Pragmatic Language Ability in Autism

Louise Malkin

School of Psychology

University of Kent

A thesis submitted for the degree of Ph.D. in the Faculty of Social Sciences at the University of Kent

July 2020

Declaration

I declare that this thesis is my own work carried out under the normal terms of supervision.

Louise Malkin

Publications

Chapter 3 and Chapter 4 of this thesis have been published:

Chapter 3:

L. Malkin, K. Abbot-Smith, D. Williams (2018). Is verbal reference impaired in autism spectrum disorder? A systematic review. *Autism & Developmental Language Impairments, 3*, 1-24. doi: 10.1177/2396941518763166

Chapter 4:

Malkin, L., Abbot-Smith, K., Williams, D., & Ayling, J. (2018). When do children with Autism Spectrum Disorder take common ground into account during communication?. *Autism Research, 11*(10), 1366-1375. doi: 10.1177/2396941518763166

Acknowledgements

First and foremost, I must thank my amazing supervisor, Dr Kirsten Abbot-Smith for taking a chance on a new mum who had been out of academia for over a decade. I have valued your patience, kindness, attention to detail and willingness to help me develop my own research interests at every step of this nearly five-year journey. I could not have done this without your guidance.

This research would also not have been possible without the parents, local schools and most importantly children who took part. Thank-you also to my KCDU family, especially the many under-graduate and master's students who helped with data collection. Even when it meant driving long distances or coming in at the weekend you always kept smiling.

I am grateful to the wider School of Psychology team. Thanks to Adam Britcher for the endless technical support and Frank Gasking for lending expertise with PsychoPy. Too many fellow PhD students have helped me on this journey to name them all, but Julia Landsiedel deserves a special mention. Thank-you for the chocolate, tea and friendship.

A big thank-you must go to my dear parents, especially mum for donating endless hours of unpaid childcare and emotional support. We would be lost without you. Finally, I wish to thank my husband Will for unreservedly supporting yet another of my hare-brained schemes and celebrating every success with me. I am incredibly lucky to have you in my life.

Elina and Zoe, my darling daughters, this thesis is dedicated to you. As you grow up hopefully it will be reminder (as Michelle Obama's mother said to her) that 'If it can be done, you can do it'…. and now we can enjoy some uninterrupted time together.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Difficulty with certain areas of pragmatic language, for example engaging in back and forth conversation, are part of the diagnostic criteria for autism. That said, not all autistic children display every symptom described within these criteria. In addition, pragmatic language comprises a relatively heterogeneous set of skills. Hence, an impairment may not be present across all pragmatic language abilities for autistic individuals. Where difficulty does exist, the associated cognitive underpinnings are unclear. This thesis mainly focuses on one key pragmatic language skill, *verbal reference*, that is, the ability to identify the entity that is currently of interest between interlocutors via spoken language. In Chapter 2, broader pragmatic ability in autistic children is also examined using a parent report measure (the CCC-2). The findings of chapter 3, 4 and 5 indicate that autistic children are less able than well-matched typically developing peers to *produce* verbal reference that is tailored to listener needs. Chapters 3 and 4, however, suggest aspects of verbal reference *interpretation* are intact. Finally, chapters 2, 4 and 5 indicate that pragmatic language difficulties in autism, in particular the production of verbal reference, may be more related to executive functioning than social cognition ability. In sum, pragmatic language impairment might not be uniform within all areas of pragmatic language in autism. Categorising pragmatic abilities based upon the cognitive skills they recruit may improve understanding of areas of autistic strength and weakness in this diverse skill set.

CHAPTER 1. GENERAL INTRODUCTION

1.1.Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a developmental neurological condition. Early attempts to describe and define the features of ASD highlighted difficulty in social situations and 'insistence on sameness' (Kanner, 1943). Both features remain integral to current diagnostic criteria. The American Psychiatric Association's (APA) Diagnostic Statistical Manual (DSM) forms a guide to diagnoses of mental disorders. In the latest revision of its diagnostic criteria for ASD (DSM-5, APA, 2013) impairments must be present across two domains; namely social communication / interaction *and* restricted, repetitive patterns of behaviour, interests, or activities. Three criteria comprise the domain of social communication and interaction; firstly, 'social-emotional reciprocity' includes traits such as abnormal social approach, failure of normal back-and-forth conversation, reduced sharing of interests and failure to initiate or respond to social interactions; secondly, 'nonverbal communicative behaviours for social interaction', include difficulties with appropriate eye contact, body language, gestures and facial expressions; thirdly, 'developing, maintaining, and understanding relationships' includes difficulties adjusting behaviour to a given social context, sharing imaginative play and absence of interest in peers. Impairments must be present across all three criteria for a diagnosis of ASD to be considered. The domain of Restrictive / Repetitive Behaviours, meanwhile, includes four criteria; firstly, stereotyped or repetitive motor movements, use of objects, or speech; secondly, insistence on sameness and inflexible adherence to routines; thirdly, highly restricted, fixated interests and, finally, hyper- or hypo-reactivity to sensory input and unusual interest in sensory aspects of the environment. Two of these four criteria must be met for a diagnosis of ASD.

Non-verbal IQ (NVIQ) varies significantly between autistic individuals and an uneven profile of cognitive abilities within individuals is also common (Joseph, 2011). Hence, the

DSM-5 diagnostic criteria for ASD does not include or exclude individuals based on their NVIQ. Structural language abilities (phonology, semantics, and syntax) are similarly heterogeneous in ASD, with many individuals possessing structural language impairment and a substantial sub-group having little or no language, even into adulthood (Loucas, et al. 2008; Boucher, 2012). Structural language difficulties are not unique to autism, but they remain an important area of consideration in any account of ASD (Tager-Flusberg, 2000).

Whilst NVIQ and structural language will be within typical range for some autistic individuals, one area considered to be globally impaired in ASD is the social *use* of language, or 'pragmatics' (Tager-Flusberg, Paul & Lord, 2005). Although changes in diagnostic criteria have rendered the view obsolete, at one time pragmatic impairment was even suggested as unique and specific to ASD (Tager-Flusberg 1996).

2.1. Pragmatic Language

Defining Pragmatic Language

The term 'pragmatics' comprises skills that enable successful communication and thus affect quality of life (Hatton, 1998). Despite the importance of pragmatic ability for psychosocial outcomes (Whitehouse, Watt, Line & Bishop, 2009), defining pragmatic language has proven a contentious issue (see e.g. Ariel, 2010). It is generally agreed that pragmatics involves the use of context to establish meaning within communicative exchanges (Matthews, 2014). Some theorists restrict this definition to the *linguistic* context of an utterance, that is the interpretation of language based on prior knowledge and experience (e.g. Leech 1983; Levinson, 1983; Sperber & Novek, 2004). This 'component' view identifies pragmatics as a sub-domain of language, separable from semantics and grammar, and defines a reasonably narrow set of abilities as 'pragmatic'. Skills typically included in this approach include implicature (where meaning is conveyed indirectly e.g. using the phrase 'I'm cold' to

communicate the desire for a window to be shut; Horn, 2006), presupposition (where an utterance allows assumptions to be drawn e.g. 'I have stopped eating chocolate every day' presupposes I used to eat chocolate every day; Stalnaker, 1977), speech acts (identifying the purpose of an utterance e.g. to request, promise or apologise; Searle, 1969), deixis (where the meaning of an utterance is dependent on the context e.g. 'here / there' 'I / you', Levinson, 1983) and reference (the process by which conversational partners identify the focus of attention; Yule, 1997). Pragmatic language has, thus, often been defined via a list of loosely related topics, rather than a single definition (e.g. Levinson, 1983). Presumably, this is partly due to the difficulty of defining such a complex and fluid skill set.

Other theorists include *non-linguistic* as well as linguistic forms (e.g. the use and interpretation of facial expression, gesture, eye-gaze and posture) under the umbrella of 'pragmatic' ability (e.g. Perkins, 2010). In this approach, rather than a separable component of linguistic theory, pragmatics is viewed to incorporate cognitive, social and cultural perspectives on linguistic phenomena (Morris, 1938; Verschueren, 1999). Whilst this 'perspective' view benefits clinicians supporting individuals with minimal linguistic capacity, and bolsters interdisciplinary co-operation that may aid real-life problem solving (Perkins, 2005), such a holistic approach has been criticised for being too broad to be of theoretical significance (Huang, 2014). Indeed, the field of pragmatics has been described to comprise 'a large, loose, and disorganised collection of research efforts' (Horn, 1988). This lack of clarity in defining pragmatics is a potential barrier to effective assessment, identification of need and provision of intervention.

A notable challenge in defining pragmatic language concerns the divide between 'semantics', which involves decoding the conventional 'dictionary' meaning of words, and 'pragmatics', which requires additional inference to decipher speaker meaning (Levinson, 1983; Ariel, 2010). A key pragmatic theory, 'Relevance Theory' (Sperber & Wilson, 1986)

highlights the importance of context in interpreting a speaker's meaning. Here, lexical semantics alone is insufficient to interpret the intended meaning of an utterance. Rather, the listener must search for the contextual 'relevance' of an utterance in order to decipher its meaning. Whilst dividing semantic from pragmatic requirements is theoretically useful, separating them in practice is challenging (Manor, 2001). This is highlighted by differences in categorisation of certain language abilities as 'semantic' or 'pragmatic'. To illustrate, the use and interpretation of figurative language (e.g. metaphor, similes) is considered a semantic skill by Landa (2000), but a pragmatic skill by others (e.g. Baird & Norbury, 2016; Kalandadze, Norbury, Nærland & Næss, 2018). Differences in opinion on which skills should be categorised as 'pragmatic' or 'semantic' creates clear challenges for measurement.

Aiming to delineate and categorise pragmatic abilities to aid assessment Landa (2000, 2005) describes an organisational framework where pragmatics involves tailoring and interpreting language within the social demands of the linguistic and non-linguistic context. Three domains are defined within the framework; understanding and expressing communicative intentions (including identifying whether a word, phrase or gesture is intended to greet, comment or reject), presupposition (including considering listener knowledge to provide an appropriate amount of information or including situational factors such as appearance to select correct level of formality) and discourse management (including conversation and story-telling, including topic maintenance and repairing communication breakdowns). Whilst this framework usefully outlines abilities that might be considered 'pragmatic', of equal importance is understanding the cognitive skills associated with pragmatic competence.

In a move away from pragmatic language as the application of 'rules' that govern language use (e.g. Bates, 1976) towards understanding the processes that underlie pragmatic competence, O'Neill (2012) developed a framework that seeks to organise pragmatic abilities

alongside the knowledge potentially required for each. Aiming to structure the vast range of abilities that might be considered 'pragmatic' and potential factors related to their mastery, the framework proposes three core components of pragmatic skill. Firstly, social pragmatics (requiring social knowledge) concerns understanding conventional expectations of how to interact with people or groups, for example, using polite forms when interacting with a senior or simpler language with a child. Next, mindful pragmatics (requiring social-cognitive knowledge) concerns using knowledge about mental states of *individuals* whilst in conversation with them. Skills include the development of joint attention through to understanding humour, sarcasm, and irony. Finally, cognitive pragmatics (requiring cognitive knowledge) requires consideration of situational factors not related to the conversation partner, such as the correct use of deictic terms (e.g. here and there, this and that). Both O'Neil (2012) and Landa's (2000, 2005) frameworks consider linguistic *and* non-linguistic (gesture, facial expression) means of achieving these outcomes as pragmatic skills.

While effective communication unarguably requires a vast range of skills, concerns have been expressed that pragmatics risks becoming the 'waste basket' where any skill that does not fit within other domains of communication is filed (McTear and Conti-Ramsden, 1992). Norbury (2014) suggests skills that may be influenced by societal norms such as turn taking in conversation, the use of non-verbal cues (gesture, facial expression) and humour are better labelled 'social communication' than 'pragmatic' skills. Given that the inclusion of non-verbal communication when defining pragmatic language is contentious, the empirical studies in this thesis will primarily focus on verbal communication. This does not, however, deny the importance of cognitive, social, and linguistic knowledge in achieving pragmatic competence as outlined by O'Neill (2012). We now briefly discuss the challenges associated with the measurement pragmatic language abilities.

Measuring Pragmatic Language

Alongside the difficulty of agreeing which skills should be labelled 'pragmatic', there exists the challenge of how such dynamic and context dependant skills can be measured in a parsimonious fashion (Landa, 2000, 2005; Adams, 2002). There are relatively few standardised tests of pragmatic ability (O'Neill, 2014). Those that do exist generally fall into one of several types. Firstly, direct structured assessment. The Test of Pragmatic Language-2 (TOPL-2; Phelps-Terasaki & Phelps-Gunn, 2007), for example, requires children to view pictures of common social situations and produce an appropriate response for a character in the scene. Another structured assessment, 'The Comprehensive Assessment of Spoken Language' (CASL; Carrow-Woolfolk, 1999), includes a 'pragmatic judgement' subtest that requires children to listen to vignettes then judge the appropriateness of the language used by those in the story or suggest language appropriate for the given situation. These two tests are found to correlate with structural language ability (Akbar, Loomis & Paul, 2013), bringing into question the extent to which they reflect pragmatic language ability per se. It has also been argued that performance in such structured tasks might not accurately reflect real-life pragmatic competence (Norbury, 2014).

A second type of structured assessment involves eliciting narratives, for example, based on picture books (e.g. The Bus Story, Renfrew, 1991) which are then coded for the presence of certain pragmatic features. Whilst the cognitive demands of narrative production are arguably closer to real-life speech, an effective narrative might be produced without necessarily considering listener perspective (Arnold, Bennetto and Diehl, 2009). Other measures of pragmatic ability use direct observation, for example during semi-naturalistic conversation (e.g. The Pragmatic Protocol, Prutting & Kirchner, 1987). These measures have clear advantages in terms of ecological validity but may also be more time consuming to administer and score reliably.

A final category of measurement are behavioural checklists, which are completed by an adult who know the individual well (e.g. the Children's Communication Checklist, CCC, Bishop, 2003). Such measures have the advantage of sampling real-life pragmatic language proficiency but rely on the ability of the scorer to make accurate judgements. It is likely that the most effective assessment of pragmatic ability would involve a combination of direct and indirect measurements. Finally, a global difficulty with measures of pragmatic language is that whilst they might reveal differences between individuals or clinical groups, they do little to reveal the cognitive underpinnings of such difficulties (Wilson & Bishop, 2019). This is an issue to which we return shortly.

1.3. Pragmatic Language Ability in ASD

Pragmatic language impairment is generally accepted as a hallmark of ASD (Frith, 1989; Martin & McDonald, 2003; Tager-Flusberg, Paul & Lord, 2005; Eigsti, de Marchena, Schuh & Kelley, 2011). That said, given the expanse of skills considered 'pragmatic', studies examining pragmatics in ASD cover a diverse range of abilities that are often poorly defined and measured inconsistently (Landa, 2000). This creates difficulties in synthesising the findings of multiple studies to form a coherent picture of pragmatic ability in ASD.

Reviewing the literature on the interpretation of pragmatic language in autism, Geurts, Kissine & van Tiel (2018) conclude that pragmatic impairment in this group is 'neither global nor uniform'. Whilst social communication must unarguably be impaired to a degree to meet the diagnostic criteria for ASD, the three areas within the 'social communication' domain of the DSM-5 diagnostic criteria are broadly defined. Accordingly, a deficit in social communication could be assigned based on the presence of a limited number of example behaviours, rather than every one of a finite set of criteria (Lord & Bishop, 2015). Hence an

impairment across *all* areas of social communication and pragmatic ability is not necessarily required to receive a diagnosis of autism. Whilst pragmatic impairment is a key feature of ASD, it thus remains unclear whether all domains of pragmatic ability are equally impaired. We now briefly consider the current literature on pragmatic ability in autism.

Conversation Skills

Rather than the blanket of impairment that might be anticipated, one recent review of conversational skill in ASD reported some inconsistencies in performance between areas of pragmatic ability across the eighteen studies included (Ying Sng, Carter and Stephenson, 2018). ASD group performance did show a clear pattern of impairment in some areas of conversation; *topic management* e.g. difficulty extending a topic / offering new and relevant information (Capps, Kehres & Sigman, 1998) and appropriately shifting to a new topic (Bauminger-Zviely, Karin, Kimhi & Agam-Ben-Artzi, 2014); *initiation* e.g. initiating an exchange less frequently or being more likely to ignore bids for interaction (Jones & Schwartz, 2009) and *commenting* e.g. making 'bizarre', irrelevant or overly detailed comments (Capps, Kehres & Sigman, 1998; Paul, Orlovski, Marcinko, & Volkmar,2009). In contrast, ASD and TD groups appeared equally able to take conversational turns (Adams, Green, Gilchrist, & Cox, 2002; Bauminger-Zviely et al., 2014) and recognise communication breakdowns and the need for conversational repair (Capps et al. 1998, Volden, 2004). In the turn-taking studies, however, the ASD group responded less frequently overall. Thus, seemingly appropriate turn taking behaviour in the ASD group may have been driven by a desire to minimise interaction, rather than an ability to appropriately 'pass the floor' to a conversational partner. Unfortunately, as the quality of conversational turns (i.e. what the

autistic individuals said during their conversational 'turn') was not reported in these studies, it is not possible to assess the likelihood of this explanation.

The conclusions drawn from the communicative repair studies were similarly problematic. Whilst Volden (2004) noted that initial requests for clarification received similar responses across groups, as further clarification requests were made, ASD group responses became increasingly bizarre e.g:

Adult: What do you do with those toys?

Child: I put things into the yellow and red to make another

Adult: What?

Child: To make a people to live

Regarding requests for clarification, Capps et al. (1998) reported a similar number of clarification requests were made by ASD and TD children. As the number of clarification requests per group were not reported, however, it is unclear whether the lack of difference was driven by an appropriate frequency of clarification requests in the ASD group, or by low requests for clarification from both groups. It is therefore unclear if a lack of ASD impairment in these skills can confidently be concluded.

Figurative Language

A more consistent picture of impairment was reported in a review of twenty studies examining non-literal (figurative) language comprehension and inference in ASD (Loukusa & Moilanen, 2009). Inference is viewed by many to represent a key aspect of pragmatic comprehension as it requires the use of context to deduce meaning (e.g. Sperber & Wilson, 2002; Wilson & Bishop, 2019). Autistic children were less able to use context to answer questions about visual scenes (Loukusa, et al. 2007), to make inferences from social scripts,

understand metaphors and produce appropriate speech acts (Dennis, Lazenby, & Lockyer, 2001) or make inferences to aid story comprehension (Norbury & Bishop, 2002). Idiom comprehension was also found reduced (Kerbel & Grunwell, 1998). Difficulties were too noted in explaining non-literal utterances, such as jokes, sarcasm, and persuasion (Happe, 1994). The ability to detect irony based on contextual information was also reduced in ASD, though the use of prosodic cues to detect irony appeared intact (Wang, Lee, Sigman, & Dapretto, 2006).

Despite this pattern of impairment, Loukusa & Moilanen, (2009) reported a reduction rather than absence of non-literal language comprehension and inference ability in ASD. Many of the included studies measured pragmatic ability using tasks that arguably require structural (syntax and grammar) language ability. Whilst the inclusion criteria for their review required studies to include ASD participants with intelligence 'in normal range', it was not always clear whether groups were matched on structural language ability. Moreover, the vocabulary based verbal IQ measures favoured by many of the studies might problematically overestimate language ability in individuals with ASD (Mottron, 2004). Finally, few studies measured language *comprehension* which is likely of importance for inference. Hence, reduced non-literal language comprehension and inference in the ASD group may have been related to reduced structural language ability, rather than an autism specific deficit. Supporting this possibility, in a meta-analysis examining the comprehension of figurative language (metaphor and verbal irony) in ASD, once language ability was accounted for, no significant between-group differences remained (Kalandadze, Norbury, Nærland & Næss, 2018). This suggests that, at the very least, difficulties understanding figurative language are not necessarily unique to ASD, but rather relate to difficulties in structural language that could be experienced by any clinical population.

Scalar Implicature

Scalar implicature relates to the ability to make inferences based on logical terms relating to quantity such as 'some' and 'or' (Horn, 2006). For example, the statement 'I ate *some* of the sweets' may be understood as 'I did not eat *all* of the sweets, whilst the statement 'Wendy *or* Sue cooked a lasagne' suggests they did not *both* cook a lasagne. Interestingly, several studies have suggested that autistic children and adolescents are as likely as well-matched typically developing peers to correctly derive scalar implicatures (Pijnacker, Hagoort, Buitelaar, Teunisse, & Geurts, 2009; Chevallier, Wilson, Happé, & Noveck, 2010; Andrés-Roqueta & Katsos, 2020). The potential reasons for this apparent pocket of pragmatic ability in autistic individuals are explored in section 6.3 of the general discussion chapter.

Referential Communication - a key pragmatic skill

Across approaches and definitions, one skill consistently considered 'pragmatic' is the ability to provide enough information to allow a listener to accurately identify which entity is the current focus of attention in a conversation, without providing redundant information. This skill is defined as 'referential communication'; to avoid communication breakdown the listener must be able to identify to what the speaker refers (Yule, 1997). This ability has clear links with 'The Co-operative Principle' (Grice, 1975), which is an influential pragmatic theory seeking to describe how interlocutors achieve successful conversation. The Co-operative Principle proposes that speakers conform to a set of four conversational maxims. First, 'quantity' i.e. utterances should be as informative as required, second, 'quality' i.e. utterances should be truthful; third, 'relation' i.e. utterances should be relevant to the current context; and fourth, 'manner' i.e. utterances should be clear, brief and orderly. Appropriate referential communication contributes vitally to the ability to achieve these aims of producing clear and coherent utterances during conversation (Clark & Brennan, 1991).

Unlike skills that might not typically be expected in early childhood such as understanding of irony and sarcasm (Glenwright & Pexman, 2010), referential communication is of practical importance from the moment children begin communicating. The accurate use of a referring term enables a child, for example, to gain access to specific item; 'Want *teddy*' (rather than 'Want *toy*') in a context where there is a toy teddy and a toy doll available or 'I want the *green* apple' when there is a red and green apple available. Referential communication ability is also a key component of successful conversation. It is used, for example, to accurately identify the focus of conversation 'I went to visit *Aunt* Mary' when a friend at school is also called Mary, thus preventing communicative breakdown. Aside from being a key pragmatic ability, referential communication is a skill that can be experimentally isolated, controlled and measured with relative ease (Yule, 1997). For these reasons, examining referential communication has provided useful insights into pragmatic development in both typical and atypical development (see e.g. Graf & Davis, 2014). Referential communication ability is, therefore, the focus of most studies in this thesis.

Referential Communication in ASD

Similar to research examining conversational skills in autism, studies examining referential communication in this population are relatively scant. Across the approximately twenty studies to examine referential communication in autism, findings regarding the presence and nature of an impairment in this population is mixed, and the reasons for the discrepancy between studies not immediately evident. Given the important role that referential communication plays in facilitating effective communication, better understanding this ability in autism represents a key puzzle piece in advancing current understanding of

pragmatic ability in this population. Hence, Chapter 3 of this thesis presents a systematic literature review of verbal reference ability in autism.

Pragmatic Language and ASD Summary

The term 'pragmatics' is not easily defined and encompasses a potential multitude of component skills. Whilst social communication forms part of the diagnostic criteria for ASD, the extent to which pragmatic language impairment is universal in autism is unclear. Firstly, a question remains regarding whether all areas of pragmatic ability are equally affected in ASD. A second question concerns whether all abilities within individual areas of pragmatic ability are uniformly reduced, or if some pockets of ability may exist. Equally pressing for the development of effective interventions, however, is better understanding the factors that contribute to pragmatic difficulty in ASD. There is a clear need to elucidate the cognitive factors that contribute to pragmatic difficulties in this population (Matthews, Biney, Abbot-Smith, 2018).

1.4. Potential Underpinnings of Pragmatic Difficulties in Autism

Whilst the factors contributing to pragmatic difficulty in autism are unclear, some of the difficulties associated to ASD might offer insight regarding their cognitive basis.

Social Cognition

Social cognition concerns how cognitive processes such as perception, attention and memory relate to behaviour in a social context (Frith, 2008). This includes the social interaction and communication processes required to understand the effect of one person on another during dyadic interaction (Frith & Blakemore, 2006). A vast array of social abilities including social motivation, emotion recognition, empathy, social attention, and theory of mind have been named components of social cognition (Happe, Cook & Bird, 2017). Many of these skills could relate to pragmatic ability, for example, selecting an appropriately

informative statement may require consideration of the listener's knowledge state, whilst selecting an appropriate conversational topic might require an ability to monitor a listener interest. Given that individuals with autism are recognised as having difficulties related to social cognition (see e.g. Happe & Frith, 2014), this could provide one possible explanation for the pragmatic impairments observed in this population. Here, three areas of social cognition and their potential contributions to pragmatic difficulties in autism are examined, namely, theory of mind, shared intentionality, and common ground.

Theory of Mind

Theory of Mind (ToM), also known as mindreading, is the ability to represent the mental states of others, including understanding another's intentions, thoughts, desires, and beliefs (Baron-Cohen, 1999). A division is proposed between first-order ToM (the understanding that different people can have different thoughts about the same situation) and second-order ToM (understanding embedded mental states, e.g., what John thinks that Mary thinks) (Baron-Cohen, 2001). Given that effective communication requires consideration of conversational partner knowledge, interests and intentions, an inability in either area of ToM could conceivably restrict social communication (Astington & Jenkins, 1995). Sperber and Wilson (2002) even suggest pragmatic language might be best viewed as a sub-component of mindreading.

Difficulties with ToM have long been associated with autism (Baron-Cohen, Leslie & Frith, 1986; Frith, Morton & Leslie, 1991; Happe, 1994; Tager-Flusberg, 1996). A deficit in ToM might therefore offer a parsimonious explanation of pragmatic language difficulties in autism. However, the universality of ToM impairment in ASD has been questioned. A popular measurement of first-order ToM is the Sally-Anne task (Baron-Cohen, Leslie, & Frith, 1985). In this task, the participant is introduced to two characters 'Sally' and 'Anne'.

Sally hides an item (e.g. a marble) in one location (e.g. a box), then leaves the setting. During her absence, Anne moves the item from the original location to a new one. Sally then returns to the setting and the participant is asked where Sally will look for the item she left behind. This task measures the ability to identify that Sally will falsely believe that the item is in its original location, crucially recognising that Sally's knowledge differs from their own. Hence, these tasks are often referred to as 'false belief' tasks. Typically developing children generally pass this task by 4-years of age. While some autistic individuals are found to fail this task, leading to the conclusion that ToM is lacking in ASD, others complete it successfully (Happe, 1993).

One potential explanation for autistic individual's success in the Sally-Anne task is that, rather than using mentalising abilities, they draw on their language skills to 'hack out' the correct response. Indeed, performance in false belief tasks is related to better structural language ability in autistic children (Happe, 1995). Thus, passing a false belief task is not necessarily an indication of intact ToM capacity. In a version of the Sally-Anne task, Senju, Southgate, White and Frith (2009) measured the eye movements of adults with Asperger's syndrome and found that their attentional focus did not follow that of the actor's false belief. The same group passed two standard false belief tasks. Senju et al. (2009) concluded that mental states were not attributed spontaneously, but rather through compensatory mechanisms.

Autistic individuals also demonstrate difficulty with more complex ToM tasks and those where 'real-life' reasoning about mental states is required. In the 'animations task' (Heider & Simmel, 1944) autistic adults were less likely to attribute social meaning when creating narratives of geometric shapes enacting a social plot (Klin, 2000). For example, assigning fewer cognitive (knowledge, belief, desires) and affective (anger, jealousy, fear) mental state terms to shapes than a well matched typically developing group. Similarly, in the 'Strange

Stories' task (Happe, 1994) which requires the intentions of individuals to be surmised based on short vignettes, autistic adults were less able to provide context appropriate mental state explanations for the utterances of story characters. In conclusion, when ToM measures broader than standard false belief tasks are employed, a deficit in ToM in autism is apparent.

A deficit in ToM might then provide a viable explanation for the difficulty autistic individuals face with pragmatic language. Indeed, there is some experimental evidence that ToM and certain areas of pragmatic ability are related in autism e.g. inference (Le Sourn-Bissaoui, Caillies, Gierski & Motte, 2009) and discourse contingency (Hale and Tager-Flusberg, 2005), though the direction of the relationship is unclear (Tager-Flusberg, 2000). Problematically, however, some studies find no significant relationship between ToM and pragmatic ability in autism once structural language ability is accounted for (e.g. Capps et al. 1998; Pellicano, 2013). Further, ToM ability alone was not found to sufficiently account for the pragmatic abilities of typically developing children, namely the ability to recognise and repair communicative failures (Bosco & Gabbatore, 2017). Such findings suggest that, whilst ToM may play a role in pragmatic ability, it may not be solely responsible for communicative success. A recent review of the relationship between pragmatic and ToM ability in typical and atypical populations concluded that, while they are related, pragmatic ability is separable from and influenced by abilities in addition to ToM (Bosco, Tirassa & Gabbatore, 2018). Hence, whilst a deficit in ToM presents a potentially parsimonious explanation for the difficulties experienced in pragmatic ability in ASD, the nature of the relationship, whether the relationship holds across all pragmatic abilities independent of structural language ability, and the role of other cognitive abilities, remains unclear.

Shared Intentionality

Shared intentionality refers to collaborative interactions in which at least two individuals have the desire and ability to 'act as we' in coordinating their actions to pursue a shared goal (Tomasello, Carpenter, Call, Behne & Moll, 2005). To illustrate, two individuals agree to work together to attach a painting the wall; one partner holds the painting and the other hammers in the nails. To be successful, both partners must recall their shared goal and continue to fulfil their role until it is achieved. Engaging in conversation might similarly be considered to require shared intentionality (Tomasello, 2018). Conversational partners work together, for example, to construct the topic of conversation, with each partner building on and expanding the previous utterance. If difficulties in forming shared intentions are evident in autism, this might then provide a plausible candidate for pragmatic language difficulties in this population.

