. was reported as possessing precocious abilities in digit span (p.44) and arithmetic (p. 45), while Harro L. was described by Asperger as a "bright boy" with a "mastery of numbers", (p. 56). In contrast, Ernst K. apparently showed no such skills, and Hellmuth L. was reported as having "very uneven" school knowledge and "very poor" arithmetic (p. 66). Thus, though the most able individuals described by Asperger had no equivalent in Kanner's case reports, this did not mean that all of Asperger's cases had very high intellectual ability (Frith, 2004).

Although writing at the same time as Kanner, Asperger's reports were little used in English-language research and practice until the early 1980's (following the work of Lorna Wing and colleagues), and were not published in English until their translation by Uta Frith in 1991 (Frith, 1991). The label of "Asperger's Syndrome" (now simply Asperger Syndrome) comes from Wing (1981), who chose the term in preference to "autistic psychopathy" in order to avoid overt associations with sociopathic behaviour (p. 115).

In her 1981 paper, Wing presented 34 cases of children and adults that she considered to be affected by similar difficulties to those that Asperger described. Nineteen cases exhibited a "typical" history of AS and 15 showed AS-like behaviour at the age of presentation, but without the characteristics of precocious early development described by Asperger. Describing the group in general, Wing states that "the child usually begins to speak at the age expected in normal children, whereas walking may be delayed" (p. 116).

Again, some good language skills were described, such as strengths in vocabulary and grammar, although Wing departed from Asperger in the extent to which she interpreted this as evidence of genuine language ability:

"Despite the eventual good use of grammar and a large vocabulary, careful observation over a large enough period of time discloses that the content of speech is impoverished and much of it is copied inappropriately from other people or books." (p.116)

Moreover, Wing argued that the presence or absence of language delay was not necessarily significant to the kind of difficulties seen by her AS cases at the age of presentation. The 15 cases with an atypical AS history were those who had early problems with language, but nevertheless ended up with an Asperger-like profile in later years.



Further descriptions followed from researchers in Sweden (e.g. C. Gillberg, 1985) and Canada (e.g. Szatmari, Bartolucci, Finlayson, & Krames, 1986), and by the early 1990s Asperger Syndrome came to be codified in the official diagnostic criteria for ICD-10 (WHO, 1993) and DSM-IV (APA, 1994). Despite Wing's more nuanced view of language in AS, the definitions of the syndrome set by both bodies stated that there must be no "clinically significant" delay in structural language development, defined as the presence of single words by age 2, and phrases by age 3. Subsequent application of the diagnosis has been inconsistent, and there has been continuous debate as to the relevance of the language delay criterion, but this has continued to be the most "workable distinction" (Frith, 2004, p. 674) between AS and others forms of autistic spectrum disorder<sup>42</sup>.

### **Research comparing AS and autism**

Since the initial clinical accounts of Asperger and Wing many studies have attempted to assess the similarities and differences between AS and autism in behaviour and cognitive skills. Aside from studies on social cognition (which will not be discussed here), much of this work concerns the general cognitive profiles of each group, their language skills, and executive functioning abilities.

#### *General abilities and cognitive profile*

Higher levels of general cognitive ability in AS, particularly for verbal IQ, have been a commonly reported finding in the literature. After the the DSM-IV codification, one of the first studies to compare the cognitive profiles of AS and HFA was conducted by Szatmari, Archer, Fisman, Streiner, & Wilson (1995), who examined the behavioural and cognitive characteristics of 21 children with Asperger Syndrome and 47 children with autism aged 4-6. The children with autism, though all within the high-functioning range, scored significantly lower than the AS children on the Leiter International Performance Scales, a measure of non-

---

<sup>42</sup> It is important to acknowledge that further differences in social interaction style and motor development have also been proposed to exist between AS and autism. Asperger's original accounts suggested relatively intact language development in contrast to delayed or clumsy motor development (Asperger, 1944), while Wing (1981) suggested that her AS cases displayed an active but odd social interaction style in contrast to the more "aloof" interaction of individuals with autism. The validity and consistency of these differences are not discussed here, not because they lack importance, but because they would not appear to directly relate to cognitive skills specifically.

verbal reasoning (Levine, 1982), and the verbal reasoning subtest of the Stanford Binet (Thorndike, Hagen, & Sattler, 1986).

These findings supported earlier work by the same research group that reported significant differences between individuals with AS and autism for all of the WAIS and WISC subtests except digit span (Szatmari, Tuff, Finlayson, & Bartolucci, 1990). Interpretation of this earlier study is tricky as it precedes the DSM-IV, and the criteria used for AS did not include a specific requirement of intact early language. The later study, however, did require AS participants to be speaking in phrases by 36 months, in line with DSM-IV criteria (Szatmari, et al., 1995).

Further studies of the AS cognitive profile supported Szatmari et al.'s (1995) findings of general IQ differences between AS and autism (Ehlers et al., 1997; Manjiovana & Prior, 1999; see Kugler, 1998, for a review of early studies). Ehlers et al. (1997) reported greater scores on all subtests of the WISC-R except Block Design and Object Assembly in a sample of 5-15 year-old AS children ( $n = 40$ ) compared to children with autism, while Manjiovana & Prior (1999) reported significant differences in overall WISC estimates of FSIQ, VIQ and NVIQ in a similarly-sized sample of children with AS and autism.

Clinical reports by Klin et al. (1995) and Eisenmajer et al. (1996) supported Szatmari et al.'s (1995) findings for verbal intelligence, although they did not always observe advantages in other areas. Compared to a group of children and young people with autism ( $Age_M = 16:0$ ), Klin et al. (1995) reported greater VIQ abilities and a greater VIQ-NVIQ discrepancy in a sample of 21 adolescents with AS. In contrast, performance (i.e. non-verbal) IQ was reported to be higher in the autistic participant group, and there were no differences in overall full-scale IQ (Klin et al., 1995). Similarly, Eisenmajer et al. (1996) reported higher scores for verbal mental age in children diagnosed with AS than children with autism ( $Age_M = 10:0$ ), as measured on the Peabody Picture Vocabulary Test - Revised (L. Dunn, 1981) and the British Picture Vocabulary Scale (BPVS; L. Dunn, Whetton, & Pintilie, 1982), but no advantage in other areas.

One problem with the above studies, however, is variation in the definition of AS used. Similar to Szatmari et al.'s (1995) use of the DSM-IV, Klin et al. (1995) followed the ICD-10 research diagnostic criteria in requiring that their AS participants showed no delay in language. Ehlers et al. (1997) followed earlier criteria set by Gillberg & Gillberg (1989), but stated that at least 32/40 of their AS group qualified met the ICD-10 criterion of intact language development.

In contrast, Manjiovana & Prior (1999) based their diagnostic groups on clinical expert opinion, but also ran a secondary analysis that split their participants according to the presence or absence of language delay. When this was done, the differences in overall cognitive ability between groups were in fact *less* prominent, with discrepancies in VIQ ( $p = .052$ ) and FSIQ ( $p = .070$ ) only approaching significance. Similarly, Eisenmajer et al. (1996) did not select their groups based on language delay, but analysed the features that discriminated those given an AS or an autism diagnosis by clinicians prior to the DSM-IV criteria being created. They reported significantly higher rates of language delay in the autism group (as would be expected), but also evidence of language delay in over 40% of the AS group. Furthermore, though language delays roughly discriminated the two groups, Eisenmajer et al. argued that all of the AS group actually qualified for a diagnosis of autism by demonstrating communication impairments in the first three years of life (Eisenmajer et al., 1996).

These findings highlight two important issues. First, individuals with AS may, on average, appear more high-functioning than those with autism, but the possession of an AS diagnosis does not necessarily reflect fully intact language development in the early years. In some cases, individuals with an AS diagnosis may have been diagnosed based on their current level of functioning, and not their early development profile.

The danger of this practice is that it does not necessarily pick out separable differences that are indicators of an etiologically distinct syndrome. In this scenario, those who are more intellectually able get a diagnosis of Asperger Syndrome, while those with a lower level of functioning get a diagnosis of autism. But this means that AS as a label is just acting as a by-word for “high-functioning”, and experimental comparisons between the two groups become circular; those with AS will necessarily end up doing better on cognitive tasks by virtue of their diagnosis (Macintosh & Dissanayake, 2004; Ozonoff, South, & Miller, 2000).

A second issue, highlighted by Eisenmajer et al. (1996), is that individuals with AS may in many cases actually qualify for a diagnosis of autism under the DSM-IV codification. This is because the criteria are, to a certain extent, contradictory. The criteria for AS may require intact structural language development prior to three years of age, but the criteria for autism only requires problems in at least one of a) social interaction, b) language as used in social communication or c) symbolic or imaginative play, prior to three years. If these are met, the diagnosis of autism is supposed to take priority - but most AS children will indeed meet these criteria, as even those with good expressive language skills will have problems with a)

and c), and are likely to use language in an atypical way for b). This observation has been made a number of times in reports on AS, leading some to suggest it is actually impossible to receive an AS diagnosis under the DSM-IV criteria (Mayes, Calhoun, & Crites, 2001; Tryon, Mayes, Rhodes, & Waldo, 2006). As an apparent flaw in the diagnostic criteria, it has inevitably contributed to the decision to change guidelines in DSM-V.

Subsequent studies have either sought to adhere much more closely to the DSM-IV requirements of language delay, and only deploy an AS diagnosis where *no* problems are seen prior to three years, or they have used the presence of phrase speech by three years as a specific discriminator of the two groups, while still counting those with pragmatic language problems as AS<sup>43</sup>. Using these criteria, researchers have investigated any general ability differences that still remain when a language cut-off is used, or any underlying differences in VIQ-NVIQ profile or IQ subtests that may exist between the groups.

An example of using the three-year cut off is seen in Ghaziuddin & Mountain-Kimchi (2004), who compared 22 children and young people with AS and 12 children with HFA ( $Age_M = 12.3$ ) on the WISC and WAIS test batteries. No significant differences were observed for non-verbal IQ, but the groups differed significantly in overall VIQ ( $p = .001$ ) and more marginally in FSIQ ( $p = .06$ ). Underlying the overall differences, specific advantages for AS participants were observed in the verbal subtests for Information and Vocabulary, and the non-verbal Arithmetic subtest. In addition, 18/22 of the AS participants scored higher for VIQ than NVIQ (10 of whom showed a discrepancy of 10 points or over), compared to six of the 12 HFA participants (three of those with a 10 point difference). Thus, even with the age 3 cut off, participants with AS showed differences in general abilities, specific IQ subtests, and cognitive profile.

In an earlier study that strictly followed the DSM-IV criteria, Ozonoff, South and Miller (2000) compared cognitive profile and early language development reports of 12 children with AS and 23 HFA children who were matched for age and gender. On the WISC-III (Wechsler & Corporation, 1991) the two groups only significantly differed in their scores for the Verbal Comprehension subtest, with HFA participants scoring lower than AS participants. For Ozonoff et al. (2000), the two groups differed in level of autistic severity, but little else.

---

<sup>43</sup> This is essentially the same as the autism/AS distinction drawn in ICD 10 (WHO, 1993): children with intact structural language skills prior to three years of age who also have problems with pragmatics are still diagnosed with AS rather than autism.

Similarly, a study by Mayes & Calhoun (2001) reported no significant differences at all in general levels of ability for groups of ASD children with ( $n=23$ ) and without ( $n=24$ ) language delay. Using a mixture of IQ measures (the Stanford-Binet and WISC, specifically), their study observed no differences in full-scale, verbal or non-verbal IQ, or any IQ profile differences, between those who would and would not qualify for a diagnosis of AS based on the 3-year cut off .

A slightly more nuanced picture was suggested by Spek et al. (2009), who compared AS and HFA adults on the WAIS-III (Wechsler & Corporation, 1997). No significant differences in VIQ, NVIQ, or VIQ-NVIQ discrepancy were observed between the two groups, who were differentiated based on the DSM-IV definitions alongside additional questions from ICD-10. However, the authors suggested that relative peaks and troughs across IQ subtests still distinguished the two groups. For example, AS participants showed difficulties in digit span but a strength in comprehension (as indicated by significant within-subjects contrasts), while HFA participants were poor on digit symbol coding, and strong on matrix reasoning (Spek et al., 2008). Thus, though ability levels and overall profile did not differ, adults in the two groups appeared to present with different strengths and weaknesses.

In summary, while earlier studies supported the general notion that individuals with AS will show greater general ability levels than those with autism, particularly for verbal IQ, drawing clear conclusions has been difficult due to variation in diagnostic criteria and practice. Where groups have been diagnosed strictly according to DSM-IV criteria, or just on the presence of language delay before 3, general differences in IQ level or cognitive profile have not always been found. There may be some differences in language –related IQ subtests (Vocabulary, Comprehension and Information specifically) and relatively different strengths and weaknesses between AS and autism, but the evidence does not suggest considerable cognitive ability differences between the two groups.

### *Language skills in AS and autism*

As may be expected, early studies on language skills tended to report greater ability levels in AS individuals compared to those with autism, both in early and later childhood<sup>44</sup>. For example, in Szatmari et al. (1995) over a third of the 4-6 year olds with autism studied still

---

<sup>44</sup> There is a vast literature on language abilities in autism spectrum disorders; see Boucher (2012) and Groen et al. (2008) for recent reviews on the topic. For the purposes of this chapter, only those studies that have explicitly compared language skills in groups of AS and HFA participants are discussed.

had no functional language, whereas all of the AS group did. Scores for receptive language (Reynell & Huntley, 1987), grammatical understanding (Newcomer & Hammill, 1988) and vocabulary (McCarthy, 1972) were all higher in the AS group, who also demonstrated better social and communicative scores on the Vineland Adaptive Behavior Scales (Sparrow, Balla, Cicchetti, & Doll, 1984). These differences were observed even when IQ was controlled for and only those with functional language in the autism group were analysed (Szatmari, et al., 1995). Similarly, in Klin et al. (1995), where no overall differences in full-scale IQ were evident, teenagers with AS participants rated higher on scores of vocabulary, verbal output and articulation than HFA counterparts.

As for cognitive ability, though, subsequent studies have not always supported the presence of general language differences between the two groups. This seems especially to be the case when IQ levels are more comparable, and where differences in language skills have been seen, they have tended to be more specific than general. For example, in the study by Ozonoff, South & Miller (2000), where groups of AS and HFA adolescents did not differ in FSIQ, AS participants scored higher on expressive language scores on the CELF-III (Semel, Wiig, & Secord, 1995), but not receptive scores.

In a similar vein, Verté et al. (2006) reported significant differences (AS>HFA) specifically for speech output, syntax and coherence subscales of the Children's Communication Checklist (Bishop, 1998) in a comparison of two large groups of AS ( $n = 47$ ) and HFA ( $n = 57$ ) children who were matched for age and IQ. No other differences, however, were observed in CCC subscales or overall composite score.

Sometimes, barely any differences in language skill are observed between AS and autistic participants. As in their analysis of cognitive ability differences, Mayes & Calhoun (2001) reported no differences at all in language abilities between ASD children with and without language delay. (Language abilities were assessed using a composite measure of expressive skills including VIQ and rates of specific language characteristics). Likewise, in a study of ASD adolescents and adults, Seung (2007) observed no differences between HFA and AS participants on a general measure of language skill (the Test of Language Competence, Wiig & Secord, 1989). Scores for language construction on a narrative task were also very similar, with AS participants only scoring marginally higher than HFA participants for some grammatical constructions (Seung, 2007).

Finally, in a comparison of adults with AS ( $n = 42$ ) and autism ( $n = 34$ ) matched for age and non-verbal IQ, Howlin (2003) reported marginally higher scores for AS participants on

expressive ratings for the Expressive One Word Picture Vocabulary Test (EOWPVT; Gardner, 1983) but no difference in age equivalent scores on the same task. On a similar measure, the BPVS, no group differences were observed in expressive ratings or age equivalent scores (Howlin, 2003). Based on these data, Howlin argued that language differences between AS and autism, while evident in childhood and adolescence, actually recede in adulthood.

Crucially, in each of the above studies AS participants were distinguished from HFA participants based on the presence of language delay<sup>45</sup>. When this happens, and groups are clearly comparable for other intellectual abilities, overall language scores do not necessarily show up significant differences between those with Asperger Syndrome and those with autism.

Subtler differences may be present, however, in the ways in which language is used to support cognitive processes in both groups. For example, two early studies reported greater verbal memory performance in AS individuals, even when matching or controlling for FSIQ (Klin et al., 1995) or verbal IQ (Ozonoff, Rogers, & Pennington, 1991). Similarly, Iwanaga, Kawasaki, & Tsuchida (2000), in a FSIQ-matched comparison of 4-7 year olds with AS or autism, reported significant differences (AS>HFA) on two verbal subtests (Following Directions and Sentence Repetition) of the Japanese Miller Assessment for Preschoolers (J-MAP; Tsuchida, Sato, Yamada, & Matsushita, 1989). HFA participants have also been reported to be less sensitive to semantic lures in false memory processes (Kamio & Toichi, 2007), and produce less imaginative features in their story-telling (Craig & Baron-Cohen, 2000) than AS participants matched for age and IQ. Thus, the apparent “catching up” of autistic participants, following language delay, does not appear to preclude the existence of persistent atypicalities in language processing – atypicalities that are not necessarily seen in AS.

Before moving on to discuss executive functioning skills, a final issue to consider is the possibility of there being other early language indicators that may be relevant to understanding AS and autism. If the three year cut-off does not always reliably discriminate outcomes within the spectrum, that does not necessarily mean that there are not other

---

<sup>45</sup> In the study by Verté et al. (2006) no specific mention of language delay is made in the outline of how participants were diagnosed: all participants are described as being diagnosed by a multidisciplinary panel with diagnoses being validated with use of the ADI-R (Lord, Rutter & Le Couteur, 1994). However, as the authors note for the subscales of the ADI-R, the ADI-R and DSM-IV largely overlap in their distinction of AS and autism.

indicators in early to mid-childhood that may be better predictors of outcome for ASD individuals.

A study conducted by Bennett et al. (2008) highlighted this issue very well, in one of the few projects to examine the effects of language delay on ASD longitudinally. Rather than comparing groups of ASD participants with and without a history of language delay prior to age 3, Bennett and colleagues assessed the language skills of 19 children with AS and 45 children with HFA every two years from the age of 4-6, up until 15-17. Group diagnosis and presence of specific language impairment (SLI) at either time 1 (4-6 years) or time 2 (6-8 years) were then analysed for their predictive value for later outcomes in autistic symptomatology and adaptive skills. The authors found that the presence of language difficulties at time 2 - defined here specifically as problems with grammar and syntax - predicted later skills more reliably than either the original diagnosis, or time 1 scores. This was especially the case for predicting outcomes in late adolescence (13-17 years). That is, skills at 6-8 years *but not earlier* were the best predictors of eventual outcome (Bennett et al., 2008)<sup>46</sup>.

Bennett et al.'s (2008) findings are limited to specific aspects of structural language skill (i.e. syntax) that do not immediately relate to the kinds of skills that one might expect to be important on a task like the TQT (i.e. semantics). But they are at least suggestive that a categorical approach to early language delay may not be the most useful discriminator of early language differences, and that there might be other time-points for early language skills that are important to measure also.

### *Executive functioning in AS and autism*

Few differences have been observed between AS and HFA individuals in executive functioning skills. In an early study contrasting AS and HFA adolescents, Ozonoff, Rogers, and Pennington (1991) compared composite data on executive functioning abilities based on performance on the Tower of Hanoi and Wisconsin Card Sorting Test (WCST; Heaton, 1993). No group differences were observed on either the composite measure or any of the individual outcome metrics from each of the tasks, while both groups performed worse than matched typically-developing controls. This finding was broadly consistent with the results

---

<sup>46</sup> See also Eisenmajer et al. (1998) for a similar finding concerning the short term predictive value of early language indicators, c.f. Szatmari, Bryson, Boyle, Streiner, & Duku (2003).



of Szatmari et al. (1990), who observed some small differences in perseveration and errors committed on the WCST (with AS outperforming HFA) but not enough to reliably discriminate the two groups. This suggested that the presence of executive impairments was evident in both autism and AS (Ozonoff, Rogers, et al., 1991).

Subsequent studies on planning and cognitive flexibility have largely supported this picture (Manjiviona & Prior, 1999; Miller & Ozonoff, 2000; Ozonoff, et al., 2000; although see Rinehart, Bradshaw, Moss, Brereton, & Tonge, 2001). In their neuropsychological profiles of autism and AS, Manjiviona and Prior (1999) tested participants on the Tower of London (ToL) test and found no significant group differences. In two studies by Ozonoff and colleagues (Miller & Ozonoff, 2000; Ozonoff, et al., 2000), no group differences in planning were observed on either Tower of Hanoi or the Tower of London task from the CANTAB (Robbins, et al., 1994), nor were they evident for two measures of cognitive flexibility, the WCST and the Intradimensional/Extradimensional Test. In a concurrent study to their analysis of CCC scores in autism and AS, Verté, Geurts, Roeyers, Oosterlaan, & Sergeant (2006) compared AS and HFA participants on a large EF battery, including both the WCST and the ToL. Again, no group differences were observed in EF skills.

For other executive tasks a more mixed picture has been observed. Rinehart and colleagues have argued (Rinehart, et al., 2001, 2008; Rinehart, et al., 2002) that individuals with HFA show deficits in inhibition and selective attention that are not seen in Asperger Syndrome. Rinehart et al. (2001) compared groups of 6-20 year-old participants with autism or AS each with separate groups of typically-developing age- and FSIQ-matched controls on a hierarchical stimulus task (similar to a Navon Hierarchical Figures task: Navon, 1977). The task, along with measuring perception of local-global stimuli, required participants to switch between either local or global responses. Compared to their controls, HFA participants were slower to switch from a local response to a global response, whereas no such difference was seen between AS participants and their control counterparts (Rinehart, et al., 2001). Using a similar matching methodology (i.e. no direct AS/HFA comparison, but concurrent comparisons with matched TD controls), Rinehart and colleagues have reported autism-specific difficulties on a Stroop-style inhibition task (Rinehart, et al., 2002) and an inhibition-of-return measure of selective attention (Rinehart, et al., 2008).

These findings come from one research team and are based on measures that have not been widely deployed in studies by other groups: research using different measures of attention and inhibition has not always supported Rinehart et al.'s claims. A small study by Kleinhans, Akshoomoff, & Delis (2005) that used the DKEFS (Delis, et al., 2001) battery to

compare EF skills in adolescents and adults with AS ( $n = 6$ ) and HFA ( $n = 6$ ) reported better visual-scanning performance in AS participants on the DKEFS Trail-Making Test.

However, Verté et al.'s (2006) much larger EF study, which included a number of measures that rely on intact inhibition and spatial attention skills (e.g. the Change task; De Jong, Coles & Logan, 1995; and the Self-Ordered Pointing Task; Petrides & Milner, 1982 ) found no differences between autistic and AS participants.

Moreover, it may be important when talking about such impairments to distinguish between different types of selective attention and inhibition (as noted in chapter 1): reviews of studies on the topic suggest that selective attention, but not prepotent inhibition, is problematic in particular for ASD individuals (Adams & Jarrold, 2012; Christ, et al., 2011). Specific comparisons of this kind between HFA and AS individuals have not – to this author's knowledge – been made to date.

Very few studies have reported specifically on problem-solving skills in AS. The problem-solving task used here, Twenty Questions, has been thus far only been applied to groups of ASD participants who have a diagnosis of autism. In Minshew et al. (1994) and Minshew et al. (2002), all the participants were reported to have possessed a diagnosis of autism, while in Alderson-Day & McGonigle-Chalmers (2011) and Alderson-Day (2011) only one participant in each study possessed a diagnosis of AS. The AS participants in both of the latter cases were either the top or second best performer in the ASD group, suggesting good problem-solving skills, but these data are clearly insufficient to draw any firm conclusions about the wider AS population.

Other studies on AS and HFA groups have deployed measures that are referred to as verbal problem-solving tasks, but the tasks used do not necessarily tap similar processes to the TQT, and have produced mixed results. For example, in Szatmari et al.'s (1990) early clinical profile of an AS group, participants completed the Children's Word Finding Test (Pajurkova, Orr, Byron, & Finlayson, 1976), a task where participants must infer a common missing word from a set of incomplete sentences. No significant differences were observed between AS participants and a group of HFA individuals, suggesting comparable problem-solving abilities, and AS individuals performed significantly worse than typically-developing counterparts (Szatmari et al., 1990).

Aside from the problem that a language delay criterion was not deployed in this particular study (as noted above), a more problematic issue is the requirement to utilise sentence contexts to identify the missing words. Problems with using contextual cues in language and

other domains are thought to be present across the autistic spectrum, and have been a commonly cited finding in support of a weak central coherence deficit (Happé & Frith, 2006). As such, while nominally a verbal problem-solving measure, this task is possibly better interpreted as a measure of contextual processing.

Other verbal problem-solving measures that have been used include the Verbal Absurdities test and the Problem Situations Test of the Stanford Binet battery. Manjiviona & Prior (1999) in their comparison of the neuropsychological profiles of AS and HFA individuals, reported group differences on the latter but not the former, with AS individuals performing better on the Problem Situations task. On both measures, linguistic cues and descriptions must be used to identify an answer, and in this respect, a problem must be solved verbally. Unlike classic problem-solving, though, there is no generation of strategy or selection of possible steps or moves towards solving a problem – or at least, not in the same way as the TQT. Thus, while the findings of Manjiviona & Prior (1999) may broadly support the expectation of an AS/autism difference on the TQT, they only do so in a very general way.

On related tasks, a few studies have examined generative and concept formation abilities in AS as compared to autism. Word generation has been found to distinguish between AS and HFA participants in one recent study: Spek, Schatorje, Scholte, & van Berckelaer-Onnes (2009) reported impaired letter and semantic fluency in a group of HFA adults (compared to controls), but relatively intact performance in a matched group of AS adults. However, previous studies that have tested AS and HFA participants on fluency have not always observed consistent differences (Kleinhans, et al., 2005; Manjiviona & Prior, 1999; Verté, Geurts, Roeyers, Oosterlaan, et al., 2006). In other studies, Klin et al. (1995) reported that AS participants were exhibited greater difficulties with non-verbal concept formation than HFA participants, while Craig, Baron-Cohen and Scott's (2001) observed that children with AS were better than HFA participants at combining unreal categories (e.g. a metal horse) and spontaneously transforming pictures on different types of drawing tasks. There are no measures of generativity or concept formation, though, that have been used in more than one study and have shown consistent AS/HFA differences.

---

In summary, while a number of studies have compared individuals with Asperger Syndrome and autism in terms of their cognitive, linguistic and executive profiles, few conclusive differences have been found. Differences are sometimes seen for general cognitive abilities (in particular VIQ), for specific IQ subtests (such as Vocabulary), and for cognitive tasks

that rely heavily on language skills. Gross differences in cognitive and linguistic abilities are often not evident when the cut-off of age 3 is deployed, but there may be other early language indicators that are more powerful discriminators of subgroups within the autistic spectrum. Executive functioning skills appear to be largely similar in autism and AS for planning and cognitive flexibility, and they may differ for aspects of inhibition, problem-solving and generativity, but further evidence is needed to support specific conclusions about generalisable group differences. Particular components of language processing also need to be examined more specifically before predictions can be made about problem-solving performance in AS and autism; namely, semantics, and how semantic organisation may differ between these two groups.



## **Chapter 7: Problem-solving in Asperger Syndrome and autism**

While few general differences between Asperger Syndrome and autism appear to exist, the preceding review highlights important ways in which the cognitive and linguistic skills of AS and autistic individuals *could* differ in an experimental comparison between the two groups. For the measurement of problem-solving skills, this highlights possible confounds that need to be controlled for, and provides a basis for predictions to be made about TQT performance.

### **Controlling for the right factors**

First, it is clearly important that any comparison between participants will involve groups with comparable levels of full-scale, verbal, and non-verbal IQ. Although individuals with AS may in general have greater strengths in VIQ, it is important to show that any advantage in problem-solving exists over and above differences in VIQ abilities (such as vocabulary). Greater VIQ skills, or a relatively stronger VIQ to NVIQ profile may be taken to reflect the developmental path of AS participants. But only matching or controlling for VIQ would show that, with the same cognitive resources, AS participants are able to approach a verbal problem more efficiently than HFA participants.