In some respects, shared intentionality appears to overlap with ToM ability; both arguably require consideration of another's mental state. That said, the formation of shared intentions may require skills other than, or in addition to, mind reading. The ability to understand another's goals and intentions (Tomasello & Hamann, 2012), the ability to jointly commit to engage in an activity with another (Gilbert, 2009), and the motivation to engage in both of these (Michael, Sebanz & Knoblich, 2016; Michael & Székely, 2018) have all been named key skills in this regard. While children with autism may be able to understand another's goals (Carpenter, Pennington & Rogers, 2002; Hamilton, 2012) they are less likely to form *shared* goals with others (Liebal, Colombi, Rogers, Warneken & Tomasello, 2007). They additionally appear to possess reduced desire to share intentions (Colombi et al., 2009) and reduced social motivation more generally (Chevallier, Kohls, Troiani, Brodkin, & Schultz, 2012; although see also Jaswal & Akhtar, 2018 for a competing account). Each of these factors might result in a failure or inability to form shared intentions, and thus impact pragmatic ability in autism.

Shared intentionality is also proposed to require acting together with others to attend to things, that is, to engage in 'joint attention' (Tomasello, 2018). Joint attention requires social partners coordinate attention in order to share an awareness of objects or events (Mundy et al., 1986). Conversation has been described to involve 'joint attention to mental states' (O'Madagain & Tomasello, 2019). Joint attention is notably reduced in autism (Dawson et al., 2004; Colombi et al., 2009) and this might affect the ability to form shared intentions (among a range of other skills). Hence, pragmatic difficulties observed in autism might be due at least in part to a failure to form shared intentions. That early joint attention predicts later performance in false-belief tasks (Brink, Lane, Wellman, 2015; Sodian & Kristen-Antonow, 2015) suggests shared intentionality might even underpin certain aspects of ToM.

Whilst the ability to form a shared intention may be a key component of communicative success, pragmatic language likely requires wider consideration of one's communicative partners perspective than intention sharing or recognising false beliefs. Successful pragmatic ability also requires conversational partners to consider the knowledge they share.

Common Ground

Common ground is knowledge shared between individuals, where each communicative partner is aware of the knowledge state of the other (Clark, 1996). Clark and Bangerter (2004) identify two broad types of common ground. Firstly, 'communal common ground' regards the general assumptions that can be made about the knowledge of a population (e.g. assuming an English adult would correctly identify the referent 'Big Ben'). Secondly, 'personal common ground' encompasses perceptual (or visual) common ground, that is, the immediate surroundings in which the exchange takes place (e.g. both partners can

see a vase on the table) and linguistic (or social) common ground, that is, the information that has been exchanged between conversational partners (e.g. once a partner states their brother is called Paul, they might refer to him as 'Paul', or with the pronoun 'he' later in the conversation).

The link between common ground and pragmatic ability appears relatively intuitive. Considering the knowledge shared with a listener could enable utterances to be both appropriately produced and interpreted (Clark & Marshall, 1981). Use of social common ground might enable speakers to adhere to Grice's maxim of quantity (providing neither too little nor too much information when producing an utterance). For example, when addressing an unfamiliar partner more information might be required; 'We went on holiday to *my parent's villa in Spain*' versus a close friend 'we went to *the villa*'. Meanwhile, the phrase 'Please pass *the glass*' might require speaker or listener to consider that the speaker has visual access to the water glass on the table, but not the wine glass in the cupboard (i.e. take into account *visual* common ground). We will now consider the evidence that these abilities are impaired in autism and thus might contribute to pragmatic difficulties.

Visual perspective taking can require understanding what lies within someone else's line-of-sight (level 1 VPT) or mentally adopting someone else's spatial viewpoint (level 2 VPT) (Flavell, Green, Flavell, Watson & Campione, 1986). Level 2 VPT is considered more challenging that level 1 VPT, as it recruits both spatial and social skills (Pearson, Marsh, Roper & Hamilton, 2016). Whilst level 1 VPT is generally accepted to be intact in autism (e.g. Baron-Cohen, 1989), there is evidence of difficulty with level 2 VPT that is unrelated to spatial rotation ability (Hamilton, Brindley & Frith, 2009). Strengthening the argument for an impairment, those studies that report intact VPT2 in autism have suffered methodological failings, for example, ceiling effects masking group differences (e.g. Tan & Harris, 1991) or a lack of matching for cognitive ability across TD and ASD groups (Hobson, 1984).

Moreover, when autistic children do succeed in level 2 VPT tasks, there is evidence this is achieved via different routes than TD children (Pearson, Marsh, Roper & Hamilton, 2016). In sum, this suggests impaired, or at least atypical, level 2 VPT in autism. This discrepancy in ability between level 1 and 2 VPT in autism might be explained by the need for mentalising in the latter, but not the former (Hamilton, Brindley & Frith, 2009). That said, there have been failings in tasks that nominally require only level 1 VPT.

Despite consensus that level 1 VPT is intact in autism, performance on some pragmatic tasks requiring this skill have shown reduced performance relative to TD controls. An example is the 'Director Task' (Krauss & Glucksberg, 1977). In this task an object must be accurately described or identified based on the array of items visible to one's conversational partner. This involves level 1 VPT as assessment of what one's communicative partner can or cannot see is required. Whilst autistic individuals accurately comprehend referring expressions during this task (e.g. Begeer, Malle, Nieuwland and Keysar, 2010) they make more errors relative to TD controls in the production of appropriately informative referring expressions (Nadig, Vivanti and Ozonoff, 2009; Fukumura, 2016). That level 1 VPT is found intact in autism suggests difficulties in these tasks are not caused by an inability to take the visual perspective of another during the task per se. One candidate for the difficulties in this task is the potential need for ToM. The director task has, in some cases, been utilised as a measure of ToM ability (e.g. Begeer, et al., 2010). A debate exists, however, regarding whether mentalising is required to succeed in the director task (Santiesteban, Shah, White, Bird and Heyes, 2015; Rubio-Fernández, 2017). It is also unclear why a failure in ToM might affect production but not comprehension in this task.

Whilst VPT may be relevant for the formation of Clark's (1996) perceptual common ground, not all pragmatic abilities require VPT. Arguably, it is more likely that conversation

will consist mainly of common ground that is established via linguistic rather than visual co-presence. For successful conversation, whilst conversational partners might sometimes build mutual knowledge based on what was is perceptually shared, it is more likely that knowledge of our shared interactive experiences plays a pivotal role in conversational success (i.e. social common ground).

In the sense that social common ground is formed via joint experiences and requires interaction between partners (Clark & Marshall, 1981; Clark, 1996), whether any study to date has adequately manipulated *social* common ground in autism is arguable. De Marchena & Eigsti (2016), for example, had autistic and non-autistic adolescents recount the events of a cartoon to a researcher who was either present or absent during the original viewing. Some awareness of narrative adaptation, reportedly based on common ground, was evident in the autistic group. However, as the researcher merely sat next to the participant, this common ground was arguably *visual* rather than *social* in nature. Further, Clark (1996) describes the formation of perceptual (visual) common ground to require a gestural indication (e.g. pointing), partner activity (e.g. as a communicative partner is looking a painting commenting 'That is by Picasso') or salient perceptual event (e.g. uttering 'What was that?' upon hearing a fire alarm). As the participant and researcher sat side by side, but did not interact during the cartoon, the criteria for 'perceptual co-presence' might also be considered absent in this study. Moreover, no study has attempted to examine the ability to use *social* common ground in the *comprehension* of utterances in autism. Given the potential for pragmatic difficulties in autism to be related to a failure to form or use social common ground, better understanding these processes in this population is clearly a priority. This is explored in Chapter 4.

Given that common ground formation requires consideration of shared knowledge states, it is unsurprising that parallels with ToM ability have been drawn. Whilst the potential overlap between these two abilities has drawn theoretical interest, the nature of the

relationship, and the extent to which the two abilities are separable, is unclear. The ability to establish common ground knowledge is suggested to be related to the development of ToM (Apperly, 2018), with both abilities recruiting perspective taking ability. Some theorists consider the ability to identify another's beliefs key to developing common ground understanding (Clark, 1996; Lee, 2001), whilst others suggest common ground understanding enables the development of perspective taking (Brown-Schmidt & Duff, 2016). Bohn & Köymen (2018) argue common ground understanding might be gained via mechanisms other than mindreading. Namely, direct social engagement (e.g. engaging in a ball game with a play partner) may be sufficient to arrive at common ground understanding (selecting the ball previously played with when a request for 'the ball' is made) without the need for mindreading. Similarly, Moore (2014) proposes that interactions might be interpreted via monitoring behavioural expressions (e.g. facial expression, gaze, attention) rather than consideration of mental states per se. This suggests there may be some benefit to considering ToM and common ground understanding as separate abilities.

Summary: Social Cognition and Pragmatics in Autism

There are robust theoretical reasons why deficits in the social cognition skills of ToM, shared intentionality and common ground might impact pragmatic abilities. In addition, there is compelling evidence that these skills may be diminished in autism. They are therefore strong candidates to explain pragmatic difficulties in autism. That said, there are instances where these skills appear at least superficially intact and yet difficulties in pragmatic ability remain, and relationships have not always been found between certain aspects of social cognition and pragmatics. It is therefore possible that other cognitive mechanisms influence pragmatic difficulties in autism independently, or in tandem with, social cognition ability.

Executive Functioning

Executive functions (EF) are a group of cognitive abilities that together support goal-directed behaviour (Diamond, 2013). Three separable (yet moderately correlated) functions are proposed to comprise EF ability (Miyake et al. 2000). Firstly, *mental set shifting* (also referred to as cognitive flexibility) concerns shifting back and forth between multiple tasks or mental sets (Monsell, 1996). Secondly, *information updating and monitoring* (closely linked to working memory) is the process whereby incoming 'new' information is incorporated into, or replaces, current or 'old' information (Morris & Jones, 1990). Finally, *inhibition* relates to the ability to suppress a more dominant or automatic response (Stroop, 1935).

All three areas of EF might aid pragmatic proficiency. Cognitive flexibility, for example, could enable a speaker to switch between their own and another's perspective, alter how they label entities within a conversation or move between conversation topics. Updating (or working memory), meanwhile, might enable a listener to simultaneously incorporate information they receive from a conversation partner as they formulate an appropriate response. Finally, inhibition could allow conversational partners to supress their own perspective in order to consider the knowledge state of their partner, or to avoid talking exclusively about their own interests. Indeed, relationships between EFs and pragmatic ability have been observed in typical development. Inhibition skills have been linked with decreased talkativeness (i.e., providing an appropriate amount of information) (Blain-Brière, Bouchard & Bigra, 2014) and better comprehension of referring expressions (Nilsen & Graham, 2009). Better working memory has been found to relate to the production of more fluid utterances and contingent conversational responses (Blain-Brière et al. 2014), whilst both inhibition and working memory ability related to the production of more appropriately informative referring statements in a referential communication task (Nilsen, Varghese, Xu, & Fecica, 2015). In a similar referential communication task, children with better cognitive

flexibility repaired statements more efficiently following feedback (Bacso & Nilsen, 2017). It is therefore plausible that difficulties in EF could result in poorer pragmatic language ability.

Despite some heterogeneity in EF performance in autism (Geurts, Sinzig, Booth & Happé, 2014) EF deficits are widely reported in this population (Lai et al. 2016; Demetriou et al. 2018). Given mounting evidence for the link between EF and pragmatic language in typical development, impaired EF presents a promising explanation for pragmatic difficulties in autism. Several studies report a relationship between EF and pragmatic ability in this group. It should be noted, however, that some of the measures of pragmatic ability utilised in these studies contain a broader range of social communication skills alongside more 'pure' pragmatic language abilities. In autistic children, for example, better working memory was related to improved comprehension in a discourse task (Schuh, Eigsti & Mirman, 2016) and fewer difficulties with expressive and receptive communication (Pugliese et al. 2015), as measured by the communication sub-scale of The Vineland Adaptive Behaviour Scales (VABS; Sparrow, Cicchetti & Balla, 2005), a parental report. Direct measures of dual attention and working memory were also related to social reciprocity and social interaction respectively, based on scores from the Autism Diagnostic Observation Schedule (ADOS; Lord et al. 1999) and Autism Diagnostic Interview (ADI-R; Lord et al.1994) (Kenworthy, Black, Harrison, Della Rosa, and Wallace, 2009). Utilising parent reports, relationships have been reported between both inhibition and conversation abilities (Hutchison, Müller & Iarocci, 2019) and cognitive flexibility and the socialisation subscale of the VAB (Pugliese et al. 2015), which includes items relating to interpersonal relationships.

Whilst associations are apparent in autism, none of these studies have identified relationships between *all* aspects of EF and pragmatics. Rather, relationships between specific areas of EF and pragmatic ability appear somewhat unpredictable. Akbar, Loomis &

Paul (2013), for example, found a relationship between working memory and the pragmatic judgement subtest of the Comprehensive Assessment of Spoken Language (CASL; Carrow-Woolfolk, 1999), but none between inhibition or cognitive flexibility and the same measure of pragmatic ability. Joseph and Tager-Flusberg (2004) meanwhile failed to find a relationship between any individual component of EF and communication / social interaction scores on the ADOS, although a task that required multiple cognitive operations (The NEPSY Tower, Korkman, Kirk & Kemp, 1998) did relate to communication scores. The variety of measurements used for both EF and pragmatic ability alongside the range of ages in these studies may be responsible for the disparity in findings.

Adding further complexity to the link between EF and pragmatic ability, there exists a strong association between EF and ToM in typical development (Hughes & Ensor, 2007) and autism (Pellicano, 2007). Regarding the direction of this relationship, longitudinal studies have indicated EF predicts ToM overtime in both typical development (Doenyas, Yavuz, & Selcuk., 2018; Lecce, Bianco, Devine & Hughes, 2017) and autism (Pellicano, 2010). Hence, EF may contribute to the development of ToM. Given that both EF and ToM have been reported to associate with pragmatic language (and social communication more broadly) the nature of these relationships has received some research interest. In autistic children, EF but not ToM predicted scores on the ADOS at both a 3-year (Pellicano et al. 2013) and 12-year follow-up (Kenny, Cribb & Pellicano, 2019). That is, whilst ToM and EF were both related to future social communication ability, only EF contributed unique variance once verbal and non-verbal ability were accounted for. ToM was, however, measured exclusively by false belief tasks in this study which are notoriously related to structural language ability in autism (Happé, 1995). Jones et al. (2018), meanwhile, used a wider range of ToM measurements in a cross-sectional study of adolescents with autism and found, whilst ToM and EF skills were significantly associated, only ToM (not EF) directly associated with social communication

ability. Jones et al. (2018) suggest EF might have an indirect effect on social communication via ToM. This notion fits Nilsen & Fecica's (2011) model of communicative perspective taking, where EF enables the *use* of mentalising skills in communication.

Finally, of note is the finding that structural and pragmatic ability predicted working memory in autistic children (Akbar, Loomis & Paul, 2013). This may be due to the need for inner language strategies such as verbal rehearsal in these tasks (Russell, Jarrold & Hood, 1999). This raises further questions regarding the nature of the relationship between language (including pragmatic language) and certain aspects of EF. These findings highlight a key difficulty with correlational designs; they do not indicate the causal direction of relationships between variables. In combination with the possibility that ToM may also be related to EF, this makes the relationship between EF and pragmatic language even harder to interpret. Experimentally manipulating variables may reveal more about the specific role of EF in pragmatic language.

In summary, whether a global relationship between EF and pragmatic ability exists, and the nature of this relationship, is far from clear. One possibility not fully explored in the literature to date is that discrete EF functions may impact specific areas of pragmatic ability in autism. Examining this possibility via the experimental manipulation of EF demands within pragmatic tasks, rather that via correlational designs, will importantly clarify the nature of any effects observed. Identifying whether specific areas of EF influence pragmatic language ability in this way is pertinent to intervention selection (Berger, Aerts, Spaendonck, Cools & Teunisse, 2003), and thus a research priority.

Structural Language

Structural language (grammar and vocabulary) has been considered relatively spared in autistic individuals who acquire speech (e.g. Tager-Flusberg, 1981). There is, however,

mounting evidence that structural language difficulties are frequent in autism (Boucher, 2012). These deficits might contribute to, or moderate, pragmatic difficulties in this population (Volden, Coolican, Garon, White & Bryson, 2009). In fact, a recent review of the cognitive abilities that associate with children's pragmatic ability in typical and atypical populations found the most strong and consistent relationship with structural language (Matthews et al., 2018).

Regarding the nature of this relationship between structural and pragmatic language, there is evidence that structural language ability might mediate ToM performance (Milligan, Astington & Dack, 2007) and EF skill (Akbar, Loomis & Paul, 2013; Botting et al. 2017). Given that ToM and EF may both be related to pragmatic language ability, it is possible that the relationship between pragmatic and structural language is indirect. Whilst it therefore seems reasonably uncontentious to expect structural language to impact on pragmatic ability, whether it exerts an influence above and beyond other factors that might be considered 'autism specific' remains unclear. Despite uncertainty regarding the precise nature of the relationship between these measures, given that structural language proficiency contributes to pragmatic language, ToM and EF, it should at the very least be taken into account when assessing the nature of relationships between these abilities.

Finally, it should be noted that the direction of the relationship between structural and pragmatic language development is unclear. That is, better structural language ability may enable better pragmatic language skills to develop (Zufferey, 2016). Meanwhile, skills such as joint attention that are viewed by some as early pragmatic language abilities may enable the development of structural language (Clark, 2014). Whilst the precise developmental nature of these two skills is uncertain, it remains important to consider structural language ability when examining pragmatic language skill.

1.5. Summary

Evidence of a broad impairment in pragmatic ability in ASD is plentiful, yet there is some indication that the global impairment might not be as 'global' as it first appears. A first question unanswered by the existing literature is whether difficulties are equal across and within all areas of pragmatic skill for autistic children. The systematic literature review in Chapter 3 of this thesis explores this question for one sub-set of pragmatic skill, namely referential communication. A second question remains regarding the factors that contribute to the reported difficulties in pragmatics reported in ASD. The empirical studies in this thesis (Chapter 2, 4 & 5) explore several candidate abilities in autistic children with language and non-verbal IQ in the typical range; the understanding of social common ground, theory of mind, structural language proficiency, and finally, an aspect of executive functioning; cognitive flexibility. The findings reported within this thesis suggest that there may be some pockets of ability within pragmatic language for autistic individuals, and, where difficulties exist, these may be more related to failures in executive functioning than social cognition.

CHAPTER 2: THE RELATIONSHIP BETWEEN PRAGMATIC LANGUAGE AND THEORY OF MIND ABILITY IN CHILDREN WITH AUTISM

2.1. Introduction

The relationship between autism spectrum disorder (ASD) and pragmatic language difficulties are well documented (Landa, 2000), yet the mechanisms related to pragmatic proficiency remain unclear across populations (Matthews, Biney & Abbot-Smith, 2018). One key ability linked with pragmatic development and recognised as a core difficulty in ASD is theory of mind (ToM) (Baron-Cohen, Leslie & Frith, 1986; Martin & McDonald, 2003). Defined as the ability to attribute mental states to oneself and others (Premack & Woodruff, 1978), there are sound theoretical reasons why ToM abilities may be linked to pragmatic performance (Baron-Cohen, 1988). The scope of abilities defined as 'pragmatic', however, is wide and it is unclear whether ToM is related to all evenly in ASD. This is an important consideration when developing pragmatic language interventions for this group.

Despite ample speculation on the relationship between ToM and pragmatics in ASD, relatively few studies have measured both abilities. Those that do so, generally examine only single areas of pragmatic ability. Significant relationships have been reported, for example, between inferencing ability and second-order ToM (Le Sourn-Bissaoui, Caillies, Gierski & Motte, 2009) and the production of referring terms and first order ToM (Dahlgren & Dahlgren Sandberg, 2008). In the latter study referential communication task performance correlated significantly with structural language ability and non-verbal IQ, which were not controlled for. In a referential communication task requiring comprehension, Schuh, Eigsti, and Mirman (2016) analysed eye tracking data and found ToM (as measured by the NEPSY-

II ToM subtest) accounted for significant variance in the speed at which the correct referent was fixated.

Several studies have explored the potential role of ToM in the interpretation of non-literal language in ASD. Measuring ToM with the 'Reading the Mind in the Eyes' (RTME; Baron-Cohen, Jolliffe, Mortimore & Robertson, 1997) and 'Strange Stories' tasks (Happe, 1994), Whyte, Nelson & Scherf (2014) found a significant relationship with idiom comprehension, independent of language ability. RTME has also been found to significantly relate to the non-literal language sub-test of the Comprehensive Assessment of Spoken Language (CASL) (which includes items related to metaphor, sarcasm, indirect requests), but surprisingly not the pragmatic judgement sub-test (Whyte & Nelson, 2015). Examining metaphor and sarcasm individually, Adachi et al. (2004) found that performance on a first order ToM task was significantly related to interpretation of sarcasm but not metaphor, which was related instead to age and IQ. Norbury (2005) similarly found factors other than first and second order ToM predicted metaphor interpretation, in this case, semantic knowledge.

A mixed picture also emerges when considering the role of ToM in maintaining a topic of conversation and adding relevant, new information, known as 'discourse contingency'. Hale & Tager-Flusberg (2005) coded the natural conversation of 4-13-year-olds with ASD for contingent utterances and administered a battery of tasks designed to span a developmental range of ToM. This measure of ToM remained a significant predictor of discourse contingency independent of age, structural language ability and IQ. In an earlier study, Capps, Kehres and Sigman (1998) found a relationship between discourse contingency in children with ASD and first order ToM was no longer significant once structural language ability was accounted for. Hale & Tager-Flusberg (2005) suggest this discrepancy may be due to their study's larger sample size and broader measurement of ToM. Examining a wider range of communication skills, Pellicano (2013) also found that first and second order ToM

at time 1 failed to contribute unique variance to social communication (as measured by the ADOS-G) at time 2 once other abilities, including verbal and non-verbal IQ, were accounted for. These findings highlight the need to take account of potentially confounding variables when examining relationships between ToM and pragmatic ability.

Another method to elicit continuous speech is the production of narratives, for example, to tell a story from pictures. This allows specific areas of language ability, including pragmatics, to be examined more precisely (Botting, 2002). In a group of ASD, ADHD and TD children, Kuijper, Hartman, Bogaerds-Hazenberg & Hendricks (2017) measured 'discourse pragmatics' within narrative elicited from a wordless story book. Encompassing referential coherence, narrative cohesion, and interruptions, this measure correlated with performance on first and second order false-belief tasks. As structural language ability was again not accounted for, the exact nature of this relationship is unclear. Indeed, previous studies examining narrative ability and ToM, (Capps, Losh & Thurber, 2000; Tager-Flusberg & Sullivan, 1995) found relationships did not remain once structural language ability was accounted for. In narrative produced from pictures depicting a simple series of events, Kuijper, Hartman and Hendriks (2015) found first and second order ToM related to appropriate pronoun selection when a character was reintroduced. Only second order ToM remained predictive in a multivariable analysis. Losh and Capps (2003), meanwhile, found no correlation between measures of second order and advanced ToM with a range of narrative measures including story length, grammatical complexity, evaluation and structure in a group of children with high-functioning ASD. In summary, the link between ToM and pragmatic ability in ASD appears inconsistent. Adding further ambiguity; some studies that report a significant relationship fail to control for age, language and IQ.

The studies described thus far have measured specific pragmatic abilities in structured situations. One method of obtaining a wider picture of real-life pragmatic language abilities

in a time efficient manner is with ratings given by an adult who knows the child well. The Children's Communication Checklist Second Edition (CCC-2; Bishop, 2003) is a 70-item questionnaire frequently used for this purpose. It comprises ten sub-scales relating to structural and pragmatic language, alongside traits associated with ASD (A. Speech, B. Syntax, C. Semantics, D. Coherence, E. Inappropriate initiation, F. Stereotyped language, G. Use of context, H. Non-verbal communication, I. Social Relations, J. Interests). These can be combined into composite scores to compare pragmatic and structural language ability (Norbury, Nash, Baird & Bishop, 2004), though debate exists regarding the classification of some sub-scales (Volden & Philips, 2010; Ash, Redmond, Timler & Kean, 2017).

Two studies have utilised the CCC-2 to examine the relationship between ToM and pragmatic language ability. Losh, Martin, Klusek, Hogan-Brown & Sideris (2012) compared teacher ratings on the CCC-2 and scores on the pragmatic judgement sub-test of the Comprehensive Assessment of Spoken Language (CASL), with performance on a battery of ToM tasks. ToM correlated significantly with the CASL and the initiation sub-scale of the CCC-2, though language and non-verbal IQ were not controlled for. In addition to individual sub-scales from the CCC-2, Baixauli-Fortea, Miranda Casas, Berenguer-Forner, Colomer-Diago & Roselló-Miranda (2017) examined relationships between ToM and a 'pragmatic composite'. This was the sum of the initiation, stereotyped language, context and non-verbal sub-scales. ToM was measured using The Theory of Mind Inventory, (Hutchins, Prelock & Bonazinga, 2012) a 60-item parent questionnaire designed to tap a wide range of social cognitive understanding. Regression analysis indicated that both ToMI scores and a 'structural language' composite from the CCC-2 (speech, syntax, semantics and coherence sub-scales) were predictive of the pragmatic composite. Baixauli-Fortea et al. (2017) note that as both measures relied on parent ratings a 'bandwagon effect' may be responsible for correlations between measures. Additionally, their structural language composite included

the CCC-2 ‘coherence’ sub-scale, which may recruit both structural and pragmatic abilities and thus also account somewhat for the relationship between structural and pragmatic language.

As structural language ability correlates significantly with both ToM performance (e.g. Tager-Flusberg & Sullivan, 1994; Norbury, 2005) and pragmatic language ability (e.g. Whyte & Nelson, 2015). The potential moderating effect of structural language on both pragmatic and ToM skill may ambiguate the association between these two abilities. Given that difficulties in structural language are found in autism (Boucher, 2012) any investigation of the relationship between ToM and pragmatics in this population clearly requires structural language ability be accounted for.

The current study examines the relationship between ToM and pragmatic language ability in children with autism using the CCC-2. Our study differs from Losh et al. (2012) and Baixauli-Fortea et al. (2017) in several ways. Firstly, we compare parent (rather than teacher) ratings of pragmatic ability with a *direct* measure of ToM (first order, second order and advanced). Parent ratings on the CCC have been suggested to possess enhanced validity, being likely to relate to the child’s diagnostic status (Bishop & Baird, 2001). Combining this with a direct measure of ToM removes the possibility of a ‘bandwagon’ effect. Whilst ToM encompasses more than false belief understanding, these tasks were selected to avoid the risk of ‘circularity’ that may occur when more socio-communicative ToM tasks (e.g. joint attention, social referencing) are used in ASD populations (Hughes & Leekham, 2004; Pellicano, 2013). Secondly, to investigate whether ToM is associated more strongly with specific areas of pragmatic development we examine the relationship between our direct ToM measures and both individual CCC-2 sub-scales and global composite scores. Thirdly, we account for the potential role of structural language and NVIQ in moderating the relationship between ToM and pragmatic ability, using direct measures of these abilities.

To better understand the pragmatic abilities of primary-schooled aged children with autism, and the cognitive processes relating to these skills, the present study addresses three questions. First, we examine the nature of the language and communication impairment in our sample of autistic children, as assessed by the CCC-2. That is, the extent to which this group are reported by their parents to display pragmatic and/or structural language difficulties. Second, we examine whether, and how, pragmatic and structural language ability (as measured by the CCC-2) and ToM performance are associated with chronological age, direct measures of expressive and receptive language, and non-verbal IQ. Finally, our key question concerns whether a relationship exists between pragmatic language (as measured by the CCC-2), ToM, and a measure of autistic symptomology, namely The Social Responsiveness Scale (SRS) (Constantino, 2005).

2.2. Method

Participants

Twenty-two children (18 boys and 4 girls) with a diagnosis of ASD were recruited through a developmental lab database, four schools (two mainstream and two specialist ASD) and local parent support groups all in South-East England. All children had received their diagnosis from either a psychologist or a paediatrician. Twenty-one of these children also took part in the study outlined in Chapter 4 of this thesis. Both receptive (the *Sentence Structures sub-test)* and expressive language (the *Formulated Sentences* sub-test) were measured by the Clinical Evaluation of Language Fundamentals® - Fifth Edition (CELF®-5, Wiig, Semel & Secord, 2013). To assess non-verbal IQ, the *Pattern Construction* and *Matrices* tasks from the British Ability Scales Third Edition (BAS-3, Elliot & Smith, 2011) were administered. Parents of children additionally completed the Social Responsiveness Scale (SRS), a questionnaire designed to screen for ASD, with a cut-off T-score of 59 (Constantino & Gruber, 2005). The demographics for the sample (N = 22) is shown on Table 1 below.

Table 1: *Group Characteristics*

	Mean (SD)	Minimum	Maximum	% below typical range
Chronological age	6;7	5	7;11	-
Expressive Language Ability[1]	11 (3.22)	4	16	9
Receptive Language Ability[2]	10.73 (2.83)	6	16	5
Non-Verbal IQ[3]	48.32 (10.34)	26	64.5	18
SRS (t-score)	79.73 (12.28)	50	90	-

[1] CELF formulated sentences sub-test scaled score

[2] CELF sentence comprehension sub-test scaled score

[3] Mean t-score BAS pattern construction and matrices sub-tests

Measures

Children's Communication Checklist – Second Edition (CCC-2)

As a measure of children's pragmatic and structural language ability, parents completed the CCC-2 (Bishop, 2003). This is a 70-item checklist designed for children aged 4 to 16;11 years who produce multiword utterances. Responders rate statements regarding structural, pragmatic and ASD related behaviours by recording the frequency with which that behaviour is typically observed. Statements are both strength and weakness based, and each relates to one of 10 sub-scales (A. Speech. B. Syntax. C. Semantics. D. Coherence. E. Inappropriate initiation F. Stereotyped language. G. Use of context. H. Non-verbal communication. I. Social Relations. J. Interests). Scaled scores for each sub-scale range from 1 to 19, with higher values indicating communicative strength.

The sub-scale scores may be combined to yield two composite scores. Firstly, the 'General Communication Composite' (GCC) (sum of scaled scores A-H) is designed to identify children with a communication impairment (structural or pragmatic or both) indicated by a score below 55. Secondly, the 'Social Interaction Deviance Composite' (SIDC) (the sum of scaled scores E+H+I+J minus A-D) is designed to discriminate between children with a profile suggesting specific language impairment (SLI) versus those where social communication deficits outweigh structural language difficulties. A negative SIDC score in conjunction with a GCC score of less than 55 is deemed indicative of ASD. The SIDC replaces the original CCC (Bishop, 1998) pragmatics composite (D-H) which was not found to adequately discriminate between subgroups of communication impairment. In the CCC-2, the coherence (D), stereotyped language (F) and context (G) subscales were removed from the pragmatic composite as they were deemed to recruit structural language ability. The CCC-2 reports a sensitivity value of .89 and a specificity value of .97 for identifying children with autistic symptomatology and pragmatic social impairment (Bishop, 2006).

As the current study examined relationships between ToM, pragmatic and structural language, two additional composites were calculated. Firstly, the 'pragmatic composite' mirrored that of the original CCC (D-H) (see Ash, Redmond, Timler & Kean, 2017, for discussion of this grouping's suitability to capture pragmatic ability due to close alignment with the DSM-5 criteria for social (pragmatic) communication disorder). Secondly, the 'structural composite' was the sum of three scales; two originally used to measure structural language ability in the CCC (A & B) and the CCC-2 semantics sub-scale (C).

Theory of Mind

Children completed tasks designed to assess first and second order false belief and advanced theory of mind. All children first completed a measure of second-order theory of mind; Coull, Leekham, and Bennett's (2006) 'Robot Story' with accompanying pictures.

This received a maximum of two points. Children who failed to answer the questions correctly completed two measures of first-order theory of mind, the 'Sally-Anne' task (Baron-Cohen, Leslie & Frith, 1985) and the 'smarties task' (Perner, Frith, Leslie & Leekham, 1989). Each child received a maximum score of two points for 'First order Theory of Mind', one point per task. Children who passed the robot story were automatically assumed to pass first-order Theory of Mind (see Charman et al., 2011, for a similar step-wise procedure). These children proceeded to advanced theory of mind tasks, utilising two items from Happe's (1994) 'Strange Stories' ('lies to persuade' and 'double-bluff'). The former was similar to Happe's (1994) original 'kittens' story, whereby difficult vocabulary items were simplified. The 'double-bluff' story was re-written to revolve around a sister hiding biscuits from her brother (See Appendix 1). All stories were accompanied by pictures. The maximum possible score for 'Strange Stories' was four points. Cronbach's alpha indicated good levels of reliability ($\alpha = .614$) for a composite Theory of Mind score (maximum = eight points), which combined all ToM measures.

In addition to the measures which were included in the Theory of Mind composite, we followed Happe's (1994) procedure in first administering a practise story (a story about pretending to be a ghost) on which children received feedback. After all Theory of Mind measures were complete, we also administered the 'mountains story' (White, Hill, Happe & Frith, 2009), designed to control for inferencing ability per se rather than mental state inferencing. All participants passed this task, suggesting any difficulties with the remaining tasks were not due to text and sentence level comprehension.

Procedure

Parents completed the CCC-2 in a quiet space in the laboratory while children completed experimental tasks, or in the home environment if their child competed activities

at school. Children were administered theory of mind, language and non-verbal IQ measures by an experimenter in a quiet room in either a laboratory or school setting. Tasks were interspersed between a range of experimental tasks that formed part of a wider study. All responses were audio-recorded to allow transcription and coding.

Data Analysis

Statistical analyses were performed with the Statistical Package for the Social Sciences (SPSS) software, version 25.00 (SPSS Inc., Chicago, IL USA). Differences between mean scaled scores for the structural and pragmatic language composites of the CCC-2 were examined using paired *t*-tests. Pearson's correlation analyses were also performed, firstly, to assess relationships between both CCC-2 composite and sub-scale scores and ToM with potential confounding variables (age, expressive and receptive language, non-verbal IQ) and, secondly, to investigate relationships between ToM, pragmatic and structural language (CCC-2), and autistic symptomology (as measured by the SRS).

2.3. Results

CCC-2

Internal Reliability

Cronbach's alpha indicated that reliability for CCC-2 as a whole was excellent (α = .91). The 8 scales comprising the GCC (α = .80), 5 scales comprising the Pragmatic Language composite (α = .94), and 3 scales comprising the Structural Language composite (α = .81) also had either good or excellent reliability.

Research Question 1: What is the nature of the language and communication impairment in our sample of autistic children, as assessed by the CCC-2?

Table 2 shows a summary of all CCC-2 composite scores. Scores are compared with the population norm, as we did not administer the CCC-2 to a group of typically developing children in this study. Communicative difficulties are common, with only two children obtaining a GCC score in the typical range (>55). The mean scores for all composites fall below the typical range, indicating that communicative difficulties are pervasive. Paired *t*-tests were computed to examine differences between mean scaled scores for each composite. The structural language score (M = 5.50, SD = 2.88) was significantly higher than the pragmatic language score (M = 4.16, SD = 2.14) ($t\ (21) = 2.51$, $p = .02$) and marginally higher than the ASD traits score (M = 4.22, SD = 2.37) ($t\ (21) = 2.04$, $p = .06$). Scaled scores on the pragmatic and ASD trait composites did not differ significantly ($t\ (21) = 0.23$, $p = .82$). Communicative difficulties are therefore not confined to pragmatic language, but this area appears most problematic at group level.

When a GCC score falls in the impaired range, the SIDC may be used to identify whether difficulties are primarily pragmatic or structural in nature (Norbury, Nash, Baird & Bishop, 2004). Of the 20 children with a GCC score of less than 55, only one child scored zero on the SIDC, indicating equal pragmatic and structural difficulties for this individual. 11 children (55%) had negative SIDC scores, indicative of pragmatic language difficulties that outweigh structural language difficulties. Unexpectantly, 8 (40%) obtained a positive SIDC score, indicative of greater structural than pragmatic language impairment. Of the 8 with a positive SIDC score, one child obtained a GCC below 55 and an SIDC above 9, deemed to indicate specific language impairment (SLI).

Table 2: *Summary of CCC-2 Composite Scores*

Composite/ Scale	Mean (SD)	Minimum	Maximum	Typical Mean (SD)	% below population norms	
					1 SD or more	2 SD or more
GCC	37.32 (17.03)	12	86	80 (24)	90.91	31.82
Prag Lang	20.82 (10.72)	8	56	50 (15)	90.91	54.55
Struct Lang	16.50 (8.65)	2	31	30 (9)	63.64	27.27
ASD Traits	8.45 (4.74)	1	21	20 (6)	81.82	54.55
SIDC	-3.59 (8.03)	-19	12	0	-	-
Percentage < 0	55					

Sub-scales

To investigate patterns of strength and weakness in communicative ability, mean scaled scores for each sub-scale were calculated (Table 3). Each scale has a mean of 10 and SD of 3. A scaled score of 6 or more indicates typical functioning, corresponding approximately to a percentile score of 15 or above (Bishop 2003). The only sub-scale with a mean score in the typical range was 'Speech', which relates to the physical quality of speech sounds, that is, whether words are correctly articulated.

Relationships between structural and pragmatic language measures

To examine relationships between structural and pragmatic language ability, Pearson's correlations were calculated between the structural and pragmatic language composites of the CCC-2 and also direct measures of structural language and the CCC-2 pragmatic composite. There was a significant relationship between the pragmatic and structural language composites of the CCC-2 ($r(22) = .541$, $p = .009$), but not between the

pragmatic composite of the CCC-2 and direct measure of structural language (CELF formulated sentences and sentence comprehension mean score) (r (22) = .121, p = .592).

Table 3: *Summary of CCC-2 sub-scales*

				% below population mean	
Sub-Scale	**Mean (SD)**	**Min**	**Max**	**1 SD or more**	**2 SD or more**
A. Speech	6.50 (3.69)	0	13	36	27
B. Syntax	4.50 (3.31)	0	12	73	36
C. Semantics	5.50 (3.13)	2	16	77	27
D. Coherence	4.00 (1.98)	2	9	86	50
E. Inappropriate initiation	5.18 (1.84)	1	10	82	14
F. Stereotyped language	4.41 (3.08)	0	14	86	50
G. Use of context	3.95 (2.85)	1	13	91	55
H. Non-verbal communication	3.27 (1.91)	1	10	95	64
I. Social relations	3.14 (3.01)	0	10	82	68
J. Interests	5.32 (2.23)	1	11	82	9

ToM

ToM Task Performance

Table 4 shows mean scores for each ToM task. Most participants demonstrated first-order ToM ability; 17 participants (77%) passed both first-order ToM tasks and only 3 participants (14%) failed both. Performance on the second order ToM task was more mixed, with 9 participants (41%) achieving a maximum score and 7 (32%) scoring zero. Advanced ToM posed most difficulty for this group; of the 15 participants who progressed to this stage

in the ToM battery 11 (73%) scored zero for both tasks, and no participant made sufficient reference to the protagonist's mental state to obtain a maximum score on either story.

Table 4. *Summary of ToM task scores*

Measure	Mean (SD)	Min	Max	Total Possible
First Order	1.64 (.73)	0	2	2
Second Order	1.09 (.87)	0	2	2
Advanced	.18 (.39)	0	1	4
Total	3 (1.69)	0	6	8

Research question 2: Relationships with control variables

To assess relationships between our measures of interest (CCC-2; composite and sub-scale scores and ToM; first and second order, advanced and overall task performance) with potential confounding variables (age, expressive and receptive language, non-verbal IQ), Pearson's correlations were calculated.

CCC-2

General Communication Composite scores did not correlate significantly with any control variable (all $r < .250$, all $p > .256$). Similarly, there were no significant relationships between the pragmatic language, structural language and ASD behaviours composites with any control variable (all $r < .282$, all $p > .203$). Regarding individual sub-scales, the relationship between expressive language ability (measured by the formulated sentences sub-test of the CELF) and the initiation sub-scale was approaching significance, but the effect size of this relationship was small (r (22) = .377, p = .084). The same was true of the relationship between receptive communication ability (as measured by the sentence comprehension sub-test of the CELF) and the syntax sub-scale (r (22) = .387, p = .075).

Relationships between all other sub-scales and control variables were non-significant (all $r <$ .337, all $p >$ .125).

ToM

The only control variable to significantly relate to measures of ToM was receptive language ability (as measured by the sentence comprehension sub-test of the CELF). Receptive language and; 1st order ToM (r (22) = .412, p = .057), 2nd order ToM (r (22) = .495, p = .019), Total ToM (r (22) = .527, p = .012). There were no other significant associations between any measurement of ToM and any control variable (all $r <$ 352., all p >.109).

Research Question 3: Relationship between CCC-2, ToM and SRS

To investigate relationships between ToM, pragmatic and structural language, and autistic symptomology, Pearson's correlations were calculated (see Table 5). Receptive language (sentence comprehension; CELF) was controlled for due to its significant relationship with ToM. Structural language was the only CCC-2 composite to significantly associate with any ToM task (second-order ToM, p = .02; Total ToM, p = .03). A significant correlation existed between the ToM total score and the CCC-2 speech sub-scale (r (22) = .527, p = .012) and syntax sub-scale (r (22) = .460, p = .031), but not the semantic sub-scale (r (22) = .198, p = .377). There were no significant relationships between any pragmatic language sub-scale of the CCC-2 and any measure of ToM (all $r <$ 343., all p >.118). All CCC-2 composites related significantly to SRS scores, with higher autistic symptomatology relating to lower communicative ability. Finally, no relationship was found between any ToM measure and SRS scores (all $r <$ .236, all p >.290).

		ToM				SRS
		Total	1st Order	2nd Order	Advanced	
CCC-2	GCC	*r* = .200	*r* = .110	*r* = .294	*r* = .033	*r* = -.678***
	Structural composite (A + B + C)	*r* = .473* (*r* = .417, *p* = .060)	*r* = .364	*r* = .482* (*r* = .429, *p* = .053)	*r* = .209	*r* = -.453*
	Pragmatic composite	*r* = -.063	*r* = -.119	*r* = .079	*r* = -.116	*r* = -.763**
	Autism composite	*r* = -.113	*r* = -.143	*r* = -.022	*r* = -.072	*r* = -.840**

Table 5: *Correlation Coefficients for CCC-2, ToM and SRS*

Notes: *p < .05, **p < .01. ***p<.001

Partial correlations adjusting for receptive language ability (CELF; Sentence Comprehension) shown in parentheses

2.4. Discussion

Summary of Findings

Utilising a parent report of pragmatic and structural language ability (the CCC-2), alongside direct measures of ToM, receptive and expressive language and non-verbal IQ, this study had three aims. Firstly, to examine the pattern of communicative impairments in

everyday life of autistic children; secondly, to examine how parental CCC-2 ratings and direct measurements of ToM relate to chronological age, direct language measures and non-verbal IQ; and thirdly, our key question of interest was whether a relationship exists between pragmatic ability (as measured by the CCC-2) and theory of mind (as measured by a battery of false belief and advanced ToM tasks) in children with autism. In relation to the first aim, despite all but two children scoring within the typical range on direct measures of structural language, below average pragmatic *and* structural language ability were apparent in many parent ratings. As anticipated, at group level pragmatic ability was rated worse than structural language ability. Unexpectantly, however, for nearly half of the individuals in our sample, structural language was rated lower than pragmatic language. Regarding the second aim, no significant correlations were found between any composite or sub-scale of the CCC-2 and age, direct structural language measures or non-verbal IQ. ToM, meanwhile, related to a direct measure of receptive language, but not expressive language, age or non-verbal IQ. In relation to our third, and primary question of interest, no relationship was found between pragmatic ability (as measured by the CCC-2) and ToM performance. The only significant relationship to emerge was a positive correlation between the structural language composite of the CCC-2 and ToM performance. This indicates that ToM may not play a pivotal role in the real-life pragmatic ability of the autistic children in this group.

The nature of communicative difficulties in autism

Our first finding, that parents report their autistic children demonstrate poor pragmatic and structural language in everyday life, is in line with the suggestion that pragmatic language difficulties are present in autism (Frith, 1989; Martin & McDonald, 2003; Tager-Flusberg, Paul & Lord, 2005), alongside potential structural language difficulties (Boucher, 2012). That said, the vast majority of children scored in the typical range on direct measures of structural language (the CELF), and there was only a weak relationship between this

measure and the structural language composite of the CCC-2. A discrepancy between CELF and CCC-2 scores has been reported in previous studies (e.g. Antoniazzi, Snow & Dickson-Swift, 2010). In the current study, differences between these two measures may reflect the objective nature of parent ratings (King & Palikara, 2018), or alternatively the failure of the CELF to detect some real-life structural language difficulties (Bishop & McDonald, 2009). Given the diverse range of skills included in the CCC-2 (Wilson & Bishop, 2019), it is also possible that the CCC-2 structural language composite and sentence comprehension / formulated sentences sub-tests of the CELF measured slightly different skills. Our second finding, that no CCC-2 composite or sub-scale correlated significantly with chronological age, direct measures of structural language or non-verbal IQ, suggests that, at least for these children who mostly scored in the typical range on these measures, general intelligence and structural language ability are not key factors in their pragmatic language difficulties. The finding that ToM was related to receptive language ability as measured by the CELF is returned to shortly.

The relationship between ToM and pragmatic language in autism

Regarding the third, and central question, the present study did not find evidence that pragmatic language difficulties in autistic children are related to their ability to consider another's mental state. These findings contrast with two previous studies that examined relationships between ToM and pragmatics as measured by the CCC-2. Firstly, Losh et al., (2012) found a significant relationship between the *inappropriate initiation* sub-scale of the CCC-2 and a battery of ToM tasks. Participants differed from those in the present study, having below average mental age, and ranging from 4-12 years in chronological age. Further, Losh et al., (2012) did not control for age, language and non-verbal IQ, hence it is unclear whether the relationship between ToM and pragmatic ability would exist independent of these factors. Finally, the only significant relationship between ToM and the CCC-2 in

Losh et al's study was with the *initiation* sub-sale. It is unclear why ToM proficiency may be more related to this CCC-2 sub-scale than others defined as 'pragmatic' in their study (*Stereotyped language, Context, Non-verbal communication, Social relations* and *Interests*).

The second study to report a relationship between ToM and the CCC-2 was Baixauli-Fortea et al., (2017). They found ToM ability (as measured by the ToMI) to predict scores on a 'pragmatic' composite from the CCC-2 (*Inappropriate initiation, Stereotyped language, Context* and *Non-verbal communication*, but not '*Coherence*', which was included within a structural language composite). As both pragmatic language and ToM measures were obtained from parent ratings, this finding may be due to parent judgements tending to be internally consistent, rather than demonstrating a genuine relationship between ToM and pragmatic ability. Indeed, in the current study our parent report measures (CCC-2 / SRS) were significantly related to one another, whilst the CCC-2 score for structural language did not correlate with a direct measure of this skill (CELF formulated sentences/ sentence comprehension). This latter finding indicates that the findings of Baixauli-Fortea et al., (2017) may indeed have been caused by a 'bandwagon' effect, as it suggests parents tend to rate their children reliably across measures, but these ratings may not relate with direct measurement of the same skills. In sum, the two studies reporting a relationship between pragmatic language as measured by the CCC-2 and ToM suffer from methodological difficulties (failure to control structural language ability, Losh et al. 2012; potential bandwagon effects, Baixauli-Fortea et al., 2017) that render the relationship reported between these abilities uncertain.

The role of structural language and other potentially mediating abilities

Controlling for structural language ability is important when drawing conclusions regarding a relationship between ToM and pragmatic language. Firstly, structural language

and ToM are related in autism (Happe, 1995; Tager-Flusberg, 2000). Indeed, in the present study, ToM performance was significantly related to the CCC-2 structural language composite and CELF sentence comprehension. Structural language skills may have aided the comprehension of ToM stories and/or supported the production of adequately detailed responses. Alternatively, better structural language might support children's development of ToM (de Villiers & de Villiers, 2014; Milligan, Astington & Dack, 2007) or vice-versa (Tager-Flusberg & Sullivan, 1994). Secondly, structural language ability is also strongly associated with pragmatic language performance (Matthews, Biney & Abbot-Smith, 2018). In the current study, a relationship was evident between parent ratings of pragmatic and structural language (CCC-2), but limited relationships existed between parent ratings of pragmatic language and direct measures of structural language (a marginally significant relationship between the CELF formulated sentences and initiation sub-scale of the CCC-2). We explore reasons for this weak relationship between structural and pragmatic language in the general discussion of this thesis. Nonetheless, given the potential inter-relationships between structural language, ToM and pragmatic language, drawing the conclusion that ToM is related to pragmatic ability is problematic if structural language is not controlled for.

Only three previous studies report a relationship between ToM and pragmatic language in autism once structural language is controlled. Firstly, in a comprehension task involving identification of a target referent, Schuh, Eigsti & Mirman (2016) reported a relationship between gaze target fixations and ToM as measured by the NEPSY-II. This measure included 1st order ToM, interpretation of non-literal language and inferring a character's internal state from context. It might be argued that differences in ToM measurement contributed to the disparate findings between this and the current study. That said, no group difference was found on the NEPSY-II ToM by Schuh et al., (2016), making these findings harder to interpret. Further, working memory ability did differ significantly

between groups, and associated with performance on both the reference comprehension and ToM tasks. Given that working memory also correlates with ToM in typical development (Gordon & Olsen, 1998; Carlson, Moses & Breton, 2002) it may have mediated the relationship between theory of mind and reference comprehension for these autistic individuals.

The second study to find a relationship between an area of pragmatic ability and ToM, after controlling for language, is Whyte Nelson and Scherf, 2014. They found a relationship between idiom comprehension and ToM as measured by RTME and strange stories. Whilst both ToM and structural language were uniquely related to idiom comprehension for autistic children, structural language was related to performance on both idiom comprehension and ToM tasks. Additionally, whilst autistic children's idiom comprehension was worse than a NVIQ and chronological age matched TD group, it was similar to a chronologically younger group matched for syntactic ability. The nature of the relationships between ToM, structural language and idiom comprehension in this study, therefore, remains unclear. Finally, Hale and Tager-Flusberg (2005) found a relationship between contingency in conversation and a battery of ToM measures. The CCC-2 includes multiple items relevant to conversational ability e.g. 'Doesn't explain what he/she is talking about to someone who doesn't share his/her experiences' and 'Talks repetitively about things no one is interested in', but it is possible that ToM is particularly relevant for following the topic of one's conversational partner. The wider range of ToM measured by Hale and Tager-Flusberg (2005), and indeed Whyte et al., (2014) and Schuh et al., (2016), may contribute towards differences between findings from these and the current study, a possibility that we now consider.

ToM as a multi-component ability

Two component subsystems of ToM have been suggested in the literature; a social-perceptual (affective) component allowing, for example, emotional state to be deduced, and a social-cognitive component enabling more complex social reasoning (Tager-Flusberg, 2001; Joseph & Tager-Flusberg, 2004). Hale and Tager-Flusberg (2005) included ToM tasks that might be considered to tap into 'affective' aspects of this skill (e.g. the 'Desire Task'; Wellman & Woolley, 1990, where a puppets actions are judged based on their goal; the 'Sticker-hiding task'; Devries, 1970, which includes a measure of the participants ability to engage in deceptive strategies; and the 'Traits Task'; Yuill, 1992, where a characters intent is judged based on their personality traits). As previously described, the NEPSY-II ToM subtest utilised by Schuh et al., (2016) also includes tasks that appeal to both social-perceptual and social-cognitive elements of ToM ability. Finally, Whyte et al. (2014) utilised the RMTE task which intuitively involves affective ToM. That said, a relationship was also evident between the strange stories task (a social-cognitive task) and idiom comprehension in this study. In contrast to these three studies, our study utilised exclusively social-cognitive tasks. Incorporating a social-perceptual measure may have revealed a relationship between ToM and pragmatic language as measured by the CCC-2. Whilst a degree of circularity might exist between many measures of pragmatic language and measures of social-perceptual ToM (Pellicano, 2013), it is possible that different pragmatic tasks require different types of ToM, a possibility we return to in the general discussion within this thesis (Chapter 6).

Future directions

In sum, the three studies to report a link between ToM and pragmatic ability in autism highlight a need to clarify the nature of this relationship. First, the skills that may mediate the relationship between these abilities (e.g. structural language and working memory) require attention. To disentangle the nature of relationships between ToM, pragmatic and structural language ability, for example, a measure of ToM that does not burden structural language

would be valuable. This has proved particularly challenging for more advanced measures of ToM (Osterhaus, Koerber & Sodian, 2016). Second, given that the current study did not detect a relationship between social-cognitive aspects of ToM and pragmatic language, there is a need to better understand whether, and how, different aspects of ToM might relate to specific areas of pragmatic language, or pragmatic language as a whole.

Limitations

Regarding the present study, it may be questioned whether our sample size affected findings. An a priori power analysis revealed that a sample size of 85 would be required to detect a medium size effect with a power of .80 (Cohen, 1992). Our sample size of 22 was relatively small, but as effect sizes were small for all ToM and pragmatic relationships, the modest sample size may not have been instrumental in our findings. Variability in ToM scores was also present, despite deriving largely from the second-order task. Thus, a lack of variance was unlikely to have contributed to the absence of a significant relationship between measures. Finally, autistic children were mostly in the typical range for direct measures of structural language but not on the 'syntax' and 'semantics' CCC-2 sub-scales. We will return to this issue in the general discussion (Chapter 6).

Summary

In summary, the current study did not find evidence for a relationship between first order, second order, or advanced ToM and pragmatic language. In fact, only three previous studies find relationships between ToM and specific areas of pragmatic language independent of structural language ability, and even in these studies the nature of the relationship is unclear. These findings bring into question the role of ToM in pragmatic impairment in autism. Whilst drawing conclusions based on a failure to detect a correlation is problematic, our findings suggest it may be beneficial to examine alternative factors contributing to

pragmatic impairment in autism. As previously described, examining relationships with broader areas of awareness of other's minds, such as common ground understanding may be beneficial. This possibility is explored further in Chapter 4 of this thesis. There is also growing evidence, for example, that executive functioning is diminished in children with autism (Lai. et al., 2016). Impairments in these cognitive abilities that enable goal-orientated behaviour might contribute to pragmatic language difficulties in this population (Norbury, 2014). Obtaining a clearer understanding of the skills associated to pragmatic ability in autism is key to develop effective support strategies. Turning empirical attention to relationships with a wider range of cognitive demands, such as executive functioning, may well advance progress towards this goal, and this is explored further in Chapter 5 of this thesis.

CHAPTER 3: IS VERBAL REFERENCE IMPAIRED IN AUTISM SPECTRUM DISORDER? A SYSTEMATIC REVIEW

3.1. Introduction

Autism spectrum disorder (ASD) is defined by persistent deficits in social-communication alongside restricted, repetitive patterns of behaviour, interests or activities (Diagnostic and Statistical Manual of Mental Disorders 5th ed [DSM-5]; American Psychiatric Association, 2013). Social use of language, or 'pragmatics' (social verbal communication), is considered a central impairment in ASD (Landa, 2000). Difficulties in this area hinder the ability to establish and maintain reciprocity in conversation and impair the successful exchange of relevant information necessary for collaboration, negotiation and daily interaction (Tager-Flusberg, Paul & Lord, 2005). A deficit in pragmatics has additionally been linked to mental health difficulties in a number of populations (e.g. Helland, Lundervold, Heimann & Posserud, 2014). Whether all domains of pragmatics are equally impaired or indeed impaired in all individuals with ASD, however, remains unclear (Simmons, Paul & Volkman, 2014).

A core component of pragmatics is reference, that is, the ability to denote an entity, person or event with sufficient clarity for one's interlocutor. While theorists from a semiotic perspective give equal weight to verbal and non-verbal means (see Perkins, 2005, for discussion), in the current review, we follow Norbury (2014) among others in distinguishing pragmatic language from social communication more broadly and therefore we focus solely on 'verbal reference'. Impaired verbal reference is highly likely to have a severe detrimental effect on conversational flow (and thus on social relationships) and on the ability to collaborate with others (e.g. Murphy, Faulkner & Farley, 2014), which would also have educational consequences. While the appropriate production and interpretation of verbal reference is often included in speech and language assessment and intervention for

individuals with ASD (e.g. Adams, Gaile, Freed & Lockton, 2010), the extent to which this area of pragmatic ability is universally impaired in this population has not been examined to date.

Forms of verbal reference vary greatly in complexity, as illustrated in Table 6, from highly specific referring expressions such as 'the red ball on top of the cupboard' to pronominal referring expressions, such as 'she' or 'it'. Bare noun phrases (definite: 'the ball'; indefinite 'a ball') may be considered intermediate in terms of specificity (e.g. Gundel, Hedberg & Zacharski, 1996). Correct use and interpretation of verbal reference requires children to acquire the network of potential forms used in his or her language. Pronominal (and equivalent) systems, for example, show great cross-linguistic variation in their level of acquisition difficulty, depending on whether forms vary regarding factors such as the gender, plurality or honorific status of the referent and whether case needs to be marked. Children learning Germanic and Romance languages usually acquire pronominal form by the mid preschool years (e.g. Girouard, Ricard & Decarie, 1997; Chiat, 1986). The form of complex referring expressions can include, for example, adjectival phrases ('the red ball') or modifying phrases such as prepositional phrases ('the ball on the shelf') or relative clauses ('the ball that Daddy bought you'). Syntactic forms used for complex referring expressions are generally mastered by the late preschool years (see e.g. Brandt, 2011; Kidd & Cameron-Faulkner, 2008; Berman & Slobin, 1994).

Table 6. Forms of verbal reference

***Forms* of verbal reference**		
Pronominal and related forms of verbal reference	Pronoun Zero form	He / she / it / they Laura yawned and ___ fell asleep
Deixis		That / this one / that girl
Simple / bare noun phrase	Common noun Proper noun	The / a girl Laura
Complex referring expression	With adjective With modifying phrase	The tall girl The girl on the bike The girl that you met yesterday.

An adult-like mastery of verbal reference not only requires acquisition of form, but also the ability to vary the level of complexity in accordance with context. Referring terms are therefore matched to the informational needs of a specific interlocutor. The appropriate use of verbal reference is often described by appealing to Grice's (1975) theory of communication, in particular the cooperative principle specifying the maxims of quantity and manner. These specify that a speaker should provide sufficient information for the listener to determine reference, but also be concise (i.e. speakers should not be over-informative). Although many researchers have since pointed out major problems for the Gricean account (e.g. Sperber & Wilson, 1995; Gergely & Csibra, 2005; Moore, 2014; Horn, 1988, see also Levinson, 1989), the framework still provides a useful means of conceptualising the types of skills required to carry out and understand acts of reference.

Regardless of theoretical perspective, the match (or mismatch) between a particular form of verbal reference (e.g. pronominal, bare noun phrase or complex referring expression) and a particular context can be judged as 'correct' or 'incorrect'. Here, 'context' can include the information that specific interlocutors know one another to share; for example, if a child knows that his father is well acquainted with his friend Jamie, then it would be over-informative to use a complex referring such as 'the Jamie that came to my birthday party'

every time the referent ‘Jamie’ is introduced into the conversation. ‘Context’ can also include whether there are competing referents in the visual context. To illustrate, if there is only one brush in the vicinity, then asking a listener to pass ‘the brush’ may be sufficient. In contrast, if more than one brush is present, then the speaker may need to specify ‘the brush with the brown handle’. Finally, the relevant ‘context’ would also include how recently a referent has been mentioned. That is, if a referent has just been mentioned in dialogue or narrative, the speaker can usually (depending on whether there are competing referents) reduce the specificity of the referring term further by using pronouns, since speaker and listener can use their knowledge of the shared common ground to determine which referents are likely to be most salient (Sperber & Wilson, 1995) and / or activated in working memory, which is usually considered to be the component of short-term memory used to manipulate and update concurrently incoming material (see Baddeley & Hitch, 1974).

For the purpose of this review, we focus on whether referring terms are appropriately based on the informational needs of the listener, given ‘context’ as defined above. Thus, when we say ‘appropriate’, this is not a qualitative judgement; a participant who says, ‘Give me the duck’ in a context in which he or she can see that the addressee can see two ducks should arguably receive a score equating to ‘incorrect’ for this particular request. However, since even typical adult speakers do not perform at ceiling in these types of tasks (Keysar, 2007) we use the terms ‘appropriate’ vs. ‘inappropriate’ throughout, to describe the match or mismatch between the form and context.