A related factor to control for is language ability. The studies reviewed previously show that gross language differences are largely not evident when IQ is properly controlled for, and it is important to note that VIQ assessments, while nominally measures of cognitive ability, will also tap skills covered by standardised language measures, such as the CELF (Semel, et al., 1995). For example, the WASI measures Vocabulary and Similarities both rely upon a sufficient level of expressive and receptive language proficiency. Nevertheless - especially when one considers Asperger's original descriptions of chatty, verbose children - it is important to bear in mind other ways in which a group of AS participants may seem more linguistically able than a group of HFA participants. It is at least plausible that participants in both groups may possess the same vocabulary skills, for instance, and yet AS participants be more fluent in their use and production of such vocabulary. If so, asking questions on a TQT could just be easier for more fluent ASD participants. Thus, the inclusion of a measure of verbal fluency is needed, to control for this confound.

Equally important, though, is to check to what extent participants actually showed delays in their language development in early life, as a diagnosis of either AS or HFA does not necessarily indicate intact or delayed language development respectively. Therefore, any analysis based on diagnosis must be followed by one based on the presence or absence of language delay. Alongside this, additional metrics of early language skills at different ages may also contribute important information on the relationship between language development and later cognitive skills.

Finally, although socio-emotional functioning has not been discussed in detail so far, it is important to ensure that AS and HFA individuals have at least a comparable level of autistic behaviour. If one group showed impairments on the tasks described above but also demonstrated a much more severe clinical profile, any differences in group performance could just result from problems with handling the social-interactive demands of the testing session. Clear conclusions about underlying cognitive differences would therefore be problematic. One way to do this is to utilise a measure of autistic tendencies and behaviour – such as the Autism Spectrum Quotient (Baron-Cohen, Hoekstra, Knickmeyer, & Wheelwright, 2006) – to check the two diagnostic groups do not greatly differ in the degree of their symptomatology.

Bearing these factors in mind, an experiment was planned that compared three groups of adolescents: those with Asperger Syndrome (AS), those with High-Functioning Autism (HFA), and those with a history of typical development (TD). Participants completed the Twenty Questions Task, the Question Discrimination and Plan Construction Tasks, and the Semantic Decision Task. The Vocabulary, Similarities and Matrix Reasoning subtests of the WASI were used to match and otherwise control for cognitive ability levels<sup>47</sup>. Where possible, participants were matched for age, although given that the HFA group all necessarily had a history of language delay, it was expected that older participants may need to be recruited in order to match VIQ levels for this group in particular.

To provide an additional check on diagnoses, the families of participants were asked to complete a short questionnaire on their child's language history. Families were asked to

---

<sup>47</sup> To enable the inclusion of the new tasks in the testing session, the Block Design subtest of the WASI was not deployed in this study. Because of the nature of the TQT, VIQ subtests were prioritised, and of the two NVIQ subtests, MR is related to TQT performance more consistently than Block Design. In addition, MR but not Block Design is used in the two-test short form of the WASI (Wechsler, 1999).

indicate the age of their child's first word and first phrase (following questions on the ADI; Lord et al., 1994). They were also asked to rate their children's overall language competence at 3, 5, 7 and the age they are now. To address potential differences in fluency, two measures of verbal fluency were deployed (the Letter Fluency and Semantic Fluency subtests from the ACE-R; Mioshi, Dawson, Mitchell, Arnold, & Hodges, 2006). To control for possible differences in autistic tendencies, participants' families also completed the adolescent-appropriate version of the AQ (Baron-Cohen, et al., 2006).

### **Making predictions about AS vs HFA performance on the TQT and related tasks**

As outlined in the previous chapter, specific data on verbal problem-solving skills in Asperger Syndrome are relatively scarce, and no studies to date have included an AS-only group in experiments with the TQT. Predictions about problem-solving performance can be made, however, by drawing on two sources: the patterns of data already seen in experiments 1-3, and the findings from studies on related cognitive skills that were reviewed in the previous chapter.

In each of the above experiments, the key outcome on the TQT that differentiated typical and atypical performance was question quality, or the proportion of items that were guaranteed to be eliminated per question. This outcome, above all others, would appear to pick out the similarity in performance seen in deaf and ASD participants, and both the strong and weak arguments propose that this similarity in performance is the product of atypical language development. If these arguments are correct, then participants with AS should not show this difficulty, because of their more typical language history. Therefore, the main outcome on which AS and HFA participants should be expected to differ on the TQT is question quality: if HFA participants show reduced question quality compared to controls, AS participants should not. If typical language development is required for questioning efficiency, AS participants should, if anything, be very similar to their typically-developing peers.

This represents what may be considered the primary outcome of the TQT, but there are also a number of secondary outcomes on the task and other measures that may indicate differences between AS and HFA participants. For these, one can look to the underlying processes required on the TQT, such as executive functions (namely planning, selective attention and working memory), and semantic skills (categorisation and semantic organisation).



Concerning planning, few differences have been evident in studies comparing the abilities of AS and HFA groups. Studies by Ozonoff et al. (1991), Ozonoff et al. (2000), and Manjiviona & Prior (1999) all found no differences in Tower task performance between AS and autistic participants, suggesting similar levels of classically-measured planning abilities. In those studies, both groups showed difficulties in planning, whereas in the first experiment reported here, only less able ASD participants struggled with constructing an effective plan for the TQT (here defined as a plan that systematically narrows down possibilities). As AS participants are likely to be at the more-able end of the spectrum, and there is no reason to assume that their planning abilities differ from able HFA participants, this means that both groups should be able to produce narrowing plans.

Where AS and HFA may be expected to differ, however, is in the general efficiency of their planning. In experiment 1, ASD participants selected less efficient questions on the TQT *and* on the the planning task. This appeared to indicate a common problem with concept selection on the two tasks. Therefore, if AS participants were expected to be more efficient than HFA participants in their questioning on the TQT, they should also show advantages in the expected question quality of their plans - irrespective of any similarities the two groups may have in their use of narrowing<sup>48</sup>.

Previous research suggests no clear differences in selective attention abilities between AS and HFA individuals, although it has been suggested that inhibitory skills may be more likely to be impaired in people with autism (Rinehart, et al., 2002). This may imply that individuals with AS are less likely to impulsively guess at answers on Twenty Questions than HFA participants. Beyond, this, however, it is not clear that selective attention abilities in the two groups would lead to diverging performance on the TQT.

A potentially more important difference is in verbal working memory skills. If participants with AS have greater levels of ability for cognitive processes that have to draw on language, then their response to TQT trials where elimination is prohibited may differ. Specifically, they may be expected to show less of a decrement in performance, as compared to the increased number of questions participants with autism use when faced with an unchanging set (Alderson-Day, 2011). Therefore, it was predicted that HFA participants would perform worse when elimination is prohibited, but AS participants would not.

---

<sup>48</sup> For question discrimination, based on the previous ASD data from experiment 1, intact performance was expected in HFA participants. Accordingly, there was no reason to assume otherwise for an AS-only group.

Thus, in addition to the hypothesised difference in question quality, specific predictions of how AS and HFA participants may differ can be made for performance on the planning tasks and other outcomes on the TQT itself. Predicting responses on the semantic decision task is more difficult, however. First, the measure was new for both AS and HFA participants; in contrast to the planning tasks, this particular tool had not been used before with an ASD group. Second, the studies on language skills that were discussed in chapter 6 only compared AS and autistic participants in a very general way. Making specific predictions about SDT performance required a more focussed exploration of semantic processing in individuals in autism, and how this compares to the same sorts of skills in Asperger Syndrome. As such, before outlining the main hypotheses about this task, the following section provides a brief review of semantic categorisation and organisation skills in autism and AS.

### **Semantic processing in autism**

In contrast to other structural language skills, semantic processing has historically been an area of language and cognition that is thought to be specifically disrupted in autism spectrum disorders (e.g. Hermelin & O'Connor, 1970; Kjelgaard & Tager-Flusberg, 2001). Semantic development has been proposed to be delayed or atypical in autism by various researchers, who have suggested that the early difficulties in social interaction and communication shown by autistic children also disrupt the process by which they acquire new word meanings (e.g. Hobson, 1995).

Problems with language acquisition for young autistic children are consistent with such an idea. But it is also important to show that any such atypicalities in semantics endure in later years for autistic children, rather than simply reflecting a delay in semantic vocabulary. Young children with autism who show poor semantic skills are likely to just have language delay. In contrast, children whose language abilities catch up (in terms of vocabulary for example), but *still* show differences in underlying semantic processes (such as encoding or organisation) point to an enduring difference that reflects their language history.

Research on categorisation skills in autism appeared to suggest that at least basic semantic abilities are not necessarily impaired in autism: sorting and identification of basic and superordinate categories in autistic participants have been observed to be intact (Tager-Flusberg, 1985a, 1985b; Ungerer & Sigman, 1987). However, later studies also highlighted problems with more abstract or complex semantic sorting (Shulman, et al., 1995), slowed responses to atypical exemplars of categories (Gastgeb, et al., 2006) and preferences for

concrete over abstract criteria in sorting tasks (Ropar & Peebles, 2007). This pattern of findings arguably implicates abnormalities in underlying semantic processes, even when basic categorisation abilities are intact.

Further evidence of this kind of residual semantic atypicality was reported in a study by Kelley, Paul, Fein, & Naigles (2006), that compared the language abilities of 14 children with an ASD (age-range 5-9) and 14 typically developing age and sex-matched peers. The ASD children were described as being an “optimal outcome” group in the sense that they were all a) within the normal range for IQ, b) placed within mainstream schools, and c) considered to be functioning at a level comparable to their fellow students. While the ASD group was mixed in terms of prior diagnosis (mostly PDD-NOS, but also AS and autism), re-assessment by the study authors at the time of testing confirmed that most actually qualified for autistic disorder or PDD-NOS, and all but two showed evidence of delays in language skills in the first 36 months. On standard measures of expressive vocabulary, no differences were observed between ASD and TD participants. ASD participants also showed relative strengths in grammatical understanding and phonology. However, ASD participants were observed to be worse than their counterparts in their lexical semantic skills, as measured on tasks where a new category term must be inferred by participants. This suggests that ASD children may retain good word knowledge without a fuller understanding of semantic structures. As put by the authors, the ASD participants in their study

“knew many words, thus demonstrating understanding of the identification function of words and concepts...[but] What they may have had difficulty with are other conceptual functions, such as that which motivates the relations between concepts and that which allows for the induction of hidden properties” (Kelley, et al., 2006, p. 821) .

What is notable about the study by Kelley and colleagues is that it examined a very able group of autistic children and assessed them in-depth on a range of language measures. In this group, assessment with standardised language measures alone did not highlight significant differences with TD participants – showing what can clearly “catch up” following language delay – but further assessment of complex and subtle linguistic skills highlighted differences in their semantic skills.

Evidence from electrophysiology research supports the existence of underlying semantic abnormalities in autism. In a study using electroencephalography (EEG), Dunn & Bates (2005) compared electrophysiological responses to auditorally presented words in 18 children with autism and 18 typically-developing controls, matched for age and NVIQ, and

split into younger (age 8) and older (age 11) age groups. The paradigm consisted of participants listening to a series of words that were either a member of a particular semantic category or not. While younger autistic participants showed an abnormal, early response to the words (N1), this was not seen in older ASD participants (aged over 11). However, compared to controls both younger and older participants showed reduced moderation of the N400 response (an indicator of semantic congruency) for in vs out-category words. This result replicated the findings of an earlier experiment by the same research group (M. Dunn, Vaughan, Kreuzer, & Kurtzberg, 1999), and was more prominent for older ASD participants, suggesting a persistent, rather than declining, atypicality in semantic processing.

Individuals with autism also appear to draw upon different cognitive resources to support their semantic skills. For example, a study by Toichi and Kamio (2001) used a word priming paradigm to examine the semantic associations of a group of adolescents and young adults with autism compared with age and IQ-matched typically developing controls. In the task participants were presented with a prime word for two seconds (printed on a card) and then a partially complete target word for 15 seconds, after which they had a minute to provide the correct word (there was only one possible answer each time). Primes and targets were either semantically related (e.g. bus-train) or unrelated (TV-tie). In both HFA and TD groups, performance was more accurate for semantically related pairs of words than unrelated pairs, suggesting similar semantic associations for common words. However, correlations with task performance in each group appeared to differ: the performance of HFA participants was significantly related to scores on the Raven's Progressive Matrices and other measures of NVIQ, a relation that was not seen for TD participants. The authors speculated that "...the two groups might have employed different strategies. For example, individuals with autism may be more dependent on nonverbal strategies, such as visual imagery, which results in manipulating language differently from individuals without autism" (Toichi & Kamio, 2001, p. 488).

Use of other cognitive resources in semantic processing was also documented in a study by Gaffrey et al. (2007) using functional magnetic resonance imaging (fMRI). The study examined the neural correlates of semantic decision making in a mixed group of 10 adults with autism spectrum disorder (8 autism, 2 AS), and 10 age and NVIQ-matched controls. The semantic decision task (which provided the basis for the measure used in chapter 5) consisted of a series of basic exemplar words, presented consecutively. Participants had to indicate for each word whether they were part of a given semantic category, such as tools. In contrast to controls, ASD participants were significantly less accurate for judging colour and

feeling words, but not tools. More importantly, though, ASD participants appeared to utilise different cortical regions in their semantic processing, drawing particularly on extrastriate, visual cortex areas of the brain when classifying different words<sup>49</sup>.

While these findings appear to converge to suggest atypical processing of semantic information in autism, this is not always observed. In contrast to the above results, Whitehouse, Mayberry & Durkin (2007) argued that semantic skills were typical in individuals with autism for encoding of new semantic information. Using a word recall paradigm, the authors examined recall for semantically and phonologically related words in 20 children with autism and 20 typically-developing children matched for verbal age and reading ability. No differences were observed for the number of items recalled in either condition, suggesting that semantic relations were typical enough to support recall processes in ASD participants.

### **Direct comparisons of semantics in autism and Asperger Syndrome**

Very few studies have directly compared semantic processing in autism and Asperger Syndrome, perhaps because language skills in general are often assumed to be better in the latter. In fact, if the link between social interaction and semantic development is assumed to be a valid one (and a number of developmental theorists would make that assumption, e.g. Baldwin & Tomasello, 1998; Gopnik & Meltzoff, 1992), an important conclusion follows: children across the autistic spectrum should show evidence of atypical semantic development, whether they have autism, Asperger Syndrome or other forms of ASD. That is, one should not expect semantics to necessarily be “normal” in children with Asperger Syndrome, because the quality of their early social interactions will still likely differ from typically-developing children<sup>50</sup>.

Where they might be expected to vary is in the extent of that atypicality, as some children may be able to engage in the interactions necessary for acquiring semantics more readily than others, because of their additional structural language skills (for an example of this, see

---

<sup>49</sup> See also Harris et al. (2006), for an example of irregular responses to abstract vs concrete semantic terms in ASD individuals. That study is not discussed here because it used a mixed ASD group with a high proportion of AS participants (5/14), without reporting on within-group differences. Nevertheless, its findings are broadly consistent with those discussed above.

<sup>50</sup> This is essentially Lorna Wing’s (1981) point: children with AS may show greater vocabulary than in autism, but the semantic knowledge they develop is not quite the same as that seen in the lexicons of typically developing children, because their difficulties with social interaction will affect word acquisition and use.

McGregor & Bean, 2012). Comparisons between Asperger Syndrome and autism, in terms of semantics, may be likely to be one of degree: semantic atypicalities in both, but only significantly so in the case of children with autism.

The majority of categorisation studies discussed previously have only examined groups of individuals with autism, or have used mixed groups with the generic label of autism spectrum disorder. One exception is Verté et al. (2006), in which the SON-R Categories test was applied to groups of AS, HFA, PDD-NOS and typically developing participants. However, in this case no significant group differences were observed.

In some cases, the same researchers have deployed similar paradigms across separate studies with AS and HFA groups. In a follow-up study to Toichi & Kamio (2001), Kamio, Robins, Kelley, Swainson, & Fein (2007) examined whether AS individuals show similar differences in their approach to semantic tasks. The study compared a mixed group of AS and PDD-NOS children ( $n = 11$ ), all of whom showed no delays in their language development, to a group of age and IQ matched TD children on a lexical decision task. At the time of writing, the authors considered this to be “the first [study] to address automatic lexical/semantic abnormalities in individuals with HFASD without a history of early language delay” (Kamio, et al., 2007, p. 1120).

As in Toichi & Kamio (2001), the paradigm depended on the presentation of a prime (presented for 250ms), and then a response that was either semantically related or unrelated to that prime. In this study, the response depended on judging whether the target word (presented for 4s) was an actual word (e.g. bread) or a non-word (e.g. boarb). Whereas control participants exhibited a facilitation effect for semantically-related pairs of prime and target words, AS/PDD-NOS participants demonstrated no such effects, suggesting irregular associative processes in their response to words. Kamio et al. (2007) used this lack of an effect to suggest that AS individuals, just like language delayed autistic individuals, exhibit abnormal semantic processing.

When compared to the typically-developing individuals in their study, this conclusion is reasonable: the AS group in Kamio et al. (2007) were no less accurate or slower in their responses to words, but did not show the benefit that control participants did when words were semantically-linked. This suggests an underlying difference in how responses to words

were arrived at, rather than a specific problem with word comprehension, for example, or a more general difference in speed of processing.

The findings do not, however, suggest that associative processes were necessarily identical across the ASD groups reported on in Toichi & Kamio (2001) and Kamio et al. (2007). First, the HFA participants in Toichi & Kamio (2001) *did* show semantic priming effects, whereas the AS/PDD-NOS participants in Kamio et al. (2007) did not. Second, the paradigms utilised in each study differed notably; whereas in Toichi & Kamio (2001) participants viewed a prime for two seconds, primes were displayed for 250 *milliseconds* in Kamio et al. (2007): as noted by the authors, the former engages explicit processing, whereas the latter engages much more automatic, implicit processes. Finally, the tasks differed in what participants were asked to do: in Toichi & Kamio (2001), the identification of a possible word to fit the partially-completed target is a processing-heavy, explicit search of possibilities. In contrast, the lexical decision paradigm employed by Kamio et al. (2007) asks for a much quicker yes/no decision about the validity of a presented word. As such, while the two studies both point to potential abnormalities in ASD semantic processing, it is not at all clear that they speak to exactly the same effects or processes.

A recent study that compared HFA and AS participants on the same paradigm was conducted by Speirs et al. (2011). Rather than examine semantic priming, Speirs and colleagues used a lexical decision task to measure priming based on either phonological or orthographic similarity; that is, the sound of words as compared to the actual form of the words themselves. Primes that were identical to their target necessarily contained both types of similarity (e.g. blue-BLUE), while other primes were similar in one respect more than another: for example, the pair “blue-BLEW” provided a full phonological prime without a full orthographic prime, whereas “blue-BLOG” contained some orthographic association, but less of a phonological cue for the target word. Primes appeared for 54ms and targets for 500ms. In typically-developing individuals, the identical prime was expected to produce significantly faster responses than primes that are only similar, but not identical, to their target, reflecting an automatic processing response.

In Speirs et al. (2011), 11 participants with AS, 11 with HFA and 11 typically-developing controls, matched for age and IQ, were compared on the lexical task. TD participants showed the expected effect, with identical primes producing significantly quicker responses than orthographically similar primes. The AS group showed a similar effect, suggesting easy

access to the target words when primed with the identical word. In contrast, the HFA group showed no difference between identical pairs of words and orthographically similar pairs of words (both of which were responded to quicker than baseline unrelated pairs). This, the authors suggested, implied an immature or at least differential process by which words were being accessed in the autistic lexicon (Speirs et al., 2011).

In contrast to the studies by Kamio and colleagues, the paradigm used by Speirs et al. (2011) does not assess semantic knowledge or semantic organisation, and the ability to generalise their findings is limited by the small sample of participants studied. Despite these caveats, its strength is that it allows for direct comparisons between matched AS, HFA and TD participants. In this particular case, access to the lexicon in AS individuals would appear closer to typical development than autism.

### **Making predictions about semantic decision performance in AS and HFA**

To this author's knowledge, no studies to date have directly compared AS and HFA individuals on the same measures of word association (e.g. Marschark et al., 2004), semantic priming (e.g. Toichi & Kamio, 2001) or semantic decision (e.g. Gaffrey et al., 2007). In the absence of such studies, predicting the comparative performance of AS and HFA participants on the semantic decision task used in the preceding chapter is speculative. Bearing these caveats in mind, some initial predictions could at least be made.

First, it was hypothesised that participants with autism would show atypical performance on the semantic decision task. If the strong argument outlined in previous chapters is taken to be correct, then it followed that atypicality would be evident as a discrepancy in reaction times for basic and superordinate category judgements on the SDT (as it was in deaf participants). Furthermore, if this was to serve an explanatory purpose, then this discrepancy should predict their performance on the Twenty Questions Task.

Second, it was thought that AS participants were also likely to show atypicalities on the semantic decision task. However, this was predicted to be less evident than in autistic participants, and to either not correlate or be related to a lesser degree to TQT performance (assuming that problem-solving performance is intact).



## **Experiment 4: Comparing AS and autism on the Twenty Questions Task and Semantic Decision Task**

### **Method**

#### *Participants*

Fifteen children with Asperger Syndrome (14 M: 1F; ages 9-16) and 15 children with autism (14 M: 1F, ages 9-18) were recruited from the Edinburgh and Lothians area. As in experiment 1, participants' original diagnoses were made by local clinical services, who use a range of gold-standard diagnostic measures, including the ADI-R (Lord, Rutter & Le Couteur, 1994) and the ADOS (Lord et al., 2000). A subset of the participants ( $n = 5$ ) had also had their diagnosis reconfirmed within the past three years by a trained researcher using the ADI-R. Following recruitment, parents were asked to confirm any prior diagnoses and complete the questionnaire on their child's language history (see below). Exclusion criteria included the presence of any other neurological conditions, language impairments or specific reading difficulties that could prejudice the results, with reference specifically to the language demands of the tasks (e.g. dyslexia). Participants with other developmental disorders known to be comorbid with ASD (such as ADHD) were not excluded in the first instance, although in the event only one ASD participant had a dual diagnosis of autism and ADHD<sup>51</sup>. All ASD participants fell within the high-functioning range (FSIQ >70). Fifteen typically-developing children (10 M/5 F; ages 9-18) were recruited from the Edinburgh and York areas to provide a neurotypical control group.

#### *Recruitment & settings*

Participants with an ASD were primarily recruited via mail-outs from a local charity for autism and parent groups from the region. Families interested in the research were asked to return a reply slip attached to an information letter explaining the research. Following this, families that were known to the university due to involvement in prior research were contacted via email or telephone to complete the participant groups. Typically-developing participants were recruited from a database of local families that had been involved in research with the department before. Testing sessions were conducted either in a quiet room at the university, or, where it was more convenient for families, in the homes of participants.

---

<sup>51</sup> This participant was marked for later analysis, just in case their data were unrepresentative of the HFA group.

### *Matching*

Table 9 shows the age and IQ data for the three participant groups. IQ estimates were based on the Vocabulary and Similarities (VIQ) and matrix reasoning (NVIQ) subtests from the WASI. Although not individually matched, the three groups did not significantly differ for overall full-scale, verbal and non-verbal IQ, based on scores on the WASI (Wechsler, 1999). However, mean differences in FSIQ and VIQ were clearly approaching significance between HFA and TD participants. No significant pairwise differences were observed for verbal mental age ( $AS_{M(SD)} = 17.13 (6.05)$ ,  $HFA_{M(SD)} = 15.22 (5.28)$ ,  $TD_{M(SD)} = 19.08 (6.71)$ ) or non-verbal mental age ( $AS_{M(SD)} = 13.20 (2.91)$ ,  $HFA_{M(SD)} = 13.01 (2.80)$ ,  $TD_{M(SD)} = 11.65 (3.22)$ ), although a trend was observed for TD participants to have greater VMA than HFA participants ( $t = -1.758$ ,  $df = 28$ ,  $p = .091$ ). Attempts were made to match the groups for age, although in order to recruit sufficient numbers of HFA participants with comparable ability levels to the AS group, the eventual HFA group was significantly older than the group of AS participants ( $HFA > AS$ ,  $t(28) = 2.157$ ,  $p = .040$ ).

One point to note here is that the HFA participants included in this study were of a generally higher level of ability than those who took part in experiment 1, particularly for VIQ. The two groups differed both in terms of mean VIQ (Ex1: 95.55 vs Ex4: 104.40) and range (Ex1: 65-141 vs Ex4: 74-132).

### *Design*

A between-groups (AS vs HFA vs TD) design was deployed, comparing participants on the TQT and subsequent control tasks. Following this, within-group analysis was conducted to assess the relationship between TQT performance and other task scores.

### *Materials & procedure*

#### *- The Twenty Questions Task (TQT)*

The first task attempted was the TQT. The stimuli, format and instructions for the TQT deployed was identical to those used with young deaf participants in the preceding experiment. Participants completed three games of Twenty Questions using a 24-item set: the first two trials allowed for elimination of items during search while on the last trial elimination was prohibited. Alongside the game board a Dell Inspiron 17 “Laptop was used to provide the “random selector” animation and audio-visual feedback during the game. Prior to starting participants were asked if they had played Twenty Questions or similar games (such as *Guess Who?*) before.

**Table 9. Age and IQ scores for AS, HFA and TD participants.**

	AS ( <i>n</i> =15)		HFA ( <i>n</i> =15)		TD ( <i>n</i> =15)		<i>Sig.</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Age	12.93	2.09	14.74	2.49	14.05	2.72	HFA>AS*
FSIQ	107.40	14.48	102.73	13.71	110.40	10.54	(TD>HFA)
VIQ	111.00	18.05	104.40	16.36	113.80	12.62	(TD>HFA)
NVIQ	102.20	11.67	99.87	12.11	105.33	16.72	n.s.

\*  $p < .05$ . Differences in brackets indicate contrasts approaching significance ( $p < .1$ )

AS = Asperger Syndrome, HFA = High-Functioning Autism, TD = Typically Developing.

- *Question Discrimination (QD) and Plan Construction (PC)*

Following the TQT, all participants attempted the question discrimination and plan construction tasks from experiment 1. For QD, participants completed 10 forced-choice question discriminations (i.e. is it better to ask X or Y first?). For PC, participants were asked to select five questions that would be useful to use “if we were to play the game (Twenty Questions) again in a moment”. Once five questions were selected, participants were then asked to order them in terms of which question they would ask first, which they would ask next, *etc.* In a departure from experiment 1, however, participants did not then attempt further TQT trials.

- *Semantic Decision Task (SDT)*

The semantic decision task from the preceding experiment was deployed. As with deaf participants, the first trial of the first round (identifying “LOOT”) was used as a practice trial and excluded from later analyses.

- *Letter and Semantic Fluency*

To assess verbal fluency abilities the Letter and Semantic Fluency subtests from the Addenbrooke ‘s Cognitive Examination – Revised (ACE-R; Mioshi, et al., 2006) were administered. For letter fluency, participants were given the following instructions:

*“I’m going to give you a letter of the alphabet and I’d like you to generate as many words as you can beginning with that letter, but not the names of people or places. Are you ready? You’ve got a minute, and the letter is P”*

Following this, participant completed the semantic fluency task for the category “animals”.

- *AQ and Language Questionnaires*

As outlined above, the families of participants were asked to complete the AQ-Adolescent Version (Baron-Cohen, et al., 2006), a measure of autistic traits. The AQ consists of 50 items describing specific behaviours, to which respondents indicate their level of agreement. Each item has four response options (*Definitely Agree, Slightly Agree, Slightly Disagree, Definitely Disagree*).

The language questionnaire was a bespoke measure consisting of six questions, all to be completed by the participants’ family. The first two concerned early language milestones.

1. *What age was your child when they said their first word?*

2. *What age was your child when they used their first phrase of three or more words?*

Parents were asked to indicate, where possible, to the nearest month, or provide their best estimate if unsure.

The following four questions asked parents to rate their children's overall language competence at ages 3 and up.

*Compared to other children, what were your child's language skills like*

*3. At three years of age?*

*4. At five years of age?*

*5. At seven years of age?*

*6. Compared to other children, what are your child's language skills like now?*

Parents indicated their ratings on a 5-point Likert scale (*Much worse, A bit worse, About the same, A bit better, Much better*).

#### *Scoring & analysis*

Scoring and analysis for the TQT, QD, PC and SDT tasks were the same as reported in previous chapters. The primary outcome upon which problem-solving was assessed was again mean question quality. Unless otherwise stated, parametric statistics (primarily analysis of covariance) were used to compare the three groups on the main task outcomes. Covariate analysis, using age, VIQ and NVIQ as covariates, was used to assess influences of age and general ability. In addition, matched subsets of participants were used to follow up the covariate analysis (as in previous chapters).

The ACE-R fluency tests can be scored on a scale from 0-7 depending on the number of words generated. To maximise variance in participants' scores, however, raw scores were used instead to compare the three participant groups.