The first aim of the current paper was therefore to carry out a systematic review to determine whether an impairment in the appropriate usage / interpretation of verbal reference is a global feature of ASD (or whether verbal reference is only impaired in individuals with ASD with comorbid intellectual or formal language difficulties). To this end, our focus was not on whether individuals with ASD used the same forms of reference (e.g. whether they use

the same proportion of pronouns within, for example, a conversation as do typically-developing controls). Rather, our focus was on whether individuals with ASD are atypical in their understanding of the 'fit' between reference form and context.

If we found that some studies did not report an impairment in verbal reference in ASD, our second research goal was to investigate the extent to which this might be due to either to the methodology used or to the modality in which proficiency with verbal reference was measured. Finally, we also wished to investigate whether studies including individuals with ASD provide evidence regarding the cognitive underpinnings of verbal reference ability.

To determine our key search terms, we first attempted to pinpoint the types of tasks typically used to assess verbal reference; that is, naturalistic interaction, narrative or the 'director task' / referential communication paradigm (see Graf & Davies, 2014, for a review). We also attempted to identify the key concepts most commonly associated with verbal reference in the literature. One such concept is that of 'listener needs' or 'audience design'; as previously described, to be optimally informative, a referring term should provide sufficient information without being over-informative. This type of adaption to the informational needs of the listener is considered appropriate audience design (Clark & Murphy, 1982). Successful audience design may be achieved through consideration of the information listener and speaker share, or 'common ground' (Clark & Marshall, 1981).

3.2. Criteria for current review

Systematic searches were conducted in two databases: PsychINFO and Web of Science for all dates up until March 2016. Our search terms were entered into the 'keyword' field as follows: (a) autis* AND narrative, (b) autis* AND referen* AND communicat*, (c) autis* AND common ground, autis* AND audience design, (d) autis* AND listener needs,

(e) autis* AND director task. Given that these two search engines are imperfect, it is inevitable that this review will not be exhaustive. Indeed, we found and included one study which met our search engine criteria (Kuijper Hartman & Hendriks, 2015), but which was detected by neither search engine. Nonetheless, this review should constitute an accurate representation of literature in this topic to date.

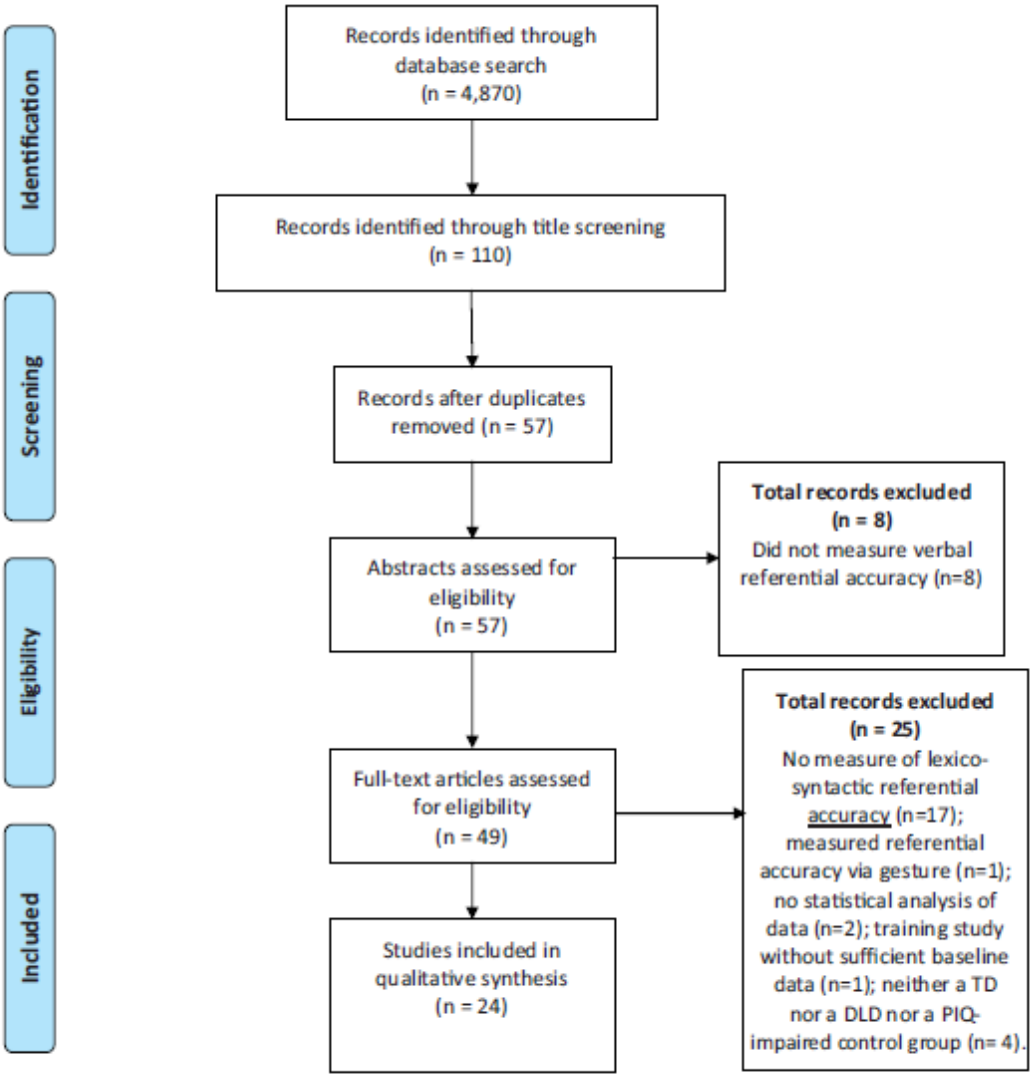


Figure 1. PRISMA diagram of study identification and selection.

An initial review of titles and abstracts excluded studies that were clearly not related to the key topics of interest, such as articles on literature or politics. The remaining full

articles were then examined and our inclusion criteria were applied as follows. To be included the study was required to (1) include participants with a diagnosis of either autism spectrum disorder (ASD), Asperger syndrome (AS) or pervasive development disorder-not otherwise specified (PDD-NOS), (2) include a measure of the appropriacy of the match between verbal (lexico-syntactic) reference and context, (3) contain quantitative data which were analysed statistically and (4) to include a control group consisting of either a) typically-developing individuals, b) individuals with Developmental Language Disorder (Specific Language Impairment) or c) individuals with an impairment in non-verbal (performance) IQ. Without one of these control groups it is difficult to conclude whether or not individuals with ASD are impaired in referential communication. A study was additionally excluded if: (a) it was a case study with a single participant (due to issues of generalizability), (b) it was a training study which did not contain sufficiently detailed baseline data for conclusions regarding impairment in verbal reference to be drawn or (c) it exclusively examined non-verbal communication such as gesture, facial expression and eye-contact or solely considered prosody (rather than lexico-syntactic form). Studies exclusively examining non-verbal communication were excluded because we were primarily interested in verbal reference. One reason for this, is that when investigating the role of cognitive underpinnings, it is likely that the role of formal language in particular would play quite a different role in relation to non-verbal reference than in relation to verbal reference.

The total number of studies considered for inclusion, and those excluded at each stage in the search process are shown in Figure 1. The 24 studies that met our inclusion criteria are listed in Tables 7-12. First, studies comparing the production of verbal reference in ASD and TD groups are summarised. Next, studies that compare comprehension of verbal reference in ASD and TD groups are summarised. Finally, evidence for the potential cognitive underpinnings related to successful verbal reference is reviewed.

3.3. Production of referring expressions

We first review studies that have used methodologies which most closely map onto naturalistic usage of verbal reference in daily life. We then review studies that have measured appropriate verbal reference during more structured narrative tasks, and finally the most structured elicitation technique, the referential communication task.

Production of referring terms during conversation

Though most closely mirroring real-life interaction, only one study with at least one control group with clearly defined characteristics returned in our search measured the appropriacy of verbal reference use by individuals with ASD during conversation (see Table 7). Baltaxe and D'Angiola (1996) examined the use of pronominal, demonstrative (e.g. here/there) and comparative (e.g. bigger/ smaller) reference during an hour-long interactive play session. Children with ASD (M = 7;9yrs, n=10) were matched on language ability with a chronologically younger TD group (M = 3;5yrs, n=8). Use of ambiguous reference (e.g. saying 'it' when the reference is unclear) was never found in the TD group (p.252). This study also compared the ASD group to a group with Developmental Language Disorder (DLD), (n=8) matched on receptive language and Mean Length of Utterance (MLU). DLD is a diagnosis of language impairment in the absence of a known biomedical condition (Bishop, Snowling, Thompson, & Greenhalgh, 2016). In comparison to the DLD group, the ASD group used more ambiguous personal pronouns, though this failed to reach significance. This is presumably in part due to the extremely small sample size. Nonetheless, as we will see, the finding of a tendency towards ambiguity and the finding of deficits relative to children with DLD will be a recurrent theme throughout this review.

Table 7. Production - *Conversation/ Personal Narrative*

Author, year	*Participants*				*Matching*			*Measure of Reference*	*RC impairment in ASD group*
	Sample size (n per group)	*Mean age (range) per group*	*Language*	*ASD diagnosis*	*Structural Language*	*Age*	*Non-Verbal IQ*	*Statistical Analysis / Inter-Rater Reliability*	
Baltaxe and D'Angiola (1996)	*n* = 10 (ASD) *n* = 8 (TD) *n* = 8 (SLI)	ASD: 7;9 TD: 3;5 SLI: 7;7	English	Met DSM-III criteria for autism	✓ MLU: ASD/TD: d=0.69 ASD/SLI: d=0.14 Receptive Vocabulary (PPVT): ASD/TD: d=0.19 ASD/SLI: d=0.02 Language Comprehension (TACL): ASD/TD: d=0.2 ASD/SLI: d=0.29	× ASD/TD: d=-2.28 ASD/SLI: d=-0.03	- Not reported	Pronominal, demonstrative & comparative references, including ambiguous use. ANOVA % reliability	✓

Notes. ASD: Autism Spectrum Disorder; MLU: Mean Length of Utterance; PPVT: Peabody Picture Vocabulary Test; TACL: Test for Auditory Comprehension of language; SLI: Specific Language Impairment; TD: Typically Developing.

Overview

Narrative tasks usually require the participant to generate or retell a story based on a picture book or film. They therefore constitute a monologue, rather than a reciprocal interaction. Therefore, narrative tasks might seem quite far removed from naturalistic verbal interaction. Nonetheless, narrative measures have been found to correlate strongly with standardised measures of pragmatic language more broadly, such as the Test Of Pragmatic Language (e.g. Manolistsi & Botting, 2011).

Since we are interested in the degree to which individuals understand the function of verbal reference, our focus is on measures which assess whether the lexico-syntactic form is appropriate given the context. In the sentence 'Laura went to the shop and she bought some bread', for example, the initial reference 'Laura' is appropriate as it introduces a new character. The third-person subject pronoun 'she' is also appropriately unambiguous, referring the listener back to a character 'Laura' recently established as the focus of the conversation. The use of 'she' in this way is an example of 'anaphoric reference'. Errors may be in the direction of over-informativity (e.g. if the full noun 'Laura' were used throughout). When two potential referents have recently been mentioned, conversely, an anaphoric reference may be under-informative (e.g. 'I saw Laura and Karen and she looked upset').

Our survey of studies, which quantitatively measured referential accuracy within narrative, is organised in terms of the elicitation method employed. First, we review narrative generation studies, in which narrative is elicited from a stimulus (generally pictures depicting a story) without a prior model. Then, we review narrative re-tell studies, in which events are witnessed either in picture or video format and then retold either with reference to the original stimuli or from memory.

Narrative Generation

We first discuss narrative generation studies where groups were not well-matched (i.e. not matched for formal language ability, non-verbal IQ and chronological age), followed by those that were (see Table 8). The five less well-matched studies which examined narrative generation all included primary-school aged children (Norbury & Bishop, 2003; Norbury, Gemmell & Paul, 2014; Colozzo, Morris & Mirenda, 2015; Tager-Flusberg, 1995; Mäkinen et al., 2014). One of these studies (Norbury et al., 2014) also included young adolescents. Three of the five studies elicited narratives using the 24-page wordless story book 'Frog, where are you?' (Mayer, 1969) which has been used extensively to compare narrative generation cross-linguistically in TD children (see Berman & Slobin, 1994). One study found that children with ASD showed awareness of the rules for introducing and re-introducing characters in a narrative (Norbury & Bishop, 2003). That is, they were as likely as a TD group to use a noun phrase, rather than a pronoun, to reintroduce a character when was not the most recently mentioned. All five studies, including Norbury and Bishop (2003), however, found evidence that children with ASD were more likely to use ambiguous reference than were typical controls.

Three of these studies also compared the ASD group to a group of individuals with DLD. Norbury and Bishop (2003) found that their ASD sample used more ambiguous nouns than did a DLD group matched for chronological age and language ability. The same pattern of results was found by Colozzo et al. (2015) study, whereby here the ASD group possessed superior formal language ability than the DLD group but still used a higher number of ambiguous character references than did the DLD group. Finally, in Norbury et al. (2014), the difference between the ASD group and the DLD group did not reach significance. However, there was a moderate effect size (d=.47) despite the fact that the ASD group in fact had significantly better formal language skills than the DLD group.

Only one of these five studies reported no differences between an ASD and TD group in production of ambiguous character reference (Mäkinen et al., 2014). This study was carried out with Finnish children aged 5-10 years . Groups were well-matched for chronological age but the TD group scored higher for formal language and memory. One reason for the lack of a between-groups difference in this study might be that the participants were Finnish-speaking and the authors note that in Finnish, TD children tend not to master accurate reference until 8 years of age.

All of the above five narrative generation studies were not well-matched to typical controls. Four additional studies examined narrative generation in comparison to well-matched controls. Three of these studies found impairments in individuals with ASD in comparison to typical controls. One of these studies focussed on adults and used the Mayer (1969) story 'Frog, where are you?' (Colle, Baron-Cohen, Wheelwright, & van der Lely, 2008), which was told to an experimenter who did not have visual access to the story pictures. Their sample with ASD used more ambiguous references to the dog and non-protagonist characters than did a TD group, despite being told that the listener had no previous knowledge of the story and that they should therefore be 'as clear as possible'. The other three studies tested upper-primary school aged children and adolescents. Two elicited narratives using a 29-page wordless picture book called 'Tuesday' by Weisner, (1991). These two studies reported that children and adolescents with ASD were more likely to use ambiguous reference in comparison to well-matched TD groups (Suh et al., 2014; Banney, Harper-Hill & Arnott, 2015). The third study required participants to tell four stories to an experimenter, who did not have visual access to the pictures. Each story had two characters of the same gender and which were specifically constructed to examine reference selection for character introduction, character maintenance and character reintroduction (Kuijper et al., 2015). In this study, there were no significance differences in appropriacy of reference

selection between the ASD and TD groups, despite large sample sizes (and despite the fact the ASD group scored significantly lower on the WISC ‘Vocabulary’ measure). However, the stories were much simpler than those used by the majority of narrative studies, both in terms of length (as each consisted solely of six pictures) and in terms of the amount of detail in each picture. This might have reduced both the working memory load of the task and the degree to which the individuals with ASD were likely to be distracted by irrelevant information. The potential issue of stimuli-dependent performance is one which will also emerge in next section.

Table 8: Production - *Narrative Generation (Stimulus Present)*

Author, year	*Participants*					*Matching*		*Task*	*Measure of Reference*	*RC impairment in ASD group*
	Sample size (n per group)	*Mean age (range) per group*	*Language spoken*	*ASD diagnosis*	*Structural Language*	*Age*	*Non-Verbal IQ*			
Not well-matched										
Norbury and Bishop (2003)	*n* = 12 (ASD) *n* = 18 (TD) *n* = 17 (SLI) *n* = 21 (PLI)	ASD: 8.8 TD: 8.56 SLI: 9.33 PLI: 8.92 (all within 6-10 yrs)	English	SCQ/ ADOS	✗ Comprehension: BPVS: ASD/TD: d=1.47 ASD/SLI: d= 0.40 TROG or CELF $_{RLC}$: ASD/TD: d=1.64 ASD/SLI: d= 0.52 Expressive Language $CELF_{RS}$: ASD/TD: d=2.83 ASD/SLI: d=0.07	✓ ASD/TD d=2.20 ASD/SLI d= 0.45	✓ RCM ASD/TD d= 0.31 ASD/SLI d= 0.58	Frog where are you?	Character reference (introduction, re-introduction & maintenance), including ambiguity.	✓
Makinen et al. (2014)	*n* = 16	HFA: 7;7 (5-10) TD: 7;5 (5-10)	Finnish	ADOS/ ADI-R	✗ TTFC d= 1.25	✓ d= 0.11	✗ NEPSY-II d= 1.21	The cat story	Character reference (including ambiguity)	✗
Norbury et al. (2014)	*n* = 22	ASD: 11:1 (6:5-15;71) TD: 9;8 (6;7-15;1) LI: 10;7 (6;5-15;3)	English	Met criteria for ASD based on DSM-IV	CELF ✓ ASD/TD: d=0.315 ✗ ASD/LI: d=4.02	✗ ASD/TD: d=0.47 ASD/SLI: d=0.12	✓ ASD/TD: d=0.52 ASD/SLI: d=0.22	A boy, a dog and a frog	Character reference (including ambiguity)	✓

Table 8. (Continued)

	Sample size (n per group)	*Mean age (range) per group*	*Language spoken*	*ASD diagnosis*	*Structural Language*	*Age*	*Non-Verbal IQ*	*Task*	*Measure of Reference*	*RC impairment in ASD group*
Colozzo et al. (2015)	*n* = 12	ASD: 8;5 (6-10) TD: 8;6 SLI: 8;5	English	Fulfilled diagnostic criteria for ASD according to experienced community-based clinicians	✗ Only ASD performance reported	✓ ASD/TD: d=0.07 ASD/SLI: d=0.04	✗ Only ASD performance reported	Late for School (wordless story)/ Aliens (single picture) from: Test of narrative language	Character, object and place reference.	Character ✓ Object ✗ Place ✗
Well-matched										
Colle et al. (2008)	*n* = 12	ASD: 27;5 (n/r) TD: 27;2 (n/r)	English	ADI/ ADOS	✓ WAIS d= 0.01	✓ d=0.02	✓ WAIS d= 0.01	Frog where are you?	Character (introduction, re-introduction, maintenance) and temporal references, including ambiguity.	✓
Rumpf et al. (2012)	*n* = 11	AS: 10;5 (8-12;11) TD: 9;11 ADHD: 9;9	German	Met DSM-IV criteria for autism	✓ MLU d=0.78	✓ d=0.41	✓ SD (n/r)	Tuesday	Character reference, overall use of nouns and pronouns.	✓
Suh et al. (2014)	*n* = 15	ASD: 12;9 (10;5-15;7) TD: 13 (9;9-15;6)	English	ADOS	✓ d= 0.66	✓ d=*0.16*	✓ d= 0.98	Tuesday	Character reference (ambiguous pronouns)	✓
Banney et al. (2015)	*n* = 11 (ASD) *n* = 17 (TD)	ASD: 11;6 (9-15) TD: (9-15;2)	English	ADOS	✓ CELF d= 0.63	✓ d= 0.27	✓ RCM d= 0.37	Tuesday	Character reference (introduction, re-introduction & maintenance), including ambiguity.	✓

Table 8. (Continued)

	Sample size (n per group)	*Mean age (range) per group*	*Language spoken*	*ASD diagnosis*	*Structural Language*	*Age*	*Non-Verbal IQ*	*Task*	*Measure of Reference*	*RC impairment in ASD group*
Kuijper et al. (2015)	*n = 46 (ASD)* *n = 37 (ADHD)* *n = 38 (TD)*	ASD: 9;3 (6;1-12;10) ADHD: 8;9 (6;1-12;10) TD: 9;0 (6;1-12;10)	Dutch	ADOS ADI-R	✓ PPVT ASD/TD d= 0.36 ASD/ADHD d= 0.28 × WISC Vocabulary ASD/TD d= 1.73 ASD/ADHD d= .30	✓ ASD/TD d= 0.14 ASD/ADHD d= 0.29	✓ WISC Block Design ASD/TD d= 0.38 ASD/ADHD d= 0.45	4 stories of 6 pics developed by Hendriks, Koster & Hoeks (2014	Specifically designed to measure appropriacy of character reference during a) introduction b) maintenance (where no competing reference vs. competing reference) and c) reintroduction	×
Matching not fully reported										
Tager-Flusberg (1995)	*n* = 10	ASD: 12;1 TD: 7;9 MR: 11;3	English	Met DSM-III-R criteria for autism	✓ PPVT ASD/TD d= 0.56 ASD/MR d= 0.06	ASD/TD× ASD/MR✓	- Not reported	Frog where are you?	Character reference (introduction & maintenance), including appropriateness of nouns/ pronouns.	✓ for character introduction × for 'referential style' n/s trend for TD to opt for more appropriate anaphoric style

Notes. AS: Asperger Syndrome; ASD: Autism Spectrum Disorder; BPVS: British Picture Vocabulary Scale; CELF: Clinical Evaluation of Language Fundamentals; $CELF_{RLC}$: Receptive Language Composite; $CELF_{RS}$: Recallling Sentences; HFA: High-Functioning Autism; HFA-F: female: HFA-M: Male; ITPA: Illinois Test of Psychological Abilities; MLU: Mean Length of Utterance; MR: Mentally Retarded; PLI: Pragmatic Language Impairment; PPVT: Peabody Picture Vocabulary Test; RCM: Raven's Coloured Matrices; SLI: Specific Language Impairment; TD: Typically Developing; TD-F: Female; TTFC: Token Test for Children; WAIS: Wechsler Adult Intelligence Scale; WISC: Wechsler Intelligence Scale for Children.

Narrative Re-tell

In contrast to narrative generation, narrative re-tell instead involves listening to and/ or viewing a story unfold, then retelling events from memory. Though not as widely used as generation tasks, four studies meeting our inclusion criteria utilised a re-tell method (see Table 9).

In the study by Arnold, Bennetto, and Diehl (2009), children and adolescents with and without ASD watched a Sylvester and Tweetie cartoon and then re- told this from memory to a confederate who feigned ignorance of the story. There was no narrator dialogue in the video clip. Instead participants simply watched events unfold. Each character reference was coded for recency of mention of the antecedent. If a referent was mentioned no more than two clauses back, the children with ASD (9;8 -12;9) used a significantly higher proportion of noun phrases (as opposed to pronouns) than did the typical controls, which the authors interpreted as over-informativity in this context. In contrast, the adolescents with ASD (13;1 - 17; 8) did not differ from a well-matched TD group in any of the measures used. However, Arnold et al. (2009) did not in assess the appropriacy per se of verbal reference selection; the latter does not solely depend on how many clauses back the antecedent was but rather, whether a referential alterative (e.g. Tweety Bird / Sylvester) was also recently mentioned and of course whether the pronouns are gender marked.

In addition, it is possible that the particular elicitation method / stimuli used partly accounts for discrepant findings between studies. To examine the extent to which elicitation method influences performance, Novogrodsky (2013) compared a narrative retell and generation task, analysing the ambiguity of third person subject pronouns. The same data was reanalysed by Novogrodsky and Edelson (2016), whereby they extended their analysis to include subject and object pronouns. The retell task was the ‘Bus Story’ task (Renfrew,

1991), which requires the child to retell a story, which has first been told to the child, about a bus that escapes from its driver. In this task, participants can look at the pictures as they retell the story. In the generation task, children told 'Frog, where are you?' (Mayer, 1969) from pictures, without an initial model. Whilst ASD and TD group performance did not differ in terms of ambiguous pronominal reference during the retold Bus Story, in the generation task the ASD group used significantly more ambiguous pronouns.

Unfortunately, due to the design of this study, there are many potential reasons why results may have differed depending on the particular elicitation paradigm. First, since the children had just heard the administrator tell the Bus Story, those with good auditory recall (which is often a relative strength in ASD) might simply have been able to select appropriate forms of verbal reference by recalling this ad verbum. Second, the 'Bus Story' only consists of 12 pictures and thus it could be that the relative simplicity of the story allowed more accurate use of reference.

The fourth study of narrative retell was conducted by de Marchena and Eigsti (2016), who used 60-second cartoon clips. This study differs from the other narrative studies in that listener informational needs were specifically manipulated by having two within-subjects conditions; 'shared' (the listener watched a short preview of the clip with the participant) and 'private' (the listener was not present during any part of the clip). Some aspects of de Marchena and Eigsti's (2015) data indicate that the adolescents with ASD considered listener information needs to a degree; there was a significant difference in communicative quality ratings (i.e. a rating of how easy the story was to follow) between narratives produced by the ASD group in the shared versus the private condition. However, for the key dependent variable, the degree of referential shortening, there was a between-groups difference. The authors argue that the referential shortening effect is a measure of whether participants take audience needs into account. The argument is that, if speakers take audience needs into

account, their narratives should be shorter when retelling in the shared as opposed to the private condition. This effect was seen for the typical control but not the ASD group, indicating that the latter had difficulty adapting to listener information needs. However, de Marchena and Eigsti's analysis rests on the assumption that a longer narrative would contain a greater number of full noun phrases or indeed noun phrases with modifying phrases (see Table 6). This is of course not necessarily the case since a proper noun (e.g. Laura) is usually highly informative and yet does not differ in word length from a pronoun. Conversely, not all modifying phrases provide sufficient differentiating information. The extent to which reference selection was appropriate for a given context was not examined.

Nonetheless, there was a significant relationship between the referential shortening effect and symptom severity as measured by the Social Responsiveness Scale (Constantino & Gruber, 2007) in the ASD group, whereby those more likely to demonstrate the effect showed less ASD traits. This supports the authors' conclusion that the referential shortening effect taps some of the social communicative deficits which are diagnostic for ASD. Older children with ASD were also more likely to show the referential shortening effect than those who were younger, tying in with Arnold et al.'s (2009) finding that selection of appropriate reference may improve with age in the ASD population.

Table 9 Production - *Narrative Re-tell*

Author, year	*Participants*				*Matching*			*Task*	*Measure of Reference*	*RC impairment in ASD group*
	Sample size (n per group)	*Mean age (range) per group*	*Language*	*ASD Diagnosis*	*Structural Language*	*Age*	*Non-Verbal IQ*		*Statistical Analysis / Inter-rater Reliability*	
Arnold, Bennetto & Diehl (2009)	*n* = 10 (ASD-Y) *n* = 13 (ASD-O) *n* = 10 (TD-Y) *n* = 13 (TD-O)	ASD-Y: 11;6 (9;8-12;9) ASD-O: 15;1 (13;1-17;8) TD-Y: 11;6 (9;8-12;9) TD-O: 14;6 (13;1-17;8)	English	ADOS/ ADI-R	✓ VIQ: WISC/WAIS ASD-Y/TD-Y d= 0.25 ASD-O/TD-O d= 0.20 Receptive Language: PPVT: ASD-Y/TD-Y d= 0.24 ASD-O/TD-O d=0	✓ ASD-Y/ TD-Y d= 0.53 ASD-O/ TD-O d= 0.44	✓ WISC/WAIS ASD-Y/TD-Y d=0 ASD-O/TD-O d= 0.10	Sylvester & Tweety Cartoon Re-tell video (no audio) from memory	References to characters: noun (Sylvester, the cat) pronoun (he, it), zero (. . .and ϕ ran) Number of clauses since the most recent mention of the same character (1, 2, 3 clauses back, or no prior mention). ANOVA % reliability	Younger: ✓ Older: ×
Novogrodsky (2013)	*n* = 23 (ASD) *n* = 17 (TD)	ASD: 10 (6;1-14;3) TD: 9;9 (5;11-14;4)	English	ADOS/ ADI-R	✓ Woodcock-Johnson III d= 0.30	✓ d= 0.11	Not reported	1. Re-tell: Bus Story (after having heard an adult tell story) 2. Generation: Frog where are you?	Third person subject pronouns, including ambiguity.	Re-tell × Generation ✓
Novogrodsky & Edelson (2016): re-analysing data from Novogrodsky (2013)	*n* = 24 (ASD) *n* = 17 (TD)	ASD: 10 (6;1-14;3) TD: 9;9 (5;11-14;4)	English	ADOS/ ADI-R	✓ Woodcock-Johnson III d= 0.30	✓ SD not reported	Not reported	1. Re-tell: Bus Story 2. Generation: Frog where are you?	Use of subject, object and possessive pronouns, including ambiguity. ANOVA IRR not reported	Re-tell × Generation (subject & possessive ambiguity) ✓ (object) × n.b. relatively low occurrence of object pronouns for both groups

Table 9(Continued)

	Sample size (n per group)	*Mean age (range) per group*	*Language spoken*	*ASD diagnosis*	*Structural Language*	*Age*	*Non-Verbal IQ*	*Task*	*Measure of Reference*	*RC impairment in ASD group*
De Marchena & Eigsti (2015)	*n* = 19	ASD: 14;10 (12;7-16;11) TD: 15;4 (12;2-17;11)	English	ADOS/ SCQ/ SRS	✓ Receptive Vocabulary PPVT p=.09 *(marginal)* d= 0.55 Expressive: Stanford-Binet Verbal d=0	✓ d=*0.35*	✓ Stanford-Binet NV d= 0.49	Cartoon clips Re-tell video (no audio) from memory	Referential shortening ANOVA Excellent IRR	✓

Notes. ASD: ADI-R: Autism Diagnostic Interview-Revised; ADOS: Autism Diagnostic Observation Schedule; Autism Spectrum Disorder; ASD-O; Older; ASD-Y; Younger; MLU: Mean Length of Utterance; PPVT: Peabody Picture Vocabulary Test; SCQ: Social Communication Questionnaire; SRS: Social Responsiveness Scale; TD: Typically Developing; TD-O; Older; TD-Y: Younger; VIQ: Verbal IQ, WAIS: Wechsler Adult Intelligence Scale; WISC: Wechsler Intelligence Scale for Children.

Referential Communication Tasks

Overall narrative elicitation methods, the overwhelming tendency indicates an impairment in the appropriate usage (production) of verbal reference. However, it might be argued that the difficulties individuals with ASD experience with narrative tasks are not related to deficits in the production of appropriate verbal reference per se, but instead are related to extraneous demands required by these tasks. Individuals with ASD may be particularly hindered in narrative tasks by the need for episodic memory (e.g. Lind, Williams, Bowler, and Peel, 2014), executive functioning (e.g. Geurts, Verté, Oosterlaan, Roeyers, & Sergeant, 2004) imagination (e.g. Lind et al., 2014), and central coherence (e.g. Happé & Frith, 2006). Therefore, an elicitation method which does not burden episodic memory and which mirrors the back and forth nature of conversation might be better able to reveal underlying latent ability in individuals with ASD to use verbal reference appropriately. One such method is the referential communication task.

Referential communication tasks allow both the production and comprehension of referring terms to be measured. Here, we first review the results of studies where referential communication tasks have been used to examine the production of referring terms. Studies which examined the interpretation of reference are discussed later in this review. Our search returned five studies involving a type of referential communication paradigm to examine the appropriacy of referring terms selected in production.

A very frequently used type of referential communication task that allows visual common ground to be manipulated is the ‘Director Task’ (Keysar, Barr, Balin, & Brauner, 2000). Typically, an array of items or pictures is presented to a speaker, or ‘director’ who provides relevant information to a listener to enable them to select a target referent. The inclusion of items that differ in only one aspect e.g. ‘big cup’ versus ‘small cup’ creates the

need to alter the specificity of the referring terms selected. The visual perspective of the listener can be manipulated to either match or be discrepant with that of the speaker by blocking visual access to some referents in the array. Two potential conditions are therefore possible: A 'shared' condition, in which all potential referents can be viewed by both parties (i.e. visual perspective taking is not required), and a 'privileged' condition, in which the speaker has visual access to referents which are blocked from the view of the listener (i.e. visual perspective taking is required). In the shared condition, the use of a size adjective (e.g. 'big cup') is appropriate as the listener has visual access to two possible referents, whilst in the privileged condition a bare noun (e.g. 'cup') is appropriate as visual access to the competing cup is blocked from the listener. Table 10 shows studies utilising this paradigm to examine production of referring terms meeting our inclusion criteria.

Using a director task, Nadig, Vivanti and Ozonoff, (2009: Exp 1) found that children with ASD aged 9-14-years used proportionally fewer appropriate referring terms to identify objects than a well-matched TD group (both groups: n=17). In the privileged condition, participants with ASD tended towards over-informativity, inappropriately using a specific referring term (e.g. 'big cup' when there was only one cup available from the listener's visual perspective) significantly more frequently than the TD group ($p < .01$). In the shared condition, the ASD group more frequently failed to use a complex referring term when two competing referents were visible, though this group difference was only of marginal significance ($p = .08$, effect size $r = 0.24$). These findings reflect the simultaneous over and under-informativity in reference use by individuals with ASD which was also the general finding from narrative and conversational studies.

Fukumura (2015) used a similar director task whereby she directly compared the 'privileged' and 'shared' perspective conditions. The dependent variable was the percentage of complex referring expressions (e.g. "the small door") as opposed to unmodified nouns

(e.g. “the door”). Thus, if individuals with ASD were taking listener informational needs into account, there should be significantly less complex referring expressions used in the ‘privileged’ condition, since this would be over-informative from the addressee’s perspective. For both 6-10-year-olds (Exp 1) and 11-16-year-olds (Exp 2), there was a group by condition interaction, indicating that the typical controls were significantly more likely to make this audience design distinction than were the individuals with ASD.

Nadig et al. (2009) and Fukumura’s (2016) studies indicate that when the speaker and listener perspectives differ, individuals with ASD have difficulty selecting referring expressions appropriate to their listener’s perspective. However, even when participants know that their listener can see the same visual array as themselves, Nadig et al. (2009) found diminished performance for ASD groups. This latter observation is reflected in the findings of two referential communication studies which did not manipulate listener perspective. Both used adaptations of the original reference communication paradigm developed by Glucksberg and Krauss (1967) in which the participant and a confederate play a version of the ‘Guess Who?’ game. Volden, Mulcahy, and Holdgrafer (1997) asked adolescents and adults with ASD to provide information to identify a target from one of two circles which varied on one of four possible attributes (colour, shape, pattern and position of a small black dot). Whilst individuals with ASD never failed to provide the distinguishing feature in their description, they were more likely to include redundant information that did not uniquely identify the target referent.

Using a similar paradigm, Dahlgren and Dahlgren Sandberg (2008) asked children with ASD to provide descriptions to identify a given face from a selection of 16. They measured how many of the features mentioned were ‘relevant’ (appropriately discriminated between pictures), ‘irrelevant’ (common to all pictures e.g. ‘has a mouth’) and ‘redundant’ (already a given). Children with ASD produced significantly fewer relevant features than did

TD controls and they also included proportionally more irrelevant than relevant features than the TD group.

Whilst director tasks have been used to manipulate visual common ground knowledge, social common ground (namely the ability to determine the knowledge what one shares with a specific interlocutor; Moll & Kadipasaoglu, 2013) is arguably the skill used more often when selecting referring terms in everyday conversation. In a 'referential pact' paradigm, Nadig, Seth, and Sasson (2015) examined whether adults with ASD engaged in lexical entrainment - the process by which interlocutors come to agree on mutual referring terms. Participants provided information to enable their listener to identify one of an array of abstract forms (tangrams). Individuals tended to alter referential descriptions in co-operation with the listener over successive trials (e.g. pairs may agree to call a shape 'the elephant' after initially describing it as 'a four legged or two legged animal facing the right… The head is a parallelogram and its back leg is a rectangle and the front legs look like paws'). To investigate whether this alignment of referring terms was due merely to priming or if social common ground was utilised, the game continued with either the original or a new listener. If common ground was considered, the agreed referring terms should be used with the original, but not a new listener. In the 'new listener' condition, the ASD group were marginally ($p = .05$, $r = .37$) less likely than the TD group to change the referring expression (referential pact) they had agreed with the original listener.

Table 10 Production - *Referential Communication Tasks*

Author, year	*Participants*				*Matching*			*Task*	*Measure of Reference*	*RC impairment in ASD group*
	Sample size (n per group)	*Mean age (range) per group*	*Language*	*ASD diagnosis*	*Formal Language*	*Age*	*Non-Verbal IQ*		*Statistical Analysis / Inter-Rater Reliability*	
Nadig, Vivanti & Ozonoff (2009)	*n* = 17	ASD: 11;3 (9;6-14;6) TD: (8-14)	English	SCQ/ADOS	✓ CELF d= 0.24	✓ d= 0.36	✓ PIQ WASI d= 0.26	Director task with objects	Use of size adjective in shared/privileged condition Mann-Whitney U No IRR	✓
Fukumura (2016)	*n* = 20	ASD-Y: 8;8 (6;8-10;9) ASD-O: 13;7 (11;8-16;3) TD-Y: 8;3 (6;3-10;5) TD-O: 13;7 (12;1-16;4) TD-Adult: 21 (18-23)	English	SCQ	✓ WASI Vocabulary d=0.26	✓ d=0.37	✓ WASI Matrix reasoning d= 0.23	Director task with pictures	Use of size adjective in shared/privileged condition Logit mixed effects modelling. No IRR	‘Shared’ condition: × ‘Privileged’ condition: Younger: ✓ Older: × (marginal p=.07)
Volden et al. (1997)	*n* = 10	ASD: 18;8 (13;6-24;4) TD: 18;8 (13;1-24;4)	English	Met DSM-III-R diagnostic criteria for autism	Reported as ‘similar’	✓ d=0	Not reported	‘Guess Who’ type task: provide description to identify one of two circles. VPT not required	Contrastive (i.e. correct), redundant (distinguishing feature + at least one other), uninformative (did not include distinguishing attribute) t-test, Fisher’s Exact % reliability	✓

Table 10. (continued)

	Sample size (n per group)	*Mean age (range) per group*	*Language spoken*	*ASD diagnosis*	*Structural Language*	*Age*	*Non-Verbal IQ*	*Task*	*Measure of Reference*	*RC impairment in ASD group*
Dahlgren & Sandberg (2008)	*n* = 30	ASD: 10.06 (7.58-14.6) TD: 9.55 (7-13.92)	English	Met DSM-III-R diagnostic criteria for autism	✓ WISC VIQ d= 0.35	✓ d=0.28	× FSIQ d=0.60 ✓ WISC PIQ d=0.27	'Guess Who' type task: Provide description to identify one of 16 faces. VPT not required.	Relevant/ irrelevant/ redundant features mentioned Mann-Whitney / Wilcoxon / Spearman's No IRR	✓
Nadig, Seth & Sasson (2015)	*n*=13	ASD: 22;2 (18-29) TD: 21;2	English	ADOS/ SCQ	✓ WASI VIQ d= 0.21	✓ d=0.33	Not reported	Referential pact paradigm with tangrams	Lexical alignment with original/ new partner ANOVAs Cohen's k	Initial lexical entrainment × Adaptation to new partner ✓

Notes. ADOS: Autism Diagnostic Observation Schedule; AS: Asperger Syndrome; ASD: Autism Spectrum Disorder; ASD-O; Older; ASD-Y; Younger; CELF: Clinical Evaluation of Language Fundamentals; DSM: Diagnostic and Statistical Manual of Mental Disorders; FSIQ: Full Scale IQ; PIQ: Performance IQ; SCQ: Social Communication Questionnaire; TD: Typically Developing; VIQ: Verbal IQ, WAIS: Wechsler Adult Intelligence Scale; WISC: Wechsler Intelligence Scale for Children.

Studies comparing ASD with other neuro-developmental disorders

Some studies returned in our search compared groups of individuals with ASD to groups of children with other neuro-developmental disorders. For five studies the disorder concerned was DLD. Such comparisons between ASD and DLD can help elucidate the degree to which formal language (lexical and morpho-syntactic skills) might be a contributing factor in proficiency with verbal reference.

One such study is that of Manolitsi and Botting (2011) (Table 11), who found that children with ASD were significantly poorer in pronominal reference use than were children with DLD during a narrative re-tell task (‘Peter and the Cat, Leitão & Allan, 2003). Both groups were Greek-speaking children matched on expressive language, chronological age, gender and non-verbal IQ. Participants were categorised on a scale of 0 to 3. A child categorised a 0 would generally show great difficulty with appropriate selection of referring expressions whereas a child categorised as 3 would be a proficient user of verbal reference. There was a significant difference between the ASD and DLD group in this regard, with children with DLD showing better performance. Unfortunately, the ASD group also had lower receptive language than did the group with DLD, which makes any between-group differences in measures of verbal reference difficult to interpret.

The remaining four DLD-comparison studies have already been mentioned above, since they also included a typical control group (Baltaxe & D’Angiola, 1996; Norbury & Bishop, 2003; Colozzo et al., 2015; Norbury et al., 2014). All four found significant difficulties (or effect size indicative of a difference, Norbury et al., 2014), whereby the ASD group performed worse than the DLD group. This is particularly striking in the case of the two DLD comparison studies in which the group with ASD had higher formal language scores than the group with DLD (Colozzo et al., 2015; Norbury et al., 2014). If the ASD

group still showed significantly greater difficulties in verbal reference, this provides somewhat stronger evidence that formal language is unlikely to be the main cause of these pragmatic language difficulties.

In sum, the results of all five studies which compared ASD to DLD suggest that, although referential accuracy poses a challenge for children in both groups, deficits in referential communication are more pronounced in individuals with ASD, even when the latter have superior lexio-syntactic abilities. Thus, referential communication deficits in ASD are unlikely to be solely attributable to difficulties with formal language.

In addition to cross-syndrome comparisons with DLD, two studies returned in our search compared the use of reference in ASD and Down Syndrome (DS), which is a neuro-developmental disorder associated with intellectual disability (see Table 11). The first study is that of Loveland, McEvoy, Tunali, and Kelley (1990) who tested children with ASD and children with DS (n=16 in each group) matched on verbal mental age. Children were asked to re-tell a story depicted via a video or puppet show to a naïve listener. In each group an equal proportion of children made ambiguous references to characters. Whilst Loveland et al. (1990) aimed for participants to have 'similar' non-verbal IQ and chronological age, the ASD group had marginally higher mean IQ scores and chronological age than the DS group.

In a less structured task, again comparing reference use in ASD and DS groups, Loveland, Tunalia, Mcevoy, and Kelley (1989) asked participants to provide information to a naïve listener (E2) about how to play a board game. Participants were helped by E1 to provide adequate information using a gradient of prompts from more general, for example, 'Tell me about these things here' to more specific, for example, 'Tell me where to start the game'. The ASD group produced significantly less 'adequate' descriptions than did the DS group at the most 'general' level of prompting and they also required a higher level of

Table 11 Production - *Studies comparing ASD with another clinical group (where no typically-developing group was included in the study)*

Author, year	*Participants*				*Matching*			*Task*	*Measure of Reference*	*RC impairment in ASD group*
	Sample size (n per group)	*Mean age (range) per group*	*Language*	*ASD diagnosis*	*Structural Language*	*Age*	*Non-Verbal IQ*		*Statistical Analysis* *Inter-Rater Reliability*	
Loveland et al. (1989)	*n = 13 (ASD)* *n = 14 (DS)*	ASD: 16;2 (8-27) DS: 15;2 (12;3-18;7)	English	Not reported	✓ McCarthy's Scale/ PPVT d= 0.26	✗ d= 0.20	✗ LIPS d= 1.07 *(ASD higher)*	Explain rules of 'zoo game' to partner	Amount of appropriate information produced and level of prompting required to elicit this t-tests % agreement	✓
Loveland, Mcevoy, & Tunali (1990)	*n* = 16	ASD: 13;6 (5-27) DS: 13;3 (5-27)	English	Vineland adaptive behaviour scales	✓ McCarthy's Scale/ PPVT d= 0.19	✓ d= 0.05	✗ LIPS d= 0.81 *(ASD higher)*	Retell narrative (presented as video or puppets) to 'naïve' addressee	Character and event reference, including ambiguity. Chi-square % agreement	✗
Manolitsi and Botting (2011)	*n* = 13	ASD: 7;2 (4;2-13) SLI: 7;4 (5-13)	Greek	Diagnosed by child psychiatrist following American Psychiatric Association (2000) criteria	✗ CELF-Receptive d= 0.97 ✓ CELF-Expressive d= 0.56	✓ d= 0.18	✓ d= 0.44	Story: *Peter and the cat* (Leitao and Allan, 2003)	Character reference ANOVA Cohen's k	✓

Notes. ASD: Autism Spectrum Disorder; CELF: Clinical Evaluation of Language Fundamentals; DS: Down's Syndrome; LIPS: Leiter International Performance Scale; PIQ: Performance IQ; PPVT: Peabody Picture Vocabulary Test; SLI: Specific Language Impairment; TD: Typically Developing.

specific prompts than did the DS group to provide the an adequate amount of information. This was the case even though the two groups were matched on verbal age and although the ASD group tended towards higher non-verbal IQ scores than the DS group. Given that the ASD group had overall higher IQ in both studies, yet exhibited difficulties equal to, or more pronounced than, a DS group, these studies suggest that the ability to develop appropriate usage of verbal reference may not be due solely to latent non-verbal intellectual difficulties.

Production Summary

Overall, our search returned 22 studies of verbal reference production in ASD. There were seven studies which compared a group with ASD to a group with another neuro-developmental disorder (either DLD or DS) and all but one of these studies found indications of poorer performance by the ASD group.

Eleven studies compared a group with ASD to a group of typical controls, whereby groups were either not well-matched for formal language and / or nonverbal IQ ability, or this was not reported (Baltaxe & d'Angiola, 1996; Norbury & Bishop, 2003; Norbury et al., 2014; Colozzo et al., 2015; Tager-Flusberg, 1995; Mäkinen et al., 2014; Novogrodsky, 2013; Novogrodsky & Edelson, 2016; Volden et al., 1997; Dahlgren & Dahlgren Sandberg, 2008; Nadig et al., 2015). Nonetheless, it is noteworthy that 10 out of these 11 studies reported that individuals with ASD performed significantly worse than typical controls on at least one reference measure.

Finally, eight studies did compare a group with ASD to well-matched controls and all except one (Kuijper et al., 2015) of these well-matched case-control studies found evidence of a deficit in comparison to the typical group in terms of appropriacy of verbal reference usage (Colle et al., 2008; Suh et al., 2014; Banney et al., 2015; Arnold et al., 2009; de

Marchena & Eigsti, 2015; Nadig et al., 2009; Fukumura, 2016). These latter seven studies include a range of age groups. They also include a range of elicitation methods, namely narrative generation (Colle et al., 2008; Suh et al., 2014; Banney et al., 2015), narrative re-tell (Arnold et al., 2009), the 'director' task (Nadig et al., 2009; Fukumura, 2016) and interlocutor-specific perspective-taking (de Marchena & Eigsti, 2015). Therefore, it is safe to conclude that there is very good evidence for a clear impairment in appropriate reference selection (production) in ASD.

3.4. Comprehension of Referring Expressions

In contrast to the ample number of studies examining the production of referring terms in ASD, we found only three studies that compared ASD to another group in terms of comprehension of the same phenomena (see Table 12). All three suggest that the pattern of ability differs considerably between the production and comprehension of referring terms. Although not examining interpretation of referring expressions per se, Volden et al. (1997) examined the ability of the adolescents and adults with ASD to judge whether the addressee in a referential communication paradigm had sufficient information to be able to correctly identify the referent. The authors argue that their ASD group performed at ceiling on this meta-pragmatic judgement task. In fact the ASD group were correct on average 87% of the time but the typical group were correct 100% of time (with an SD of zero), with the result that statistical analyses were not carried out. Moreover, since the Glucksberg and Krauss (1967) paradigm was used (where the participant and the confederate are aware that they are viewing identical sets of cards), it could be argued that the ability to take another's perspective was not necessary for this task since the participant merely has to judge whether a confederate's instruction is informative from his or her own perspective.

The final two studies did in fact investigate performance in reference interpretation where the participant's perspective differed from that of the speaker. Both used the director task. In contrast to the other 'director' studies already outlined, here the participants were in the role of the addressee. In key (ambiguous) condition trials, each participant is instructed to pick up an object (e.g. spoon) for which the participant (but crucially, not the 'director') can see a referential alternative (e.g. another spoon). One dependent variable is thus the number of egocentric errors made; i.e. the number of trials on which a participant selects the object which is occluded from the director's view and thus cannot be the intended referent. A second dependent variable is typically response latency. That is, the longer a participant takes to select the correct object is an indication of the degree to which he or she (egocentrically) considered the referential alterative as a possible target.

In the first such study, Begeer, Malle, Nieuwland, and Keysar (2010) examined the ability of adolescents with ASD to interpret referring expressions (e.g. 'the cup' versus 'the big cup') when responding to instructions in a shared and a privileged condition. Across both groups, participants made egocentric errors on 39% of trials in the key (ambiguous) condition and their response latencies were also significantly longer in the ambiguous than in the neutral condition, indicating that they considered the referential alternative prior to making correct selections. However, crucially, there were no between-group differences for either of these dependent variables. This indicates that the ASD group were as able as the TD group to use visual perspective-taking to interpret verbal reference, at least when the visual perspective is as simple as determining whether the interlocutor can see a particular object. Santiesteban, Shah, White, Bird, and Heyes (2015) carried out a computerised version of the director task with adults. Similarly to Begeer et al., (2010), they found no between-group differences. Moreover, they found that adults with and without ASD were equally successful in completing the task when a human addressee (avatar) was replaced with a camera.

Table 12: Comprehension - *Referential Communication Task*

Author, year	*Participants*				*Matching*			*Task*	*Measure of Reference*	*RC impairment in ASD group*
	Sample size (n per group)	*Mean age (range) per group*	*Language*	*ASD diagnosis*	*Formal Language*	*Age*	*Non-Verbal IQ*		*Statistical analysis / Inter-rater reliability*	
Volden et al. (1997)	*n* = 10	ASD: 18;8 (13;6-24;4) TD: 18;8 (13;1-24;4)	English	Met DSM-III-R diagnostic criteria for autism	Reported as 'similar'	✓ d=0	Not reported	Listen as description given of circle referent	Ability to judge when sufficient information given to identify referent No statistical analyses carried out .(Individuals with ASD correct 87% of time, TD individuals correct 100% of time, SD = 0). % agreement	×
Begeer et al. (2010)	*n* = 34	ASD: 16;7 TD: 16;8	English	Met DSM-IV-TR diagnostic criteria for autism	✓ WISC/WAIS Vocabulary d=0.14	✓ d=0.02	✓ WISC/WAIS Arithmetic/ Block Design/ Picture Arrangement d=0.14	'Real life' version of the Director Task, VPT manipulated	Egocentric errors on ambiguous trials Response latency ANCOVA Perfect IRR	×

Table 12 (continued)

	Sample size (n per group)	*Mean age (range) per group*	*Language spoken*	*ASD diagnosis*	*Structural Language*	*Age*	*Non-Verbal IQ*	*Task*	*Measure of Reference*	*RC impairment in ASD group*
Santiesteban et al. (2015)	*n* = 20 (ASD) *n* = 18 (TD)	ASD: 36 TD: 41	English	ADOS	✓ WAIS d=0.15	✓ d=0.39	✓ WAIS d=0.15	Computerised version of the Director Task with person/ camera control. VPT manipulated.	Egocentric bias in object selection Number of 100ms fixations on the competitor object ANOVA No IRR	×

Notes. ADOS: Autism Diagnostic Observation Schedule; ASD: Autism Spectrum Disorder; DSM-III- Diagnostic and Statistical Manual of Mental Disorders, 3rd Ed; DSM-IV- Diagnostic and Statistical Manual of Mental Disorders, 4th Ed; TD: Typically Developing; WAIS: Wechsler Adult Intelligence Scale; WISC: Wechsler Intelligence Scale for Children.

Thus, suggesting that the ability to take another's perspective to accurately interpret reference is relatively spared in ASD.

3.5. Potential Cognitive Underpinnings

The picture emerging from studies on the interpretation of verbal reference is that this is not an area of impairment in individuals with ASD (Begeer et al., 2010; Sanstieban et al., 2015). This stands in stark contrast to the overwhelming finding that individuals with ASD are impaired relative to both typical peers and peers with neurodevelopmental when reference production is examined. One possible reason for the apparent discrepancy between an impairment in the selection of an appropriate referring expression (production) and an intact ability to take another's perspective to interpret a referring term might be the differing cognitive underpinnings of each skill. We now therefore survey studies which explicitly examined relations between referential communication in ASD and the potential cognitive underpinnings of this skill.

We begin by examining studies which have examined relationships between proficiency with verbal reference, on the one hand and either formal language (lexical or syntactic) proficiency and / or non-verbal IQ on the other in the samples of individuals with ASD.

Non-verbal IQ

Three studies (all discussed above) examined the relationship between non-verbal IQ and the appropriacy of reference selection. Nadig et al. (2009) found that performance in the shared perspective condition correlated with non-verbal IQ; that is, those with higher non-verbal IQ used more adjectives when they (and their interlocutor) could see two referential alternatives (e.g. two ducks) than when only one potential referent was present. However, in the privileged perspective condition there was no relationship with non-verbal IQ, which

makes the first finding difficult to interpret. In line with this latter finding, both Dahlgren and Dahlgren Sandberg (2008) and Fukumura (2016) did not find any evidence for a relationship between non-verbal IQ and any measures of reference production in their ASD groups. Thus, on the whole, these findings - when considered together with the studies outlined above comparing children with ASD to children with DS – indicate that non-verbal IQ is unlikely to play a primary causal role in difficulties with verbal reference in ASD (although analyses in future studies should certainly control for non-verbal IQ).

Formal language

Three studies (all discussed above) examined the relationship between formal language and the appropriacy of reference selection. All three studies used a referential communication paradigm. Dahlgren and Dahlgren Sandberg (2008) found that verbal IQ correlated in the ASD group (but not in the TD group) with the number of relevant features mentioned and their measure of referential efficiency. However, since they did not manipulate the distinction between the participant's and the interlocutor's perspectives, it is unclear whether this indicates that formal language is important for the appropriacy of reference selection or whether it merely suggests that a more advanced mastery of formal language leads to a greater complexity of referring expressions.

The latter interpretation is supported by Fukumura (2016), who found no relationship between formal language and performance in the privileged ground condition in her ASD groups. Rather, the only relationships with formal language (British Picture Vocabulary Scale and WASI vocabulary) were with the number of adjectives produced by the ASD group in the shared ground condition. That is, those children with ASD who had larger vocabularies tended to produce more adjectives in the shared ground condition. Since the shared ground condition does not differentiate the participant's own perspective from that of the

interlocutor, this finding merely indicates that those individuals with ASD who have larger vocabularies tend to find it easier to produce complex referring expression. In contrast, Nadig et al. (2009) found that formal language ability (CELF) correlated with appropriately informative verbal reference by participants with ASD in the 'privileged view' condition of their director task, i.e. the condition which required participants to take the addressee's perspective, since it differed from their own. That is, those with higher scores on the CELF were more able to curtail the usage of complex referring expressions when this would be over-informative. In sum, it seems likely that formal language contributes to difficulties with the production of appropriate referring expressions in children with ASD. However, given that comparisons with DLD indicated that difficulties in verbal reference production are more marked in ASD, despite better formal language skills in the latter group, it appears likely that other factors may contribute to the observed impairment.

Theory of Mind and Executive Functioning

Traditionally, difficulties with appropriate verbal reference selection have been linked to difficulties with Theory of Mind (ToM), which is the ability to represent others' mental states including their beliefs, emotions, and desires (e.g. Baron-Cohen, Leslie & Frith, 1985). However, it is equally plausible that a failure to provide an appropriate level of information (i.e. under- and over-informativity) could be due to a failure to differentiate between old and new information during a verbal interaction (e.g. Baltaxe, 1977). Such difficulties may be caused in part by an impairment, for example, in working memory. Working memory is usually considered one component of executive functioning, which comprises a set of highly correlated, but separable, aspects of memory, inhibitory control and cognitive flexibility needed for considering consequences to actions (e.g., Miyake, et al., 2000; see also Pennington & Ozonoff, 1996 for an overview of EF domains and measurement methods).

Given that several reviews and meta-analyses report clear evidence for impairments in all domains of executive functioning bar inhibitory control in ASD (e.g. Hill, 2004; Russo et al. 2007; Lai et al., 2016) and there is evidence of a link between EF and verbal reference in TD populations (e.g. Brown-Schmidt, 2009), it is somewhat surprising that only two studies examined relationships between EF and usage or interpretation of verbal reference in ASD. Both studies are also the only two to examine the relationship between the appropriacy of reference selection and Theory of Mind in ASD. The first study, carried out by Dahlgren and Dahlgren Sandberg (2008), included only two tasks which might plausibly be considered a measure of executive functioning and both of these measured short memory. The first was 'verbal free recall' (11 lists of words with 10 words in each) and the second was 'object free recall' (in which the child is shown 10 sets of 10 objects and is required to verbally recall them). In both memory tasks, once all items had been presented, the child was asked to repeat as many words (or objects, respectively) as he or she could remember and in any order. Relationships were found between both memory measures and certain aspects of verbal reference, namely the number of relevant features mentioned and the 'efficiency' of reference usage, that is, the extent to which descriptions were optimally informative (for the comparison group this was only significant for verbal free recall). The authors interpret this as indicating that working memory impacted on the number of referential alternatives which a child could hold in mind and possibly also on the ability to verbally encode the relevant distinguishing information

Dahlgren and Dahlgren Sandberg (2008) also directly examined the relationship between ToM and the usage / interpretation of verbal reference. To measure ToM they used a first-order change-of-location task (Baron-Cohen et al., 1985) as well as Baron-Cohen's (1989)'s second order false belief 'ice-cream' task. For the second-order ToM measure no significant relationships were found with any verbal reference measures for either group. For

the children with ASD there was a relationship between first order ToM and the same aspects of verbal reference used (number of relevant features mentioned and the 'efficiency') that correlated with free recall. They note, however, that the correlation with first order ToM is based only on five children in the ASD group who failed the first-order theory of mind task (whereas 25 children with ASD passed). More problematically, Spearman's rho was used for all correlational measures, when a point-biserial correlation is appropriate for the first order ToM task which was essentially a pass / fail measure.

The other study which examined relationships in ASD between appropriacy of reference selection, on the one hand, and either EF or ToM, on the other, is Kuijper et al., (2015). They used the Stop Signal Reaction Time Task (Van den Wildenberg & Christoffels, 2010) to measure inhibitory control and the n-back task to measure working memory. First and second order Theory of Mind was assessed in a scale consisting of eight stories (Hollebrandse, Van Hout & Hendriks, 2014). For the 'reintroduction of character in a narrative' condition (where a noun and not a pronoun would be appropriate), the authors found in a multivariate model relationships between reference usage and both second order ToM and working memory. Unfortunately, language measures were not entered into the analysis and the results were conflated over three groups, which included a group with Attention Deficit Hyperactivity Disorder, making this finding difficult to interpret.

Based on the studies included in this review, it appears there is insufficient evidence to determine whether the development of verbal reference usage in ASD is underpinned by ToM and / or EF. The degree to which ToM and EF underpin the development of verbal reference in ASD is complicated by the fact that these two areas tend to be inter-correlated with each other (e.g. Pellicano, 2013) and also with formal language (e.g. Milligan, Astington & Dack, 2007, for a review). However, considering the evidence for relationships between both EF and ToM and other areas of pragmatics (see e.g. Matthews, Biney & Abbot-Smith,

2018) further exploration of the cognitive underpinnings related to comprehension and production of referring terms is clearly a priority. In chapter 5, we explore the role of one component of EF, namely, cognitive flexibility, upon pragmatic language ability in autistic and non-autistic children. See also section 6.3 of the general discussion for a lengthier discussion of the cognitive underpinnings of pragmatic language in autism more broadly.