The AQ provides a total score, and subscale scores for Social Skills, Attention to Detail, Attention Switching, Communication, and Imagination. In the original AQ (Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) and AQ-Adolescent (Baron-Cohen, et al., 2006), items were scored 0 or 1 depending on the presence or absence of an autistic trait (i.e., positive and negative options are collapsed into two categories), on a scale from 0 to 50. However, later developments of the AQ (Auyeung, Baron-Cohen, Wheelwright, & Allison, 2008), have scored items on a four point scale (0-3), reflecting the four options that respondents can choose, and providing scores from 0-150. The present study used the latter scoring in order to maximise the amount of information provided by participants.

## Results

Prior to the main analysis, potential confounds of game experience, prior study participation and AQ levels were checked. All but one participant (a control) had prior experience of Twenty Questions in some form. No significant differences in question quality were observed between participants who had ( $n = 19$ ) and had not ( $n = 26$ ) attempted the TQT before ( $t(43) = 0.126, p = .900$ , two-tailed t-test used). Despite repeated requests, one family in the HFA group did not provide a completed AQ or language questionnaire. In the remaining data, both AS ( $AQ_{M(SD)} = 105.73(19.08)$ ) and HFA ( $AQ_{M(SD)} = 101.50(23.33)$ ,  $n = 14$ ) participants scored higher than TD participants ( $AQ_{M(SD)} = 46.13(22.42)$ ). Crucially the two ASD groups did not significantly differ from one another ( $t(27) = 0.537, p = .596$ , all subtests  $p > .50$ ).

### *Twenty Questions Task*

Table 10 displays the main task outcomes for the TQT. A univariate analysis of covariance was used to compare mean question quality scores in the three groups, using age, VIQ and NVIQ as covariates. A significant main effect of group was observed ( $F(2,39) = 4.858, p = .013, \eta^2_p = .199$ ), alongside positive contributions for age ( $F(1,39) = 4.033, p = .052, \eta^2_p = .094$ ) and VIQ ( $F(1,39) = 10.283, p = .003, \eta^2_p = .209$ ; all other effects *n.s.*). Bonferroni-corrected pairwise comparisons indicated that the questions of HFA participants were significantly less efficient than those of AS participants ( $p = .021$ ) and TD participants ( $p = .041$ ). No difference was observed between AS and TD participants ( $p = 1.0$ )<sup>52</sup>.

A less marked but nonetheless significant difference was also observed for the mean number of questions taken on each trial. A similar ANCOVA was run, producing a main effect of group ( $F(2,39) = 4.310, p = .020, \eta^2_p = .181$ ) and significant effect of Age ( $F(1,39) = 12.670, p = .001, \eta^2_p = .245$ ; all other effects *n.s.*). HFA participants were likely to require

---

<sup>52</sup> The HFA participant with a dual diagnosis of ADHD scored well within the range for his group (Mean QQ = 0.29). When the analysis was rerun without his data, almost identical results were produced.

**Table 10. Task performance by AS, HFA, TD participants.**

	AS		HFA		TD		Sig.
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
<b>TQT</b>							
QQ	0.33	0.05	0.27	0.07	0.33	0.50	HFA<AS,TD* HFA>TD*, (HFA>AS)
QT	5.05	1.02	5.53	1.41	4.62	0.84	
Grouping (mean %)	66.35	13.76	59.92	15.85	65.07	15.23	<i>n.s.</i>
Guessing (mean %)	14.76	16.41	10.55	12.92	6.38	7.36	<i>n.s.</i>
<b>QD</b>							
	7.86	2.17	7.13	1.73	8.13	0.99	<i>n.s.</i>
<b>PC</b>							
QQ1	0.30	0.14	0.27	0.15	0.33	0.12	<i>n.s.</i>
QQ2	0.22	0.12	0.21	0.10	0.26	0.12	<i>n.s.</i>
QQ3	0.20	0.08	0.13	0.10	0.21	0.11	(HFA< AS, TD)
QQ4	0.24	0.13	0.20	0.10	0.20	0.10	<i>n.s.</i>
QQ5	0.17	0.10	0.16	0.10	0.19	0.13	<i>n.s.</i>
Mean QQ	0.23	0.06	0.19	0.07	0.24	0.04	<i>n.s.</i>
<b>SDT</b>							
Accuracy 1	27.40	2.74	27.07	3.81	27.73	2.71	<i>n.s.</i>
Accuracy 2	27.13	2.36	28.33	2.35	29.00	1.13	AS<HFA, TD*
Accuracy 3	25.87	2.13	26.53	2.59	27.40	2.26	<i>n.s.</i>
Mean RT1 (ms)	653.15	231.51	657.27	291.12	563.14	131.97	<i>n.s.</i>
Mean RT2 (ms)	879.36	279.15	870.08	313.20	734.39	139.61	<i>n.s.</i>
Mean RT3 (ms)	1009.75	410.11	1011.88	347.32	853.44	205.02	<i>n.s.</i>
<b>Verbal Fluency</b>							
Letter (P)	9.20	5.27	9.73	4.25	12.40	3.98	TD>AS*, (TD>HFA)
Semantic (Animals)	17.60	6.56	17.53	5.66	21.40	4.22	TD>AS*, (TD>HFA)

\*  $p < .05$ . Differences in brackets indicate contrasts approaching significance ( $p < .1$ )

AS = Asperger Syndrome, HFA = High-Functioning Autism, TD = Typically Developing, TQT = Twenty Questions Task, QD = Question Discrimination, PC = Plan Construction, SDT = Semantic Decision Task, QQ = Question Quality, QT = Questions used per Trial.

significantly more questions than TD participants ( $p = .026$ ), and, to a lesser degree, AS participants ( $p = .083$ )<sup>53</sup>.

Non-parametric statistics were used for the rates of grouping and guessing. The use of grouping questions was high (60-65%) in all groups, and on average outright guesses were used twice as much in AS and HFA participants compared to TD participants. However, Kruskal-Wallis ANOVAs indicated no significant differences between the groups (all  $p > .400$ ).

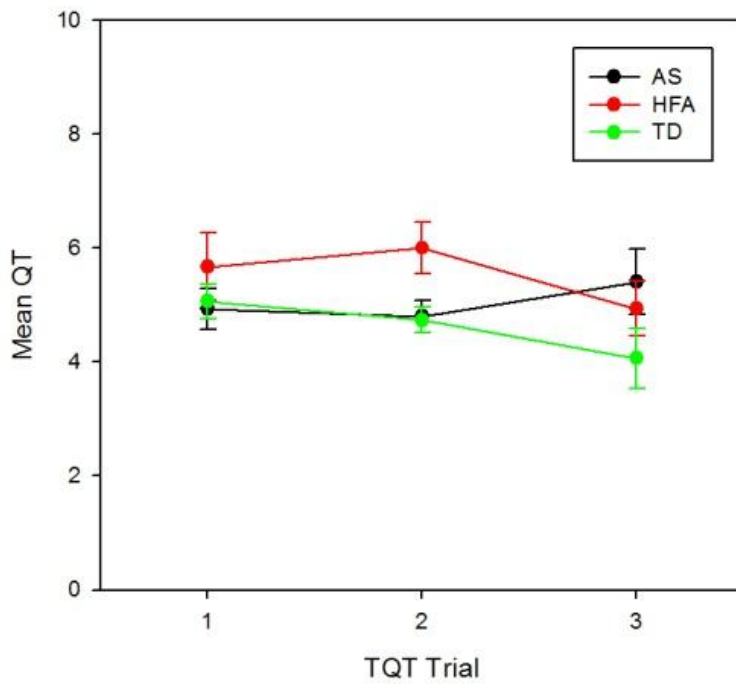
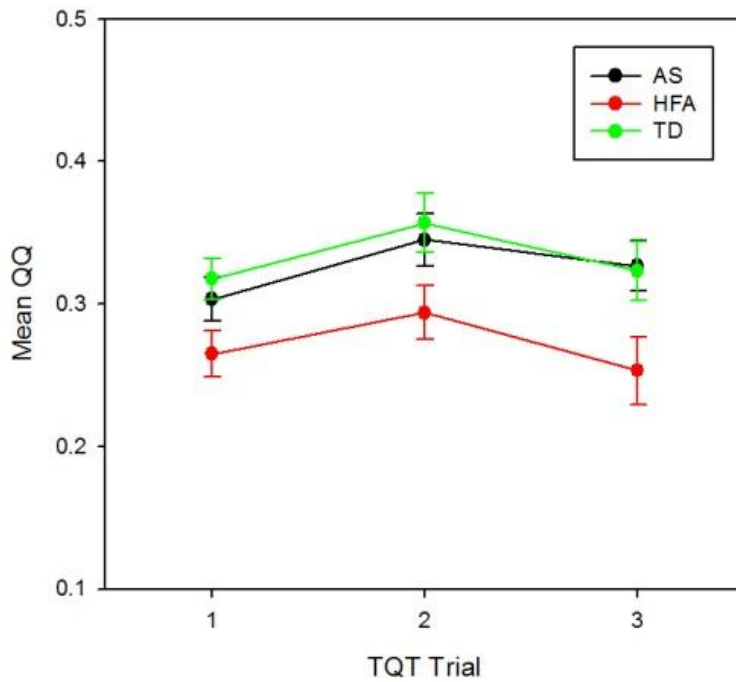
Mixed ANCOVAs were used to assess changes in efficiency across the three task trials. Despite the switch from allowing elimination (trials 1 & 2) to prohibiting it, no significant Group\*Trial interactions were evident for question quality (all  $p > .300$ , all  $\eta^2_p < .060$ ) or the number of questions taken (all  $p > .290$ , all  $\eta^2_p < .065$ ). As figure 24 shows, this suggests that overall group differences on these variables were fairly consistent across trials. Changes in grouping and guessing were assessed non-parametrically using individual Friedman's ANOVAs in each group. These analyses indicated no significant changes in the use of guesses, but changes for grouping in AS ( $X^2(2) = 5.216, N = 15, p = .075$ ) and HFA ( $X^2(2) = 7.509, N = 15, p = .020$ ) participants. AS participants generally improved in their grouping before dropping on trial 3, whereas HFA participants used grouping less with each trial (see figure 25).

Finally, the group differences in question quality were re-analysed using matched subsets of participants. It was possible to match 12 participants in each group to within 10 points for verbal IQ (AS-12  $VIQ_{M(SD)} = 108.92 (14.45)$ ; HFA-12  $VIQ_{M(SD)} = 110.33 (11.82)$ ; TD-12  $VIQ_{M(SD)} = 110.58 (11.60)$ ; all pairwise contrasts  $p > .750$ ). The groups also did not differ on NVIQ (AS-12  $NVIQ_{M(SD)} = 102.25 (11.41)$ ; HFA-12  $NVIQ_{M(SD)} = 103.50 (17.47)$ ; TD-12  $NVIQ_{M(SD)} = 103.50 (11.60)$ ; all  $p > .780$ ), although some age differences between the groups were still apparent (AS-12  $Age_{M(SD)} = 155.25 \text{ months} (25.71)$ ; HFA-12  $Age_{M(SD)} = 176.25 (33.08)$ ; TD-12  $Age_{M(SD)} = 168.50 (35.22)$ ; AS<HFA:  $t(22) = -1.736, p = .096$ ).

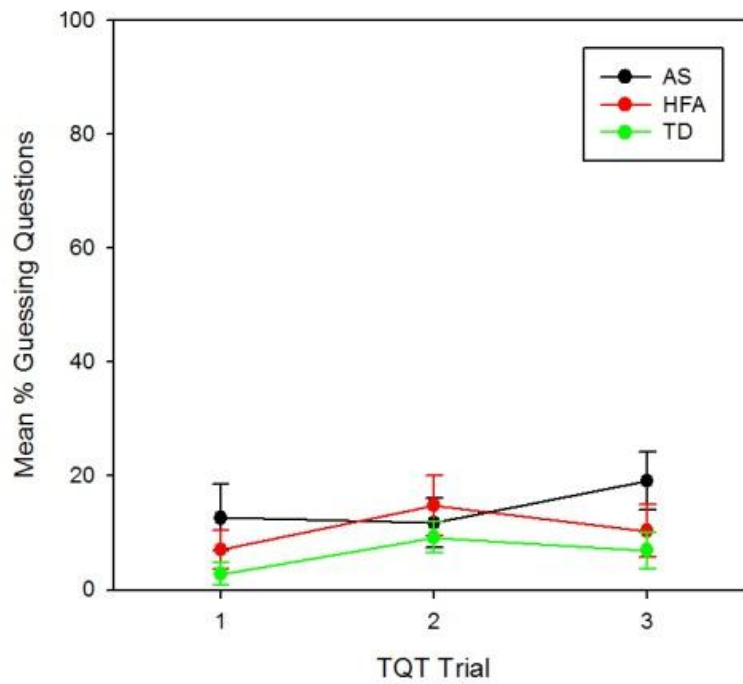
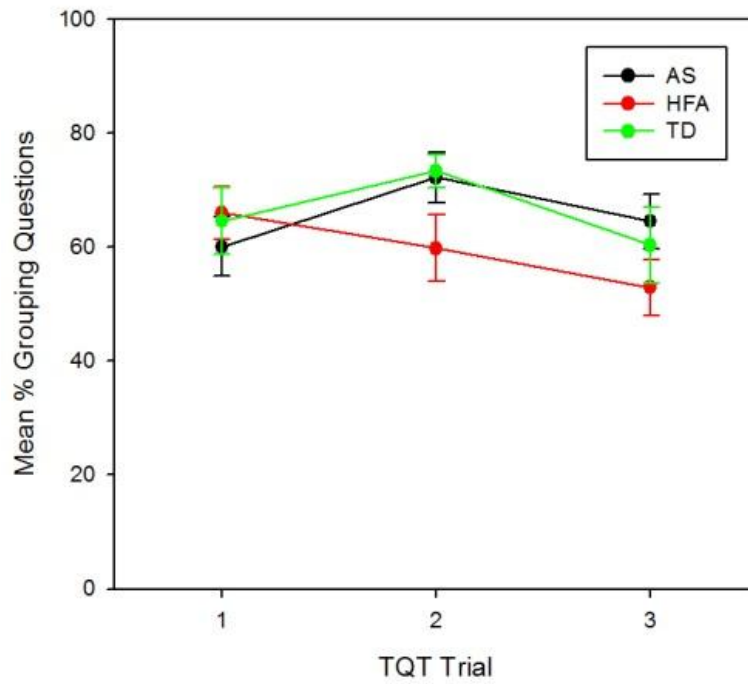
---

<sup>53</sup> When VMA was included in this analysis as an alternative covariate, very similar results were returned for both QQ (group  $F(1,41) = 3.931, p = .027, \eta^2_p = .161$ ) and QT (group  $F(1,40) = 3.426, p = .042, \eta^2_p = .146$ ).





**Figure 24. Changes across Twenty Questions trials in question quality (top) and questions used per trial (bottom) in each group.** TQT = Twenty Questions Task, QQ = Question Quality, QT – Question used per trial, AS =Asperger Syndrome, HFA = High-Functioning Autism, TD = Typically-Developing.



**Figure 25. Changes across TQT trials in grouping questions (top) and guess questions (bottom) in each group.** TQT = Twenty Questions Task, AS = Asperger Syndrome, HFA = High-Functioning Autism, TD = Typically-Developing.

A univariate analysis of covariance, this time only including Age as a covariate, was applied to the three subsets of participants for their mean QQ scores. This produced a main effect of group ( $F(2,32) = 3.684, p = .036, \eta^2_p = .187$ ), with significant pairwise contrasts between HFA participants and the other two groups (HFA < AS,  $p = .015$ ; HFA < TD,  $p = .048$ )<sup>54</sup>. In this analysis, AS participants were actually the best performers (QQ<sub>M(SD)</sub> = 0.328 (0.040)), followed by TD participants (QQ<sub>M(SD)</sub> = 0.320 (0.046)) and HFA participants (QQ<sub>M(SD)</sub> = 0.282 (0.050)). Thus, the subset analysis largely supported the initial analysis of the whole group, especially for differences in AS and HFA performance<sup>55</sup>.

### *Question Discrimination*

A three-way univariate ANCOVA was used to compare question discrimination scores in AS, HFA and TD participants. No significant main effect of group or any specific covariate effects were observed (Group:  $F(2,39) = 0.886, p = .420, \eta^2_p = .043$ , n.s.; all other comparisons also n.s.), suggesting comparable question discrimination skills in the three groups.

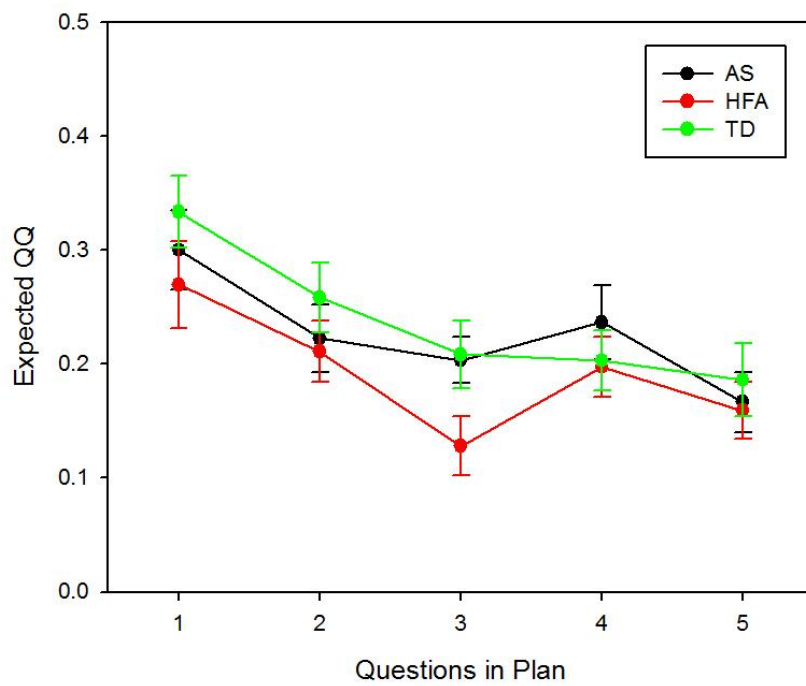
### *Plan Construction*

Separate Friedman's ANOVAs were used within each group to assess the structure of participant plans. Each group showed evidence of moving from high QQ to low QQ questions across their five questions, with effects that were either significant or approaching significance (AS:  $X_2(4) = 9.333, N = 15, p = .053$ ; HFA  $X_2(4) = 7.509, N = 15, p = .013$ ; TD:  $X_2(4) = 12.222, N = 15, p = .016$ ). As figure 26 shows, the plans constructed by participants were largely similar across groups. The questions selected for question 3 were on average slightly lower for HFA participants than AS participants (Mann-Whitney  $U =$

---

<sup>54</sup> The actual contribution of Age in this model was non-significant ( $F(1,32) = 0.594, p = .447, \eta^2_p = .018$ , n.s.), but its inclusion seemed to make a difference to the Group effect observed (presumably because it removed any age-related noise in the data). When age was not included, the main group effect was less marked, but still significant ( $F(2,33) = 3.435, p = .044, \eta^2_p = .172$ ).

<sup>55</sup> Again, very similar results were observed when mental age was taken into account. No differences were evident between the subgroups for VMA (AS-12<sub>M(SD)</sub> = 16.17 (5.74); HFA-12<sub>M(SD)</sub> = 16.46 (5.19); TD-12<sub>M(SD)</sub> = 18.34 (6.17); all  $p > .300$ ). For NVMA, AS ( $M(SD) = 13.61 (2.97)$ ) and HFA ( $M(SD) = 13.67 (2.70)$ ) participants scored significantly higher than TD ( $M(SD) = 10.97 (2.88)$ ) participants (AS > TD  $p = .038$ ; HFA > TD  $p = .027$ ). When NVMA was included as a covariate in the analysis of mean QQ scores, a significant group effect was observed ( $F(2,32) = 4.242, p = .023, \eta^2_p = .210$ ) alongside pairwise comparisons that were approaching significance for AS > HFA ( $p = .05$ ) and TD > HFA ( $p = .059$ ).



**Figure 26. Mean efficiency of plans in AS, HFA and TD participants.** QQ = Question Quality, AS =Asperger Syndrome, HFA = High-Functioning Autism, TD = Typically-Developing.

95.00,  $N = 30$ ,  $p = .004$ ) and TD participants ( $U = 61.50$ ,  $N = 30$ ,  $p = .032$ ), but it should be noted that this difference would not survive a correction for multiple comparisons.

Three-way univariate ANCOVAs indicated no significant differences for mean question quality (Group:  $F(2,39) = 0.993$ ,  $p = .380$ ,  $\eta^2_p = .048$ , n.s.) or plan gradient (Group:  $F(2,39) = 0.126$ ,  $p = .882$ ,  $\eta^2_p = .006$ ). VIQ was observed to significantly contribute to mean QQ of plans (VIQ:  $F(1,39) = 4.550$ ,  $p = .039$ ,  $\eta^2_p = .104$ ; all other effects *n.s.*). For plan gradient, effects were observable for Age ( $F(1,39) = 2.872$ ,  $p = .098$ ,  $\eta^2_p = .069$ ), VIQ ( $F(1,39) = 4.540$ ,  $p = .039$ ,  $\eta^2_p = .104$ ) and NVIQ ( $F(1,39) = 4.486$ ,  $p = .041$ ,  $\eta^2_p = .103$ ), suggesting that older and more able participants tended to narrow their plans to a greater degree.

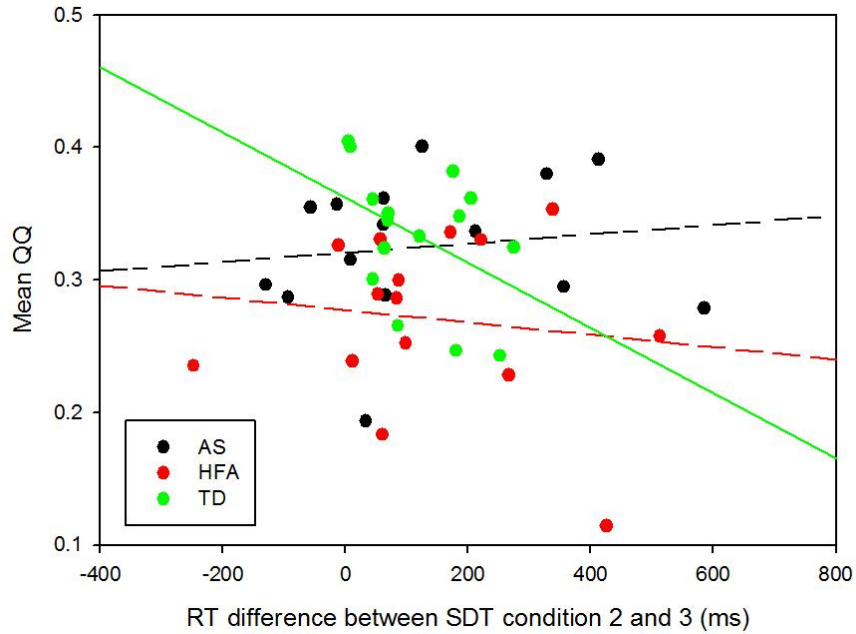
### *Semantic Decision Task*

Accuracy rates were high for participants in each group across all three conditions of the semantic decision task. Kruskal-Wallis tests indicated no significant differences between the groups for condition 1 (word identification:  $X^2(2) = 0.266$ ,  $N = 45$ ,  $p = .875$ ) and condition 3 (basic>super:  $X^2(2) = 4.295$ ,  $N = 45$ ,  $p = .117$ ), but a significant contrast for condition 2 (super>basic:  $X^2(2) = 8.462$ ,  $N = 45$ ,  $p = .012$ ). Pairwise Mann-Whitney U-tests indicated that this difference lay primarily in the performance of AS participants, who scored slightly lower for accuracy than HFA ( $U = 62.50$ ,  $N = 30$ ,  $p = .030$ ) and TD ( $U = 49.50$ ,  $N = 30$ ,  $p = .007$ ) participants.

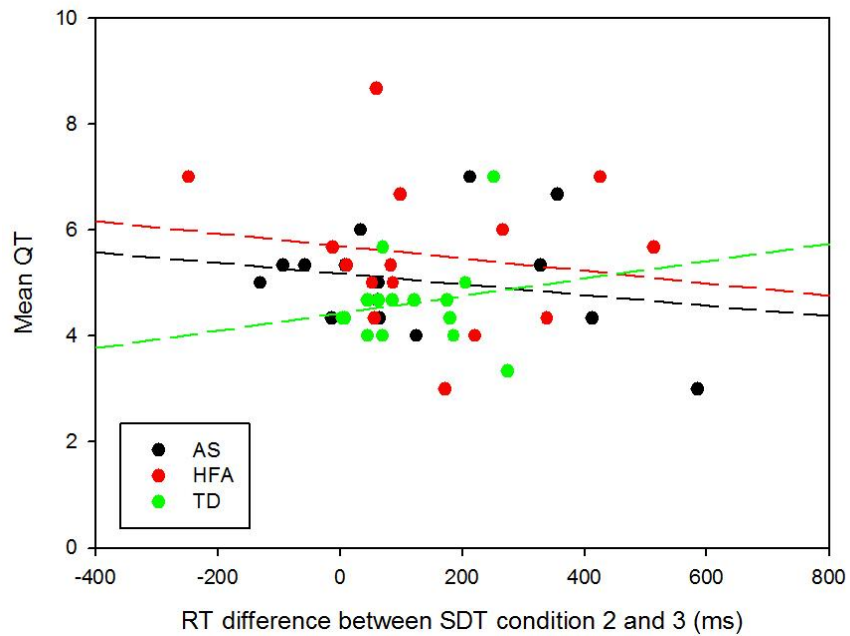
Mean RTs for participants on the SDT were log10 transformed to enable a repeated measures of analysis of covariance to be run on the data<sup>56</sup>. A linear effect approaching significance was observed ( $F(1,39) = 3.476$ ,  $p = .070$ ,  $\eta^2_p = .082$ ), reflecting the general increase in RTs with each condition (see table 10). However, no significant group effect ( $F(2,39) = 0.455$ ,  $p = .637$ ,  $\eta^2_p = .023$ , n.s.) or Group\*Condition interaction effect ( $F(2,39) = 1.024$ ,  $p = .369$ ,  $\eta^2_p = .050$ , n.s.) was observed, indicating that RTs were comparable across the groups, for each condition. Faster performance overall was significantly related to age ( $F(1,39) = 8.967$ ,  $p = .005$ ,  $\eta^2_p = .187$ ) and VIQ ( $F(1,39) = 5.881$ ,  $p = .020$ ,  $\eta^2_p = .131$ ).

---

<sup>56</sup> Such a procedure could have also been applied to the data from the Plan Construction task. However, log transforming that data did not significantly normalise the data distributions, making non-parametric statistics the most appropriate.



**Figure 27. The relationship between question quality during Twenty Questions and reaction differences in semantic decisions for AS< HFA and TD groups.** QQ = Question Quality, SDT = Semantic Decision Task, AS = Asperger Syndrome, HFA = High-Functioning Autism, TD = Typically-Developing. Significant relationships ( $p < .05$ ) are shown with solid lines, non-significant with dashed lines.



**Figure 28. The relationship between questions used per trial during Twenty Questions and reaction differences in semantic decisions for AS< HFA and TD groups.** QT = Question used per Trial, SDT = Semantic Decision Task, AS = Asperger Syndrome, HFA = High-Functioning Autism, TD = Typically-Developing.

For comparison with the data from experiment 3, the difference between mean RTs for conditions 2 and 3 was also calculated. The mean difference was highest in the HFA group (141.80ms), followed by the AS group (130.40ms) and the TD group (119.06ms), but a univariate ANCOVA found no significant effect of group on this outcome ( $F(2,39) = 0.036$ ,  $p = .965$ ,  $\eta^2_p = .002$ , n.s.)<sup>57</sup>.

As in experiment 3, backwards regression analyses were used to assess for relations between SDT and TQT performance. For mean question quality, the difference in RTs between semantic conditions was a significant predictor in the model returned for typically-developing participants ( $R^2 = .482$ ,  $F(2,12) = 5.584$ , stan. beta =  $-.773$ ,  $p = .008$ ) but not AS (stan. beta =  $.147$ ,  $p = .625$ ) or HFA participants (stan. beta =  $.002$ ,  $p = .989$ ; see figure 27).

For the number of questions used on the TQT, none of the models returned included semantic asymmetry as significant predictor of performance for any group (all stan. beta  $< .450$ , all p-values  $> .098$ ). Thus, as figure 28 shows, the specific relation observed between SDT and TQT performance in deaf children was not replicated in any of the groups tested.