3.6. Summary and discussion

The current systematic review found 19 studies which met our criteria and in which verbal reference production by a group of individuals of ASD was compared to that used by a typically-developing control group. Seventeen of these 19 studies found that the group with ASD were impaired in at least one measure in terms of the appropriacy of match between context and the form of verbal reference. While many of these studies had various methodological issues, this pattern of results also held for seven of the eight studies in which the typical group were matched to the ASD group in terms of chronological age, non-verbal IQ and formal language (Colle et al., 2008; Suh et al., 2014; Banney et al., 2015; Arnold et al., 2009; de Marchena & Eigsti, 2015; Nadig et al., 2009; Fukumura, 2016). This stands in stark contrast to the findings from the three studies of verbal reference comprehension, in which individuals with ASD were observed to show typical understanding/interpretation of verbal reference. , This was even true for the two studies in which the perspective of participants differed from that of the speaker, and, thus, required a shift in mental perspective) (Begeer et al., 2010; Sanstieban et al., 2015).

However, this apparent discrepancy between production and comprehension measures may be an artefact of certain characteristics of the existing studies rather than an actuality. The first characteristic of the data that prevents us from drawing firm conclusions is that the participants of these three comprehension studies were all adults or older adolescents; there are some indications that proficiency with the production of verbal reference may ameliorate

to some degree during adolescence (e.g. Arnold et al., 2009; de Marchena & Eigsti, 2015). However, improvement over development seems unlikely to be the full story for the difference between production and comprehension studies since two production studies with adults with ASD did find evidence of impairment in comparison to well-matched controls (Nadig et al., 2015; Colle et al., 2008).

Another possibility is that the apparent discrepancy between comprehension and production is due to task-related differences across studies. For example, the majority of production studies used narrative elicitation (for which there is no obvious comprehension-task counterpart), whereas all three comprehension studies used a referential communication paradigm (Volden et al., 1997; Begeer et al., 2010; Sanstieban et al., 2015). Indeed, one commonality amongst comprehension tasks used in all three studies (Begeer et al., 2010; Volden et al., 1997; Sanstieban et al., 2015) is that the dependent variable is binary forced choice, which is certainly far from the case for most production dependent variables. That said, five production studies (two of which were methodologically well-controlled) also used a referential communication paradigm, where the dependent variable could possibly be considered binary forced-choice, and all found impairments in the ASD group relative to the typical control group (see Table 13). Therefore, it seems unlikely that the dichotomy found between comprehension and production studies can be attributed to the fact that the comprehension paradigms are binary forced-choice.

To unpick the cognitive underpinnings of this discrepancy, this field needs much more fine-grained task analysis of the processes involved in the appropriate selection of referring expressions and of the processes involved in using the interlocutor's perspective to interpret referring expressions It is tempting to suggest that production of verbal reference is inherently more burdensome to executive functioning than is interpretation of verbal reference. In production of verbal reference, the speaker requires, for example, working

memory to hold information relevant to the listener whilst a sentence is formulated and executed. If the specific syntactic form of the target referring expression (e.g. simple noun phrase versus complex noun phrase) differs between trials, this may also place additional demands on mental set-shifting, that is, the ability to switch back and forth between multiple trials (see e.g. Sikora, Roelofs, Hermans & Knoors, 2016). That said, even comprehension variants of the visual perspective referential communication task have been found to tap various aspects of EF in the typical population (e.g. Brown-Schmidt, 2009; Lin et al., 2010; Cane, Ferguson & Apperly, 2016; Nilsen & Graham, 2009). Thus, we clearly need a more precise mechanistic model of the fine-grained steps required for comprehension and production and how this might differ depending on the specific tasks used for each.

Whatever the explanation for the discrepancy between performance on laboratory-based measures of verbal reference interpretation and laboratory-based measures of verbal reference production, there is a further overarching issue that needs to be considered when drawing conclusions about these abilities in ASD. Even if individuals with ASD are unimpaired in interpretation using the referential communication tasks, this does not mean that they are necessarily unimpaired in interpretation of verbal reference in everyday life. This is because all referential communication tasks to date in this field have essentially manipulated only level one visual perspective , which is an individual's understanding that the content of what they see may differ from the content of what another sees in the same physical position, (e.g. Salatas & Flavell, 1976). This requires the ability to follow another person's line of sight and draw conclusions about whether person's perception of an object is occluded, yet one need not have a very deep understanding of mental states to determine this (e.g. Moll & Kadipasaoglu, 2013; see also Sanstieban et al., 2015, for a sub-mentalising account).

In everyday life, in contrast, the interpretation of reference is often dependent on "social" perspective-taking / common ground; that is, an understanding of what a specific interlocutor knows, or is likely to find interesting or salient.). This often depends on a consideration of which particular information or experiences we have shared with which specific interlocutors. The only study meeting our criteria which investigated this is Nadig et al. (2015), who found that adults with ASD were less likely than typical adults to take discussion shared with a particular interlocutor (via a referential pact) into account when selecting a referent term. Of course, there are numerous divergent ways in which social common ground can be established with a specific interlocutor. One way is through sharing a particular collaborative experience (e.g. painting an action figure) with a certain interlocutor. To date this has only been explored to a degree in a couple of very small-scale production study pilots without control groups (Geller, 1988; Rosenthal Rollins, 2014). No studies have investigated whether individuals with autism can use social common ground to interpret verbal reference.

3.7. Conclusions

To move this field forward, we need studies which manipulate the role of social perspective-taking and compare this using comprehension and production variants of the task in the same sample of individuals with ASD. This is explored further in Chapter 4 of this thesis. We also need the field to shift away from an over-reliance on narrative paradigms. In addition to some issues with narrative paradigms outlined above, narrative is problematic here because verbal reference can be used appropriately in narrative without a real consideration of the listener's perspective, by simply tracking whether the form used for introduction or maintenance of reference is appropriate from one's own perspective (see e.g. Arnold, 2008, for a discussion of 'narrator-oriented' use of verbal reference). Finally, we

need a more detailed account of how deficits revealed in experimentally elicited production of verbal reference link to pragmatic language impairments in naturalistic dialogue. To that end, it is striking that to date there exists only one case-controlled study of reference production in conversation (Baltaxe & d'Angolia, 1996) and this study had highly problematic methodological issues. We need to empirically document in more detail the degree to which an impairment in reference usage hinders real-life verbal interaction, and to demonstrate more precisely the potential links that such an impairment has with difficulties in peer interaction and/or mental health difficulties.

CHAPTER 4: WHEN DO CHILDREN WITH AUTISM SPECTRUM DISORDER TAKE COMMON GROUND INTO ACCOUNT DURING COMMUNICATION?

Abstract

One feature of autism spectrum disorder (ASD) is a deficit in verbal reference production; i.e., providing an appropriate amount of verbal information for the listener to refer to things, people, and events. However, very few studies have manipulated whether individuals with ASD can take a speaker's perspective in order to interpret verbal reference. A critical limitation of all interpretation studies is that comprehension of another's verbal reference required the participant to represent only the other's *visual* perspective. Yet, many everyday interpretations of verbal reference require knowledge of *social* perspective (i.e., a consideration of which experiences one has shared with which interlocutor).

We investigated whether 22 5;0- to 7;11-year-old children with ASD and 22 well-matched typically developing (TD) children used *social* perspective to comprehend (Study 1) *and* produce (Study 2) verbal reference. Social perspective-taking was manipulated by having children collaboratively complete activities with one of two interlocutors such that for a given activity, one interlocutor was Knowledgeable, and one was Naïve. Study 1 found no between-group differences for the *interpretation* of ambiguous references based on social perspective. In Study 2, when *producing* referring terms, the ASD group made modifications based on listener needs, but this effect was significantly stronger in the TD group. Overall, the findings suggest that high-functioning children with ASD know with which interlocutor they have previously shared a given experience and can take this information into account to steer verbal reference. Nonetheless, they show clear performance limitations in this regard relative to well-matched controls.

4.1. Introduction

As described in Chapter 1, the diagnostic criteria for Autism Spectrum Disorder (ASD) include deficits in verbal social communication (APA, DSM-5, 2013) (see section 1.1). A distinctive characteristic of verbal social communication among individuals with ASD is an impairment in 'audience design' (Volden, 2002), the ability to provide an appropriate amount of information for the listener's informational needs (Clark & Marshall, 1981). For example, if a speaker wants a particular loaf at the bakery, then a complex referring expression such as 'the biggest loaf' could be appropriately informative, whereas requesting 'the bread' would be under-informative. However, at dinner with only one loaf of bread, asking for 'the big loaf' would be over-informative. A deficit in audience design is likely to contribute to difficulties with peer popularity (see e.g. Rubin, 1972; Gottman, Gonso & Rassmussen 1975) and reciprocal conversation (e.g. Fine, Bartolucci, Szatmari & Ginsberg, 1994). Use of under-informative referring expressions, in particular, leads to breakdowns in the ability to make successful requests (e.g. Matthews, Lieven & Tomasello, 2007).

The degree to which an expression is appropriate depends on consideration of the knowledge shared between speaker and listener, referred to as 'common ground' (Clark, 1996). Moll and Kadipasaoglu (2013) distinguish between visual common ground, which requires a consideration of whether the interlocutor can see particular items, and social common ground, which requires a consideration of the experiences the individual has previously shared with the interlocutor. They argue that visual perspective-taking may not require an in-depth consideration of the contents of the interlocutor's mental states; any interlocutor positioned behind a barrier will not be able to see a particular object. In contrast, social common ground is interlocutor-specific. Prior joint engagement and collaboration are argued to be crucial in determining social common ground for typically-developing (TD)

children (Kern & Moll, 2017; Rakoczy, 2017; Tomasello, Carpenter, Call, Behne & Moll, 2005). For example, if a speaker and interlocutor have been collaborating by using a specific hammer to put a toy together, then the simple referring expression in 'Pass the hammer' is appropriately informative even when other hammers are visually co-present (e.g. Schmerse, Lieven & Tomasello, 2015).

As highlighted in Chapter 3, compared to TD groups which are well-matched for chronological age, non-verbal IQ and structural language, ASD groups usually under-perform in the *production* of appropriately informative referring expressions during narrative tasks (Colle et al., 2008; Suh et al., 2014; Banney, Harper-Hill & Arnott, 2015; Arnold, Bennetto & Diehl, 2009). The same has been found in production studies that have used interactive experimental manipulations of the participant and interlocutor's shared visual perspective (e.g. Nadig, Vivanti & Ozonoff, 2009; Fukumura, 2016; Dahlgren & Dahlgren Sandberg, 2008; Volden Mulcahy & Holdgrafer, 1997; Nadig, Seth & Sassoon, 2015). That said, one study which used a very simple narrative elicitation task (Kuijper, Hartman & Hendriks, 2015) found no differences between children and adolescents with ASD and TD controls. Moreover, some studies have found – alongside evidence of impairments relative to controls – indications that children and adolescents with ASD show an awareness of listener information needs (Makinen et al., 2014; Arnold et al., 2009; Norbury & Bishop, 2003).

In contrast to research on the *production* of contextually-appropriate verbal reference in ASD, there are very few studies manipulating the role of speaker perspective on the *interpretation* of verbal reference, which have included typical controls. In a 'Director Task' where speaker and listener visual perspectives matched (common ground condition) or were disparate (privileged ground condition), Begeer, Malle, Nieuwland and Keysar (2010) and Santiesteban, Shah, White, Bird and Heyes (2015) found no differences in either accuracy, response latency or gaze fixations for adults with ASD relative to typical controls. Schuh,

Eigsti & Mirman (2016) also found no differences between 13-year-olds with ASD and TD controls on accuracy measures in a referential communication task requiring participants to recall which referents were known to the speaker. The ASD group made more gaze fixations to the objects unknown to the speaker (privileged ground condition) than did the TD controls, when the task involved a high working memory burden (a larger number of shapes). Therefore, it may be that individuals with ASD possess the necessary socio-cognitive competence to use common ground to interpret verbal reference but fail to consistently apply this competence/knowledge depending on extraneous factors which contribute to task difficulty.

That said, it would be surprising if individuals with ASD really possessed an underlying conceptual competence in verbal reference interpretation per se, because this ability is usually considered to rely on the Theory of Mind (ToM)/mentalising ability that is known to be impaired in ASD (e.g. White, Hill, Happé & Frith, 2009). There are several reasons to be cautious when drawing conclusions from the findings of existing studies of verbal reference interpretation in ASD. First, all studies of verbal reference interpretation by individuals with ASD have only included adolescents or adults. Therefore, individuals with ASD might 'outgrow' an initial impairment in the ability to take common ground into account while interpreting verbal reference. Alternatively, it might be that older adolescents and adults with ASD employ alternative, compensatory strategies to "hack out" solutions to tests of verbal reference interpretation, in a manner similar to that reported for classic ToM tasks (e.g., Frith, Morton & Leslie, 1991). If, however, young children with ASD can in principle demonstrate the ability to take the knowledge-state of an interlocutor into account to drive verbal reference interpretation or production, then arguments regarding maturation or compensatory strategies are difficult to sustain.

Second, previous studies only assessed the interpretation of verbal reference in the context of level one *visual* perspective taking, namely whether the interlocutor can see the referential alternative[1]. This is problematic, because level one visual perspective taking could be carried out by simply using a non-mentalising heuristic, such as following the physical line of gaze without giving any in-depth consideration to the content of the gazer's mind. Evidence supporting this suggestion comes from Sanstieban et al.'s (2013) study, that found no differences for either group between the experimental and a control condition, in which the interlocutor was replaced with a camera. Therefore, the results from the studies of Begeer et al. (2010), Sanstieban et al. (2015) and the low working memory condition in Schuh et al., (2016) cannot be taken as definitive evidence that an ability to use personal common ground to interpret verbal reference is spared in ASD. To test this, we need to investigate whether individuals with ASD can interpret verbal reference by taking into account *social* common ground, i.e. by considering which previous joint action they have shared with which particular interlocutor (Rakoczy, 2017; Kern & Moll, 2017).

We therefore carried out the first study to manipulate social perspective-taking in the context of verbal reference production or interpretation by children with ASD. Indeed, ours is the first study of verbal reference interpretation by children with ASD. The mean age of our sample is also younger than any study which has investigated verbal reference production in ASD; we included children with ASD aged 5;0 to 7;11-years and TD children matched for age, non-verbal IQ, and formal language. For our interpretation experiment (Experiment 1) we adapted a social common ground paradigm originally developed for TD children by Liebal, Behne, Carpenter and Tomasello (2009). For each experimental trial, each participant first collaborated in one activity (e.g. constructing an electric circuit) with one experimenter, who then left the room ostensibly to find a missing object (here: a battery) needed to complete that activity. Each participant then collaborated in another activity (e.g. putting

together a remote-controlled car) with *another* experimenter, who then also left the room ostensibly to find a missing object – the same missing object – needed to complete that activity. For the final component of each trial, one of the two experimenters (the Requester) returned to the room, offered the participant the missing object and said "Here it is! Now you can do it". The measure was whether the participant inserted the object into the activity in which the participant had engaged with the Requester. To control for the possibility that children might select the correct referent using a heuristic such as 'go to the activity the person making the request was co-present with' (e.g. O'Neill, 2005) we also had a *co-presence only* control condition.

Our verbal reference production experiment (Experiment 2) was elicited by asking the same children to re-tell how to construct toys, that they had made as part of the interpretation experiment, in two within-subject conditions. One required the participant to re-tell to the adult who had completed the task with them (Knowledgeable Listener). The other condition required each participant to explain the toy construction to an adult who had not been present for this activity (Naïve Listener). If participants engaged in audience design, they should use complex referring expressions (e.g. 'the long elastic band') for the Naïve Listener but simple referring expressions (e.g. 'the elastic band') for the Knowledgeable Listener.

If autistic difficulties with audience design found in naturalistic interaction derive primarily from a competence deficit in the ability to take social common ground into account, this would predict that our sample of children with ASD should not differ from chance in our interpretation experiment. A competence deficit account would also predict no differences between the Knowledgeable and Naïve conditions in the production experiment, since this production measure specifically assesses whether participants tailor their selection of referring expressions based on shared knowledge

If, however, naturalistic audience design difficulties are primarily a performance issue, this would allow for above-chance performance in the ASD group in the interpretation experiment and a difference between the Knowledgeable and Naïve conditions in production. Importantly, however, this account would predict an interaction whereby ASD children should show a lesser distinction between the two conditions than would the TD children, indicating a reduced ability to demonstrate their competence. To investigate the role of mentalising ability, we also examined whether successful reference interpretation or audience design in production related to performance on a Theory of Mind scale.

4.3. Experiment 1: Using social common ground to interpret referring expressions.

Methods

Participants. Twenty-two TD children were recruited through a developmental lab database and mainstream schools. Eleven children with ASD attended specialist provisions for children with ASD. The remaining children with ASD were recruited through parent support groups. When recruiting children with ASD, parents and teachers were told that the inclusion criteria included verbal fluency and knowledge of colour and size adjectives. TD and ASD children were matched on chronological age, formal language and non-verbal IQ. To assess formal language, we carried out both receptive (the *Sentence Structures* sub-test*)* and expressive language (the *Formulated Sentences* sub-test) measures from the Clinical Evaluation of Language Fundamentals® - Fifth Edition (CELF®-5, Wiig, Semel & Secord, 2013). Non-verbal IQ was assessed using the *Pattern Construction* and *Matrices* tasks from the British Ability Scales (BAS, Elliot & Smith, 2011). Children in the ASD group had been diagnosed either by a clinical psychologist or a paediatrician. Parents of children in both groups completed the Social Responsiveness Scale (SRS, Constantino & Gruber, 2005). Two children with ASD were tested but excluded because their SRS T-scores were below the clinical cut-off and two

TD children were tested but excluded because they scored above this threshold. The demographics for the final sample are shown in Table 13 below, which also includes the scaled scores for the CELF subtests, the average t-score for non-verbal IQ and between-groups differences for tests of First Order Theory of Mind and Advanced Theory of Mind tests (see table legend and also Appendix 1).

Table 13. *Means (SD in brackets) for participant characteristics*

	ASD (*n* = 22; 18 male)	TD (*n* = 22; 18 male)		
	Mean (*SD*)	Mean (*SD*)	*p*	*d*
Chronological Age in Months	78.86 (11.46)	78.77 (11.19)	.98	0.01
Sentence Comprehension CELF®-5 Scaled Score	10.64 (2.82)	10.05 (1.81)	.41	0.25
Formulated Sentences CELF®-5 Scaled Score	11.14 (3.20)	11.50 (1.87)	.65	0.14
Non-verbal IQ: British Ability Scale T-score[1]	48.59 (10.02)	50.30 (8.93)	.54	0.18
Social Responsiveness Scale T-score	82.86 (9.14)	45.38 (6.14)	<.001	4.81
Theory of Mind First order[3]: max score = 2.0	1.68 (0.65)	1.95 (0.21)	.07	0.56
Theory of Mind 2nd order[4]: max score = 2.0	1.05 (0.9)	1.5 (0.67)	.06	0.50
Theory of Mind Advanced[5]: max score = 4.0	0.18 (0.40)	1.09 (1.19)	.002	1.03

1..This is the mean T score of the 'Pattern Construction' (visuo-spatial pattern matching) and the 'Matrices' (non-verbal fluid reasoning ability) sub-tests of the British Ability Scales. 3. This is the combined score of 'Contents False Belief (Wellman & Liu, 2004; Perner, Leekam & Wimmer, 1987) and the 'Change of location' (Baron-Cohen, Leslie & Frith, 1985). 4. From Coull, Leekam & Bennett, 2006 – see Appendix 1.1. 5. Two of Happe's (1994) 'Strange Stories items ('Kittens' and 'Double bluff') with vocabulary adapted for younger children – see Appendix 1.3.1 & 1.3.2).

Overall procedure. The Experiment 1 trials were presented as 'breaks' from the standardised (non-verbal IQ and formal language) and ToM tests, and were interspersed between them (See Appendix 2). Each trial (experimental and control) involved one pair of activities requiring a unique common item to enable completion. There were six pairs of activities in total (see Table 14). Each participant completed four experimental trials and two control trials. The six activity pairs were counterbalanced across participants to appear as control trials on an even number of occasions. At the end of the experiment, to verify that participants could remember with whom they shared social common ground, each child was asked to tell either a parent or teacher which activities they had completed with which experimenter. All children successfully remembered this.

Procedure. For each trial (for both experimental and control trials), the child first engaged in constructing one activity (e.g. putting together the piece of wood for a slingshot) and then engaged in the construction of a different activity (e.g. folding and gluing a paper woodpecker). Each construction was completed jointly with one experimenter in the absence of the other experimenter. At the end of each construction activity, the experimenter with whom the child was constructing the activity declared that a key item was missing (here: an elastic band) and left the room to find it. The two experimenters were never in the room at the same time during both control and experimental trials.

Table 14: *Construction activity pairs and common missing items (comprehension measure)*

Activity Pair	Activities	Key 'missing'/common item
1	Electric circuit Battery operated toy car	Battery
2	Paper aeroplane Certificate	Glue
3	Tumbling bunny Marble run	Marble
4	Woodpecker Slingshot	Elastic Band
5	Telephone Flying fish	String
6	Spinning bird Blow fish	Straw

For both experimental and control trials, one experimenter ('Requester') returned to the room and stood in a position equidistant to each activity, offering the child an object (e.g. an elastic band) saying *'Here it is. Now you can do it'*. If the child failed to respond within 5 seconds, or asked the 'requester' a question, the requester proceeded with a prompt / response as outlined in Table 15. An additional prompt was provided every 5 seconds until the child either selected an activity, or the requester reached the final prompt and named the correct activity. The experimenter only looked between the child and the object (e.g. the elastic band) until the child inserted it in an activity. The visual status of the two activities was equated for both child and the Requester during the request (e.g. both activities occluded in opaque boxes).

In the interaction between Requester and participant prior to the test trial request, the Requester had only seen one of the two activities. Thus, if the Requester had previously (i.e. prior to leaving the room) only seen the slingshot (and not the woodpecker), the slingshot was designated the 'target' and the woodpecker the 'foil' for that trial. If a child asked for clarification, the Requester's responses were scripted.

For both experimental and control trials, we counterbalanced within and across participants which of the two experimenters was the Requester, whether the target was

located to the child's right or left and whether the activity associated with Requester was the most recently completed one for the child. We also counterbalanced across participants the order of the activities, which activity of a pair was the target, and whether an activity was assigned to the experimental or control condition.

In the *experimental* trials, the target activity (e.g. slingshot) had been co-constructed with the Requester and the foil activity (e.g. woodpecker) had been co-constructed with the other experimenter. In the *control* trials, in contrast, the target and foil activities were both completed with the non-requesting experimenter one after the other (e.g. non-requesting E constructs telephone, leaves room to find string, returns having failed to locate string with materials for 'flying bird, leaves room to look again for string). Importantly, in the control trials, the Requester had not interacted with the target construction activity. Rather, they had entered the room in the absence of the non-requesting experimenter and merely sat next to either the first or second construction activity (counterbalanced) whilst completing a puzzle with the child. The Requester then left the room, before later returning with the missing item. In these control trials the target was the construction activity which the Requester had sat next to. Both control trials were carried out after all experimental trials were completed.

Table 15*: Order and form of responses to clarification requests for the interpretation trials (both experimental and control)*

Clarification Request number	Response by E
1	'You can do it now. Go for it!'
2	'You can finish it now'.
3	'Go ahead and do it'.
4	E names toy. This is scored as 0 (i.e. not selecting the activity associated with the requesting E).

Scoring

In both experimental and control trials, a score of 1 was awarded for selection of the target activity (i.e. associated with the Requester) and a score of 0 was awarded for selection of the foil. As there were four trials in the experimental condition, the maximum raw score

per child was 4. Inter-coder agreement for the interpretation experiment was 100% for both experimental and control trials.

Bayesian analyses

According to Jeffreys' (1961), Bayes factors (BF_{10}) > 3 provide firm evidence for the alternative hypothesis and values under 1 provide evidence for the null (with values < 0.33 providing firm evidence). A BF_{10} of 3 suggests the alternative hypothesis is three times more likely to be true than the null hypothesis.

Results

Co-presence only control trials. The proportion of times that the co-present activity was selected in the control trials did not differ between the ASD (M = 0.39, SD = 0.31) and TD (M = 0.52, SD = 0.29) groups, $t(42) = 1.52$, $p = .14$, $d = 0.47$, $BF_{10} = 0.74$. This is important because it indicates that co-presence alone is insufficient for children in either group to interpret ambiguous reference.

Experimental trials. For the experimental there was no significant between-group difference in the proportion of trials on which children chose the activity that corresponded to the one they had undertaken with the experimenter making the ambiguous request (ASD: M = .85, SD = 0.23; TD: M = .94, SD = 0.15), $t(42) = -1.56$, $p = .13$, $d = 0.47$, $BF_{10} = 0.78$). Thus, the two groups did not differ from one another in terms of the likelihood that participants would correctly interpret ambiguous verbal reference by selecting the activity associated with a specific interlocutor, namely the interlocutor with whom they had shared that particular activity. Importantly, both groups showed above-chance performance in this experimental condition (ASD group: $t(22) = 7.28$, $p < .001$; TD group: $t(22) = 13.59$, $p < .001$).

4.4. Experiment 2: Do children with ASD use audience design in their selection of referring terms?

Participants. Participants were the same as for Experiment 1.

Relationship between Experiments 1 & 2. Four construction activities from Experiment 1, namely 'electric circuit', 'tumbling bunny', telephone' and 'woodpecker' were used in the production task. Experiment 2 trials always followed the Experiment 1 trial that involved the relevant construction activity. Thus, for example, immediately after engaging in the 'telephone' activity as part of the interpretation study, the child was asked to tell the addressee how to construct the telephone for the production study.

Design. Both groups participated in two within-subject conditions. In the *Knowledgeable* condition, each participant explained how to construct a toy to the experimenter with whom s/he had made a toy. In the *Naïve* condition each participant explained how to construct a toy to the experimenter who had not been present during the activity. Two construction activities were re-told in each condition (counterbalanced), with three target (binary-coded) referents for each, resulting in a possible raw score of 6 per condition (see Table 16). Both experimenters were Knowledgeable for two activities and Naïve for the other two activities for each participant.

Table 16: *List of referents and their competitors (production measure)*

Activity	Referent	Target	Distractor	Total referents
Electric Circuit	Wires	White (with clips)	Black (no clips)	3 (colour, size, colour)
	Battery	Small (fits)	Big (does not fit)	
	Battery Holder	Black (works)	Yellow (broken)	
Tumbling Bunny	Bunny	Big	(Too) Small	3 (size, size, length)
	Marbles	Small	(Too) Big	
	Ramp	Long	(Too) Short	
Telephone	Cups	White (pencil will go through)	Pink (solid-pencil will not go through)	3 (colour, size, length)
	Blu-tac	Big	(Too) Small	
	String	Long	(Too) Short	
Woodpecker	Woodpecker	Green (on card)	White (on paper)	3 (colour, size, length)
	Sellotape	Small (sticky)	Big (with fluff on, will not stick)	
	Elastic Band	Long	(Too) Short	

Procedure: Each child instructed how to construct an overall total of four toys. During the shared experience phase for each Experiment 1 activity, the experimenter who constructed each toy with the child referred to each Experiment 2 target referent and its distractor (e.g. ‘we have to use the black battery holder because the yellow one is broken’, showing the child the broken item) as they constructed the given toy. For the activities which were used in the production study, one experimenter then told the child that the other experimenter (addressee) would write down what the child said and then use those notes to make a replica toy that could be taken home by the child.

To avoid the use of gestures, children provided instructions via a web-cam, whereby the child could see the addressee, but it was explained that the addressee could not see the child. Simple pictures were provided to remind the child of the steps required in the activity. These pictures were not visible to the addressee, who sat in a separate room. Both the addressee and the child could see (for each activity) three items (e.g. cup, blu-tac, string) which had referential alternatives.

Each mention of a target referent (e.g. cup, blu-tac, string) was coded as follows. Use of complex referring (e.g. 'the small battery'; 'the battery that fits') received a score of 1. Usage of a simple noun phrase (e.g. 'the battery'), a pronoun ('it') or complete omission was scored as zero. The dependent variable was the proportion of referring expressions which were complex. As there were a total of six referents in each condition (three for each construction activity) the maximum raw score for each condition was six. All uses of verbal reference during the instruction task were transcribed and coded by the first author. Eighteen per cent of the data was also transcribed and coded by an independent rater, who was blind to each child's diagnostic status, with strong agreement between the coders (Cohen's k = .95).

Post-test vocabulary check: On completion of testing, children were asked to describe how items differed (e.g. size and colour names). All children were able to use these terms accurately.

Results

To examine the extent to which children engaged in audience design during production, the mean proportion of complex referring expressions used in the Knowledgeable and Naïve listener conditions was examined (see Figure 2). For context, a proportion score of one in the Naïve Listener condition combined with a proportion score of zero in the Knowledgeable

Listener condition would indicate that participants were tailoring their use of verbal reference perfectly to the audience (listener) informational needs.

A mixed 2 (Group: TD/ASD) × 2 (Listener Knowledge: Knowledgeable/Naïve) ANOVA was conducted to compare the mean proportion of complex referring terms used. A significant main effect of Listener Knowledge reflected the use of more complex referring terms with a Naïve than a Knowledgeable listener across groups (*F*(1,42) = 30.55, *p* = <.001, η_p^2=.42 BF_{10} = 1.40). The main effect of Group was marginally significant (*F*(1, 42) = 3.66, *p* = .063, η_p^2=.08 BF_{10} = 1.40), reflecting less use of complex referring terms among ASD than comparison participants across conditions. More importantly, the Group × Listener Knowledge interaction effect was significant (*F* (1, 42) = 6.1, *p* = .02, η_p^2=.13, BF_{10} = 3.37). The TD children showed a significant difference between conditions (*t*(21) = 5.41, *p* = <.001, *d* = 1.19, BF_{10} = 1031), indicative of audience design. Importantly, the difference across conditions was also significant for the ASD group (*t*(21) = 2.27, *p* = .03, *d* = 0.57, BF_{10} = 1.83,), suggesting that they were aware of their listener's informational needs and made attempts to address these. This indicates the presence of an underlying competence in the ASD group.

However, the significant interaction shows that actual performance of the two groups clearly differed and between-participants *t*-tests in each condition separately showed that the extent of this audience design was significantly smaller among ASD than TD participants. The between-group difference in the number of complex referring terms used in the Knowledgeable Listener condition was not significant, *t*(42) = .38, *p* = .70, *d* = 0.12, BF_{10} = 0.32. However, TD children used significantly more complex referring terms than children with ASD in the Naïve condition, *t*(42) = 2.36, *p* = .02, *d* = 0.71, BF_{10} = 5.15. Thus, TD participants altered their usage of complex referring expressions according to listener needs to a significantly greater extent than did participants with ASD.

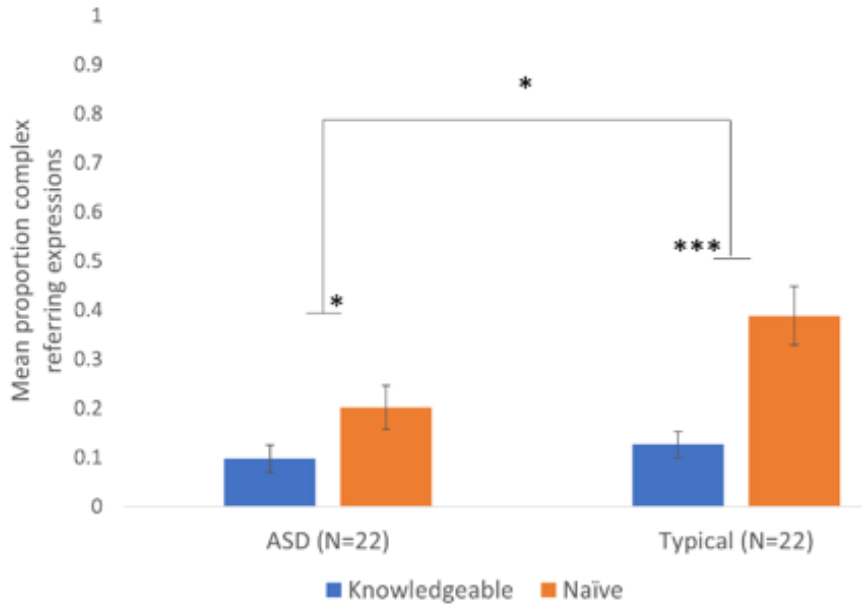


Figure 2. *Proportion usage of complex referring term with knowledgeable and naïve listener*

*p < .05, **p < .01. ***p<.001

Correlational analyses

Finally, we also examined whether any associations exist between either our interpretation or our production measures, on the one hand, and any measures of non-verbal IQ, formal language, autism traits and features, or theory of mind, on the other hand. To examine relationships with *production*, we calculated difference scores, whereby for each child the raw number of complex referring expressions used in the Knowledgeable Listener conditions (where this would be inappropriate) was subtracted from the raw number of complex referring expressions used in the Naïve Listener condition (where complex referring terms were appropriate). The larger the resulting value, the more a participant was tailoring their use of referring expressions to listener needs. This directly measures adaptation for audience design and controls for individual differences in the preference for using complex referring expressions per se (regardless of appropriateness).

A battery of theory of mind (ToM) measures was administered in a step-wise fashion. The 'Sally-Anne' task (Baron-Cohen, Leslie & Frith, 1985) and 'Smarties' task (Perner, Frith,

Leslie & Leekham, 1989) measured first order ToM (maximum two points). The 'Robot Story' (Coull, Leekham, & Bennett, 2006) measured second order theory of mind (maximum two points) and two items from Happé's (1994) 'Strange Stories' ('lies to persuade' and 'double-bluff'), measured higher-order ToM (maximum four points). 'Lies to persuade' was very similar to Happé's (1994) original 'kittens' story, with difficult vocabulary items simplified. The 'double-bluff' story was re-written to revolve around a sister hiding biscuits from her brother (See Appendix 1). The maximum score on this ToM composite was eight points.

Neither the interpretation nor the production task were significantly associated with formal language, non-verbal IQ or ASD traits in either group (all rs < .36, all ps > .10, all BF_{10}s < 0.92). Likewise, the ToM composite was not significantly associated to the production difference score in either group (all rs < .28, all ps > .11, one-tailed, all BF_{10} < 0.28). Finally, for the interpretation task, while the relationship with the ToM composite did not reach significance for the typical controls ($r(22) = .29$, $p = .10$, one-tailed, $BF_{10} = 1.05$), there was a relationship for the children with ASD ($r(22) = .47$, $p = .01$, one-tailed, $BF_{10} = 5.06$). Given that this was the only analysis to reach significance, and that three autistic children failed both first-order theory of minds tasks yet responded correctly to all four interpretation experimental trials, this result should arguably be treated with caution. We return to the relationship between ToM and referential communication in Chapter 6.

4.5. Discussion

The current study is the first to manipulate the role of social perspective-taking in the interpretation / production of verbal reference by children with ASD. This is important because a deficit in common ground understanding might not necessarily lead to impairments in audience design, if the speaker and listener's perspectives merely differ *visually* (Sanstieban et al., 2015; Moll & Kadipasaoglu, 2013). Therefore, we used collaborative tasks that manipulated the *social* common ground shared with an interlocutor. In Exp 1 (reference

interpretation) we found that 5;0- to 7;11-year-olds with ASD were above chance in the use of interlocutor-specific common ground knowledge to accurately disambiguate referring terms and did not significantly differ in this regard from a matched TD control group. In Exp 2 (production), we also found that children with ASD showed evidence of taking social common ground into account; they used significantly more complex referring expressions (e.g. 'the big marble') in the Naïve condition (where complex referring terms were necessary to avoid ambiguity) than in the Knowledgeable condition. In this sense, their production behaviour ties in with their interpretation behaviour; they have the competence to take social common ground into account to drive verbal reference. However, the interaction effect found for production indicates that children with ASD were less able than TD peers to select a referring term based on whether a specific listener had previously shared the experience of constructing that particular toy. Finally, we found no evidence of a direct association between the ToM impairments that characterised our sample of children with ASD and their diminished production of appropriate referring terms.

Could the interpretation task have been resolved via a sub-mentalising heuristic?

In Exp 1 (reference interpretation), our control task ruled out the possibility that children with ASD might in the experimental task have simply blindly associated a particular experimenter with a particular activity. In this control task, the Requester had sat next to one of the two activities, but this association between Requester and activity was not sufficient to enable children in either group to interpret the ambiguous pronoun as referring to this activity. Moreover, it also cannot be the case that, for the experimental task, the children with ASD simply tracked the ambiguous pronoun back to the noun phrase (activity) which they had most recently heard (see Arnold et al., 2009). This is because for half the experimental

trials per child, this would have led to false assumption that the target was the activity in which they had engaged with the non-requesting experimenter.

Exploring competence versus performance

Children with ASD thus can, on the one hand, demonstrate a competence for taking listener informational needs into account (i.e. significant difference between conditions for production task) but are nonetheless significantly less able to demonstrate this, in comparison to well-matched typical peers (as evidenced by the interaction effect). There are many reasons why the performance of children with ASD might be vulnerable, particularly if the task involves holding various elements in mind while planning a response, as was the case for our production task (e.g. Sikora, Roelofs, Hermans, & Knoors 2016).

Indeed, we might have found a between-groups difference if we had used a more complex interpretation task, such as one with several foils, or with several interlocutors, or a greater length of time between the shared experience and the test request. However, a between-groups difference in such a paradigm would not be due to a difference in the ability per se of the children with ASD to take in account the experience shared with a particular interlocutor; rather, it would be due to performance limitations, presumably relating to executive functioning or memory.

A good illustration of the difference between competence and performance in the interpretation of verbal reference can be seen in several studies of neurotypical adults (e.g. Cane, Ferguson & Apperly, 2017). These studies all manipulated level one visual perspective taking, for which no-one doubts neurotypical adult competence. Nonetheless, neurotypical adults make some errors in the privileged ground conditions of these 'director task' paradigms. Their 'egocentric errors' relate to inhibitory control (Brown-Schmidt, 2009) and cognitive flexibility (Lin, Eply & Keysar, 2010). Thus, while neurotypical adults have the

competence to perform these reference interpretation tasks, they also have *performance* limitations. Future studies will need to explore why individuals with ASD seem on average to be more susceptible than the average neuro-typical individual to performance issues.

The manipulation of social common ground

The aspect of social common ground which we manipulated is relatively basic; TD 18-month-olds are capable of taking this into account to interpret pointing reference (Liebal et al., 2009). Nonetheless, our interpretation task (Exp 1) required some kind of implicit consideration of the knowledge to which a *specific* interlocutor had access and moreover, this knowledge cannot be based (as in previous ASD reference interpretation studies) on level one visual perspective-taking because when the Requester entered the room in the experimental trials, she could see neither the target nor the foil. Thus, we are the first to demonstrate that 5- to 7-year-olds with ASD can take interlocutor-specific knowledge into account in order to interpret reference. Indeed, the fact that the children with ASD showed a significant difference between the Knowledgeable and Naïve Listener conditions in Study 2 (Production) shows that they also demonstrated in production the competence to consider this when selecting a referring expression.

Liebal et al. (2009) interpreted the behaviour of the infants in their original study as indicating that they were taking 'shared experience' in account to interpret reference. If our group with ASD were doing so, this could be problematic for claims by Tomasello et al. (2005) that children with ASD do not demonstrate shared intentionality.

We exercise caution in this regard. It is possible that our ASD group would still have selected the target, if they had merely observed the Requester constructing the toy. If this were the case, then the skills needed by children with ASD to 'pass' the interpretation task would merely be intention-reading (see Carpenter, Pennington & Rogers, 2002), social co-

ordination (Melis, Hare & Tomasello, 2006) and a willingness to help others (see Liebal, Colombi, Warneken & Tomasello, 2008, for autism).

Furthermore, there are many ways in which social common ground can be established and drawn upon when interpreting and producing verbal reference. In daily life, this often involves recalling shared experience over an extended period of time. It also often involves detecting not only what the interlocutor knows but also what he or she is likely to enjoy and / or find salient (e.g. Clark, Schreuder & Buttrick, 1983). This could be an area of difficulty in ASD since many individuals appear not to experience stimuli salience in a neuro-typical way (see e.g. O'Neill & Happé, 2000). It is certainly yet to be shown that individuals with ASD acknowledge that a joint commitment has been made towards a shared goal (see Tomasello & Hamann, 2012), which would be a stronger test of shared intentionality.

Summary

We have shown that children with ASD take interlocutor-specific prior experience into account in verbal reference interpretation and (to some extent) production. However, they lag behind well-matched typical peers in the degree to which they *utilise* information about prior shared experience in reference production. Future studies are needed to explore the precise components which lead to these disproportionate difficulties. This knowledge is essential for strategies to enable individuals with ASD to improve communication with family, peers and in the classroom. The following chapter explores one such component of executive functioning that may impact upon referential communication ability, namely cognitive flexibility.

CHAPTER 5: 'FLEXING THE DESCRIPTION': EXPLAINING PERFORMANCE DIFFICULTIES IN HOW AUTISTIC CHILDREN ADAPT REFERRING EXPRESSIONS FOR LISTENERS

5.1. Introduction

Referring expressions are appropriately informative when they provide neither too much, nor too little, information for the listener (Grice, 1975). Children with autism are less able to *produce* optimally informative referring expressions than their typically developing counterparts (Chapter 3). They use referential expressions that are both over-informative (providing a greater level of detail than required), and under-informative (providing insufficient information to identify a given referent) (Nadig, Vivanti & Ozonoff, 2009; Fukumura, 2016). Whilst the cognitive processes contributing to these difficulties are unclear, one candidate is reduced cognitive flexibility (CF).

Cognitive flexibility (also referred to as 'mental set-shifting' or 'task switching' in the literature) is one of three core aspects of executive function (EF) identified by Miyake et al. (2000). The term 'cognitive flexibility' encompasses a range of skills that avoid rigidity of thought. Abilities such as changing spatial perspective (for example, thinking about what an object would look like if viewed from a different angle) or interpersonal perspective (thinking about another person's point of view), switching to a novel approach during problem solving, or flexibly adapting to changes in plans, may all be included under the umbrella of cognitive flexibility' (Diamond, 2013). Cognitive flexibility has been conceptualised as a function that allows a context to be interpreted and / or responded to in multiple ways (Anderson, 2002). It is considered by some as the key measure of executive functioning, representing a culmination of abilities including inhibition and working memory (Zelazo, Müller, Frye, & Marcovitch, 2003; Dajani & Uddin, 2015).

Cognitive flexibility in autism

There is good evidence that cognitive flexibility may be problematic in autism. Firstly, poorer scores on performance-based measures of CF (experimental tasks and standardised psychometric tests) are reported in autistic individuals (Ozonoff & McEvoy., 1994; Shu, Lung, Tien & Chen., 2001). A recent meta-analysis of such measures of executive function in autistic individuals aged 6-years and above reported a broad impairment across EF domains, including CF (Demetriou et al., 2018). Lai et al. (2016) similarly conducted a meta-analysis examining performance during neurological tests of the three core areas of EF suggested by Miyake et al. (2000) (inhibition, working memory, flexibility) alongside planning and generativity in autistic children and adolescents. Whilst deficits were evident across all areas of EF, when participants with co-morbid ADHD were removed from the analysis, smaller effect sizes were reported for inhibition and planning. CF, meanwhile, remained significantly impaired relative to typically developing controls following exclusions of co-morbid ADHD and when IQ was matched across autistic and TD groups. This suggests the CF deficit observed in autism exists independent of co-morbid ADHD or general cognitive ability.

A second source of evidence for diminished CF in autism comes from measures of this ability in every-day life. The Behaviour Rating Inventory of Executive Function (BRIEF; Gioia et al., 2000) is a parent / teacher questionnaire designed to assess executive function ability during daily activities. Responses are used to calculate scores for several scales, including 'shift' which provides a measure of CF. Autistic children (5-18 years) not only score in the impaired range on the shift scale (Rosenthal et al. 2013) but score significantly higher (signally greater impairment) on 'shift' in comparison to other BRIEF scales (Granader et al. 2014). Higher 'shift' scores also extend into adulthood to a greater extent than other BRIEF scales in autism (Wallace et al. 2016). The restricted and repetitive

behaviours and interests that form part of the diagnostic criteria for autism may be related to difficulties with CF (Yerys et al. 2009). Well-documented autistic difficulties, for example, with changes in routine and transitioning between tasks, intuitively appear related to cognitive flexibility challenges. In summary, both experimental and real-world measures reveal CF difficulties in autism.

Cognitive flexibility is a recognised challenge in autism but is not the only area of EF reduced in this population. Working memory, for example, is found impaired (Kercood, Grskovic, Banda & Begeske, 2014). Whilst relationships exist between referential communication and both working and short-term memory in autism (Schuh, Eigsti and Mirman, 2016; Dahlgren & Dahlgren Sandberg, 2008), these memory components underlie the completion of multiple complex behaviours (Gathercole, Lamont, & Alloway 2006; Baddeley & Hitch, 1994). WM contributes a global influence upon a range of abilities in typical development (Alloway & Alloway, 2010), thus its effects may not be specific to referential communication. Cognitive flexibility, however, may have specific relevance for this skill.

The role of cognitive flexibility in referential communication

There are good reasons why cognitive flexibility might enable the production of appropriately informative referring terms (i.e. those that are neither under nor over-informative). In a set of physically present referential alternatives (e.g. a red and green toy car), the referring expression ‘the green car’ would sufficiently identify the intended referent. If the red car were tidied away and a larger green car introduced alongside the original smaller green car, the referring expression ‘the small car’ would now be appropriately informative. If later the same car was present with only a doll, then the referring expression ‘the car’ would suffice. Crucially, success here requires the *same* referent (the car) to be

labelled using a different referring term (the green car / the small car / the car) depending on changes in the visual context. Put another way, the child is required to flexibly 'shift' from focusing on the car's colour (green), to focusing on its size (small), to simply its type (car). Similarly, during conversation, what constitutes 'appropriately informative' naturally shifts as the discourse unfolds. For example, it might initially be necessary to identify the focus of the conversation; 'My brother Chris and his dog were on television last night'. Having established which 'Chris' is the focus of attention, it is unnecessary to continue using the full referring expression 'My brother Chris'. The conversation might instead continue; 'Chris is a police officer' or 'He was talking about how they train their dogs'. Here too, avoiding over or under-informativity involves flexibly changing the label used to refer to, in this case, a person.

Some empirical support exists for a relationship between CF and referential communication in typical development. Examining the production of unambiguous referring expressions, alongside the ability to repair misunderstandings, for example, CF predicted variance in initial descriptors during a referential communication task (independent of age and language ability) and was the only area of EF to significantly predict the number of new descriptors provided in response to feedback in typically developing children (Bacso & Nilsen, 2017). In sum, there are good reasons to expect CF and referential communication to be related.

Cognitive flexibility and referential communication in autism

Some studies point towards cognitive flexibility as a key component of pragmatic difficulties in autism. During conversation, for example, autistic children were less able to respond appropriately to requests for clarification than a typically developing group matched for language ability, despite having higher chronological age and non-verbal IQ (Volden,

2004). This may be the result of cognitive flexibility difficulties; namely reduced ability to alter the way information is provided to a listener. As no direct measure of cognitive flexibility was taken this possibility remains speculative. Autistic children's ability to generate multiple uses for items, and meanings of drawings, relates to pragmatic skill (Bishop & Norbury, 2005). These generation tasks might be considered to recruit CF, as the same stimuli must be interpreted in different ways. As the measures of pragmatic ability (CCC and the communication scale from ADOS-G) used by Bishop and Norbury (2005) were broad, it is not possible to identify whether all areas of pragmatic ability were equally related to generativity.

Regarding referential communication, producing appropriately informative referring terms may require the speaker to 'shift' flexibly between different ways of thinking about (and consequently referring to) an entity based upon a dynamic visual and / or social context (Graf & Davis, 2014). The dimensional change card sort task (DCCS; Zelazo, 2006), is a measurement of cognitive flexibility where children with autism fare less well than their typically developing peers when accuracy and response time is considered (Faja & Dawson, 2014; Dichter et al. 2010). The DCCS requires 'object-based set shifting' (Kloo & Perner, 2005), that is, the same object must be viewed in multiple ways. Autistic children may find this task difficult as it requires flexibly reframing the way in which an entity is represented (Kissine, 2012). The same could be true in referential communication; difficulty selecting an appropriately informative referring term could stem from reduced ability to switch between labels for the same referent.

Some support for autistic difficulties in switching between referring terms for the same entity comes from the referential pact literature. A 'referential pact' refers to a temporary agreement made between individuals regarding how a given referent will be labelled (Brennan and Clark, 1996). Whilst the same shoe, for example, might be named

'shoe', 'flip flop', 'footwear' or any other number of referring terms, communicative partners typically expect one another to continue using a mutually agreed term for a given referent throughout an exchange (Brown-Schmidt, 2009). Flexibility would, however, be required with a new communicative partner who may be unaware of a previously agreed referring term or refer to the same entity using a different label. Autistic children are found to converge on referring expressions during dialogue to the same extent as typically developing children (Branigan, Tosi & Gillespie-Smith, 2016). Nonetheless, they are less able than typically developing peers to interpret novel referring terms for items when a label has previously been used (Ostashchenko, Geelhand, Deliens, & Kissine, 2019). This suggests that, whilst autistic children can align with a partner on a referring term, they have a specific difficulty with flexibly comprehending a new referring term for the same entity.

Whether children with autism's difficulty in producing appropriate referring expressions is linked to a reduced ability to flexibly refer to the same entity using different terms has not been examined.

Cognitive flexibility and referential communication in autism: The present study

The current study examined whether autistic children's difficultly producing appropriately informative referring expressions is related to cognitive flexibility by manipulating cognitive flexibility demands within a referential communication task. Five- to seven-year-old children with and without autism completed a novel 'pet show' referential communication task that required them to tell an adult 'pet-show judge' which of two animals should be awarded a 'point' on each trial. All participants completed the task in two conditions: 'switch' and 'no-switch'. The key difference between conditions was whether the *same* referring expression remained appropriate for a given animal throughout the task (no-switch condition) or whether *different* referring expressions were required for the same

animal (switch condition). Hence cognitive flexibility demands were higher in the 'switch' condition and lower in the 'no-switch' condition. Our key measure of interest was the number of referring terms produced that were appropriately informative in each condition. In both conditions, an inappropriate referring expression resulted from an over-informative reference on half of all trials, and from an under-informative reference on the remaining half.

We made three distinct predictions. Firstly, as previous literature indicates a relationship between cognitive flexibility and referential communication in both typically developing and autistic children, we anticipated that both groups would perform better (i.e. produce more appropriately informative referring expressions) in the 'no-switch' than 'switch condition'. Secondly, given autistic children's difficulty producing appropriately informative referential expressions, we expected that their ability to produce appropriately informative referring terms would be poorer overall than the typically developing group. That is, autistic children were expected to produce fewer appropriately informative referring expressions than their typically developing counterparts in both the 'switch' and 'no-switch' conditions. Finally, given autistic children's notable difficulty in cognitive flexibility, we anticipated a more pronounced effect of switching in the autistic than typically developing group. That is, we expected autistic children's performance to be significantly worse than their typically developing peers in the 'switch' condition and that performance during this task would relate to cognitive flexibility ability, as measured by the DCCS. This prediction was based on the requirement to use using multiple labels for the same referent in the 'switch' condition, that is to utilise cognitive flexibility, an area recognised as problematic in autism.

5.2. Method

Participants

Thirty autistic and 30 typically developing 5- to 7-year-olds were tested, of which 23 in each group were males. Children were either recruited and tested in a quiet area of their

school (mainstream or special) or were recruited via a university child lab or autism charity and were thus tested in the child lab. Four TD children and 11 autistic children were tested in the child lab, with all remaining participants tested at their schools. All children were monolingual speakers of British English and none had hearing difficulties or comorbid Attention Deficit Disorder. All the autistic children had received a formal diagnosis from either a paediatrician or clinical psychologist within the British National Health Service. In addition, all scored above the threshold for autism (namely a T-score of 60) on the parent-completed Social Responsiveness Scale (SRS) (Constantino & Gruber, 2005). Each child was audio-recorded via Dictaphone. Ethical approval was obtained from the School of Psychology, University of Kent, UK. Written consent was obtained from parents and verbal assent from the children.

Groups were matched on chronological age, gender, non-verbal reasoning (as assessed by the 'Matrices' sub-test of the British Ability Scales (BAS, Elliot & Smith, 2011) and core language (as assessed by the 'Vocabulary Production sub-test of the *Clinical Evaluation of Language Fundamentals (CELF-4);* Semel, Wiig, & Secord, 2003, and the 'Sentence Structures' test of the CELF-5; Wiig, Semel & Secord, 2013). The demographics are shown in Table 17 below.

Table 17. *Means (SD in brackets) for participant characteristics*

	Autistic (n =30)	**TD** (n =30)		
	Mean (SD)	Mean (SD)	*p*	*d*
Chronological Age (Months)	77.37 (10.58)	76.94 (9.08)	.87	0.04
Sentence Comprehension CELF-5 Scaled Score	9.97 (2.46)	10.2 (1.94)	.68	0.10
Expressive Vocabulary CELF-4 Scaled Score	8.87 (2.70)	9.13 (1.78)	.65	0.11
Non-Verbal Reasoning: BAS T-Score	41.53 (7.82)	38.77 (7.06)	.16	0.37
Social Responsiveness Scale T-score	84.89 (7.41)	44.37 (6.16)	<.001	5.95

Procedure

Referential communication task (RCT)

The RCT was presented as a 'pet show' game that required children to provide descriptions of animals presented on a laptop to an adult 'judge'. Animals were always presented in pairs. Each group participated in two within-subject conditions; *switch* and *no-switch*.

Children sat at a table in front of a laptop computer that presented all stimuli via PsychoPy v3.0. A first experimenter (E1) sat next to the child and provided all instructions.

A second experimenter (E2) sat opposite the child at a second laptop computer with the screen facing away from the child. At the bottom of the laptop screens were a variety of animals in a 'pen' (Figure 3). Above the animals were two squares with black borders. E1 explained the animals were taking part in a 'pet competition'. Children were told their task was to help the 'judge' (E2) by telling them which animal had been awarded a point in each 'round' (trial). On each trial, two animals moved to the black squares for 'judging'. One square then changed from a black to a green outline to highlight the 'winner' (target animal), whilst the other square remained black (distractor animal) (see Figure 3 below). For each trial, the target and distractor animal differed in either lexical type (e.g. cat/ dog), size (big/ small), colour (e.g. black/pink) or pattern (spots/ stripes). This necessitated the use of either a simple or a complex referring term. For example, if a spotted frog (target) was presented with a striped frog (distractor) the term 'spotty frog' was appropriately informative. If the same frog was presented next to a cat, the referring term 'frog' was sufficient. Once the child had provided a referring term for the highlighted animal, E1 pressed a key on the laptop that returned both animals to their original position in the pen. E2 always provided the neutral response 'thank-you' across all trials, regardless of accuracy.

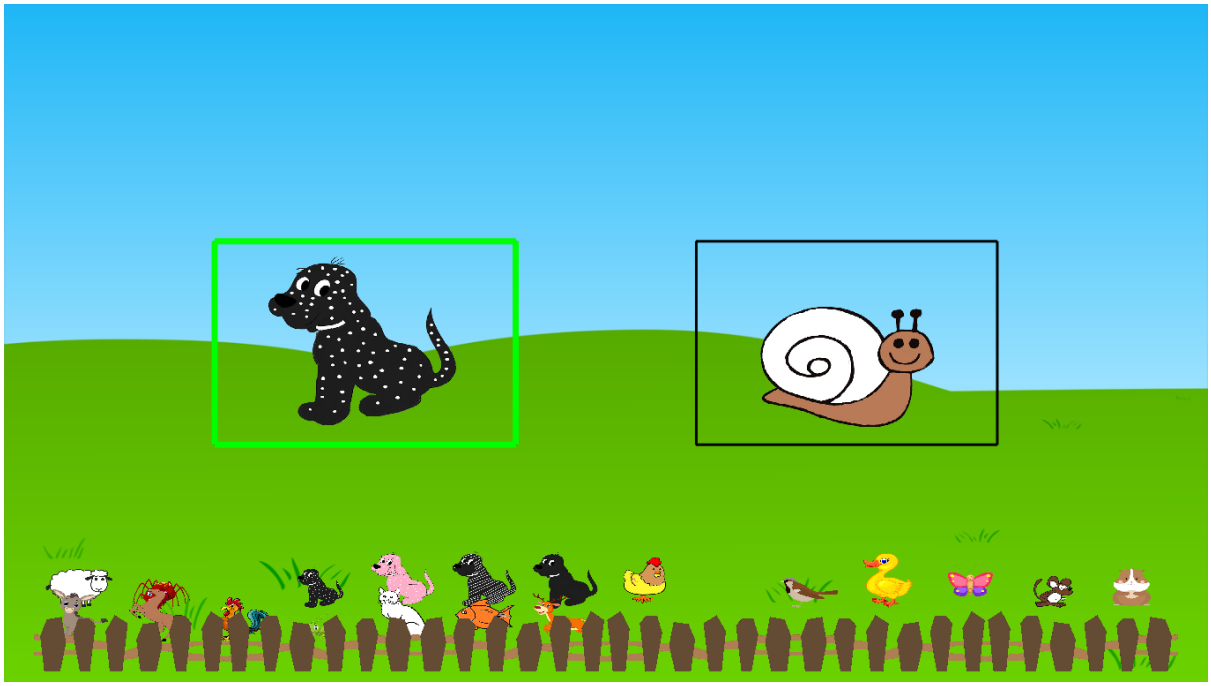


Figure 3. *Example of referential communication task stimuli presented on laptop*

Demonstration trials

Before experimental trials commenced, E1 demonstrated the task. During the demonstration there were 6 animals in the 'pen' at the bottom of the screen. Target animals in the demonstration trials were never presented as target animals in the experimental trials (see Appendix 3). The demonstration consisted of three trials. E1 modelled the correct answer on trials one and two, and on the third trial the participant was invited to identify the animal to the 'judge'. Feedback related to the accuracy of referring terms was not provided, but correction was provided if the wrong animal was described (i.e. any animal other than that appearing in the green square). Children were shown that the 'judge' (E2) would see the same two animals on their screen but would not know which had won the point. Children were also shown that the two animals on the 'judges' screen would not always be in the same position as the child's screen. This prevented children from using positional language to identify the animal. Once it was clear the child understood the task, they proceeded to the *switch* or *no switch* condition (counterbalanced).

Experimental trials

All children completed the RCT under two conditions: *switch* and *no switch*. To minimise carryover effects, children completed a range of play and test-based activities between conditions. This meant the two conditions were completed approximately 25 minutes apart.

In both the *switch* and *no-switch* conditions, 21 animals were visible in a 'pen' at the bottom of the screen (see Appendix 3 for a full list of the animal pairings included in each condition). As in the demonstration, two animals moved to the squares at the top of the screen on each trial, one turning green to signal which animal should be awarded a point. Both *switch* and *no switch* conditions consisted of 20 trials (10 test, 10 filler), meaning there were 20 animal pairs in each condition and children consequently provided 20 referring terms. An experimental trial always followed a filler trial (see Table 18). In all filler trials, the two animals were of different lexical types (e.g. chicken versus hamster), meaning a simple referring term was *always* appropriate in these trials. The target animals in filler trials were identical in both conditions (see Appendix 4). In both *switch* and *no switch* conditions, a complex referring expression was appropriate on half of the experimental trials (five trials) whilst a simple referring expression was appropriate for the remaining half (five trials). The target animals in the experimental trials never appeared as target animals in the filler trials, and vice versa in either condition.

In both the *switch* and *no-switch* condition, the first two trials were fixed. Thus, the first trial was always a filler and the second always an experimental trial, both requiring a simple referring expression. For both conditions, filler and experimental trials then alternated. The order of experimental and filler animal pairs was randomised by PsychoPy, but always within the filler / experimental pattern. The target item appeared in the square to

the left / right of the screen on an equal number of occasions for both filler and experimental trials.

Switch condition

In the *switch* condition, animal pairs in the experimental trials were presented such that different referring expressions were required for the same animal (a black dog with white spots). The same black and white spotted dog would be appropriately labelled, for example, 'spotty dog' when paired with a black and white stripped dog, 'big dog' when paired with an identical dog that is smaller in size, or 'dog' when paired with a snail (Figure 4).