Based on the lack of RT differences across the groups, post-hoc correlational analyses were conducted to explore other relations between SDT and TQT performance. Across all participants combined, mean question quality was observed to negatively correlate with mean response time on SDT condition 2 ( $R = -.302$ ,  $N = 45$ ,  $p = .044$ ), and, to a lesser extent, condition 3 ( $R = -.258$ ,  $N = 45$ ,  $p = .086$ ; Spearman's Rho used). That is, slower responses for semantic decisions were associated with lower question quality, when AS, HFA and TD groups were analysed together.

### *Verbal Fluency*

Due to skew in the performance of the AS group, Kruskal-Wallis tests were used to compare fluency performance across the groups. A difference approaching significance was observed for Letter Fluency ( $X_2(2) = 5.175$ ,  $N = 45$ ,  $p = .075$ ), alongside a significant difference for Semantic Fluency ( $X_2(2) = 6.33$ ,  $N = 45$ ,  $p = .042$ ). Pairwise Mann-Whitney tests indicated that TD participants tended to produce more words on the letter fluency task than AS

---

<sup>57</sup> This contrasts with the mean difference of 175.31ms recorded by deaf participants in the previous experiment. Direct comparisons between RT data from experiments 3 and 4 were not conducted due to considerable differences in cognitive ability between the deaf participants on the one hand and AS, HFA and TD participants on the other hand. Because the groups were barely overlapping in verbal and non-verbal IQ, ANCOVA would not have been appropriate and matching of subgroups was not possible.

participants ( $U = 62.00, N = 30, p = .035$ ) and, to a lesser degree, HFA participants ( $U = 72.00, N = 30, p = .093$ ). Similar differences were evident for semantic fluency (TD>AS:  $U = 54.50, N = 30, p = .015$ ; TD>HFA:  $U = 67.50, N = 30, p = .062$ ).

To examine what influence fluency differences may have had on TQT performance, a log10 transform was applied to fluency scores and the resulting indices were then included as covariates in an ANCOVA for mean QQ. As before, this produced a significant main effect of group ( $F(2,40) = 4.546, p = .017, \eta^2_p = .185$ ), but no significant covariate effects were observed for the log transforms of either letter fluency ( $F(1,40) = 1.542, p = .222, \eta^2_p = .037$ , n.s.) or semantic fluency ( $F(1,40) = 0.104, p = .748, \eta^2_p = .003$ ). This suggested that any differences in verbal fluency that existed between the groups did not explain the observed difference in question quality on the TQT.

#### *Reanalysis according to language ratings*

Table 11 shows the mean scores for each group on the language questionnaire. As stated above, this reflects the results from 15 AS and 15 TD but only 14 HFA participants, due to missing data from one family. Individual data points for 1 AS participant (Q6) and 3 TD participants (one Q1, three Q2) were also not available, either because of parents not being able to remember specific ages, or simply because those questions were left blank. Kruskal-Wallis omnibus tests followed by pairwise Mann-Whitney U tests indicated significantly later development of language skills for HFA participants as compared to the other two groups (all  $p < .05$ , uncorrected). For ages 3, 5 and 7 language ratings also tended to be significantly worse than either AS (all  $p < .099$ ) or TD participants (all  $p < .05$ ). For current language rating, no significant difference was evident between HFA and AS participants ( $U = 79.00, N = 28, p = .384$ , n.s.), but a difference approaching significance was observed between HFA and TD participants (HFA<TD:  $U = 62.50, N = 29, p = .058$ ).

In the AS group, 13 of the 15 participants would have qualified for a DSM-IV diagnosis of Asperger Syndrome, based on development of first words before 24 months and first phrases by 36 months. In contrast, two participants would have qualified for an autism diagnosis, based on the presence of delays in either of the above criteria. In the HFA group, 10 participants would have qualified for a diagnosis of autism. Parents for the four other HFA participants reported the reaching of these language milestones prior to 36 months, although



in at least one case this was followed by a severe regression in language skills that was only corrected many years after. Thus, at least 11 of the HFA participants fitted a diagnosis of autism based on their language history. None of the TD participants were reported to have had delays in their language skills past the 24/36 month cut-offs. In cases where parents could not remember specific ages of language milestones, or responses were left blank, there were nevertheless no concerns recalled about early language development.

Based on the parent reports, performance on the TQT was reanalysed for ASD participants with (ASD+L,  $n = 13$ ) and without (ASD-L,  $n = 16$ ) language delays. The resulting age and IQ data for these groups is presented in table 12, along with the existing data for TD participants (for ease of comparison). The only significant difference evident between the groups was for VIQ, on which TD participants scored higher than ASD participants with language delay (ASD+L < TD:  $t(26) = 2.281, p = .031$ , all other pairwise contrasts  $p > .150$ ).

Mean QQ scores of the three groups were compared using an ANCOVA that included VIQ as a covariate. A significant effect of VIQ was observed ( $F(1,40) = 8.173, p = .007, \eta^2_p = .170$ ) along with a main effect of Group that approached significance ( $F(1,40) = 3.205, p = .051, \eta^2_p = .138$ ). As in the prior analysis, this reflected a general advantage for TD ( $M_{(SD)} = 0.33 (0.05)$ ) and ASD-L ( $M_{(SD)} = 0.32 (0.05)$ ) participants over ASD+L participants ( $M_{(SD)} = 0.27 (0.07)$ ). Pairwise comparisons between the groups indicated that both of these contrasts were significant (ASD+L < TD:  $p = .021$ ; ASD+L < ASD-L:  $p = .047$ ). Thus, the reanalysis of ASD participants based on language history appeared to support the initial analysis based on current diagnosis: those with language delay asked less efficient questions than those without language delay.

Following this, post-hoc regression analyses were used to assess which early language indicators were the best predictors of TQT performance. The answers for questions 1 to 6 from the questionnaire were entered into a backwards method regression analysis, with mean QQ as the dependent variable.

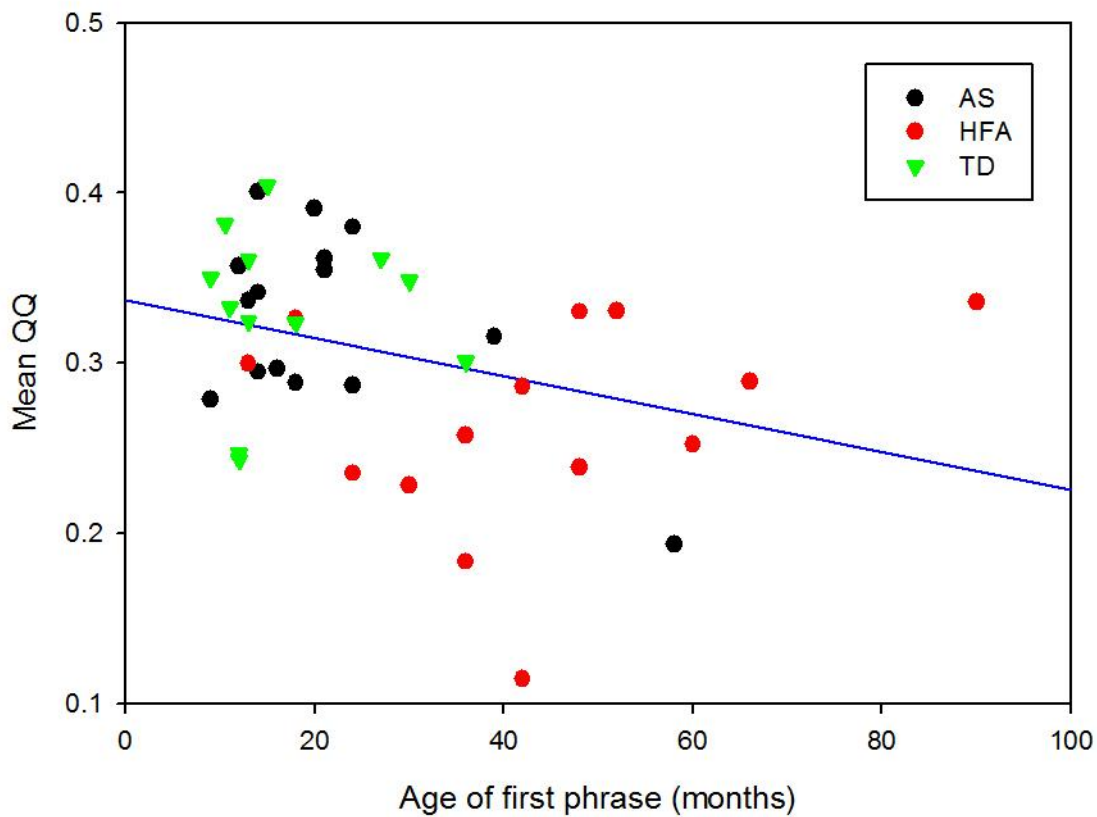
Log10 transforms were used for the analysis due to non-normal data in the questionnaire responses. In the resulting model, question 2 (age of first phrase) was the strongest predictor ( $R^2 = .272, F(3,37) = 4.615, \text{stan. beta} = -.468, p = .006$ ). As figure 29 shows, earlier ages of first phrase were associated with greater question quality scores on the TQT. Also retained in the model were question 4, language rating at age 5 ( $\text{stan. beta} = -.315, p = .095$ ), and

**Table 11. Language milestones and ratings by parents for AS, HFA and TD participants.**

	AS		HFA ( <i>n</i> =14)		TD		<i>Sig.</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Q1. Age of first word (m)	12.76	6.23	23.82	19.26	10.00	3.37	HFA>AS,TD*
Q2. Age of first phrase (m)	21.13	12.48	43.21	20.21	17.21	8.84	HFA>AS,TD*
Q3. Age 3 language rating	3.20	1.51	2.07	1.54	3.33	0.82	HFA<TD*, (HFA<AS)
Q4. Age 5 language rating	3.47	1.11	2.36	1.45	3.53	0.74	HFA<AS,TD*
Q5. Age 7 language rating	3.53	1.17	2.50	1.40	3.73	0.80	HFA<AS,TD*
Q6. Current language rating	3.50	1.18	3.04	1.37	4.00	0.86	(HFA<TD)

\*  $p < .05$ . Differences in brackets indicate contrasts approaching significance ( $p < .1$ )

AS = Asperger Syndrome, HFA = High-Functioning Autism, TD = Typically Developing.



**Figure 29. The relationship between age of first phrase and mean question quality during problem-solving.** QQ = Question Quality, AS = Asperger Syndrome, HFA = High-Functioning Autism, TD = Typically Developing,

**Table 12. Age and IQ scores for ASD participants with language delay, without language delay and typically-developing participants.**

	ASD+L ( <i>n</i> = 13)		ASD-L ( <i>n</i> = 16)		TD ( <i>n</i> = 15)		<i>Diff</i>
	<i>M</i>	( <i>SD</i> )	<i>M</i>	( <i>SD</i> )	<i>M</i>	( <i>SD</i> )	
Age	14.02	2.87	13.61	2.18	14.05	2.72	n.s.
VIQ	102.31	14.04	111.50	19.28	113.80	12.62	ASD+L < TD*
NVIQ	98.77	12.75	102.50	11.29	105.33	16.72	n.s.

\*significant at  $p < .05$ .

ASD+L = Autism Spectrum Disorder with Language Delay, ASD-L = Autism Spectrum Disorder without Language Delay, TD = Typically-Developing.

question 6, current language rating (stan. beta =  $-.352$ ,  $p = .043$ )<sup>58</sup>. None of the other language indicators were retained in the model (all stan. betas  $<.200$ , all  $p > .500$ ).

Correlational analyses were also run between mean QQ and AQ scores, for exploratory purposes. Across all participants combined, total AQ score was observed to correlate moderately with mean QQ at a level approaching significance (Spearman's Rho  $r = -.264$ ,  $N = 44$ ,  $p = .083$ )<sup>59</sup>, with greater task performance being associated with lower AQ scores. Stronger relations were observed for the Attention to Detail ( $r = -.313$ ,  $N = 44$ ,  $p = .039$ ), Communication ( $r = -.273$ ,  $N = 44$ ,  $p = .072$ ) and Imagination ( $r = -.346$ ,  $N = 44$ ,  $p = .021$ ), subscales of the AQ, although none would have been strong enough to survive an appropriate Bonferroni correction. In all cases, reports of less autistic traits were associated with greater efficiency on the TQT.

Although no significant AQ-TQT correlations were observed when groups were analysed individually (possibly due to the loss of statistical power), the directions and strengths of correlations were still informative for comparing the three groups. The correlations with mean QQ for total AQ and Communication appeared to be driven by HFA participants (AQ  $r = -.287$ , AQ-Comms  $r = -.316$ ) but not AS (AQ  $r = .050$ , AQ-Comms  $r = -.032$ ) or TD (AQ  $r = -.134$ , AQ-Comms  $r = -.039$ ) participants. In contrast, the correlations for Imagination were strongest in AS (AQ-Imag  $r = -.203$ ) and TD (AQ-Imag  $r = -.265$ ) participants, but not HFA participants (AQ-Imag  $r = -.062$ ). Attention to Detail was related to performance for TD participants (AQ-Att  $r = -.336$ ), but neither AS (AQ-Att  $r = -.032$ ) or HFA participants (AQ-Att  $r = -.062$ ). Thus, while TQT performance appeared to be related to autistic traits, these differed for HFA, AS and TD participants.

## Discussion

The main finding of the fourth experiment was that participants with high-functioning autism and participants with Asperger Syndrome significantly differ in their performance on the Twenty Questions Task. Specifically, participants with Asperger Syndrome were more efficient in their questioning than HFA participants and performed at a similar level to age- and IQ-matched typically developing children. The difference in questioning quality between the two ASD groups was significant even when general IQ abilities were taken into account either statistically via analysis of covariance, or by matching subsets of participants. The

---

<sup>58</sup> These predictors are also retained with the addition of VIQ into the model.

<sup>59</sup> All correlations two-tailed & uncorrected for multiple comparisons.

difference was also significant when controlling for verbal fluency performance, and when the AS and HFA groups were redefined according to parental reports of language history. As predicted, participants with more typical language development were better at asking efficient questions on the TQT than those who had shown some degree of language delay in their early years – a point supported by the observed relationship between age of first phrase (Q2) and mean question quality.

Alongside question quality, HFA participants took more questions on average to reach the target than TD participants, although the difference with AS participants on this outcome was less marked. On question types, HFA participants did not differ from the other two groups: all three groups used grouping for approximately two thirds of their questions. In addition, HFA, AS and TD participants did not differ in their response to changing task conditions, with all three groups showing a consistent level of performance for trials where elimination was and was not prohibited.

This suggests that the HFA group included here was in many respects a high-performing one, that certainly understood what kind of strategy was required by the task and, for the most part, deployed similar questions to AS and TD participants. It is also important to reiterate here as well that the mean IQ and range of this HFA group was higher than in experiment 1. Despite this, their efficiency in problem-solving continued to fall behind that seen for the participants in the other two groups; a finding that is consistent with the results from each of the experiments reported in previous chapters, and in accordance with the performance of the predominately HFA group of participants in Alderson-Day & McGonigle-Chalmers (2011).

Tangential to the main task outcome, it is surprising that none of the ASD participants appeared to be significantly affected by the switch from allowing elimination of items to the non-elimination trial. As reported in Alderson-Day & McGonigle-Chalmers (2011) and (to a lesser extent) in the first experiment here, this has before prompted ASD participants to require more questions to find the target, possibly because of problems with maintaining which questions have already been asked in working memory. If both HFA and AS participants are not expected to generally differ in their executive functioning skills (Macintosh & Dissanayake, 2004), then both might be expected to show some form of drop-off in performance in this condition. Alternatively, if AS participants possessed greater verbal working memory skills, one might expect only HFA participants to struggle with this stage of the task. However, neither of these possibilities occurred, as both groups were apparently unaffected by the change in task demands.

One way in which this deployment of the TQT differed from those previous experiments was in the number of trials used. For example, all participants completed at least two non-elimination trials (without written aids) in experiment 1, whereas they only attempted one trial in this way in the present experiment. As the number of questions used can be affected by lucky guesses, it may be that a single trial is insufficient to pick out changes in performance in this way.

Another possibility, given the overall level of performance, is that both HFA and AS participants possessed a strategy or schema that could support their search even when they could not eliminate items from the set. As outlined in previous chapters, deployment of questions that narrow down search in a systematic way can provide a hierarchical structure to guide ones' questions, as in the sequence LIVING>ANIMAL>PET. If such a sequence is followed, then participants only really need to remember the answer to the last question that they asked at any point in the search. If a logical sequence of questions is not followed – that is, if there is no systematic and logical link between questions – then recalling what has and has not already been asked could be much harder, as it requires recall of a number of unrelated questions (e.g. ANIMAL? FOUND IN KITCHEN? HAS WHEELS?).

Supporting this interpretation are the data from the question discrimination and plan construction tasks. As in the previous experiments, no group differences were observed for question discrimination, suggesting comparable levels of understanding for what constitutes an effective question in isolation. But in addition, all three groups showed evidence of narrowing in their planning, suggesting an appreciation that sequences of questions need to systematically eliminate possibilities from the set, moving from the general to the particular. Thus, the lack of any significant drops in TQT performance for either of the ASD groups may reflect a better understanding and ability to produce sequences of effective questions in these particular participants, in contrast to previous experiments.

Also on the topic of the plan construction task, the lack of any significant efficiency differences in planning for HFA, AS and TD participants is a slight concern that needs to be acknowledged. In contrast to the ASD participants in experiment 1 (whether analysed as a whole group or divided into high and low levels of VIQ), the HFA group in this study did not construct plans that were any less efficient in expected question quality than AS or TD participants. This being a slightly older HFA group (14.7 vs 13:7), it may be that planning skills improve with age for ASD participants, allowing for a greater or more in-depth consideration of possible questions to use in a plan. Such an interpretation would also be consistent with the lack of any major differences in planning efficiency seen for deaf adults

in experiment 2. Without further data, though, it is hard to assess whether this does reflect an age-related change in planning skills; it may just indicate a lack of reliability for the task as a measure of planning.

Finally, although some group differences were observed in accuracy on the semantic decision task, no significant differences were observed for reaction times on either of the key semantic decision conditions, or discrepancies between the two. Longer reaction times for basic-to-superordinate decisions (i.e. condition 3) compared to superordinate-to-basic decisions (condition 2) were evident in each of the three groups equally, in a similar manner to that seen for deaf participants. This suggests that an asymmetry in responses on this task is not atypical, despite the high frequency of associations included in the task (c.f. Loftus, 1973). Therefore, although semantic associations may have been related to problem-solving performance in deaf participants, they could not be said to exhibit atypical semantic organisation.

Importantly, the difference in reaction times for basic to superordinate and superordinate to basic decisions did not consistently predict TQT performance for either HFA or AS participants, in contrast to the relation observed for deaf participants. TD participants did show a negative relation between question quality and reaction time asymmetry, indicating that participants with less of a discrepancy in semantic decision performance also asked more efficient questions. But this, again, did not fully cohere with the data seen in the deaf group, who only showed TQT-SDT relations for the number of questions used, not question quality.

As in the plan construction task, it may be that the SDT lacks reliability or sensitivity to measure more subtle differences in semantic skills, or it could be that these data reflect genuinely intact levels of semantic processing in ASD participants. Either way, there are no data evident here from the SDT to suggest that semantic organisation underpins problem-solving difficulties in ASD individuals. HFA participants were less efficient than AS participants in their questioning, but apparently not because of any apparent asymmetries in their semantic relations. Indeed, the only group difference seen in semantic skills in this instance was for AS participants, not HFA participants, to be less accurate in their judgements about superordinate-to-basic relations (condition 2).

The general discussion in Chapter 8 will discuss the implications of these results, with specific reference to the strong and weak arguments outlined previously. Before doing so, it

is important to acknowledge a few caveats that limit the reliability and scope of these findings.

#### *Caveats and limitations*

Bearing in mind some of the issues discussed in chapter 6, the first limitation of these data concerns the contrast between HFA and AS participants. As in previous studies, this project was reliant on the adequacy of original diagnoses. In order to check that these do reflect language histories, parents were asked to report participants' language milestones, which may have occurred over 10 years ago. Thus, the main contrasts are dependent not only on the accuracy of previous diagnoses, but on the accuracy of parents' retrospective reports. The data provided in this way appeared to largely fit the existing diagnoses, with one or two discrepancies (13/15 AS, 10/14 HFA), something which is probably representative of the clinical population as a whole. Furthermore, the reanalysis of TQT performance according to language history largely supported the original result. Nevertheless, it still needs to be acknowledged that the method of discriminating groups deployed here is possibly less reliable or powerful than, say a full re-diagnosis of all participants according to existing diagnostic criteria, or using medical records to establish early language milestones. For practical reasons such methods were not possible in this instance, but they could be used in future work.

On a related point, while parents were asked to categorically rate previous and current language skills, this is at best a crude indicator of language proficiency. It may have been preferable to utilise a standardised measure of language skills, such as the CELF (Semel, et al., 1995): if earlier scores were available, these could be used to assess for later predictive relations with problem-solving performance, while current scores could be used to match participant groups, especially for AS and HFA participants. To allow for other experimental tasks to be used in the time allowed, it was not possible to deploy an in-depth language battery in this instance, but a larger study with an existing database of ASD participants should be able to achieve this. It is also worth noting that, while a measure like the CELF was not deployed here, the tasks that were used assessed lexical knowledge (WASI Vocabulary), categorisation skills (WASI Similarities, Semantic Decision Task) and word fluency (ACE-R Letter & Semantic Fluency). Thus, a number of language skills were covered and controlled for, even if a standardised battery was absent.



In summary, the fourth and final experiment appeared to demonstrate a difference in problem-solving performance within the autistic spectrum. HFA participants, being those with generally delayed and less typical language development, asked less efficient questions on the TQT than typically-developing participants. AS participants, without language delay and with more typical language histories, showed no such difficulty. Other group differences were generally less forthcoming, and the prediction that TQT performance in ASD individuals would be moderated by semantic decision skills was not supported. These findings, and those of the previous experiments, are discussed in the final chapter.

## **Chapter 8: General Discussion**

The main aim of this research was to explain why individuals with autism spectrum disorders have difficulty with certain types of verbal problem-solving. Using the Twenty Questions Task, this specifically was an investigation into why people with ASD who appear to have sufficient vocabulary, language knowledge, and reasoning skill, are still less efficient than typically-developing counterparts in the strategies they deploy; namely, in asking effective questions.

As outlined in the very first chapter, there is a range of theories and explanations that could be applied to this behaviour. Most pertinent to the question is Minshew and colleagues' (2002) suggestion that this reflects a "concept formation deficit", or the ability to "spontaneously form schemata or paradigms that organize information" (Minshew et al., 2002; p333). Here, inefficiency in problem-solving is a problem of coming up with a strategy unprompted, in a way that draws out the relevant information from a set of stimuli (e.g. the individual items in a TQT set) and groups them together in a meaningful and effective way. The best evidence for this from the TQT is the increased tendency of ASD participants to guess at individual items, rather than ask grouping questions (Minshew, Meyer, et al., 2002; Minshew, et al., 1994).

A related, but more general idea, is that this reflects a bias or tendency in the cognitive processing of autistic individuals, towards the local (in this case, individual exemplars) and away from the global (general semantic categories). This may be interpreted as an enhancement of local processing skills (Mottron, et al., 2006) or a bias away from global, "centrally coherent" processing (Happé & Frith, 2006). The evidence for this kind of account is largely drawn from studies of perception, rather than problem-solving or reasoning, but it fits well with the tendency of ASD participants to be restricted in the scope and generality of questions that they ask (Alderson-Day & McGonigle-Chalmers, 2011).

While both of these descriptions have their merits, the experiments included here constitute an attempt to go further in explaining what drives problem-solving performance in autism; where apparent cognitive deficits or biases come from in the development of this particular cognitive faculty. As the Twenty Questions Task is a complex measure, relying on multiple skills, it requires a detailed task analysis and close examination of a range of cognitive processes. The research conducted in Alderson-Day & McGonigle-Chalmers (2011) represented a first attempt at doing so. The studies reported here fulfil the project by testing a

range of candidate explanations, coming from both within autism research (executive functioning) and without it (deaf research).

### **The experiments**

Within recent neuropsychological research, the TQT is treated as a measure of executive functioning skills (Baldo, et al., 2004; Marshall, et al., 2003b). A number of researchers and theorists have proposed specific problems with executive functioning in autism (C. Hughes, et al., 1994; Ozonoff, Pennington, et al., 1991; Russell, 1997). Accordingly, the first experiment tested the hypothesis that children with ASD may have difficulty with the Twenty Questions Task because of problems with executive skills; namely, problems with planning questions, and problems with selectively attending to relevant information in the task array. While there was some evidence to suggest that ASD performance was moderated by being able to remove redundant items during search, this appeared to be corrected with a written reminder of questions that had been asked. This was taken to suggest that ASD participants could selectively attend to an unchanging array, and that their apparent difficulties during this condition were actually due to problems with remembering questions.

In contrast, participants with ASD appeared to differ from control participants in their planning of questions. ASD participants were good at recognising effective questions in advance, but when asked to construct a plan of questions they were less likely, as a group, to construct plans that narrowed down possibilities systematically. In addition, the questions that made up their plans tended to refer to fewer items than those of controls. When the group was divided into those with higher and lower verbal ability (VIQ), narrowing skills were evident for more able participants. Problems with selecting efficient questions, however, were apparent across the whole ASD group.

Thus, whatever was causing inefficient questioning by ASD participants on the TQT also seemed to be present at the planning stage, when they were thinking about their questions in advance. In this case participants were selecting the questions from a set of possibilities, and not having to produce those questions themselves – and yet they still selected questions that were less efficient than controls. As argued in chapter 2, this does not easily fit with Minshew et al.'s (2002) notion of a concept *formation* deficit, where the problem is one of spontaneously producing an organising concept or category. Here, participants had all the relevant concepts and categories laid out for them, in question form, and they still chose options that lacked effectiveness – despite not having to generate the strategies themselves.

Furthermore, this was consistent with the actual performance of ASD participants on the TQT: ASD participants were asking a number of grouping questions – that is, generating their own concept-based strategies – but the questions they asked were less efficient at subdividing the set. Rather than this representing a deficit in “spontaneously forming schemata” that could organise the set, this resembled a problem with *which* schemata were being deployed.

This distinction motivated comparisons with other participants groups in which access or availability of categories and concepts are thought to be atypical. In particular, the example of deaf individuals was used, as they had previously shown similar performance on the TQT (as seen in Marschark & Everhart, 1999), and often show difficulties in semantic categorisation that resemble those seen in ASD. Within deaf research, this profile has been interpreted as a consequence of their atypical language experiences and delayed language development, rather than a problem based in cognitive processes. As children with autism also show highly atypical language development, this prompted the idea that a similar explanation may apply in the case of autistic problem-solving. Such a proposal stands in contrast to what might be considered purely cognitive accounts, whether that is a concept formation deficit, a cognitive style, or specific types of executive dysfunction.

Experiments 2 and 3 sought to interrogate this possibility further by examining the deaf profile on the TQT in more detail. In experiment 2, the performance of deaf and hearing adults were compared on the TQT and a task that assessed basic and superordinate category use in everyday descriptions. The results appeared to broadly replicate previous reports of deaf adults asking less efficient questions than hearing adults, although this effect could not clearly be separated from differences in general cognitive abilities. In addition, while some differences may have been expected due to differences in category use in sign language, this was not observed on the picture descriptions task, and did not appear to relate to TQT performance.

Experiment 3 followed this by comparing the performance of deaf children on the TQT with the data from ASD children and hearing, typically-developing children collected in experiment 1. Like ASD children, the deaf participants in this study asked questions that eliminated fewer items than control children. In addition, some aspects of their performance (the number of questions they used on each trial) were predicted by a novel index of their semantic organisation: the difference between reaction times for superordinate and basic semantic judgements. Building on the theoretical proposals of Marschark et al. (2004), this

was the first demonstration that TQT performance is related to underlying semantic organisation in deaf individuals.

Finally, if the comparison between deaf and autistic problem-solving was a valid one, then ASD children with more typical language development should be expected to perform well on the TQT. The fourth experiment tested this by comparing the performance of children with Asperger Syndrome, children with autism, and typically-developing children. As predicted, children with AS were just as good at the task as control children, and showed none of the efficiency impairments seen in autistic participants – an effect evident when the groups were redefined according to parental reports of language history, and specifically related to the age that participants' first phrases appeared. However, TQT performance was not related in any clear way to semantic decision performance for either of the ASD groups. Thus, problem-solving performance appeared to be moderated by language history in ASD, but not necessarily because of underlying characteristics of semantic organisation.

### **Implications for the strong and weak arguments**

The comparison between deaf and autistic cognitive profiles in chapter 3 prompted the proposal of two hypotheses that could explain autistic problem-solving: a strong argument and a weak argument. In the strong argument, the deaf and autistic problem-solving profiles arise out of a common causal route: atypical semantic organisation, as a result of language delay. In the weak argument, the problem-solving profiles in autism and deafness are seen as similar in background, but not overlapping. Both are the product of atypical language development, but they do not necessarily take the same causal route, via atypical semantic development.

The strong argument was referred to in this way because it makes specific a route by which problem-solving impairments could arise in both deafness and autism. The barriers to language in both groups (auditory deprivation in the case of deafness, possible SLI or irregular sensory processes in autism) create an early environment in which there is reduced opportunity to interact linguistically with others. In this context, semantic categories are acquired in a piecemeal and delayed fashion, leading to a semantic lexicon with subtly atypical relations and organisational features (such as basic > superordinate asymmetries). In later years, when vocabulary and expressive language skills appear to “catch up” for both groups, semantic knowledge is intact, but categories may not be as accessible or as readily available as when they are required to support strategic, higher-order cognitive operations.

The weak argument makes no commitment to the specific route by which problem-solving difficulties arise, but it offers a broad context as a starting point: namely, atypical early language development. It is weak in the sense that it is less specific in its developmental story, and merely makes the point that autistic and deaf TQT profiles may be similar because of similar initial causes. In the scenario that the strong argument is rejected, the weak argument can still hold.

Ultimately, the data reported here cannot support the strong argument. The findings of experiment 3 were broadly in line with this argument, in appearing to support a link between semantic organisation and TQT performance in deaf children. But the lack of any comparable link between TQT performance and semantic skills in either ASD group means that the “semantic route” endorsed by the strong argument does not clearly hold in autism. In addition, the presence of asymmetry on the semantic decision task in *all* participants suggests that the differences in reaction time seen for deaf participants were not in fact atypical. Semantic organisation may be related to problem-solving performance for this group in particular, but there was no evidence to suggest that their semantic decisions were atypical in themselves (in contrast to the findings of Marschark et al., 2004).

From one perspective, this suggests that the strong argument may be wrong about both autism *and* deafness. From another point of view, it could be that deaf problem-solving performance is more closely dependent on semantic organisation in some way, but that organisation itself is not abnormal. Finally, it could be that the measures used here do not provide an adequate index of underlying semantic organisation, and that alternative tasks would highlight atypical semantic skills in both autism and deafness. In any case, there is certainly not enough positive evidence here to suggest that ASD and deaf problem-solving skills are dependent on the same underlying semantic characteristics. Therefore, the strong argument cannot be endorsed based on the present results.

In contrast, the data do still support the weak argument. The main evidence for this is the contrast between AS and autistic participants in experiment 4: those with more typical language development also asked more efficient questions. Given that the two groups of ASD participants did not differ in terms of autistic tendencies, verbal fluency or general ability levels, their difference in questioning efficiency seems to strongly support a role for language history in verbal problem-solving. Furthermore, the predictive relation of question 2, age of first phrase, supports the general picture that the weak argument outlines: verbal problem-solving difficulties in autism are in some way a reflection of, or related to early

language development. In this respect, the problem-solving skills of individuals with autism and individuals who are deaf are similar, but not necessarily overlapping.

The weak argument is committed to the idea that delays in early language development have long-term consequences for cognitive skills, and that even very able autistic individuals may show residual deficits or hangovers from early development. Such an explanation is also committed, to certain extent, to the idea of a “critical period” in language development, after which language-related abilities may develop less fluently or in an atypical way (Johnson & Newport, 1989; Lenneberg, 1967). The idea of early communicative skills in autism having a general predictive value for language skills in later childhood is a fairly common one (see, for example, Anderson et al., 2007; Anderson, Oti, Lord, & Welch, 2009; Siller & Sigman, 2008). As the review of autism and AS studies in chapter 6 outlined, long term effects on cognition can be harder to identify, and some would deny that language delay makes any difference at all to later cognitive outcomes in ASD (Mayes & Calhoun, 2001). It is important to remember, though, that the weak argument is making a claim about the etiology of verbal problem-solving in particular, and does not deny the existence of a large degree of overlap in skills between those with and without language delays. The significance of verbal problem-solving is that it involves the use of linguistic strategies to support a complex and ongoing set of cognitive processes – and it is this, in particular, that is proposed to be adversely affected by early language development.

This means that, in the end, the data here may only provide an account of a very specific set of cognitive skills. Given the eventual contrast between autism and AS, it seems unlikely that this particular kind of problem-solving performance has something general to say about typically autistic behaviours, or the autistic cognitive style, as these do not tend to greatly differ between autistic and AS individuals. At most, what the weak argument might suggest is that the autistic cognitive profile, and any apparent style that is expressed, includes separable components, each with their own developmental causes.

In addition, if the weak argument is retained there is still an explanatory gap. If semantic organisation is not the route by which verbal problem-solving difficulties come about, then there must be other intermediary mechanisms that link the early language development of ASD children and their later cognitive skills. The following section speculates upon some possible explanations, drawing on findings from neuroimaging and inner speech.

### *1. Integration of executive and language networks*

One possible explanation for later discrepancies in AS and HFA problem-solving is that they reflect underlying differences in the structure and function of neural networks, as a result of language delay. Specifically, it could be that language centres in the autistic brain, by developing late or with atypical localisation, end up lacking in integration with other cognitive processes across the cortex.

For example, on an executive task where a language-based strategy is required, people with autism may activate frontal areas associated with executive functioning (such as the dorsolateral prefrontal cortex) and temporal areas associated with language processing (such as the superior temporal gyrus), but communication between those areas may be lacking in some way. This would result in relatively intact executive or semantic skills when assessed independently, but possibly inefficient performance where information from one area (e.g. semantic categories) needs to be used as a cognitive strategy to support another network. (For an example of this kind of process from free recall research in typical individuals, see Dickerson et al., 2007).

A large number of studies have documented abnormalities in brain structure and function in autistic individuals that are thought to be related to language development, the full range of which cannot be adequately discussed here (for a review, see Groen, et al., 2008). One common idea of note is that the lateralisation of language to left hemisphere structures may be delayed or irregular in autism (see Eyler, Pierce, & Courchesne, 2012, for a recent example of this), leading to recruitment of right hemisphere areas, or reduced specialisation of typical language centres (Bigler et al., 2007). For the purposes of this argument, though, it is important to focus on studies that have directly compared individuals with AS and autism

To date, almost all the neuroimaging studies that have compared HFA and Asperger Syndrome have done so in terms of brain structure, via measurement of grey and white matter volumes (Kwon, Ow, Pedatella, Lotspeich, & Reiss, 2004; Lotspeich et al., 2004; McAlonan et al., 2009; McAlonan et al., 2008; Toal et al., 2010). For example, Kwon et al. (2004) contrasted structural MRI volumes in nine HFA adolescents, 11 AS adolescents and 13 controls. Alongside some common abnormalities for both ASD groups (decreased grey matter in right inferior frontal gyrus, entorhinal cortex and fusiform gyrus), AS participants showed reduced grey matter density in the cingulate gyrus compared to HFA and control participants.



In a study by the same research group, Lotspeich et al. (2004) compared structural MRI scans for 13 low functioning autistic participants (LFA), 18 HFA, 21 AS and 21 TD participants aged 7-18. Significant enlargements of cerebral grey matter were seen in LFA and HFA compared to controls, while the volume of AS participants fell in between HFA and TD participants. In addition, relations with task performance differed between AS and HFA participants. In HFA participants increased grey matter was associated with lower NVIQ, but this relation was not seen for AS participants. Conversely, increased white matter was associated with increased NVIQ in AS but not HFA participants. Thus, while the overall grey matter differences between groups suggested a continuum between autism, AS and typical development, the relations with behavioural data also suggested complex qualitative differences in brain-behaviour relations for AS and HFA participants.

More significant differences were evident in a recent study by McAlonan et al. (2008). The study compared grey matter volumes in 16 AS children, 17 HFA children and 55 TD children (mean age = 10). Similar abnormalities for both ASD groups compared to controls were seen across the cortex in frontal, temporal and parietal areas. Within the ASD participants, HFA children showed smaller volumes than AS children in a range of areas, including posterior cingulate, precuneus and the thalamus. Furthermore, volumes for the inferior frontal gyrus (an area commonly associated with language production) were observed to negatively correlate with age of phrase speech in HFA participants.

From the same sample, McAlonan et al. (2009) observed differences in white matter structure. HFA participants were observed to have greater white matter in the basal ganglia than AS participants, and both showed higher volumes for this structure than controls (again supporting the idea of a continuum of brain pathology). HFA participants also differed from controls in showing decreased white matter in their left hemisphere, specifically for frontal areas and the corpus callosum. In contrast, the same abnormalities were evident in AS participants, but only for the right hemisphere.

Structural differences have also been evident in groups of adults. Toal et al. (2010) compared grey matter volumes in 39 AS participants, 26 autistic participants (high and low functioning) and 33 controls. While both groups differed from controls on a range of grey and white matter measures (such as cerebellum volume), they also differed from each other, particularly in areas associated with language. Compared to AS participants, autistic participants showed increased grey matter in right superior temporal, supramarginal and inferior parietal areas, and decreased white matter in the medial temporal lobe. (For

preliminary evidence of a number of other areas differing in AS and autistic adults, see Lai et al., 2011, <https://imfar.confex.com/imfar/2011/webprogram/Paper8895.html> ).

Thus, comparisons between AS and autism have highlighted a range of differences in brain structure, in both children and adults. A number have involved language areas, which is consistent with the idea that language specialisation in the autistic brain is abnormal, but may be more regular in AS.

Ideally this picture would be added to by evidence from functional neuroimaging studies, but unfortunately there do not appear to be any studies that have directly compared groups of AS and autistic participants on a functional task (to this author's knowledge). Supporting the idea that communication between different areas of the brain in autism may be impaired, many contemporary studies have documented functional connectivity abnormalities (usually distal underconnectivity) in mixed ASD or autistic groups (Castelli, et al., 2002; Just, et al., 2007; Kana, Keller, Cherkassky, Minshew, & Just, 2009; Kana, Keller, Minshew, & Just, 2007; Kleinhans, et al., 2008; Koshino et al., 2005; Velazquez et al., 2009; Wicker et al., 2008). Some have also demonstrated this specifically for language-based tasks (Just, et al., 2004; Kana, Keller, Cherkassky, Minshew, & Just, 2006; Sahyoun, Belliveau, Soulieres, Schwartz, & Mody, 2010). Reduced functional connectivity has been observed in an Asperger-only group attempting an emotional faces task (Welchew, et al., 2005), but no studies to date have compared the connectivity of language processes or other language-dependent cognitive tasks in AS and HFA.

One study that may provide a model for testing this hypothesis was conducted by Sahyoun et al. (2010). Their experiment compared fMRI activation in 12 HFA children and 12 age- and IQ-matched typically developing children who were attempting a reasoning task. The task consisted of matrix puzzles that could either be solved based on visuospatial relations or semantic relations. Both groups deployed typical language centres when attempting the task, such as the inferior frontal gyrus and superior temporal gyrus. In contrast to controls, though, HFA participants showed reduced functional connectivity (i.e. correlated fMRI activity) between inferior frontal areas and middle temporal areas. In addition, HFA participants recruited ventral temporal and inferior parietal regions during task performance, which the authors interpreted as reflecting greater recruitment of visuo-spatial information to solve the task. Thus in this case, it appeared that a) communication in a "typical" fronto-temporal network was reduced in HFA participants, and that b) alternative visuospatial strategies were deployed to complete the reasoning task.

This is a different kind of task to the TQT, and would not seem to engage more anterior and lateral frontal areas (such as the dorsolateral prefrontal cortex) that are associated with classic executive tasks. Nevertheless, if this or another task closer to the TQT was deployed with AS participants, one prediction could be that fronto-temporal connectivity will be more intact in this group compared to HFA participants<sup>60</sup>. If this was shown to be the case, then it would provide some support for the above extension of the weak argument; namely that AS and HFA participants differ in the efficiency of their problem-solving because their language histories lead to differences in the integration of frontal and temporal networks.

One problem with this argument that needs to be noted is that the order of cause and effect is not necessarily clear. As noted by Toal et al. (2010), we do not yet know whether the differences seen in AS and autistic brains are consequences or causes of differences in language development: they may reflect “environmental, genetic or compensatory changes in the brain” (Toal, et al., 2010, p. 1178). Quite simply, differences in brain development between the groups could just reflect core, genetically-programmed differences, that both cause delayed (or intact) language development in early years, and later problems with utilising language knowledge in effective ways. If so, language delay and later verbal problem-solving difficulties in autism would correlate because of a common cause, and not because the former causes the latter.

Except for the possible use of twin study methods, it is not immediately obvious how such possibilities could be teased apart. Even longitudinal studies of brain development in the two groups would not be able to show that any difference was a specific consequence of post-natal developmental factors, rather than prenatal programming. It also seems likely (to this author) that the actual truth will be a messy combination of the two possibilities. For instance, early lack of language specialisation may be evident in early brain structure and function for autistic individuals, causing language delay. This lack of language, in turn, could reduce experiential input and practice using language networks, such that the functional connectivity of language centres and other parts of the brain becomes atypical. By adolescence or adulthood, some language centres may be well established (if not necessarily in the typical areas, or even the same hemisphere), but the communication between these

---

<sup>60</sup> In a behavioural study associated with this project, Sahyoun et al. (2009) have in fact reported differences in verbal vs pictorial strategy use in AS and HFA participants, with AS participants solving the matrix puzzles in a way that appeared to be much closer to typically-developing participants. However, an AS group was apparently not included in the fMRI experiment, so it is not clear whether such differences extended to the neural domain. Furthermore, the results from the behavioural task are possibly confounded by a notable difference in verbal IQ between AS and HFA participants in that particular study.

areas and regions responsible for other cognitive processes could be reduced – leading to reduced performance on complex cognitive tasks where language-based strategies are required.

## *2. Differential use of inner speech*

Another explanation, prompted by some old ideas in developmental psychology and new findings in cognitive psychology, concerns inner speech. Inner speech, or verbal mediation, is thought to be used by individuals on a range of tasks to support cognitive processes. A simple example of this is in short-term memory tasks, where verbal rehearsal may be used to maintain memory for a list of items.

According to Vygotsky (1987), inner speech represents the internalisation of outer dialogue, and has its roots in early communicative interactions with others. Based on this principle, it has been proposed that inner speech will be compromised in some way or develop atypically in autism, because of early and enduring problems with social interaction (e.g. Fernyhough, 1996). If this is the case, then their approach to a task like Twenty Questions could conceivably involve much less internal verbalisation of possible different strategies, questions and categories that would be effective in searching a set. In essence, their access to semantic categories and their spontaneous use of strategies could be intact, but they may lack an internal forum, provided by inner speech, in which different possibilities are weighed up and considered.

Inner speech processes can be examined in a range of ways, one of the most common methods being articulatory suppression, where participants are asked to repeat an irrelevant word or sound during a cognitive task in order to block the use of language-based strategies. A range of studies have sought to measure inner speech use in groups of participants with autism spectrum disorders, with largely mixed results (Holland & Low, 2010; Lidstone, Fernyhough, Meins, & Whitehouse, 2009; Wallace, Silvers, Martin, & Kenworthy, 2009; Whitehouse, Maybery, & Durkin, 2006; D. Williams, Happé, & Jarrold, 2008; D. Williams & Jarrold, 2010). However, a recent study by Williams, Bowler and Jarrold (2012) has made an important contribution to this debate, by arguing that autistic individuals differ in the kinds of inner speech that they deploy across different cognitive tasks.

In a study with ASD adults, Williams and colleagues observed evidence of intact inner speech use on a short-term memory task, but a lack of verbal mediation on a planning task

(the Tower of London). Following Fernyhough (1996), they argue that understanding inner speech in ASD may require a distinction between “monologic” inner speech, where verbal mediation takes a single perspective and may simply resemble a commentary or list of internal experiences, and “dialogic” inner speech, in which multiple viewpoints or perspectives are considered (akin to conversation). Thus in the case of their data, ASD participants may have deployed monologic inner speech to support their memory processes (e.g. by rehearsing items to be remembered), but did not use dialogic inner speech to consider multiple moves and possibilities on the planning task (D. Williams, et al., 2012).

A lack of dialogic inner speech in particular could affect ASD participants’ performance on the TQT, as it would theoretically limit the extent to which they could consider multiple alternatives in selecting their questions. It could also potentially explain differences within the autistic spectrum, although those differences may not necessarily map on to the autism/Asperger distinction. The point of dialogic inner speech, in coming from external interactions, is that it resembles the back and forth of communication (Fernyhough, 2009). If it comes from communication then one might expect those with greater communicative abilities to demonstrate more dialogic inner speech – and this is exactly what was seen by Williams et al. (2012) for their planning task: measures of communicative skills from the ADOS and the AQ were observed to positively correlate with greater inner speech deployment. In contrast, although there was no specific group-based comparison included in the study, differences in inner speech use were not seen between autistic and AS participants (Williams, personal communication). Thus, within-spectrum differences were evident, but in terms of communication abilities, rather than diagnosis.

One idea that could complete the weak argument, then, is that difficulties with problem-solving in ASD are a result of atypical language history via the abnormal development of inner speech processes, specifically for dialogic inner speech. In this case, though, the relevance of language history is better understood not in terms of structural language abilities, but in terms of communicative development. Here, the AS/HFA distinction may be acting as a proxy for differences in communicative abilities or opportunity to communicate with others in early years.

The scores from the AQ, collected from experiment 4, provide some preliminary insights into this possibility. Although scores for the Communication subscale of the AQ did not significantly differ between AS and HFA participants, they significantly predicted question quality performance for HFA participants: HFA participants with fewer communicative impairments also asked more efficient questions. This only provides one brief indicator of

communicative ability (in contrast to Williams et al.'s use of ADOS scores as well), but it is at least consistent with an “inner speech” explanation of TQT performance.

These are just two possible ways of completing the weak argument, based on findings from the neuroimaging of ASD individuals and studies on inner speech use in this group. These explanations are not mutually exclusive, and they are presented separately here to reflect the areas of research that these hypotheses initially come from. Indeed, with further testing of the neural correlates of inner speech they may begin to converge, and a complete account of this specific aspect of ASD development will require evidence from multiple areas. Both explanations should at least provide some means for making predictions that could be tested in future studies, in order to explain more fully how specific higher-order cognitive skills develop in ASD.

It should also be said, though, that endorsing the weak argument in this way does not represent a failure to achieve the original aim of the project; that is, to explain why problem-solving difficulties occur in ASD. The weak argument is so called because it does not make an explicit commitment to an intermediary mechanism; it does not specify the route by which early language delays lead to later problem-solving difficulties in autism. But it nevertheless makes a strong claim: difficulty with this kind of verbal, semantically-based problem-solving should be interpreted as a consequence of language delay, not a specific deficit in concept formation, and not a “cognitive style” specific to autism. In this respect, rejecting the strong argument and endorsing the weak argument makes a significant and important contribution to the understanding of problem-solving in autism.

Whether the intermediary mechanism is one best understood on a neural level, or at the level of inner speech, the weak argument itself can also provide specific predictions about cognitive skills in autism. If it is correct, other tasks that require the use of language-based strategies and heuristics to support complex cognitive processes should also produce discrepancies in performance between those with and without language delay on the autistic spectrum. The most immediate example of this could be semantic clustering in free recall. In that context, language knowledge needs to be used flexibly and efficiently to support memory processes. Therefore, according to the weak argument one may expect differences in strategy use on recall tasks between those ASD individuals with and without language delay.

On the other hand, strategy use on the TQT requires an explicit consideration of a range of alternative approaches to the task, but it is not clear that the same sort of process occurs in

the use of clustering strategies to support recall; participants can either group according to semantic associations or not. It may be that the two types of task actually require a different level of search through possible strategies, or a different amount of integration between systems responsible for executive control and language knowledge.

Nevertheless, comparisons of strategy use across tasks represent one clear avenue for further research (see the Appendix 3 for some preliminary data on this issue). The other benefit of testing such a hypothesis is that it would engage with Bowler and colleagues' ideas about the extra need for "task support" in autism (Bowler, et al., 2004), most of which are based on data from experiments on memory.

Another area in which fruitful predictions may be made is in the study of reasoning processes in autism. Tests of transitivity, defeasible reasoning, and drawing inferences about counterfactuals all require language to be used in a strategic and goal-directed way, and often require individuals to imagine or consider large sets of possible information. In this respect, they are likely to require the same top-down processes as the TQT, while also drawing on some of the same executive demands, such as planning and working memory. If the weak argument is right, then performance on such tasks is likely to also be affected by a history of language delay. It is crucial, though, that any tasks used have the complexity and sensitivity of a measure like the TQT: as the above experiments show, this is a task that produces truly rich data about a range of cognitive strengths and difficulties.

### **Practical Implications**

Before making some concluding remarks, the possibilities for assisting and improving problem-solving in ASD will be briefly discussed. While highlighting areas of difficulty in problem-solving, the data presented here also point to ways in which such difficulties could be addressed and overcome.

Much of this is based on the data collected in experiment 1. In the first experiment, ASD participants showed changes in performance following planning by attempting to ask more questions that grouped multiple possibilities at once. It could be that this change from baseline simply reflected a better understanding of the aim of the TQT among ASD participants (although if this was the case, one might also expect poorer performance on the question discrimination task). In addition, this change in questioning did not translate into

an overall improvement in efficiency, suggesting that participants still had difficulty in selecting the right question to ask.

Nevertheless it does imply that the strategies employed by ASD individuals may be amenable to change, possibly via a combination of explicit planning activities and some form of training to use semantic concepts appropriately. In the context of problem-solving, training of this kind has usually involved example-based modelling where an effective strategy is explained via a series of steps (e.g. “1. *Group the items into categories*, 2. *Put the categories in size order*, 3. *Ask about the biggest category, etc.*”). Model-based training has been associated with improvements in TQT problem-solving in typically-developing children (Denney, 1972; Denney, et al., 1973; Laughlin, et al., 1969) and children with learning disabilities (Barton, 1988; Simmonds, 1990). It may be that similar methods would work for individuals with ASD (for an example from social problem-solving see Solomon, Goodlin-Jones, & Anders, 2004), although training to utilise semantic relations in other cognitive domains has not always been very successful (Smith, Gardiner, & Bowler, 2007).

Improvement was also seen in the selective attention conditions of the Twenty Questions task: the trend observed in experiment 1 suggested that ASD participants had difficulty searching the array when items could not be removed, but this difficulty was ameliorated with the use of a written record. Thus, the use of explicit planning activities combined with written cues could be useful in facilitating problem-solving skills for ASD individuals. In educational contexts this is most relevant for multi-stage problems such as exam questions, where self-generated strategies might be required to effectively reach a solution. Problem-solving training can also have wider benefits for children with developmental disabilities, such as increased classroom participation (Agran, Blanchard, Wehmeyer, & Hughes, 2002; Glago, Mastropieri, & Scruggs, 2009), and improvements in social skills (Solomon et al., 2004).

In experiment 4, the overall performance of HFA and AS participants was generally very good, with HFA participants only performing at lower level in terms of their questioning efficiency. In many cases, difficulties in adopting the right problem-solving strategy led to subtly inefficient choices of questions, roundabout methods of finding a target and missed chances to solve the problem simply. Given the overall level of intact performance and insight into what the task demands, it may be that HFA participants would be able to improve their problem-solving via prompts or other simple reminders to always look for the best question possible.



## Conclusion

Investigating cognitive skills in individuals with autism is inevitably a complex and multifaceted process. For problem-solving this is especially the case, as it relies on the successful deployment of a number of cognitive and linguistic processes that develop atypically in autism. What this research represents is an attempt to take an apparently simple task, Twenty Questions, and delineate what drives problem-solving for people with an autism spectrum disorder; to not just describe their performance in terms of specific deficit or style, but to explain where it comes from developmentally.

The conclusion presented here is that verbal problem-solving difficulties in autism are a likely product of atypical language development, rather than a specific problem of information processing or executive dysfunction. The fact that autistic children are closer in their problem-solving skills to deaf children than they are to children with Asperger Syndrome is the best evidence for this conclusion. The process by which this occurs in autism is not necessarily the same as it is for deaf individuals, but they are argued to share broad similarities in their origin.

The comparison with deaf individuals also served to meet an additional aim concerning methodology. Here, both with reference to other neuropsychological groups in chapter 1, and the comparisons with deaf data in chapters 4 and 5, an attempt has been made to explain problem-solving performance by looking outside of autism research, and drawing on other examples of atypical cognition. Given the history and breadth of cognitive research on autism, it is always easy to develop autism-only theories and explanations that only ever reference other theories of autism and only use other developmental groups as controls that demonstrate the specificity of a deficit to autism. But if autism is understood as a multi-component disorder, with a wide range of cognitive strengths and weaknesses, then autism-only theories are not likely to be enough. There may be many potential overlaps with other developmental groups, like the comparison with deafness presented here, that could serve to inform understanding of both the etiology and presentation of the autistic cognitive profile.

As ever, further research is required to explain how exactly atypical language history in autism brings about the kinds of residual deficits seen in the case of verbal problem-solving. Unfortunately, this may be less rather than more likely to happen in future research, due to the changes in diagnostic criteria for DSM-V. At the very least, research of this kind should act as a reminder that early language skills are important and may have many subtle long-term consequences, whatever the diagnosis. Just because a child has “caught” up in terms of

their language or their cognitive skills, it does not necessarily mean that they draw on the same underlying processes as typically-developing children, or, indeed, other children on the autistic spectrum.

With the move to a more dimensional understanding of autism spectrum disorder, acknowledging individual variation within the spectrum will be just as important, if not more so, in making sure that the right sorts of support and assistance are being provided. Rather than treating ASD as singular entity, cognitive research will need to continue to interrogate how skills differ across the spectrum in order to provide a full account of autistic cognition. As in *Twenty Questions*, the most important thing is to make sure that the right questions are being asked.

## **APPENDIX 1**

**Table 1. Individual profiles of deaf adult participants**

<b>Participant</b>	<b>Sex</b>	<b>Age(y)</b>	<b>Pref Lang</b>	<b>Reason for hearing loss</b>
1	F	28	BSL	Meningitis
2	F	25	BSL	From birth/not known
3	F	22	Mixed	Meningitis
4	M	28	BSL	Genetic
5	M	24	BSL	From birth/not known
6	M	29	Mixed	Meningitis
7	M	26	Mixed	Mumps
8	M	26	BSL	Meningitis
9	M	26	BSL	From birth/not known

**Table 2. Individual profiles of deaf child participants**

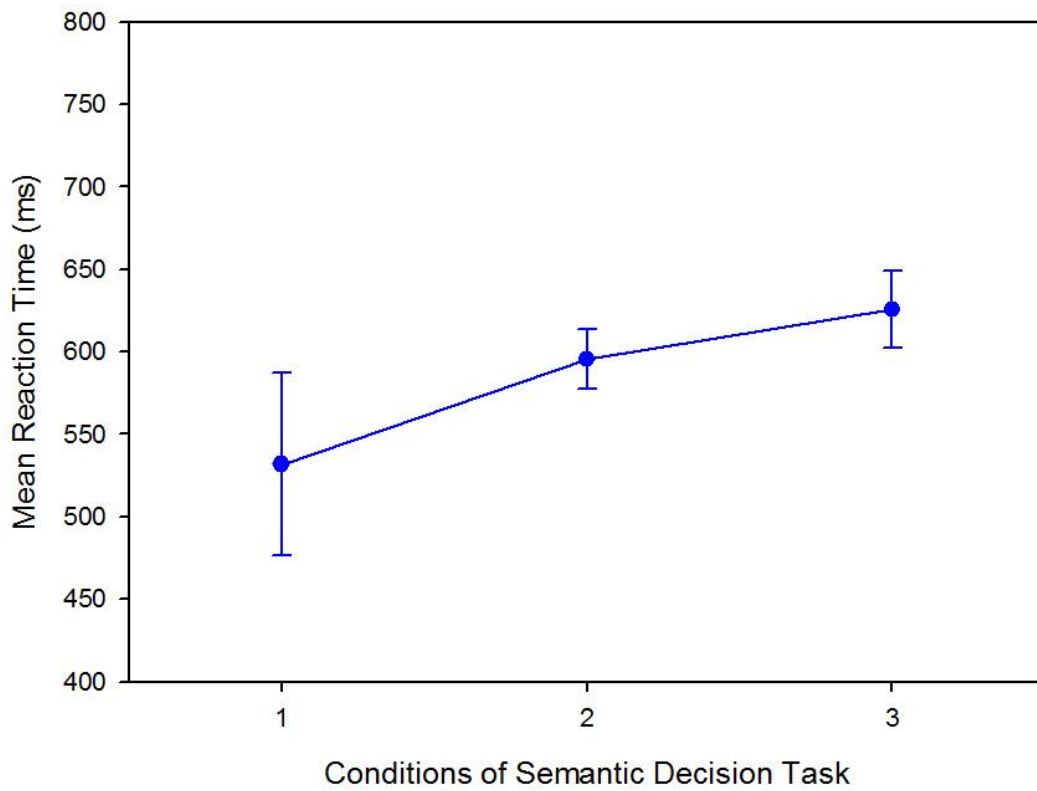
<b>Participant</b>	<b>Sex</b>	<b>Age(y)</b>	<b>Pref Lang</b>	<b>Reason for hearing loss</b>	<b>Additional diagnoses/difficulties</b>
1	M	13	Speech	CHARGE	CHARGE
2	M	13	Speech	From birth/not known	(none)
3	M	14	Speech	From birth/not known	(none)
4	M	13	BSL/SSE	Meningitis	ADHD
5	F	14	BSL/SSE	Congenital infection	(none)
6	M	15	BSL/SSE	Premature birth	Moderate LD
7	M	15	BSL/SSE	From birth/not known	“Hyperactivity”
8	F	16	Speech	From birth/not known	(none)
9	M	16	Speech	Not known, hearing deterioration at 12 m	(none)
10	M	15	Speech	From birth/not known	(none)
11	M	15	BSL/SSE	From birth/not known	(none)
12	M	14	BSL/SSE	From birth/not known	(none)
13	F	14	Speech	From birth/not known	(none)

## **APPENDIX 2**

### **Pilot analysis of the Semantic Decision Task (SDT)**

Before applying the SDT to the group of deaf children who took part in experiment 3, the task was briefly piloted with five hearing adults. The same procedure of data cleaning and analysis detailed in experiment 3 was applied to reaction times for the pilot group.

Figure A below shows the mean reaction times for this group in each of the SDT conditions. As would be expected, reaction times were fastest for basic word recognition, although there was a large degree of variability in responses. Mean RTs tended to be next fastest in condition 2 (superordinate>basic judgements) and then condition 3 (basic>superordinate judgements). However, subsequent analysis of variance indicated that the difference between conditions 2 and 3 was not significant. First, a 3x5 (Condition x Participant) ANOVA was conducted on all of the accurate SDT trials. (A repeated measures ANOVA could not be used because removal of inaccurate trials lead to differing numbers of trials in each round) This produced no significant effect of Participant ( $F(4,248) = 0.752, p = .557, \eta_p^2 = .012$ , n.s.), so the data were combined for all participants. In the ensuing three-way ANOVA, an effect for Condition that approached significance was observed ( $F(2,260) = 2.755, p = .065, \eta_p^2 = .021$ ). Pairwise comparisons (uncorrected) indicated that RTs for word recognition were significantly faster than those for basic>superordinate judgements ( $p = .023$ ), but all other contrasts were non-significant, including the crucial contrast between conditions 2 and 3 ( $p = .503$ ). This suggested that the direction of category judgements did not make a significant difference to reaction times in this pilot group.



**Figure A. Mean reaction times by pilot participants on the Semantic Decision Task.** 1 = Word Recognition, 2 = Superordinate>Basic, 3 = Basic>Superordinate.

### **APPENDIX 3**

#### **Pilot comparisons of problem-solving and free recall**

An area of research often cited in discussions of strategy use and concept formation in autism is free recall (Bowler, Gaigg, & Gardiner, 2008a; M. Cheung, et al., 2010; Minshew, Meyer, et al., 2002). Often, ASD participants have been reported as either failing to benefit from semantic relations in their recall, or using them in atypical ways; findings that would appear to be concordant with a reduced ability to utilise semantic categories effectively in problem-solving (Minshew et al., 2002).

However, no studies to date have directly compared strategy use in free recall and problem-solving performance. And while plausible, the link between strategy use in these two areas is not necessarily straightforward, not least because free recall depends on effective retrieval processes, whereas the TQT does not. In addition, a number of the studies on free recall were run with AS rather than HFA individuals (Bowler, Gaigg, & Gardiner, 2008b). If AS individuals have difficulty using semantic strategies during free recall, but intact performance on a TQT, then this suggests the engagement of different strategic processes.

In an attempt to gather some pilot data on this issue, the participants in experiment 4 all attempted a short free recall task at the end of their testing session. Participants attempted three recall trials, each consisting of 12 everyday words that split into three semantic categories (BED, CHAIR, SOFA, TABLE; ANT, BEETLE, FLY, SPIDER; BOMB, GUN, KNIFE, SWORD). The words were selected based on their presence in the top 10 for each category in the Battig and Montague norms (Battig & Montague, 1969). Using E-prime, the words appeared consecutively in a random order for 3 seconds each, after which a screen appeared, asking “How many words can you remember? Say them now!”. Before testing, participants were told that the same 12 words would appear each time, in a different order, but that they could say the words in whichever order they wanted to. The design of the task was based on a combination of free recall tasks used in Smith, Gardiner and Bowler (2007) and Cheung et al. (2010).

As table A shows, no clear differences were evident between AS, HFA and TD participants in terms of overall recall, and the lack of any group effect or group\*interaction effect on a repeated measures ANCOVA confirmed this to be the case (all  $p > .500$ , all  $\eta_p^2 < .03$ ). Neither were any group differences observed for use of semantic clustering (here simply scored as the proportion of items recalled that were placed next to another item of the same

category), when a repeated measures ANCOVA was run (Group effect:  $F(2, 39) = 0.954$ ,  $p = .394$ ,  $\eta_p^2 = .047$ ).

Thus, far from demonstrating atypical semantic strategies, whether particular to HFA participants or across both HFA and AS participants, the data here suggested intact use of semantic relations in ASD participants. If anything, the mean scores would appear to point to TD participants using clustering the least in their final attempts at recall. In addition, none of the task outcomes (total recall or clustering) were related to TQT performance in a correlational analysis.

It is not immediately clear why no group differences were observable on the task, but it may be that the measure was too brief an index to reliably assess recall abilities. Cheung et al. (2010) utilised a similar paradigm for a “learning phase”, and reported group differences between ASD and TD children, but the work of Bowler and colleagues often deploys much longer paradigms, with multiple recall trials. Showing a definite null result with regard to strategy use in this domain, compared to others, would likely require a much closer replication of other recall paradigms.

Table A	AS		HFA		TD		Sig.
	<i>M</i>	( <i>SD</i> )	<i>M</i>	( <i>SD</i> )	<i>M</i>	( <i>SD</i> )	
<b>Recall</b>							
Trial 1	7.13	1.85	7.07	1.71	7.80	1.42	n.s.
Trial 2	8.67	1.76	8.13	1.73	9.27	1.53	n.s.
Trial 3	9.27	1.67	9.20	2.31	9.93	1.39	n.s.
<b>Semantic clustering (%)</b>							
Trial 1	43.05	27.88	40.11	20.57	43.18	22.61	n.s.
Trial 2	61.90	21.52	58.21	28.25	51.12	22.35	n.s.
Trial 3	62.22	23.10	60.63	26.89	51.12	16.17	n.s.

AS = Asperger Syndrome, HFA = High-Functioning Autism, TD = Typically-Developing.

## References

- Adams, N., & Jarrold, C. (2009). Inhibition and the Validity of the Stroop Task for Children with Autism. *Journal of Autism and Developmental Disorders*, 39(8), 1112-1121.
- Adams, N., & Jarrold, C. (2012). Inhibition in Autism: Children with Autism have Difficulty Inhibiting Irrelevant Distractors but not Prepotent Responses. *Journal of Autism and Developmental Disorders*, 42(6), 1052-1063.
- Agran, M., Blanchard, C., Wehmeyer, M., & Hughes, C. (2002). Increasing the Problem-Solving Skills of Students with Developmental Disabilities Participating in General Education. *Remedial and Special Education*, 23(5), 279-288.
- Alderson-Day, B. (2011). Verbal problem-solving in autism spectrum disorders: A problem of plan construction? *Autism Research*, 4(6), 401-411.
- Alderson-Day, B., & McGonigle-Chalmers, M. (2011). Is It a Bird? Is It a Plane? Category Use in Problem-solving in Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, 41(5), 555-565.
- Altshuler, K. Z., Deming, W., Vollenweider, J., Rainer, J., & Tendler, R. (1976). Impulsivity and profound early deafness: A cross cultural inquiry. *American Annals of the Deaf*, 121(3), 331-345.
- Ames, C. S., & Fletcher-Watson, S. (2010). A review of methods in the study of attention in autism. *Developmental Review*, 30(1), 52-73.
- Ames, C. S., & Jarrold, C. (2007). The problem with using eye-gaze to infer desire: A deficit of cue inference in children with autism spectrum disorder? *Journal of Autism and Developmental Disorders*, 37(9), 1761-1775.
- Anderson, D. K., Lord, C., Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., et al. (2007). Patterns of growth in verbal abilities among children with autism spectrum disorder. *J Consult Clin Psychol*, 75(4), 594-604.
- Anderson, D. K., Oti, R. S., Lord, C., & Welch, K. (2009). Patterns of Growth in Adaptive Social Abilities Among Children with Autism Spectrum Disorders. *Journal of Abnormal Child Psychology*, 37(7), 1019-1034.
- Antia, S. D., Reed, S., & Kreimeyer, K. H. (2005). Written Language of Deaf and Hard-of-Hearing Students in Public Schools. *Journal of Deaf Studies and Deaf Education*, 10(3), 244-255.
- Anzai, Y., & Simon, H. A. (1979). The theory of learning by doing. *Psychological Review*, 86(2), 124.
- APA. (1994). *DSM-IV: Diagnostic and statistical manual of mental disorders* (4th ed.). Washington DC: American Psychiatric Association.
- APA. (2012). DSM-5 Development: A 05 Autism Spectrum Disorder. Retrieved 04/07, 2012, from <http://www.dsm5.org/proposedrevision/pages/proposedrevision.aspx?rid=94>.
- Asperger, H. (1944). Die „Autistischen Psychopathen“ im Kindesalter. *European Archives of Psychiatry and Clinical Neuroscience*, 117(1), 76-136.
- Ault, R. L. (1973). Problem-Solving Strategies of Reflective, Impulsive, Fast-Accurate, and Slow-Inaccurate Children. *Child Development*, 44(2), 259-266.
- Auyeung, B., Baron-Cohen, S., Wheelwright, S., & Allison, C. (2008). The Autism Spectrum Quotient: Children's Version (AQ-Child). *Journal of Autism and Developmental Disorders*, 38(7), 1230-1240.
- Baldo, J. V., Delis, D. C., Wilkins, D. P., & Shimamura, A. P. (2004). Is it bigger than a breadbox? Performance of patients with prefrontal lesions on a new executive function test. *Arch Clin Neuropsychol*, 19(3), 407-419.
- Baldwin, D. A., & Tomasello, M. (1998). Word learning: A window on early pragmatic understanding. In E. Clark (Ed.), *The proceedings of the twenty-ninth annual child*



- language research forum* (pp. 3-23). Chicago, IL: Center for the Study of Language and Information.
- Barnea-Goraly, N., Kwon, H., Menon, V., Eliez, S., Lotspeich, L., & Reiss, A. L. (2004). White matter structure in autism: preliminary evidence from diffusion tensor imaging. *Biol Psychiatry*, *55*(3), 323-326.
- Baron-Cohen, S. (1989). The autistic child's theory of mind: a case of specific developmental delay. *J Child Psychol Psychiatry*, *30*(2), 285-297.
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends Cogn Sci*, *6*(6), 248-254.
- Baron-Cohen, S. (2006). Two new theories of autism: hyper-systemising and assortative mating. *Archives of Disease in Childhood*, *91*(1), 2-5.
- Baron-Cohen, S., Hoekstra, R. A., Knickmeyer, R., & Wheelwright, S. (2006). The Autism-Spectrum Quotient (AQ)--adolescent version. *Journal of Autism and Developmental Disorders*, *36*(3), 343-350.
- Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or asperger syndrome. *J Child Psychol Psychiatry*, *38*(7), 813-822.
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37-46.
- Baron-Cohen, S., Ring, H. A., Bullmore, E. T., Wheelwright, S., Ashwin, C., & Williams, S. C. (2000). The amygdala theory of autism. *Neurosci Biobehav Rev*, *24*(3), 355-364.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, *31*(1), 5-17.
- Barton, J. A. (1988). Problem-solving strategies in learning disabled and normal boys: Developmental and instructional effects. *Journal of Educational Psychology*, *80*(2), 184.
- Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*(3p2), 1-46.
- Belmonte, M. K., Cook, E. H., Jr., Anderson, G. M., Rubenstein, J. L., Greenough, W. T., Beckel-Mitchener, A., et al. (2004). Autism as a disorder of neural information processing: directions for research and targets for therapy. *Mol Psychiatry*, *9*(7), 646-663.
- Bennett, T., Szatmari, P., Bryson, S., Volden, J., Zwaigenbaum, L., Vaccarella, L., et al. (2008). Differentiating autism and Asperger syndrome on the basis of language delay or impairment. *Journal of Autism and Developmental Disorders*, *38*(4), 616-625.
- Bennetto, L., Pennington, B. F., & Rogers, S. J. (1996). Intact and impaired memory functions in autism. *Child Development*, *67*(4), 1816-1835.
- Bigler, E. D., Mortensen, S., Neeley, E. S., Ozonoff, S., Krasny, L., Johnson, M., et al. (2007). Superior temporal gyrus, language function, and autism. *Developmental Neuropsychology*, *31*(2), 217-238.
- Bishop, D. V. (1998). Development of the Children's Communication Checklist (CCC): a method for assessing qualitative aspects of communicative impairment in children. *Journal of Child Psychology and Psychiatry*, *39*(6), 879-891.
- Blanton, R., & Nunnally, J. C. (1964). Semantic habits and cognitive style processes in the deaf. *The Journal of Abnormal and Social Psychology*, *68*(4), 397.
- Borys, S. V. (1979). Factors influencing the interrogative strategies of mentally retarded and nonretarded students. *American Journal of Mental Deficiency*.

- Boucher, J. (2012). Research Review: Structural language in autistic spectrum disorder – characteristics and causes. *Journal of Child Psychology and Psychiatry*, 53(3), 219-233.
- Bourne, L. (1966). *Human Conceptual Behavior*. Boston: Allyn & Bacon.
- Bowler, D. M., Gaigg, S. B., & Gardiner, J. M. (2008a). Effects of related and unrelated context on recall and recognition by adults with high-functioning autism spectrum disorder. *Neuropsychologia*, 46(4), 993-999.
- Bowler, D. M., Gaigg, S. B., & Gardiner, J. M. (2008b). Subjective organisation in the free recall learning of adults with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, 38(1), 104-113.
- Bowler, D. M., Gardiner, J. M., & Berthollier, N. (2004). Source memory in adolescents and adults with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, 34(5), 533-542.
- Bowler, D. M., Matthews, N. J., & Gardiner, J. M. (1997). Asperger's syndrome and memory: Similarity to autism but not amnesia. *Neuropsychologia*, 35(1), 65-70.
- Braden, J. (1992). Intellectual Assessment of Deaf and Hard-of-Hearing People - a Quantitative and Qualitative Research Synthesis. *School Psychology Review*, 21(1), 82-94.
- Braden, J. (1994). *Deafness, deprivation, and IQ*. New York: Plenum Press.
- Bramham, J., Ambery, F., Young, S., Morris, R., Russell, A., Xenitidis, K., et al. (2009). Executive functioning differences between adults with attention deficit hyperactivity disorder and autistic spectrum disorder in initiation, planning and strategy formation. *Autism*, 13(3), 245-264.
- Brian, J. A., Tipper, S. P., Weaver, B., & Bryson, S. E. (2003). Inhibitory mechanisms in autism spectrum disorders: typical selective inhibition of location versus facilitated perceptual processing. *J Child Psychol Psychiatry*, 44(4), 552-560.
- Brien, D., & Brennan, M. (1992). *Dictionary of British Sign Language*. London: Faber & Faber.
- Brocklehurst, R., Doherty, G., & Milbourne, A. (2002). *The Usborne Big Book Of Things To Spot*. London: Usborne.
- Bryson, S. E. (1983). Interference effects in autistic children: Evidence for the comprehension of single stimuli. *Journal of Abnormal Psychology*, 92(2), 250.
- Bühler, E., Bachmann, C., Goyert, H., Heinzl-Gutenbrunner, M., & Kamp-Becker, I. (2011). Differential Diagnosis of Autism Spectrum Disorder and Attention Deficit Hyperactivity Disorder by Means of Inhibitory Control and 'Theory of Mind'. *Journal of Autism and Developmental Disorders*, 1-9.
- Burack, J. A. (1994). Selective attention deficits in persons with autism: preliminary evidence of an inefficient attentional lens. *J Abnorm Psychol*, 103(3), 535-543.
- Burgess, P. W. (2000). Strategy application disorder: the role of the frontal lobes in human multitasking. *Psychological research*, 63(3), 279-288.
- Castelli, F., Frith, C., Happé, F., & Frith, U. (2002). Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes. *Brain*, 125(Pt 8), 1839-1849.
- Chan, A. S., Han, Y. M. Y., Leung, W. W.-m., Leung, C., Wong, V. C. N., & Cheung, M.-c. (2011). Abnormalities in the anterior cingulate cortex associated with attentional and inhibitory control deficits: A neurophysiological study on children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 5(1), 254-266.
- Channon, S., Charman, T., Heap, J., Crawford, S., & Rios, P. (2001). Real-life-type problem-solving in Asperger's syndrome. *Journal of Autism and Developmental Disorders*, 31(5), 461-469.
- Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., & Drew, A. (1997). Infants with autism: an investigation of empathy, pretend play, joint attention, and imitation. *Dev Psychol*, 33(5), 781-789.

- Cheung, C., Chua, S. E., Cheung, V., Khong, P. L., Tai, K. S., Wong, T. K. W., et al. (2009). White matter fractional anisotropy differences and correlates of diagnostic symptoms in autism. *Journal of Child Psychology and Psychiatry*, 50(9), 1102-1112.
- Cheung, M., Chan, A. S., Sze, S. L., Leung, W. W., & To, C. Y. (2010). Verbal memory deficits in relation to organization strategy in high- and low-functioning autistic children. *Research in Autism Spectrum Disorders*, 4(4), 764-771.
- Christ, S. E., Holt, D. D., White, D. A., & Green, L. (2007). Inhibitory control in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 37(6), 1155-1165.
- Christ, S. E., Kester, L. E., Bodner, K. E., & Miles, J. H. (2011). Evidence for selective inhibitory impairment in individuals with autism spectrum disorder. *Neuropsychology*, 25(6), 690-701.
- Ciesielski, K. T., Courchesne, E., & Elmasian, R. (1990). Effects of focused selective attention tasks on event-related potentials in autistic and normal individuals. *Electroencephalogr Clin Neurophysiol*, 75(3), 207-220.
- Colle, L., Baron-Cohen, S., & Hill, J. (2007). Do children with autism have a theory of mind? A non-verbal test of autism vs. specific language impairment. *Journal of Autism and Developmental Disorders*, 37(4), 716-723.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407.
- Comalli Jr, P. E., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *The Journal of Genetic Psychology: Research and Theory on Human Development*, 100, 47-53.
- Conrad, R. (1972). Speech and reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye* (pp. 205-240). Cambridge: MIT.
- Copeland, A. P., & Weissbrod, C. S. (1983). Cognitive Strategies Used by Learning Disabled Children: Does Hyperactivity Always Make Things Worse? *Journal of Learning Disabilities*, 16(8).
- Corbett, B. A., Constantine, L. J., Hendren, R., Rocke, D., & Ozonoff, S. (2009). Examining executive functioning in children with autism spectrum disorder, attention deficit hyperactivity disorder and typical development. *Psychiatry Res*, 166(2-3), 210-222.
- Courchesne, E., & Pierce, K. (2005). Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. *Current Opinion in Neurobiology*, 15(2), 225-230.
- Courtin, C. (1997). Does sign language provide deaf children with an abstraction advantage? Evidence from a categorization task. *Journal of Deaf Studies and Deaf Education*.
- Courtin, C. (2000). The Impact of Sign Language on the Cognitive Development of Deaf Children. *Journal of Deaf Studies and Deaf Education*, 5(3), 266-276.
- Courtin, C., & Melot, A. M. (1998). Development of theories of mind in deaf children. *Psychological perspectives on deafness*, 2, 79-102.
- Craig, J., & Baron-Cohen, S. (2000). Story-telling ability in children with autism or Asperger syndrome: a window into the imagination. *Isr J Psychiatry Relat Sci*, 37(1), 64-70.
- Craig, J., Baron-Cohen, S., & Scott, F. (2001). Drawing ability in autism: a window into the imagination. *Isr J Psychiatry Relat Sci*, 38(3-4), 242-253.
- D'Hondt, M., & Leybaert, J. (2003). Lateralization effects during semantic and rhyme judgement tasks in deaf and hearing subjects. *Brain and Language*, 87(2), 227-240.
- Davies, S. P. (2003). Initial and concurrent planning in solutions to well-structured problems. *Q J Exp Psychol A*, 56(7), 1147-1164.
- Davies, S. P. (2005). Planning and problem solving in well-defined domains. *The cognitive psychology of planning*, 35.
- De Jong, R., Coles, M. G. H., & Logan, G. D. (1995). Strategies and mechanisms in nonselective and selective inhibitory motor control. *Journal of Experimental Psychology: Human Perception and Performance*, 21(3), 498.

- de Villiers, P. (2005). The Role of Language Development in Theory-of-Mind Development: What Deaf Children Tell Us. In J. Astington & J. Baird (Eds.), *Why Language Matters for Theory of Mind* (pp. 266-297). New York: OUP.
- de Villiers, P., Hosler, E., Miller, K., Whalen, M., & Wong, J. (1997). *Language, theory of mind, and reading other people's emotions: A study of oral deaf children*.
- Delis, D. C., Kaplan, E., & Kramer, J. K. (2001). *Delis Kaplan Executive Function System Examiner's Manual*. San Antonio: The Psychological Corporation.
- Denney, D. (1972). Modeling and Eliciting Effects upon Conceptual Strategies. *Child Development, 43*(3), 810-823.
- Denney, D. (1973). Reflection and Impulsivity as Determinants of Conceptual Strategy. *Child Development, 44*(3), 614-623.
- Denney, D. (1974). Recognition, Formulation, and Integration in the Development of Interrogative Strategies among Normal and Retarded Children. *Child Development, 45*(4), 1068-1076.
- Denney, D. (1975). The effects of exemplary and cognitive models and self-rehearsal on children's interrogative strategies. [doi: DOI: 10.1016/0022-0965(75)90077-6]. *Journal of Experimental Child Psychology, 19*(3), 476-488.
- Denney, D., Denney, N., & Ziobrowski, M. J. (1973). Alterations in the information-processing strategies of young children following observation of adult models. *Developmental Psychology, 8*(2), 202-208.
- Dichter, G. S., Lam, K. S. L., Turner-Brown, L. M., Holtzclaw, T. N., & Bodfish, J. W. (2009). Generativity Abilities Predict Communication Deficits but not Repetitive Behaviors in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 39*(9), 1298-1304.
- Dickerson, B. C., Miller, S. L., Greve, D. N., Dale, A. M., Albert, M. S., Schacter, D. L., et al. (2007). Prefrontal-hippocampal-fusiform activity during encoding predicts intraindividual differences in free recall ability: An event-related functional-anatomic MRI study. *Hippocampus, 17*(11), 1060-1070.
- Drewe, E. (1975). Go-no go learning after frontal lobe lesions in humans. *Cortex, 11*, 8-16.
- Drumm, P., Jackson, D., & Magley, V. (1995). The primacy of superordinate-level category questions in the game of 20 questions. *Perceptual and Motor Skills, 81*(1), 4.
- Drumm, P., & Jackson, D. W. (1996). Developmental Changes in Questioning Strategies during Adolescence. *Journal of Adolescent Research, 11*(3), 285-305.
- Dunn, L. (1965). *Peabody picture vocabulary test*. Minneapolis: American Guidance Service.
- Dunn, L. (1981). *Manual for the Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Service.
- Dunn, L., Whetton, C., & Pintilie, D. (1982). *British Picture Vocabulary Test*. London: NFER-Nelson.
- Dunn, M., & Bates, J. C. (2005). Developmental Change in Neutral Processing of Words by Children with Autism. *Journal of Autism and Developmental Disorders, 35*(3), 361-376.
- Dunn, M., Vaughan, H., Kreuzer, J., & Kurtzberg, D. (1999). Electrophysiologic Correlates of Semantic Classification in Autistic and Normal Children. *Developmental Neuropsychology, 16*(1), 79 - 99.
- Ehlers, S., Nydén, A., Gillberg, C., Sandberg, A. D., Dahlgren, S. O., Hjelmquist, E., et al. (1997). Asperger syndrome, autism and attention disorders: A comparative study of the cognitive profiles of 120 children. *Journal of Child Psychology and Psychiatry, 38*(2), 207-217.
- Eisenmajer, R., Prior, M., Leekam, S., Wing, L., Gould, J., Welham, M., et al. (1996). Comparison of clinical symptoms in autism and Asperger's disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*(11), 1523-1531.

- Eisenmajer, R., Prior, M., Leekam, S., Wing, L., Ong, B., Gould, J., et al. (1998). Delayed language onset as a predictor of clinical symptoms in pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 28(6), 527-533.
- Emmorey, K., Klima, E., & Hickok, G. (1998). Mental rotation within linguistic and non-linguistic domains in users of American sign language. *Cognition*, 68(3), 221-246.
- Emmorey, K., & Kosslyn, S. M. (1996). Enhanced image generation abilities in deaf signers: A right hemisphere effect. *Brain and Cognition*, 32(1), 28-44.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Attention, Perception, & Psychophysics*, 16(1), 143-149.
- Eyler, L. T., Pierce, K., & Courchesne, E. (2012). A failure of left temporal cortex to specialize for language is an early emerging and fundamental property of autism. *Brain*, 135(3), 949-960.
- Farjardo, I., Arfé, B., Benedetti, P., & Altoé, G. (2008). Hyperlink Format, Categorization Abilities and Memory Span as Contributors to Deaf Users Hypertext Access. *Journal of Deaf Studies and Deaf Education*, 13(2), 241-256.
- Fernyhough, C. (1996). The dialogic mind: A dialogic approach to the higher mental functions. *New Ideas in Psychology*, 14(1), 47-62.
- Fernyhough, C. (2009). Dialogic Thinking. In A. Winsler, C. Fernyhough & I. Montero (Eds.), *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation* (pp. 42-52). Cambridge: Cambridge University Press.
- Friedman, J. (1985). Classification skills in normally hearing and oral deaf preschoolers: A study in language and conceptual thought. In D. Martin (Ed.), *Cognition, Education, and Deafness: Directions for Research and Instruction* (pp. 70-72). Washington: Gallaudet Univ Pr.
- Frith, U. (1991). *Autism and Asperger syndrome*. New York: Cambridge Univ Pr.
- Frith, U. (2003). *Autism: explaining the enigma* (2nd ed.). Oxford: Wiley-Blackwell.
- Frith, U. (2004). Emanuel Miller lecture: Confusions and controversies about Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 45(4), 672-686.
- Frith, U., & Happé, F. (1994). Autism: beyond "theory of mind". *Cognition*, 50(1-3), 115-132.
- Frith, U., & Snowling, M. (1983). Reading for meaning and reading for sound in autistic and dyslexic children. *British Journal of Developmental Psychology*, 1(4), 329-342.
- Furth, H. (1964). Conceptual Performance in Deaf Adults. *Journal of Abnormal and Social Psychology*, 69(6), 676-681.
- Furth, H. (1966). *Thinking without language: Psychological implications of deafness*. New York: Free Press.
- Furth, H., & Milgram, N. A. (1965). The influence of language on classification: A theoretical model applied to normal, retarded, and deaf children. *Genetic psychology monographs*.
- Gaffrey, M. S., Kleinhans, N. M., Haist, F., Akshoomoff, N., Campbell, A., Courchesne, E., et al. (2007). Atypical participation of visual cortex during word processing in autism: an fMRI study of semantic decision. *Neuropsychologia*, 45(8), 1672-1684.
- Gaigg, S. B., Gardiner, J. M., & Bowler, D. M. (2008). Free recall in autism spectrum disorder: the role of relational and item-specific encoding. *Neuropsychologia*, 46(4), 983-992.
- Gardner, M. (1983). Expressive one word picture vocabulary test (Upper Extension). *Austin, TX: Pro-Ed*.
- Garfield, J. L., Peterson, C. C., & Perry, T. (2001). Social cognition, language acquisition and the development of the theory of mind. *Mind & Language*, 16(5), 494-541.
- Gastgeb, H., Strauss, M. S., & Minshew, N. J. (2006). Do individuals with autism process categories differently? The effect of typicality and development. *Child Development*, 77(6), 1717-1729.

- Geers, A. E. (2006). Spoken language in children with cochlear implants. . In P. S. Spencer & M. Marschark (Eds.), *Advances in the spoken language development of deaf and hard-of-hearing children* (pp. 244-270). New York: OUP.
- Geschwind, D. H., & Levitt, P. (2007). Autism spectrum disorders: developmental disconnection syndromes. *Current Opinion in Neurobiology*, *17*(1), 103-111.
- Geurts, H. M., Luman, M., & Van Meel, C. S. (2008). What's in a game: The effect of social motivation on interference control in boys with ADHD and autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *49*(8), 848-857.
- Geurts, H. M., Verté, S., Oosterlaan, J., Roeyers, H., & Sergeant, J. A. (2004). How specific are executive functioning deficits in attention deficit hyperactivity disorder and autism? *J Child Psychol Psychiatry*, *45*(4), 836-854.
- Ghaziuddin, M. (2010). Brief Report: Should the DSM V Drop Asperger Syndrome? *Journal of Autism and Developmental Disorders*, *40*(9), 1146-1148.
- Ghaziuddin, M., & Mountain-Kimchi, K. (2004). Defining the intellectual profile of Asperger syndrome: Comparison with high-functioning autism. *Journal of Autism and Developmental Disorders*, *34*(3), 279-284.
- Gilhooly, K., Phillips, L. H., Wynn, V., Logie, R., & Della Sala, S. (1999). Planning processes and age in the five-disc Tower of London task. *Thinking & reasoning*, *5*(4), 339-361.
- Gilhooly, K., Wynn, V., Phillips, L. H., Logie, R., & Della Sala, S. (2002). Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual differences approach. *Thinking & reasoning*, *8*(3), 165-178.
- Gillberg, C. (1985). Asperger's syndrome and recurrent psychosis—a case study. *Journal of Autism and Developmental Disorders*, *15*(4), 389-397.
- Gillberg, I. C., & Gillberg, C. (1989). Asperger syndrome—some epidemiological considerations: a research note. *Journal of Child Psychology and Psychiatry*, *30*(4), 631-638.
- Glago, K., Mastropieri, M. A., & Scruggs, T. E. (2009). Improving Problem Solving of Elementary Students With Mild Disabilities. *Remedial and Special Education*, *30*(6), 372-380.
- Glutting, J., & Wilkinson, G. (2005). *Wide Range Achievement Test (WRAT-4)*. Austin, TX: Pro-Ed.
- Goddard, L., Howlin, P., Dritschel, B., & Patel, T. (2007). Autobiographical memory and social problem-solving in Asperger syndrome. *Journal of Autism and Developmental Disorders*, *37*(2), 291-300.
- Goldberg, M. C., Lasker, A., Zee, D., Garth, E., Tien, A., & Landa, R. (2002). Deficits in the initiation of eye movements in the absence of a visual target in adolescents with high functioning autism. *Neuropsychologia*, *40*(12), 2039-2049.
- Goldberg, M. C., Mostofsky, S. H., Cutting, L. E., Mahone, E. M., Astor, B. C., Denckla, M. B., et al. (2005). Subtle executive impairment in children with autism and children with ADHD. *Journal of Autism and Developmental Disorders*, *35*(3), 279-293.
- Goldstein, F. C., & Levin, H. (1991). Question-asking strategies after severe closed head injury. *Brain and Cognition*, *17*(1), 23-30.
- Goldstein, K., & Scheerer, M. (1941). Abstract and concrete behavior; an experimental study with special tests. *Psychological monographs*.
- Gopnik, A., & Meltzoff, A. N. (1992). Categorization and Naming: Basic-Level Sorting in Eighteen-Month-Olds and Its Relation to Language. *Child Development*, *63*(5), 1091-1103.
- Green, W. B., & Shepherd, D. C. (1975). The semantic structure in deaf children. *Journal of Communication Disorders*, *8*(4), 357-365.
- Greene, J. (1975). *Thinking and language* (Vol. 7): Methuen.
- Griffith, E. M., Pennington, B. F., Wehner, E. A., & Rogers, S. J. (1999). Executive functions in young children with autism. *Child Development*, *70*(4), 817-832.

- Groen, W. B., Zwiens, M. P., van der Gaag, R. J., & Buitelaar, J. K. (2008). The phenotype and neural correlates of language in autism: an integrative review. *Neuroscience & Biobehavioral Reviews*, 32(8), 1416-1425.
- Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66(3), 843-855.
- Happé, F. (1999). Autism: cognitive deficit or cognitive style? *Trends in Cognitive Sciences*, 3(6), 216-222.
- Happé, F., Booth, R., Charlton, R., & Hughes, C. (2006). Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: examining profiles across domains and ages. *Brain Cogn*, 61(1), 25-39.
- Happé, F., & Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 36(1), 5-25.
- Harris, G. J., Chabris, C. F., Clark, J., Urban, T., Aharon, I., Steele, S., et al. (2006). Brain activation during semantic processing in autism spectrum disorders via functional magnetic resonance imaging. *Brain Cogn*, 61(1), 54-68.
- Harris, M., & Moreno, C. (2004). Deaf children's use of phonological coding: Evidence from reading, spelling, and working memory. *Journal of Deaf Studies and Deaf Education*, 9(3), 253.
- Harris, R. I. (1978). Impulse control in deaf children: Research and clinical issues. *Deaf children: Developmental perspectives*, 137-156.
- Haug, T., & Mann, W. (2008). Adapting Tests of Sign Language Assessment for Other Sign Languages—A Review of Linguistic, Cultural, and Psychometric Problems. *Journal of Deaf Studies and Deaf Education*, 13(1), 138-147.
- Hauser, P. C., Lukomski, J., & Hillman, T. (2008). Development of Deaf and Hard-Of-Hearing Students' Executive Function. In M. Marschark & P. C. Hauser (Eds.), *Deaf Cognition: Foundations and Outcomes* (pp. 286-308). New York: Oxford University Press.
- Hayes-Roth, B., & Hayes-Roth, F. (1979). A cognitive model of planning. *Cognitive Science*, 3(4), 275-310.
- Heaton, R. K. (1993). *Wisconsin card sorting test manual*. Odessa, FL: Psychological Assessment Resources
- Heavey, L., Phillips, W., Baron-Cohen, S., & Rutter, M. (2000). The Awkward Moments Test: a naturalistic measure of social understanding in autism. *Journal of Autism and Developmental Disorders*, 30(3), 225-236.
- Henderson, H., Schwartz, C., Mundy, P., Burnette, C., Sutton, S., Zahka, N., et al. (2006). Response monitoring, the error-related negativity, and differences in social behavior in autism. *Brain and Cognition*, 61(1), 96-109.
- Hermelin, B., & O'Connor, N. (1970). *Psychological experiments with autistic children*. Oxford: Pergamon.
- Herwig, J. E. (1982). Effects of Age, Stimuli, and Category Recognition Factors in Childrens Inquiry Behavior. *Journal of Experimental Child Psychology*, 33(2), 196-206.
- Hill-Briggs, F., Dial, J. G., Morere, D. A., & Joyce, A. (2007). Neuropsychological assessment of persons with physical disability, visual impairment or blindness, and hearing impairment or deafness. *Archives of Clinical Neuropsychology*, 22(3), 389-404.
- Hill, E. L. (2004). Evaluating the theory of executive dysfunction in autism. *Developmental Review*, 24(2), 189-233.
- Hobson, R. P. (1995). *Autism and the development of mind*: Psychology Press.
- Hoemann, H. W., Andrews, C. E., & DeRosa, D. V. (1974). Categorical encoding in short-term memory by deaf and hearing children. *Journal of Speech and Hearing Research*, 17(3), 426.

- Höfler, R. (1927). Über die Bedeutung der Abstraktion für die geistige Entwicklung des taubstummen Kindes *Zeitschrift für Kinderforschung*, 33, 414-444.
- Höfler, R. (1934). Vergleichende intelligenzuntersuchung bei hoerenden und tauben mit stummen tests und ihre beziehung zun sprachbesitz. *Deutsche Gesellschaft für Sprach und Stimmheilkunde* (pp. 77-93). Berlin.
- 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(2), 369-391.
- Horton, M. S., & Markman, E. M. (1980). Developmental differences in the acquisition of basic and superordinate categories. *Child Development*, 708-719.
- Howlin, P. (2003). Outcome in high-functioning adults with autism with and without early language delays: implications for the differentiation between autism and Asperger syndrome. *Journal of Autism and Developmental Disorders*, 33(1), 3-13.
- Hughes, C., Russell, J., & Robbins, T. W. (1994). Evidence for Executive Dysfunction in Autism. *Neuropsychologia*, 32(4), 477-492.
- Hughes, J. R. (2007). Autism: The first firm finding = underconnectivity? *Epilepsy & Behavior*, 11(1), 20-24.
- Iarocci, G., & Burack, J. A. (2004). Intact covert orienting to peripheral cues among children with autism. *Journal of Autism and Developmental Disorders*, 34(3), 257-264.
- Iwanaga, R., Kawasaki, C., & Tsuchida, R. (2000). Brief report: Comparison of sensory-motor and cognitive function between autism and Asperger syndrome in preschool children. *Journal of Autism and Developmental Disorders*, 30(2), 169-174.
- Jackson, A. L. (2001). Language facility and theory of mind development in deaf children. *Journal of Deaf Studies and Deaf Education*, 6(3), 161.
- Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(1), 60-99.
- Johnston, K., Madden, A. K., Bramham, J., & Russell, A. J. (2011). Response inhibition in adults with autism spectrum disorder compared to attention deficit/hyperactivity disorder. *Journal of Autism and Developmental Disorders*, 1-10.
- Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, 16(2), 243-275.
- Jolliffe, T., & Baron-Cohen, S. (1999). The Strange Stories Test: a replication with high-functioning adults with autism or Asperger syndrome. *Journal of Autism and Developmental Disorders*, 29(5), 395-406.
- Joseph, R. M., Steele, S. D., Meyer, E., & Tager-Flusberg, H. (2005). Self-ordered pointing in children with autism: failure to use verbal mediation in the service of working memory? *Neuropsychologia*, 43(10), 1400-1411.
- Just, M. A., Cherkassky, V. L., Keller, T. A., Kana, R. K., & Minshew, N. J. (2007). Functional and anatomical cortical underconnectivity in autism: Evidence from an fMRI study of an executive function task and corpus callosum morphometry. *Cerebral Cortex*, 17(4), 951-961.
- Just, M. A., Cherkassky, V. L., Keller, T. A., & Minshew, N. J. (2004). Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain*, 127, 1811-1821.
- Kaland, N., Møller-Nielsen, A., Callesen, K., Mortensen, E. L., Gottlieb, D., & Smith, L. (2002). A new advanced test of theory of mind: evidence from children and adolescents with Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 43(4), 517-528.
- Kamio, Y., Robins, D., Kelley, E., Swainson, B., & Fein, D. (2007). Atypical lexical/semantic processing in high-functioning autism spectrum disorders without early language delay. *Journal of Autism and Developmental Disorders*, 37(6), 1116-1122.



- Kamio, Y., & Toichi, M. (2007). Memory illusion in high-functioning autism and Asperger's disorder. *Journal of Autism and Developmental Disorders*, 37(5), 867-876.
- Kana, R. K., Cherkassky, V. L., Keller, T. A., Berson, S., Minshew, N. J., & Just, M. A. (2006). Reduced corpus callosum size in autism and its impact on cortical underconnectivity. *Annals of Neurology*, 60, S128-S128.
- Kana, R. K., Keller, T. A., Cherkassky, V. L., Minshew, N. J., & Just, M. A. (2006). Sentence comprehension in autism: thinking in pictures with decreased functional connectivity. *Brain*, 129(Pt 9), 2484-2493.
- Kana, R. K., Keller, T. A., Cherkassky, V. L., Minshew, N. J., & Just, M. A. (2009). Atypical frontal-posterior synchronization of Theory of Mind regions in autism during mental state attribution. *Social Neuroscience*, 4(2), 135-152.
- Kana, R. K., Keller, T. A., Minshew, N. J., & Just, M. A. (2007). Inhibitory control in high-functioning autism: Decreased activation and underconnectivity in inhibition networks. *Biological Psychiatry*, 62(3), 198-206.
- Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous child*, 2, 217-250.
- Kaplan, E., Goodglass, H., & Weintraub, S. (1976). *Boston Naming Test, Experimental Edition*. Boston: Veteran's Administration Hospital.
- Kates, S. L., Kates, W. W., Michael, J., & Walsh, T. M. (1961). Categorization and related verbalizations in deaf and hearing adolescents. [Article]. *Journal of Educational Psychology August*, 52(4), 188-194.
- Keehn, B., Lincoln, A. J., Müller, R. A., & Townsend, J. (2010). Attentional networks in children and adolescents with autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 51(11), 1251-1259.
- Kelley, E., Paul, J., Fein, D., & Naigles, L. (2006). Residual Language Deficits in Optimal Outcome Children with a History of Autism. *Journal of Autism and Developmental Disorders*, 36(6), 807-828.
- Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychology Review*, 18(4), 320-338.
- Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: Implications for genetic subgroups. *Language and cognitive processes*, 16(2-3), 287-308.
- Kleinmans, N., Akshoomoff, N., & Delis, D. C. (2005). Executive functions in autism and Asperger's disorder: Flexibility, fluency, and inhibition. *Developmental Neuropsychology*, 27(3), 379-401.
- Kleinmans, N., Richards, T., Sterling, L., Stegbauer, K. C., Mahurin, R., Johnson, L. C., et al. (2008). Abnormal functional connectivity in autism spectrum disorders during face processing. *Brain*, 131(Pt 4), 1000-1012.
- Klin, A., Volkmar, F., Sparrow, S., Cicchetti, D., & Rourke, B. (1995). Validity and neuropsychological characterization of Asperger syndrome: convergence with nonverbal learning disabilities syndrome. *Journal of Child Psychology and Psychiatry*, 36(7), 1127-1140.
- Klouta, G., & Cooper, W. (1990). Information search following damage to the frontal lobes. *Psychological Reports*, 67(2), 411.
- Koh, S. D., Vernon, M., & Bailey, W. (1971). Free-recall learning of word lists by prelingual deaf subjects. *Journal of Verbal Learning and Verbal Behavior*, 10(5), 542-547.
- Koshino, H., Carpenter, P. A., Minshew, N. J., Cherkassky, V. L., Keller, T. A., & Just, M. A. (2005). Functional connectivity in an fMRI working memory task in high-functioning autism. *Neuroimage*, 24(3), 810-821.
- Kovattana, P. M., & Kraemer, H. C. (1974). Response to multiple visual cues of color, size, and form by autistic children. *J Autism Child Schizophr*, 4(3), 251-261.
- Krakow, R. A., & Hanson, V. L. (1985). Deaf signers and serial recall in the visual modality: Memory for signs, fingerspelling, and print. *Memory & Cognition*, 13(3), 265-272.

- Kugler, B. (1998). The differentiation between autism and Asperger syndrome. *Autism*, 2(1), 11-32.
- Kwon, H., Ow, A. W., Pedatella, K. E., Lotspeich, L. J., & Reiss, A. L. (2004). Voxel based morphometry elucidates structural neuroanatomy of high functioning autism and Asperger syndrome. *Developmental Medicine & Child Neurology*, 46(11), 760-764.
- Ladd, P. (2003). *Understanding deaf culture: In search of deafhood: Multilingual Matters Ltd.*
- Laine, M., & Butters, N. (1982). A preliminary study of the problem-solving strategies of detoxified long-term alcoholics. *Drug Alcohol Depend*, 10(2-3), 235-242.
- Landa, R. J., & Goldberg, M. C. (2005). Language, social, and executive functions in high functioning autism: A continuum of performance. *Journal of Autism and Developmental Disorders*, 35(5), 557-573.
- Laughlin, P. R., Moss, I. L., & Miller, S. M. (1969). Information-processing in children as a function of adult model, stimulus display, school grade, and sex. *J Educ Psychol*, 60(3), 188-193.
- Lavoie, M. E., Thibault, G., Stip, E., & O'Connor, K. P. (2007). Memory and executive functions in adults with Gilles de la Tourette syndrome and chronic tic disorder. *Cognitive neuropsychiatry*, 12(2), 165-181.
- Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.
- Levin, H., Culhane, K., Mendelsohn, D., Lilly, M. A., Bruce, D., Fletcher, J. M., et al. (1993). Cognition in Relation to Magnetic Resonance Imaging in Head-Injured Children and Adolescents. *Arch Neurol*, 50(9), 897-905.
- Levin, H., Song, J., Scheibel, R. S., Fletcher, J. M., Harward, H., Lilly, M., et al. (1997). Concept formation and problem-solving following closed head injury in children. *Journal of the International Neuropsychological Society*, 3(06), 598-607.
- Levine, M. N. (1982). *Leiter international performance scale: A handbook*: Western Psychological Services.
- Liben, L. S. (1979). Free recall by deaf and hearing children: Semantic clustering and recall in trained and untrained groups. *Journal of Experimental Child Psychology*, 27(1), 105-119.
- Liben, L. S. (1985). The Development and Use of Memory Strategies by Deaf Children and Adults. In D. Martin (Ed.), *Cognition, Education, and Deafness: Directions for Research and Instruction* (pp. 66-69). Washington: Gallaudet Univ Pr.
- Lichtenstein, E. (1985). Deaf working memory processes and English language skills. *Cognition, Education, and Deafness: Directions for Research and Instruction*, 111-114.
- 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(8), 1222-1225.
- Loftus, E. F. (1973). Category dominance, instance dominance, and categorization time. *Journal of Experimental Psychology*, 97(1), 70.
- Logan, K., Maybery, M., & Fletcher, J. (1996). The Short Term Memory of Profoundly Deaf People for Words, Signs, and Abstract Spatial Stimuli. *Applied cognitive psychology*, 10(2), 105-119.
- Lopez, B., Lincoln, A. J., Ozonoff, S., & Lai, Z. (2005). Examining the relationship between executive functions and restricted, repetitive symptoms of Autistic Disorder. *Journal of Autism and Developmental Disorders*, 35(4), 445-460.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205-223.

- Lord, C., Rutter, M., & Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659-685.
- Lotspeich, L. J., Kwon, H., Schumann, C. M., Fryer, S. L., Goodlin-Jones, B. L., Buonocore, M. H., et al. (2004). Investigation of neuroanatomical differences between autism and Asperger syndrome. *Archives of General Psychiatry*, 61(3), 291.
- Luftig, R. L. (1983). Variables influencing the learnability of individual signs and sign lexicons: A review of the literature. *Journal of psycholinguistic research*, 12(4), 361-376.
- Luna, B., Doll, S. K., Hegedus, S. J., Minshew, N. J., & Sweeney, J. A. (2007). Maturation of executive function in autism. *Biol Psychiatry*, 61(4), 474-481.
- Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-Flusberg, H. (2008). Language assessment and development in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 38(8), 1426-1438.
- Macintosh, K. E., & Dissanayake, C. (2004). Annotation: the similarities and differences between autistic disorder and Asperger's disorder: a review of the empirical evidence. *Journal of Child Psychology and Psychiatry*, 45(3), 421-434.
- Mackinlay, R., Charman, T., & Karmiloff-Smith, A. (2006). High functioning children with autism spectrum disorder: A novel test of multitasking. *Brain and Cognition*, 61(1), 14-24.
- Manjiviona, J., & Prior, M. (1999). Neuropsychological profiles of children with Asperger syndrome and autism. *Autism*, 3(4), 327-356.
- Marschark, M. (2003). Cognitive functioning in deaf adults and children. In M. Marschark & P. Spencer (Eds.), *Oxford handbook of deaf studies, language, and education* (pp. 464-477). New York: Oxford University Press.
- Marschark, M. (2006). Intellectual functioning of deaf adults and children: Answers and questions. *European Journal of Cognitive Psychology*, 18(1), 70 - 89.
- Marschark, M., Convertino, C., McEvoy, C., & Masteller, A. (2004). Organization and use of the mental lexicon by deaf and hearing individuals. *American Annals of the Deaf*, 149(1), 51-61.
- Marschark, M., & Everhart, V. (1999). Problem-solving by deaf and hearing students: twenty questions. *Deafness & Education International*, 1(2), 65-82.
- Marschark, M., & Hauser, P. C. (2008). Cognitive underpinnings of learning by deaf and hard-of-hearing students. *Deaf cognition: Foundations and outcomes*, 3-23.
- Marshall, R., Harvey, S. R., Freed, D. B., & Phillips, D. S. (1996). Question asking strategies of aphasic and non-brain-damaged subjects. *Clinical Aphasiology*, 24, 181-192.
- Marshall, R., Karow, C., Morelli, C., Iden, K., Dixon, J., & Cranfill, T. (2004). Effects of interactive strategy modelling training on problem-solving by persons with traumatic brain injury. *Aphasiology*, 18(8), 659 - 673.
- Marshall, R., Karow, C. M., Morelli, C. A., Iden, K. K., & Dixon, J. (2003a). A Clinical Measure for the Assessment of Problem Solving in Brain-Injured Adults. *Am J Speech Lang Pathol*, 12(3), 333-348.
- Marshall, R., Karow, C. M., Morelli, C. A., Iden, K. K., & Dixon, J. (2003b). Problem-solving by traumatically brain injured and neurologically intact subjects on an adaptation of the twenty questions test. *Brain Inj*, 17(7), 589-608.
- Marshall, R., McGurk, S. R., Karow, C. M., & Kairy, T. J. (2007). Problem-solving abilities of participants with and without diffuse neurologic involvement. *Aphasiology*, 21(6-8), 750-762.
- Mayberry, R. I. (2002). Cognitive development in deaf children: The interface of language and perception in neuropsychology. *Handbook of neuropsychology*, 8(Part II), 71-107.

- Mayes, S. D., & Calhoun, S. L. (2001). Non-significance of early speech delay in children with autism and normal intelligence and implications for DSM-IV Asperger's disorder. *Autism, 5*(1), 81-94.
- Mayes, S. D., & Calhoun, S. L. (2003). Ability profiles in children with autism - Influence of age and IQ. *Autism, 7*(1), 65-80.
- Mayes, S. D., Calhoun, S. L., & Crites, D. L. (2001). Does DSM-IV Asperger's disorder exist? *Journal of Abnormal Child Psychology, 29*(3), 263-271.
- McAlonan, G., Cheung, C., Cheung, V., Wong, N., Suckling, J., & Chua, S. E. (2009). Differential effects on white-matter systems in high-functioning autism and Asperger's syndrome. *Psychological Medicine, 39*(11), 1885-1893.
- McAlonan, G., Suckling, J., Wong, N., Cheung, V., Lienenkaemper, N., Cheung, C., et al. (2008). Distinct patterns of grey matter abnormality in high-functioning autism and Asperger's syndrome. *Journal of Child Psychology and Psychiatry, 49*(12), 1287-1295.
- McCarthy, D. (1972). *Manual for the McCarthy scales of children's abilities*. New York: Psychological Corp.
- McEvoy, C., Marschark, M., & Nelson, D. (1999). Comparing the mental lexicons of deaf and hearing individuals. *Journal of Educational Psychology, 91*(2), 312-320.
- McGeorge, P., Phillips, L. H., Crawford, J. R., Garden, S. E., Sala, S. D., Milne, A. B., et al. (2001). Using virtual environments in the assessment of executive dysfunction. *Presence: Teleoperators & Virtual Environments, 10*(4), 375-383.
- McGregor, K. K., & Bean, A. (2012). How Children With Autism Extend New Words. *J Speech Lang Hear Res, 55*(1), 70-83.
- McKinney, J. D. (1975). Problem-solving strategies in reflective and impulsive children. *Journal of Educational Psychology, 67*(6), 807-820.
- Miller, J. (1988). A warning about median reaction time. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 539.
- Miller, J., & Ozonoff, S. (1997). Did Asperger's cases have Asperger disorder? A research note. *Journal of Child Psychology and Psychiatry, 38*(2), 247-251.
- Miller, J., & Ozonoff, S. (2000). The external validity of Asperger disorder: Lack of evidence from the domain of neuropsychology. *Journal of Abnormal Psychology, 109*(2), 227.
- Minshew, N. J., & Goldstein, G. (1993a). Autism - a Distributed Neural Network Deficit. *Journal of Clinical and Experimental Neuropsychology, 15*(1), 56-56.
- Minshew, N. J., & Goldstein, G. (1993b). Is Autism an Amnesic Disorder. *Biological Psychiatry, 33*(6A), A104-A105.
- Minshew, N. J., & Goldstein, G. (1998). Autism as a disorder of complex information processing. *Mental Retardation and Developmental Disabilities Research Reviews, 4*(2), 129-136.
- Minshew, N. J., & Goldstein, G. (2001). The pattern of intact and impaired memory functions in autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*(8), 1095-1101.
- Minshew, N. J., Goldstein, G., & Siegel, D. J. (1996). Neuropsychological functioning in autism: Evidence for a generalized complex information processing deficit. *International Journal of Psychology, 31*(3-4), 1362-1362.
- Minshew, N. J., Goldstein, G., & Siegel, D. J. (1997). Neuropsychologic functioning in autism: profile of a complex information processing disorder. *J Int Neuropsychol Soc, 3*(4), 303-316.
- Minshew, N. J., Luna, B., & Sweeney, J. A. (1999). Oculomotor evidence for neocortical systems but not cerebellar dysfunction in autism. *Neurology, 52*(5), 917-922.
- Minshew, N. J., Meyer, J., & Goldstein, G. (2002). Abstract reasoning in autism: A dissociation between concept formation and concept identification. *Neuropsychology, 16*(3), 327-334.

- Minshew, N. J., Siegel, D. J., Goldstein, G., & Weldy, S. (1994). Verbal Problem-Solving in High-Functioning Autistic Individuals. *Archives of Clinical Neuropsychology*, 9(1), 31-40.
- Minshew, N. J., Sweeney, J., & Luna, B. (2002). Autism as a selective disorder of complex information processing and underdevelopment of neocortical systems. *Mol Psychiatry*, 7 Suppl 2, S14-15.
- Mioshi, E., Dawson, K., Mitchell, J., Arnold, R., & Hodges, J. R. (2006). The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *International journal of geriatric psychiatry*, 21(11), 1078-1085.
- Mitchell, R. E. (2005). National profile of deaf and hard of hearing students in special education from weighted survey results. *American Annals of the Deaf*, 149(4), 336-349.
- Mitchell, R. E., & Karchmer, M. A. (2004). Chasing the mythical ten percent: Parental hearing status of deaf and hard of hearing students in the United States. *Sign Language Studies*, 4(2), 138-163.
- Moeller, M. P., & Schick, B. (2006). Relations Between Maternal Input and Theory of Mind Understanding in Deaf Children. *Child Development*, 77(3), 751-766.
- Mosconi, M., Kay, M., D'Cruz, A. M., Seidenfeld, A., Guter, S., Stanford, L., et al. (2009). Impaired inhibitory control is associated with higher-order repetitive behaviors in autism spectrum disorders. *Psychological Medicine*, 39(09), 1559-1566.
- Mosher, F. A., & Hornsby, J. R. (1966). On asking questions. In J. S. Bruner, R. R. Olver & P. Greenfield (Eds.), *Studies in cognitive growth* (pp. 86-102). New York: Wiley.
- 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(1), 27-43.
- Müller, R.-A., Shih, P., Keehn, B., Deyoe, J. R., Leyden, K. M., & Shukla, D. K. (In press). Underconnected, but How? A Survey of Functional Connectivity MRI Studies in Autism Spectrum Disorders. *Cerebral Cortex*.
- Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, 20(1), 115-128.
- Nation, K., Clarke, P., Wright, B., & Williams, C. (2006). Patterns of reading ability in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 36(7), 911-919.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9(3), 353-383.
- NDCS. (2011). About National Deaf Children's Society. Retrieved 11th June, 2012, from [http://www.ndcs.org.uk/about\\_us/ndcs/index.html](http://www.ndcs.org.uk/about_us/ndcs/index.html)
- Nelson, D., McEvoy, C., & Schreiber, T. (1998). The University of South Florida word association, rhyme, and word fragment norms. [http. w3. usf. edu/FreeAssociation](http://w3.usf.edu/FreeAssociation).
- Neville, H. J., & Lawson, D. (1987). Attention to central and peripheral visual space in a movement detection task: An event-related potential and behavioral study. II. Congenitally deaf adults. *Brain Research*, 405(2), 268-283.
- Newcomer, P. L., & Hammill, D. D. (1988). *Test of language development-2: primary*: Pro-ed Austin, TX.
- Newell, A., & Simon, H. J. (1972). *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall.
- Nydén, A., Gillberg, C., Hjelmquist, E., & Heiman, M. (1999). Executive function/attention deficits in boys with Asperger syndrome, attention disorder and reading/writing disorder. *Autism*, 3(3), 213-228.
- O'Connor, N., & Hermelin, B. (1973). The spatial or temporal organization of short-term memory. *The Quarterly Journal of Experimental Psychology*, 25(3), 335-343.

- O'Connor, N., & Hermelin, B. (1978). *Seeing and hearing and space and time*: Academic Press.
- Oléron, P. (1951). Pensée conceptuelle et langage. Performances comparées d'enfants sourds et entendants dans des épreuves de classement multiple. *L'Année Psychologique*, 51(1), 89-120.
- Oléron, P. (1953). Conceptual thinking of the deaf. *American Annals of the Deaf*, 98, 304-310.
- Ormel, E. A., Gijssels, M. A. R., Hermans, D., Bosman, A. M. T., Knoors, H., & Verhoeven, L. (2010). Semantic categorization: A comparison between deaf and hearing children. *Journal of Communication Disorders*, 43(5), 347-360.
- Ottum, E. (1980). An Analysis of Cognitive Studies with Deaf Subjects. *American Annals of the Deaf*, 125(5), 564-575.
- Ozonoff, S. (1995). Reliability and Validity of the Wisconsin Card Sorting Test in Studies of Autism. *Neuropsychology*, 9(4), 491-500.
- Ozonoff, S., Cook, I., Coon, H., Dawson, G., Joseph, R. M., Klin, A., et al. (2004). Performance on Cambridge Neuropsychological Test Automated Battery subtests sensitive to frontal lobe function in people with autistic disorder: evidence from the Collaborative Programs of Excellence in Autism network. *Journal of Autism and Developmental Disorders*, 34(2), 139-150.
- Ozonoff, S., & Jensen, J. (1999). Brief report: Specific executive function profiles in three neurodevelopmental disorders. *Journal of Autism and Developmental Disorders*, 29(2), 171-177.
- Ozonoff, S., Pennington, B. F., & Rogers, S. J. (1991). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry*, 32(7), 1081-1105.
- Ozonoff, S., Rogers, S. J., & Pennington, B. F. (1991). Asperger's Syndrome: Evidence of an Empirical Distinction from High-Functioning Autism. *Journal of Child Psychology and Psychiatry*, 32(7), 1107-1122.
- Ozonoff, S., South, M., & Miller, J. (2000). DSM-IV-Defined Asperger Syndrome: Cognitive, Behavioral and Early History Differentiation from High-Functioning Autism. *Autism*, 4(1), 29-46.
- Ozonoff, S., & Strayer, D. L. (1997). Inhibitory function in nonretarded children with autism. *Journal of Autism and Developmental Disorders*, 27(1), 59-77.
- Ozonoff, S., Strayer, D. L., McMahon, W. M., & Filloux, F. (1994). Executive function abilities in autism and Tourette syndrome: An information processing approach. *Journal of Child Psychology and Psychiatry*, 35(6), 1015-1032.
- Pajurkova, E. M., Orr, R. R., Byron, P. R., & Finlayson, M. A. J. (1976). Children's Word-Finding Test: A verbal problem-solving task. *Perceptual and Motor Skills*, 42(3), 851-858.
- Pellicano, E. (2007). Links Between Theory of Mind and Executive Function in Young Children With Autism: Clues to Developmental Primacy. *Developmental Psychology July*, 43(4), 974-990.
- Pellicano, E., Maybery, M., Durkin, K., & Maley, A. (2006). Multiple cognitive capabilities/deficits in children with an autism spectrum disorder: "weak" central coherence and its relationship to theory of mind and executive control. *Dev Psychopathol*, 18(1), 77-98.
- Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *J Child Psychol Psychiatry*, 37(1), 51-87.
- Perner, J., Frith, U., Leslie, A. M., & Leekam, S. R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development*, 689-700.

- Peterson, C. (2004). Theory-of-mind development in oral deaf children with cochlear implants or conventional hearing aids. *Journal of Child Psychology and Psychiatry*, 45(6), 1096-1106.
- Peterson, C. (2009). Development of social cognitive and communication skills in children born deaf. *Scandinavian journal of psychology*, 50(5), 475-483.
- Peterson, C., & Siegal, M. (1995). Deafness, conversation and theory of mind. *Journal of Child Psychology and Psychiatry*, 36(3), 459-474.
- Peterson, C., & Siegal, M. (1999). Representing inner worlds: Theory of mind in autistic, deaf, and normal hearing children. *Psychological Science*, 10(2), 126.
- Peterson, C., & Siegal, M. (2000). Insights into theory of mind from deafness and autism. *Mind & Language*, 15(1), 123-145.
- Peterson, C., & Wellman, H. M. (2009). From fancy to reason: Scaling deaf and hearing children's understanding of theory of mind and pretence. *British Journal of Developmental Psychology*, 27(2), 297-310.
- Peterson, C., Wellman, H. M., & Liu, D. (2005). Steps in Theory of Mind Development for Children With Deafness or Autism. *Child Development*, 76(2), 502-517.
- Peterson, C., Wellman, H. M., & Slaughter, V. (2012). The Mind Behind the Message: Advancing Theory-of-Mind Scales for Typically Developing Children, and Those With Deafness, Autism, or Asperger Syndrome. *Child Development*, 83(2), 469-485.
- Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal-and temporal-lobe lesions in man. *Neuropsychologia*, 20(3), 249-262.
- Pettifor, J. L. (1968). The role of language in the development of abstract thinking: a comparison of hard-of-hearing and normal-hearing children on levels of conceptual thinking. *Canadian Journal of Psychology/Revue canadienne de psychologie*, 22(3), 139-156.
- Phillips, L. H. (1999). The role of memory in the Tower of London task. *Memory*, 7(2), 209-231.
- Phillips, L. H., Wynn, V., McPherson, S., & Gilhooly, K. (2001). Mental planning and the Tower of London task. *The Quarterly Journal of Experimental Psychology: Section A*, 54(2), 579-597.
- Porteus, S. D. (1950). The Porteus Maze Test and Intelligence.
- Quigley, S. P., & Paul, P. V. (1984). *Language and deafness*: College Hill Books.
- Rajendran, G., Law, A. S., Logie, R. H., van der Meulen, M., Fraser, D., & Corley, M. (2010). Investigating Multitasking in High-Functioning Adolescents with Autism Spectrum Disorders Using the Virtual Errands Task. *Journal of Autism and Developmental Disorders*, 41(11), 1445-1454.
- Rasser, P. E., Johnston, P., Lagopoulos, J., Ward, P. B., Schall, U., Thienel, R., et al. (2005). Functional MRI BOLD response to Tower of London performance of first-episode schizophrenia patients using cortical pattern matching. *Neuroimage*, 26(3), 941-951.
- Rattermann, M. J., Spector, L., Grafman, J., Levin, H., & Harward, H. (2001). Partial and total-order planning: evidence from normal and prefrontally damaged populations. *Cognitive Science*, 25(6), 941-975.
- Reiter, A., Tucha, O., & Lange, K. W. (2005). Executive functions in children with dyslexia. *Dyslexia*, 11(2), 116-131.
- Remine, M. D., Care, E., & Brown, P. M. (2008). Language ability and verbal and nonverbal executive functioning in deaf students communicating in spoken English. *Journal of Deaf Studies and Deaf Education*, 13(4), 531-545.
- Remington, A., Swettenham, J., Campbell, R., & Coleman, M. (2009). Selective Attention and Perceptual Load in Autism Spectrum Disorder. *Psychol Sci*, 20(11), 1388-1393.
- Remington, A., Swettenham, J. G., & Lavie, N. (2012). Lightening the load: Perceptual load impairs visual detection in typical adults but not in autism. *Journal of Abnormal Psychology*, 121(2), 544.

- Reynell, J., & Huntley, M. (1987). *Reynell developmental language scales manual*. Windsor: NFER-Nelson.
- Riccio, C. A., Wolfe, M. E., Romine, C., Davis, B., & Sullivan, J. R. (2004). The Tower of London and neuropsychological assessment of ADHD in adults. *Archives of Clinical Neuropsychology, 19*(5), 661-671.
- Rinehart, N. J., Bradshaw, J. L., Moss, S. A., Brereton, A. V., & Tonge, B. J. (2001). A deficit in shifting attention present in high-functioning autism but not Asperger's disorder. *Autism, 5*(1), 67-80.
- Rinehart, N. J., Bradshaw, J. L., Moss, S. A., Brereton, A. V., & Tonge, B. J. (2008). Brief report: inhibition of return in young people with autism and Asperger's disorder. *Autism, 12*(3), 249-260.
- Rinehart, N. J., Bradshaw, J. L., Tonge, B. J., Brereton, A. V., & Bellgrove, M. A. (2002). A Neurobehavioral Examination of Individuals with High-Functioning Autism and Asperger's Disorder Using a Fronto-Striatal Model of Dysfunction. *Behavioral and Cognitive Neuroscience Reviews, 1*(2), 164-177.
- Rippon, G., Brock, J., Brown, C., & Boucher, J. (2007). Disordered connectivity in the autistic brain: challenges for the "new psychophysiology". *International Journal of Psychophysiology, 63*(2), 164-172.
- Robbins, T. W., James, M., Owen, A., Sahakian, B., McInnes, L., & Rabbitt, P. (1994). Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. *Dementia and Geriatric Cognitive Disorders, 5*(5), 266-281.
- Robinson, S., Goddard, L., Dritschel, B., Wisley, M., & Howlin, P. (2009). Executive functions in children with Autism Spectrum Disorders. *Brain and Cognition, 71*(3), 362-368.
- Ropar, D., & Peebles, D. (2007). Sorting preference in children with autism: the dominance of concrete features. *Journal of Autism and Developmental Disorders, 37*(2), 270-280.
- Rosenstein, J. (1960). Cognitive abilities of deaf children. *Journal of Speech & Hearing Research, 3*, 108-119.
- Rumsey, J. M. (1985). Conceptual problem-solving in highly verbal, nonretarded autistic men. *Journal of Autism and Developmental Disorders, 15*(1), 23-36.
- Russell, J. E. (1997). *Autism as an executive disorder*: Oxford University Press.
- Saarnio, P. (1993). Question-asking strategies in abstinent alcoholics: A study of learning potential. *Alcohol and Alcoholism, 28*(5), 585.
- Sahyoun, C., Belliveau, J. W., Soulières, I., Schwartz, S., & Mody, M. (2010). Neuroimaging of the functional and structural networks underlying visuospatial vs. linguistic reasoning in high-functioning autism. *Neuropsychologia, 48*(1), 86-95.
- Sahyoun, C., Soulières, I., Belliveau, J. W., Mottron, L., & Mody, M. (2009). Cognitive differences in pictorial reasoning between high-functioning autism and Asperger's syndrome. *Journal of Autism and Developmental Disorders, 39*(7), 1014-1023.
- Scheres, A., Oosterlaan, J., Geurts, H., Morein-Zamir, S., Meiran, N., Schut, H., et al. (2004). Executive functioning in boys with ADHD: primarily an inhibition deficit? *Archives of Clinical Neuropsychology, 19*(4), 569-594.
- Schmitz, N., Rubia, K., Daly, E., Smith, A., Williams, S., & Murphy, D. G. M. (2006). Neural correlates of executive function in autistic spectrum disorders. *Biological Psychiatry, 59*(1), 7-16.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*. Pittsburgh: Psychology Software Tools, Inc.
- Scholnick, E. K., Friedman, S. L., & Wallner-Allen, K. E. (1997). What do they really measure? A comparative analysis of planning tasks. *The developmental psychology of planning: Why, how, and when do we plan*, 127-156.



- Schopler, E. (1996). Are autism and Asperger syndrome (AS) different labels or different disabilities?—VZ. *Journal of Autism and Developmental Disorders*, 26(1), 109-110.
- Scott, W. C., Austin, D. W., & Reid, D. S. (2007). Investigation of the WISC-III and WASI in Clinical Child Populations: A Framework for Further Exploration. *Canadian Journal of School Psychology*, 22(2), 249-254.
- Semel, E., Wiig, E. H., & Secord, W. A. (1995). *Clinical Evaluation of Language Fundamentals—Third Edition*. San Antonio, TX: Psychological Corporation.
- Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: an absence of spontaneous theory of mind in Asperger syndrome. *Science*, 325(5942), 883.
- Seung, H. K. (2007). Linguistic characteristics of individuals with high functioning autism and Asperger syndrome. *Clinical Linguistics & Phonetics*, 21(4), 247-259.
- Shah, A., & Frith, U. (1983). An islet of ability in autistic children: A research note. *Journal of Child Psychology and Psychiatry*, 24(4), 613-620.
- Shah, A., & Frith, U. (1993). Why do autistic individuals show superior performance on the block design task? *Journal of Child Psychology and Psychiatry*, 34(8), 1351-1364.
- Shallice, T. (1982). Specific Impairments of Planning. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 298(1089), 199-209.
- Shulman, C., Yirmiya, N., & Greenbaum, C. W. (1995). From categorization to classification: a comparison among individuals with autism, mental retardation, and normal development. *J Abnorm Psychol*, 104(4), 601-609.
- Siegler, R. S. (1977). 20 Questions Game as a Form of Problem-Solving. *Child Development*, 48(2), 395-403.
- Siller, M., & Sigman, M. (2008). Modeling longitudinal change in the language abilities of children with autism: Parent behaviors and child characteristics as predictors of change. *Developmental Psychology*, 44(6), 1691.
- Silverman, T. (1967). Categorization behavior and achievement in deaf and hearing children. *Exceptional children*, 34(4), 241.
- Simmonds, E. P. M. (1990). The Effectiveness of Two Methods for Teaching a Constraint-Seeking Questioning Strategy to Students with Learning Disabilities. [Article]. *Journal of Learning Disabilities*, 23(4).
- Sinzig, J., Morsch, D., Bruning, N., Schmidt, M. H., & Lehmkuhl, G. (2008). Inhibition, flexibility, working memory and planning in autism spectrum disorders with and without comorbid ADHD-symptoms. *Child Adolesc Psychiatry Ment Health*, 2(1), 4.
- Smith, B. J., Gardiner, J. M., & Bowler, D. M. (2007). Deficits in free recall persist in Asperger's syndrome despite training in the use of list-appropriate learning strategies. *Journal of Autism and Developmental Disorders*, 37(3), 445-454.
- Soanes, C., Stevenson, A., Pearsall, J., & Hanks, P. (2005). *Oxford dictionary of English*: Oxford University Press Toronto, Ontario, Canada.
- Solomon, M., Bauminger, N., & Rogers, S. J. (2010). Abstract Reasoning and Friendship in High Functioning Preadolescents with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*.
- Solomon, M., Goodlin-Jones, B. L., & Anders, T. F. (2004). A social adjustment enhancement intervention for high functioning autism, Asperger's syndrome, and pervasive developmental disorder NOS. *Journal of Autism and Developmental Disorders*, 34(6), 649-668.
- Soulieres, I., Dawson, M., Samson, F., Barbeau, E. B., Sahyoun, C. P., Strangman, G. E., et al. (2009). Enhanced visual processing contributes to matrix reasoning in autism. *Hum Brain Mapp*, 30(12), 4082-4107.
- Sparrow, S. S., Balla, D. A., Cicchetti, D. V., & Doll, E. A. (1984). *Vineland adaptive behavior scales: Interview edition, expanded form manual*: American Guidance Service Circle Pines, MN.
- Speirs, S., Yelland, G., Rinehart, N., & Tonge, B. (2011). Lexical processing in individuals with high-functioning autism and Asperger's disorder. *Autism*.

- Spek, A., Schatorje, T., Scholte, E., & van Berckelaer-Onnes, I. (2009). Verbal fluency in adults with high functioning autism or Asperger syndrome. *Neuropsychologia*, 47(3), 652-656.
- Stokoe, W. C. (1960/2005). Sign language structure: An outline of the visual communication systems of the American deaf. *Journal of Deaf Studies and Deaf Education*, 10(1), 3.
- Sussman, G. J. (1975). *A computer model of skill acquisition*: Elsevier Science Inc.
- Sutton-Spence, R., & Woll, B. (1999). *The linguistics of British Sign Language: an introduction*: Cambridge Univ Pr.
- Szatmari, P., Archer, L., Fisman, S., Streiner, D. L., & Wilson, F. (1995). Asperger's syndrome and autism: Differences in behavior, cognition, and adaptive functioning. *Journal of the American Academy of Child & Adolescent Psychiatry*, 34(12), 1662-1671.
- Szatmari, P., Bartolucci, G., Finlayson, A., & Krames, L. (1986). A vote for Asperger's syndrome. *Journal of Autism and Developmental Disorders*, 16(4), 515-518.
- Szatmari, P., Bryson, S., Boyle, M., Streiner, D., & Duku, E. (2003). Predictors of outcome among high functioning children with autism and Asperger syndrome. *Journal of Child Psychology and Psychiatry*, 44(4), 520-528.
- Szatmari, P., Tuff, L., Finlayson, M. A. J., & Bartolucci, G. (1990). Asperger's syndrome and autism: Neurocognitive aspects. *Journal of the American Academy of Child & Adolescent Psychiatry*, 29(1), 130-136.
- Tager-Flusberg, H. (1985a). Basic level and superordinate level categorization by autistic, mentally retarded, and normal children. *J Exp Child Psychol*, 40(3), 450-469.
- Tager-Flusberg, H. (1985b). The conceptual basis for referential word meaning in children with autism. *Child Development*, 56(5), 1167-1178.
- Tager Flusberg, H., & Joseph, R. M. (2005). How Language Facilitates the Acquisition of False-Belief Understanding in Children with Autism. In J. Astington & J. Baird (Eds.), *Why Language Matters for Theory of Mind* (pp. 298-318). New York: OUP.
- Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, 23(3), 457-482.
- Tant, J. L., & Douglas, V. I. (1982). Problem solving in hyperactive, normal, and reading-disabled boys. *J Abnorm Child Psychol*, 10(3), 285-306.
- Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in American Sign Language* (Vol. 38): Cambridge Univ Pr.
- Taylor, D., & Faust, W. L. (1952). Twenty questions: efficiency in problem solving as a function of size of group. *Journal of Experimental Psychology*, 44(5), 360-368.
- Thorndike, R., Hagen, E., & Sattler, J. (1986). *The Stanford-Binet scale of intelligence*. Riverside, Chicago.
- Thornton, S. (1982). Challenging "Early Competence": a process oriented analysis of children's classifying. *Cognitive Science*, 6(1), 77-100.
- Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The Quarterly Journal of Experimental Psychology*, 37(4), 571-590.
- Toal, F., Daly, E. M., Page, L., Deeley, Q., Hallahan, B., Bloemen, O., et al. (2010). Clinical and anatomical heterogeneity in autistic spectrum disorder: a structural MRI study. *Psychological Medicine*, 40(07), 1171-1181.
- Todman, J., & Cowdy, N. (1993). Processing of visual-action codes by deaf and hearing children: Coding orientation or M-capacity? *Intelligence*, 17(2), 237-250.
- Toichi, M., & Kamio, Y. (2001). Verbal association for simple common words in high-functioning autism. *Journal of Autism and Developmental Disorders*, 31(5), 483-490.
- Tryon, P. A., Mayes, S. D., Rhodes, R. L., & Waldo, M. (2006). Can Asperger's Disorder Be Differentiated From Autism Using DSM-IV Criteria? *Focus on Autism and Other Developmental Disabilities*, 21(1), 2.

- Tsuchida, R., Sato, T., Yamada, T., & Matsushita, N. (1989). Japanese version of the Miller assessment for preschoolers: Harcourt Brace Jovanovich, Japan.
- Turner, M. (1997). Towards an executive dysfunction account of repetitive behaviour in autism. In J. Russell (Ed.), *Autism as an executive disorder* (pp. 57–100): Oxford University Press Oxford, UK.
- Ungerer, J. A., & Sigman, M. (1987). Categorization Skills and Receptive Language-Development in Autistic-Children. *Journal of Autism and Developmental Disorders*, 17(1), 3-16.
- Upton, D., & Thompson, P. J. (1999). Twenty questions task and frontal lobe dysfunction. *Archives of Clinical Neuropsychology*, 14(2), 203-216.
- Van Horn, K. R., & Bartz, W. H. (1968). Information seeking strategies in cognitive development. [Article]. *Psychonomic Science*, 11(10), 341-342.
- Velazquez, J. L. P., Barcelo, F., Hung, Y., Leshchenko, Y., Nenadovic, V., Belkas, J., et al. (2009). Decreased brain coordinated activity in autism spectrum disorders during executive tasks: Reduced long-range synchronization in the fronto-parietal networks. *International Journal of Psychophysiology*, 73(3), 341-349.
- Vernon, M. (1967). Relationship of language to the thinking process. *Archives of General Psychiatry*, 16(3), 325.
- Vernon, M. (2005). Fifty Years of Research on the Intelligence of Deaf and Hard-of-Hearing Children: A Review of Literature and Discussion of Implications. *Journal of Deaf Studies and Deaf Education*, 10(3), 225-231.
- Verté, S., Geurts, H. M., Roeyers, H., Oosterlaan, J., & Sergeant, J. A. (2005). Executive functioning in children with autism and Tourette syndrome. *Dev Psychopathol*, 17(2), 415-445.
- Verté, S., Geurts, H. M., Roeyers, H., Oosterlaan, J., & Sergeant, J. A. (2006). Executive functioning in children with an autism spectrum disorder: Can we differentiate within the spectrum? *Journal of Autism and Developmental Disorders*, 36(3), 351-372.
- Verté, S., Geurts, H. M., Roeyers, H., Rosseel, Y., Oosterlaan, J., & Sergeant, J. A. (2006). Can the children's communication checklist differentiate autism spectrum subtypes? *Autism*, 10(3), 266-287.
- Vilkki, J. (1988). Problem solving deficits after focal cerebral lesions. *Cortex*, 24(1), 119-127.
- Villalobos, M. E., Mizuno, A., Dahl, B. C., Kemmotsu, N., & Müller, R. A. (2005). Reduced functional connectivity between V1 and inferior frontal cortex associated with visuomotor performance in autism. *Neuroimage*, 25(3), 916-925.
- Vincent, M. (1957). Sur le role du langage a un niveau elementaire de pensee abstraite. *Enfance*, 10(4), 443-464.
- Visser, M. E., Cohen, M. X., & Geurts, H. M. (In press). Brain connectivity and high functioning autism: A promising path of research that needs refined models, methodological convergence, and stronger behavioral links. *Neuroscience & Biobehavioral Reviews*(0).
- Vygotsky, L. S. (1987). *Thinking and speech. The collected works of Lev Vygotsky (Vol. 1)*. New York: Plenum Press.
- Wainwright-Sharp, J. A., & Bryson, S. E. (1993). Visual orienting deficits in high-functioning people with autism. *Journal of Autism and Developmental Disorders*, 23(1), 1-13.
- 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(12), 1735-1739.
- Ward, G., & Allport, A. (1997). Planning and problem solving using the five disc Tower of London task. *The Quarterly Journal of Experimental Psychology: Section A*, 50(1), 49-78.

- Wechsler, D. (1999). *Wechsler abbreviated scale of intelligence*. New York: Psychological Corp.
- Wechsler, D., & Corporation, P. (1991). *WISC-III: Wechsler intelligence scale for children*: Psychological Corporation, Harcourt Brace Jovanovich.
- Wechsler, D., & Corporation, P. (1997). *WAIS-III: Administration and scoring manual: Wechsler adult intelligence scale*: Psychological Corporation.
- Weeks, S. J., & Hobson, R. P. (1987). The salience of facial expression for autistic children. *J Child Psychol Psychiatry*, 28(1), 137-151.
- Welchew, D. E., Ashwin, C., Berkouk, K., Salvador, R., Suckling, J., Baron-Cohen, S., et al. (2005). Functional disconnectivity of the medial temporal lobe in Asperger's syndrome. *Biol Psychiatry*, 57(9), 991-998.
- Whitaker, S. (2008). *The merits of mental age as an additional measure of intellectual ability in the low ability range*. Paper presented at the Clinical Psychology Forum.
- White, S. J. (In Press). The Triple I Hypothesis: Taking Another ('s) Perspective on Executive Dysfunction in Autism. *Journal of Autism and Developmental Disorders*, 1-8.
- White, S. J., Burgess, P. W., & Hill, E. L. (2009). Impairments on "open-ended" executive function tests in autism. *Autism Research*, 2(3), 138-147.
- Whitehouse, A. J., Maybery, M. T., & Durkin, K. (2006). Inner speech impairments in autism. *Journal of Child Psychology and Psychiatry*, 47(8), 857-865.
- Whitehouse, A. J., Maybery, M. T., & Durkin, K. (2007). Evidence against poor semantic encoding in individuals with autism. *Autism*, 11(3), 241-254.
- WHO. (1993). *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research*: World Health Organization.
- Wicker, B., Fonlupt, P., Hubert, B., Tardif, C., Gepner, B., & Deruelle, C. (2008). Abnormal cerebral effective connectivity during explicit emotional processing in adults with autism spectrum disorder. *Soc Cogn Affect Neurosci*, 3(2), 135-143.
- Wiig, E., & Secord, W. (1989). *Test of Language Competence - Expanded Edition*. San Antonio, TX: The Psychological Corporation.
- Williams, D., 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(01), 225-239.
- Williams, D., Happé, 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(1), 51-58.
- Williams, D., & Jarrold, C. (2010). Brief report: Predicting inner speech use amongst children with autism spectrum disorder (ASD): The roles of verbal ability and cognitive profile. *Journal of Autism and Developmental Disorders*, 40(7), 907-913.
- Williams, D. L., Goldstein, G., & Minshew, N. J. (2005). Impaired memory for faces and social scenes in autism: clinical implications of memory dysfunction. *Archives of Clinical Neuropsychology*, 20(1), 1-15.
- Williams, D. L., Goldstein, G., & Minshew, N. J. (2006a). Neuropsychologic functioning in children with autism: Further evidence for disordered complex information-processing. *Child Neuropsychology*, 12(4-5), 279-298.
- Williams, D. L., Goldstein, G., & Minshew, N. J. (2006b). The profile of memory function in children with autism. *Neuropsychology*, 20(1), 21-29.
- Wilson, B. A., Evans, J. J., Alderman, N., Burgess, P. W., & Emslie, H. (1997). Behavioural assessment of the dysexecutive syndrome. *Methodology of frontal and executive function*, 239-250.
- Wilson, M., Bettger, J. G., Niculae, I., & Klima, E. S. (1997). Modality of Language Shapes Working Memory: Evidence From Digit Span and Spatial Span in ASL Signers. *Journal of Deaf Studies and Deaf Education*, 2(3), 150.

- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.
- Wing, L. (1981). Asperger's syndrome: a clinical account. *Psychological Medicine*, *11*(01), 115-129.
- Wolff, A. (1985). Analysis. In D. Martin (Ed.), *Cognition, Education and Deafness* (pp. 79-81). Washington, D.C.: Gallaudet University Press.
- Yerys, B. E., Wallace, G. L., Harrison, B., Celano, M. J., Giedd, J. N., & Kenworthy, L. E. (2009). Set-shifting in children with autism spectrum disorders: Reversal shifting deficits on the Intradimensional/Extradimensional Shift Test correlate with repetitive behaviors. *Autism*, *13*(5), 523-538.
- Yi, K., Li, D., Park, W. S., Park, K., Shim, T., Kwern, O., et al. (2011). Korean Deaf Adolescents' Awareness of Thematic and Taxonomic Relations among Ordinary Concepts Represented by Pictures and Written Words. *Journal of Deaf Studies and Deaf Education*, *16*(3), 375.
- Youniss, J., & Furth, H. (1966). Prediction of Causal Events as a Function of Transitivity and Perceptual Congruency in Hearing and Deaf Children. *Child Development*, *37*(1), 73-81.
- Yurgelun-Todd, D. (2007). Emotional and cognitive changes during adolescence. *Current Opinion in Neurobiology*, *17*(2), 251-257.
- Zelniker, T., Renan, A., Sorer, I., & Shavit, Y. (1977). Effect of Perceptual Processing Strategies on Problem Solving of Reflective and Impulsive Children. *Child Development*, *48*(4), 1436-1442.