No switch condition

In the *no switch* condition, animal pairs in the experimental trials were presented such that the same referring expression remained appropriate for each animal. A frog, for example, would be appropriately labelled 'spotty frog' when paired with a striped frog, a spotted cow as 'big cow' when paired with an identical smaller spotted cow, and a spotty rabbit as 'rabbit' when paired with a snail (Figure 4).

In sum, both conditions required participants to use a range of modifiers (size, pattern, colour) and bare nouns (e.g. 'dog') on an equal number of occasions. The key difference between conditions was whether these differing modifiers were required for the *same* animal (e.g. spotty dog, big dog, black dog, dog; *switch* condition) or a range of *different* animals (e.g. spotty frog, big cow, black pig, rabbit; *no-switch* condition). Hence cognitive flexibility demands (the need to use different labels for the same animal) were higher in the *switch* than the *no switch* condition.

Table 18. *Example of filler and experimental trial presentation in switch and no switch conditions*

Switch Condition

Trial	Target	Distractor	Appropriately informative Response	Simple or Complex
Filler	*Sheep*	Spider	Sheep	*Simple*
Exp. 1	**Spotty dog**	**Snail**	**Dog**	**Simple**
Filler	*Butterfly*	Chicken	Butterfly	*Simple*
Exp. 2	**Spotty dog**	**Stripy dog**	**Spotty dog**	**Complex**
Filler	*Snail*	Mouse	Snail	*Simple*
Exp. 3	**Spotty dog**	**Spider**	**Dog**	**Simple**
Filler	*Chicken*	Hamster	Chicken/ Hen	*Simple*
Exp. 4	**Spotty dog (black)**	**Spotty dog (pink)**	**Black dog**	**Complex**

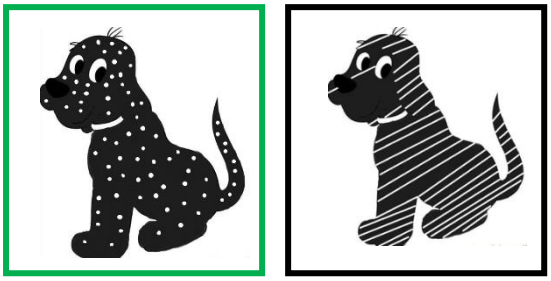
No Switch Condition

Trial	Target	Distractor	Appropriately informative Response	Simple or Complex
Filler	*Chicken*	Sheep	**Chicken**	*Simple*
Exp. 1	**Spotty cat**	**Butterfly**	**Cat**	**Simple**
Filler	*Duck*	Snail	Duck	*Simple*
Exp. 2	**Spotty snake**	**Stripy Snake**	**Spotty Snake**	**Complex**
Filler	*Chicken*	Duck	Chicken	*Simple*
Exp. 3	**Spotty rabbit**	**Butterfly**	**Rabbit**	**Simple**
Filler	*Snail*	Fish	Snail	*Simple*
Exp. 4	**Spotty pig (brown)**	**Spotty pig (pink)**	**Brown Pig**	**Complex**

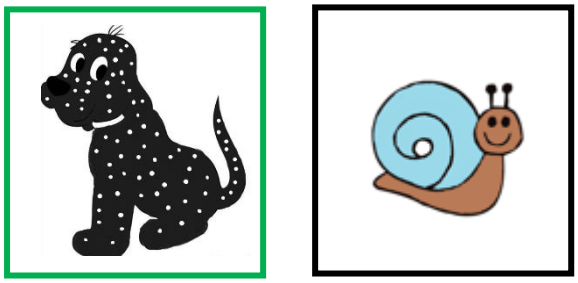
= Fixed trial order

S*witch* condition

e.g. Experimental trial 1 ‘spotty dog’ appropriately informative

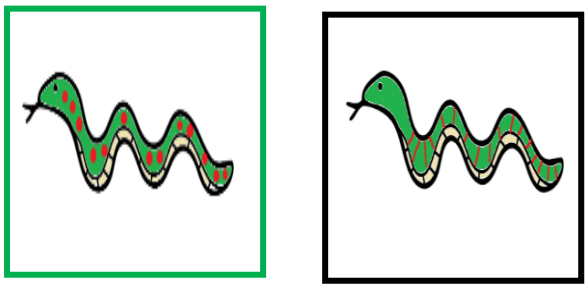


e.g. Experimental trial 2 ‘dog’ appropriately informative



No-switch condition

e.g. Experimental trial 1 ‘spotty snake’ appropriately informative



e.g. Experimental trial 2 ‘cat’ appropriately informative

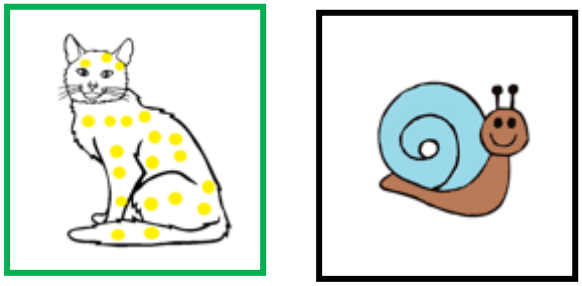


Figure 4. Example experimental trial stimuli and correlating appropriately informative referring expressions

Scoring RCT task

In each condition there were ten experimental trials. The referring expression produced on each trial was coded for appropriacy, whereby an appropriately informative referring expression was one where neither too much nor too little information was provided. Each referring expression was coded as either *appropriately informative*, *over-informative*, *under-informative* or *wrong modifier*. To illustrate, with the target animal 'black spotted dog' and distractor 'black striped dog' the referring expression 'spotty dog' would be coded *appropriately informative*, with 'black spotted dog' coded *over-informative* (the modifier 'black' is unnecessary as it does not uniquely identify the target animal), 'dog' coded as *under-informative* (this label does not adequately identify which dog) and 'dog with the black nose' coded as *wrong modifier* (a modifier was used, but not one that uniquely identifies the referent). An *appropriately informative* referring expression received a score of 1, and an *over-informative*, *under-informative* or *wrong modifier* expression received a score of 0. The key dependent variable was the proportion of referring expressions which were appropriately informative. As there were 10 experimental trails in the *switch* and *no switch* condition, the maximum score was 10 for each condition. Additionally, each referring expression was coded as either *complex (*a modifying phrase e.g. 'the spotty dog' or 'the dog with the spots') or *simple* (e.g. 'the dog'). All verbal references during the experimental task were transcribed by two third year psychology students, blind to each child's diagnostic status. Each referring expression was coded for appropriacy (appropriate vs. inappropriate) by the first author and the same two third year psychology students. The third-year psychology students also coded all references for complexity (simple vs. complex). Inter-rater reliability for appropriacy was conducted on twenty per cent of the data (6 TD and 6 ASD participants). There was strong agreement between the two coders (Cohen's k = .96).

Cognitive Flexibility

As we were interested in the role of cognitive flexibility, all children also completed a commonly used measure of *Cognitive Flexibility (CF)*, namely the Dimensional Change Card Sort (DCCS; Zelazo 2006). We developed a computer-based version administered in PsychoPy. In the Pre-Switch phase the child was asked (for each item) to sort a new card into one of two piles according to colour by clicking on the relevant pile. The items were flowers and boats coloured either red or blue. For each trial, the experimenter repeated the instruction to sort by colour. After six Pre-Switch trials had been completed, the sorting switched to object type (i.e. boat vs. flower). E1 explicitly explained the switch by saying "Now we're going to play a new game. We're not going to play the colour game anymore. We're going to play the shape game. In the shape game, all the flowers go here [pointing], and all the boats go there [pointing]. Remember, if it's a flower, put it here, but if it's a boat put it there. Okay?" Then, for each of the 'Post-Switch' trials, E1 reiterated the reminder to the child to sort by object type. Once six 'post-switch' trials had been completed, children who completed at least five post-switch trials correctly proceeded to 12 'Border Version' trials. First E1 explained orally "Now I have a more difficult game for you to play. In this game, you sometimes get cards that have a black border around it like this one [showing a red flower with a border]. If you see cards with a black border, you have to play the colour game. In the colour game, red ones go here and blue ones go there [pointing to the appropriate positions]. This card's red, so I'm going to put it right there [demonstrate placing in appropriate place]. But if the cards have no black border, like this one [show them a red flower without a border], you have to play the shape game. In the shape game, if it's a flower, we put it here, but if it's a boat, we put it there [point]. This one's a flower, so I'm going to put it right here [demonstrate correct placement]. Okay? Now it's your turn." Then, for the following 12 trials, the experimenter reiterated the

instruction “If there’s a border, play the colour game. If there’s no border, play the shape game”.

The raw total of correct trials for the ‘Pre-Switch Phase’, ‘Post-Switch Phase’ and ‘Border Trial Phase’ was calculated automatically by PsychoPy for each child. Data was scored based on Zelazo (2006): a score of 0 was assigned if the pre-switch phase was failed; a score of 1 was assigned if the pre-switch phase was passed but the post-switch phase failed; a score of 2 was assigned if both the pre- and post-switch were passed, but the border version was failed; and a score of 3 was assigned if pre-switch, post-switch and border version were all passed. Importantly, in contrast to card sort tasks used with older children, such as the Wisconsin Card Sort Task, the higher the score, the *better* the child’s cognitive flexibility.

5.3. Results

To ensure any differences observed between groups were not due to a tendency for the autistic group to use fewer complex referring terms overall, we first compared the total use of complex and simple referring expressions between autistic and TD children. The overall use of complex referring terms did not differ between the autistic (M = 11.03, SD = 5.7) and TD (M = 11.4, SD = 3.87) groups, $t(58) = .292$, $p = .77$, d = 0.08. There was also no significant difference in the overall use of simple referring terms between the autistic (M = 9, SD = 5.7) and TD (M = 8.6, SD = 3.87) groups, $t(58) = -.292$, $p = .772$, $d = -0.08$. This indicates any group differences were not driven by a difficulty using complex language in the autistic group.

Main analysis

To examine the effect of increased cognitive flexibility demands on the production of appropriately informative referring terms, we examined the number of appropriately

informative referring terms used in the Switch and No Switch conditions by both the autistic and TD group (see Figure 5). A mixed 2 (Group: TD/ASD) × 2 (Switch Demand: Switch/No switch) ANOVA was conducted to compare the number of appropriately informative referring terms used. A significant main effect of Switch Demand reflected the use of more appropriately informative referring terms in the No Switch (M = 7.10 out of 10) than the Switch condition (M = 6 out of 10) across groups ($F(1,58) = 9.44$, $p = .003$, $\eta_p^2 = .14$). The main effect of Group was also significant ($F(1, 58) = 5.77$, $p = .02$, $\eta_p^2 = .09$). In line with previous research, this reflects the production of fewer appropriately informative referring terms among ASD (M = 11.8 out of 20) than TD participants (M = 14.4 out of 20) across conditions. Contrary to our hypothesis, the Group × Switch Demand interaction effect was not significant ($F(1, 58) = .42$, $p = .52$, $\eta_p^2 = .007$).

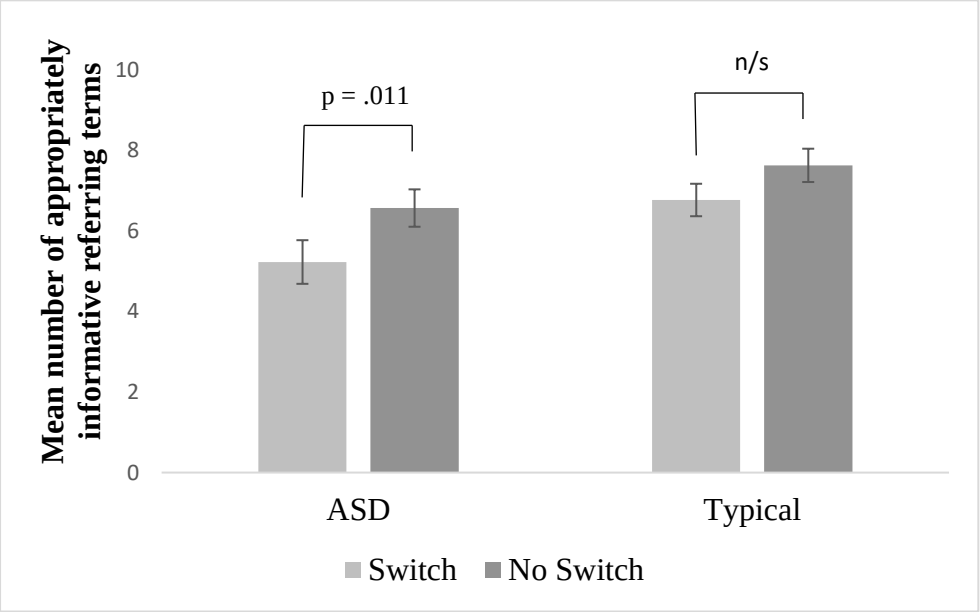


Figure. 5 Mean number of appropriately informative terms used in *switch* and *no-switch* conditions (Error bars = SE from the mean)

Secondary Analyses

Over- and under-informativity

To examine whether the effect of Group was driven by a tendency for the autistic group to be generally more over- or under-informative than the TD group, we compared the total number of occasions on which each group produced referring terms that were either over- or under-informative. There was no significant difference between the number of over-informative referring terms produced by the autistic (M = 4.67, SD = 5.16) and TD (M = 3.9, SD = 3.27) group, $t(58) = -.688$, $p = .494$, $d = 0.18$. Similarly, no significant difference was found between the production of under-informative referring terms between the autistic (M = 2.27, SD = 3.37) and TD (M = 1.17, SD = 2.44) groups, $t(58) = -1.448$, $p = .15$, $d = 0.37$. This was also true when each condition was considered separately (Table 19). This indicates that the difficulty producing appropriately informative referring terms observed in the autistic group is due to reduced ability to adapt to a listener, rather than a general tendency to be over- or under-informative.

Table 19. *Means (SD in brackets) for number of over- and under-informative referring expressions*

		Autistic (n =30)	**TD** (n =30)		
		Mean (SD)	Mean (SD)	*p*	*d*
Switch	Over-Informative	2.93 (3.52)	2.37 (2.25)	.461	0.19
	Under-Informative	1.17 (1.90)	0.60 (1.28)	.180	0.35
No-switch	Over-Informative	1.77 (2.43)	1.53 (1.93)	.628	0.11
	Under-Informative	1.13 (1.70)	0.57 (1.30)	.152	0.37

Perseveration

To examine whether the autistic group were more likely than the TD group to perseverate on referring terms during the *switch* condition, we compared the proportion of incorrect responses that were due to the re-use of the first referring expression for the target animal (the dog). The autistic group was marginally more likely to re-use the first referring term in subsequent trials than the TD group, (ASD: M = .41; TD: M = .23), $t(58) = -1.68$, $p = .099$, $d = 0.43$. This indicates that autistic children tended towards perseveration to a greater extent than the TD group in the *switch* condition. A chi-squared test compared autistic and TD group performance on the first experimental trial of the *switch* condition (where the simple referring term 'the dog' was always appropriate). Both the autistic and TD group were equally likely to provide a correct response to the first trial in the *switch* condition, X^2 (1, N = 60) = 1.92, p .165, indicating a similar ability to select an appropriate referring term for the dog when it was presented for the first time. All incorrect responses on the first trial of the switch condition (12 autistic children and 7 TD children) were coded 'over-informative', that is, a size, colour or pattern modifier was unnecessarily included by these children.

Relationships between RC Task and other variables (cognitive flexibility, language, NVIQ and autistic symptomology)

We first examined the performance of the autistic and TD groups on our direct measure of cognitive flexibility; the DCCS. There was a marginally significant difference between groups (ASD: M = 1.56, SD = 1.04; TD: M = 2.03, SD = .81), t(58) = 1.94, p = .057, d = 0.5), indicating the TD group generally possessed better cognitive flexibility than the autistic group.

To investigate relationships between performance in the referential communication task and our other measured abilities (cognitive flexibility, language, NVIQ and autistic

symptomology) Pearson's Correlations were calculated. When autistic and TD groups were combined, SRS T-scores correlated significantly with performance in the Switch (r (58) = -.284, p = .031) and No Switch conditions (r (58) = -.269, p = .041). This indicates autistic symptomology is related to poorer production of appropriately informative referring expressions overall. There were no other significant correlations between the number of appropriately informative referring terms in the switch or no-switch condition and any other measure when groups were combined (all r < .199, all p > .127|) or examined separately (ASD: all r < .310, all p > .109., TD: all r < .186, all p > .324).

5.4. Discussion

This study explored the role of cognitive flexibility in autistic children's difficulty producing appropriately informative referring terms. A novel manipulation of cognitive flexibility demand in a referential communication task was utilised, that is, the need to refer to a single referent by multiple names in the *switch* condition.

Our first hypothesis, that both autistic and typically developing children would produce fewer appropriately informative referring terms when required to switch between different phrases for the *same* object, was supported. That is, both groups produced significantly fewer appropriately informative referring terms in the *switch* than the *no switch* condition. This indicates an important role of cognitive flexibility in the production of referential communication for all children. Our second hypothesis, that autistic children would produce fewer appropriately informative referring terms than their TD counterparts in both the *switch* and the *no switch* condition, was also supported. This finding is in line with previous literature suggesting a global difficulty in the selection of appropriate referring terms in autism (e.g. Nadig, Vivanti and Ozonoff, 2009; Suh et al. 2014; Fukumura, 2016).

Contrary to our third hypothesis, however, factors relating to cognitive flexibility did not have a disproportionate impact on the autistic children. First, whilst autistic children's performance on the DCCS was marginally worse than their typically developing peers, DCCS scores did not relate to performance on the switch condition of our RC task. Most crucially, the interaction between *group* and *switch demand* was not significant, with a small effect size. That is, the effect of switching was not greater in the autistic than typically developing group. This indicates that impaired CF does not fully explain autism specific difficulties in selecting appropriate referring terms. We, therefore, now consider the factors that may associate with the production of referential communication in autism.

Investigating the role of cognitive flexibility / perseveration in autistic difficulties with referential communication

We proposed that the difficulties autistic children demonstrate with selecting context-appropriate referring expressions might be due to 'getting stuck' using the referring expression they initially used to denote a given object. In fact, not only was there no interaction between Switch Demand and diagnostic group, the latter also did not significantly differ in the degree to which they re-used their original referring terms. Despite this lack of statistical significance, the autistic group were twice as likely as their typically developing peers to re-use their original referring expression. This suggests it may be beneficial to examine differences in autistic and TD children's tendency to perseverate on referring terms with a larger sample size. Nonetheless, the small to moderate effect size of this comparison suggests that the overall poorer performance by the autistic group was not solely due to perseveration.

Indeed, despite difficulties with flexible behaviour in daily life, there is some evidence for heterogeneity in CF ability in autism. Namely, inconsistencies are reported in

performance on clinical and experimental measures in this population (see Geurts, Corbett & Solomon, 2009, for a qualitative review). If a CF deficit is not universal for autistic children, at least during more structured tasks, this may explain why autistic group DCCS scores were only marginally worse than the TD group in the current study. A key question then remains concerning the factors, aside from CF, associated with autistic children's poorer performance in both experimental conditions in comparison to TD children.

Alternative reasons for autistic difficulties with selection of appropriate referring expressions

A first candidate for autistic difficulties in producing appropriately informative referring terms is reduced structural language ability. A strong association is apparent between structural and pragmatic language in both typical and atypical development (see Matthews, Biney & Abbot-Smith, 2018, for a review). In the current study, however, groups were matched for expressive and receptive language. This reduces the likelihood that group differences are explained by structural language ability. Further, there were no significant differences between groups in the number of complex and simple referring terms they produced in each condition, indicating performance differences were unlikely due to a greater difficulties in the autistic group with using complex noun phrases (i.e. noun phrases, which included adjectives or other modifying phrases). Finally, we found no relationship between measures of structural language and performance on our RC task. In sum, structural language does not appear a likely explanation for group differences in RC production ability.

A second factor that might have impinged autistic performance is visual attention. The selection of an appropriate referring term required the comparison of a target with a competitor. Autistic children may have difficulty moving focus between visual stimuli (Keehn, Müller & Townsend, 2013; Sacrey, Armstrong, Bryson & Zwaigenbaum, 2014),

may pay attention to unusual aspects of stimuli (Roberton & Baron-Cohen, 2017) and exhibit differences in eye movements that prevent optimal processing (Kemner, Verbaten, Cuperus, Camfferman & van Engeland, 1998). These visual attention differences could increase the likelihood that the target and competitor animal are compared less accurately, resulting in fewer appropriately informative referring expressions across conditions. As the presence of visual processing difficulty in autism is controversial (e.g. Jones et al. 2011), this explanation remains speculative.

A third potential source of difficulty is the oft-cited challenge autistic children experience with perspective taking (Hamilton, Brindley & Frith, 2009). Reduced ability to consider the information a listener has access to could reduce the extent to which referring terms are appropriately tailored to their needs. In the current study, perspective taking demands were arguably minimal. Children were required to recall that the listener could not see the green square highlighting the target animal, but otherwise there was no conflict between the visual perspective of child and listener. Further, autistic children are found to perform similarly to TD children on RC comprehension tasks that require perspective taking (Chapter 2). Whilst a failure to consider the knowledge of the listener may have influenced ability to produce appropriate referring terms, it seems unlikely that this would be the sole source of group differences.

Finally, as with most daily activities, our RC task likely placed demands on working memory. In both *switch* and *no switch* conditions, children were required to hold in mind the rules of the game whilst simultaneously assessing the two referents, planning, and producing an utterance. It is possible that the difficulties with working memory reported in autism (Kercood et al., 2014; Funabiki & Shiwa, 2018) resulted in poorer performance across both conditions. Indeed, during reference comprehension, eye-tracking revealed that autistic children tended to look towards a non-target referent more frequently than the TD group

under increased working memory demands (Schuh, Eigsti & Mirman, 2016). This suggests autistic children might be disadvantaged during tasks requiring WM ability. Poor working memory may also play a role in autistic children's difficulty with appropriate reference selection during discourse (Arnold, Bennetto & Diehl, 2009; Kuijper, Hartman & Hendriks, 2015). Studies that manipulate working memory demand during the production of referring expressions are required to better understand the role of this cognitive ability. The potential role of WM in autistic difficulties with RC will be revisited in the general discussion chapter of this thesis.

In sum, a range of factors other than CF might contribute to autistic difficulties in the production of RC. Reduced ability in a number of these areas might operate in combination to affect RC, or a deficit in one area might influence RC ability via another. Understanding whether, and how, these multiple factors are interconnected, or jointly influence poorer RC production, in autistic children remains a primary question. A further question regards the factors that contributed to the poorer performance of both groups in the *switch* when compared to the *no switch* condition in the current study.

The role of switching demand in the selection of an appropriate referring expression

The requirement to switch referring terms for the same referent resulted in the production of fewer appropriately informative referring terms for both groups in comparison to the 'No Switch' condition. However, we failed to detect a relationship between our direct measure of CF (the DCCS) and performance in the *switch* condition of the RC task. This may bring into question the extent to which reduced performance in the *switch* condition is due solely to CF demands. That said, measures such as the DCCS that are designed to minimise between-participant variability (and thus reveal between-group differences) may be problematic for the study of individual differences, which require maximum between-

participant variability (Hedge, Powell & Sumner, 2017). Hence, the lack of correlation between the DCCS and *switch* condition of the RC task may be the result of the measure of CF employed, rather than reflecting performance in the *switch* condition was unrelated to CF ability.

It is clear is that the requirement to switch the referring expression previously used to denote a given animal made the RC task more difficult. However, this does not clarify which specific aspect of the requirement to switch caused the children difficulties. Perseveration is a relatively narrow measure of CF (Carroll, Blakey & FitzGibbon, 2016) and our RC task may have recruited a wider range of CF ability than that measured by the DCCS. D*istraction* errors, for example, reflect sporadic shifts in item representation that relate to neither a new nor old rule (Barcelo´ & Knight, 2002; Chevalier & Blaye, 2008; Blakey & Carroll, 2018) and these too may have exerted an influence upon *switch* condition performance.

Alternatively, or additionally, it is possible that the *switch* condition placed greater demands than the *no switch* condition on a range of cognitive processes. First, the need to repress a previously activated referring expression would likely be higher in the *switch* condition, implicating a greater tax on inhibition. Second, switching referring terms for the same animal may place greater working memory demand, with the need to hold a range of features of the dog (size, pattern, colour) in mind whilst selecting which of these is relevant to the current pairing. Whilst this is also a requirement in the *no switch* condition, working memory demand is arguably increased in the *switch* condition as different features of the same dog must be considered over successive trials. Indeed, both inhibition and working memory ability related to the ability of 9 - 12-year-old TD children to produce appropriately informative referring terms in a RC task (Nilsen, Varghese, Xu, & Fecica, 2015). Finally, following Kloo and Perner's (2005) 're-description hypothesis', the *switch* condition could be viewed as increasing the need for children to conceptualise the same object in a different

way. Thus, the need to hold multiple perspectives on the same animal in the *switch* condition may have resulted in poorer performance for both groups due to higher demands on conceptual abilities in this condition. Further research is required to unravel these possibilities.

Limitations and Future directions

Whilst we found a main effect of *switch demand* and *group*, our sample size may have limited the power of our analyses to detect an interaction between these two factors. An a priori power analysis revealed that a sample size of 64 per group would be required to detect a medium size effect with a power of .80 (Cohen, 1992). This indicates our study may have been under-powered. The Group × Switch Demand interaction, however, had a small effect size. This suggests that, even if a larger sample size had been obtained, it is unlikely an interaction would have emerged between these factors.

Disentangling the components that contribute to autistic children's difficulty in the production of appropriate referring terms, and how these interact with one another, remains a priority. One approach is to examine contributory factors in relative isolation. Perspective taking demands might be manipulated, for example, by increasing the transparency of the listener's perspective via the provision of feedback; working memory demands, by varying the number and complexity of referents; or eye-tracking utilised to assess the role of visual attention. Similarly, regarding our understanding of referential communication in both typical and atypical populations, experiments that manipulate, for example, working memory or inhibitory control demands in RC production tasks might elucidate the cognitive processes related to this ability. A potentially greater challenge, however, concerns understanding how these abilities interact with one another. It appears unlikely that an isolated mechanism is responsible for autistic difficulties in referential communication, or the development of

referential communication more generally. Instead, considering RC production within a wider model of pragmatic and communicative ability that integrates both social and cognitive factors, such as that proposed by Nilsen & Fecica (2011) or O'Neill (2012), might provide a useful framework for future research efforts.

Summary

This is the first study to experimentally manipulate the need to switch referring terms for a given referent within a RC task. Our findings indicate that the requirement to change the referring term for a given referent reduces the number of appropriately informative referring expressions used by both autistic and typically developing children. Further, switching referring terms for a referent appears harder than the requirement to switch, for example, between complex and simple referring terms. We have demonstrated that the ability to switch referring terms for a given referent is likely a key mechanism in the puzzle of why individuals do not always select referring expressions appropriately, even when they are aware of listener informational needs. Identifying the relative contribution of additional cognitive and social processes remains a key question for our understanding of the successful production of referring expressions, and why difficulties might occur.

CHAPTER 6: GENERAL DISCUSSION

6.1. Summary of Findings

The overarching aim of this thesis was to better understand pragmatic language ability in autism. For the purpose of this thesis, 'pragmatic language' was primarily considered to be the use of social context to interpret language and select which language forms to produce. A broad deficit in pragmatic language is frequently reported in autism (Baltaxe, 1977; Simmons, Paul & Volkmar, 2014) as outlined in Chapter 1, but surprisingly little attention had been paid to whether these deficits exist equally across all areas of pragmatic ability. The first aim of this thesis was therefore to explore the extent to which one pivotal aspect of pragmatic language, namely referential communication, is globally impaired within autism. The systematic literature review of verbal reference in autism reported in Chapter 3 revealed a pattern whereby autistic individuals tended towards an impairment in the production of referring expressions compared to TD controls, but no impairment in the comprehension of reference. It was speculated that this difference may be due to all previous comprehension tasks requiring only consideration of the communication partner's *visual* perspective (Volden et al. 1997; Begeer et al. 2010; Santiesteban et al. 2015). However, the experimental study reported in Chapter 4 indicated that, even when one aspect of *social* perspective taking (the ability to determine the knowledge one shares with a specific interlocutor; Moll & Kadipasaoglu, 2013) was required, autistic children remained as likely as typically developing children to correctly interpret an ambiguous referring expression. This pattern of results suggests not all areas of pragmatic language are equally impaired in autism, and some may even be comparable with typically developing individuals.

Despite the position of pragmatic language as a prominent feature of autistic difficulties in the literature, Chapter 1 and 3 highlighted that knowledge of the cognitive underpinnings of these abilities in autism is relatively sparse. A secondary aim of this thesis

was therefore to explore a range of factors that might be associated with pragmatic language difficulties in autism, such as; the understanding of social common ground, theory of mind, structural language proficiency, and an aspect of executive functioning; cognitive flexibility.

The empirical study reported in Chapter 4 unexpectedly indicated that autistic children were as able as well-matched TD peers to consider the social common ground shared with a communicative partner during both the production and comprehension of referring expressions. Further, despite poorer ToM in our autistic group, this ability did not relate to the tendency to tailor referring expressions to listener needs in this referential communication task. Similarly, no relationship emerged between ToM and wider pragmatic ability as measured by the CCC-2 parent report detailed in Chapter 2. These findings are not in line with research (e.g. Hale and Tager-Flusberg, 2005; Le Sourn-Bissaoui, Caillies, Gierski & Motte, 2009) indicating relationships between poorer social cognition and pragmatic language difficulties of autistic children who have language and non-verbal IQ in the typical range. A further unanticipated finding was the failure to detect a relationship between structural and pragmatic language across empirical studies within this thesis. One obvious reason for these null results might potentially be sample size, but other potential reasons for this are discussed in section 6.3.

Finally, whilst autistic children demonstrated awareness of the need to alter referring terms to listener needs, they were less successful at doing so than TD controls (Chapter 4). The contribution of one aspect of executive functioning to this performance deficit was examined in Chapter 5, namely, cognitive flexibility. In a RC task, fewer appropriately informative referring expressions were produced under higher cognitive flexibility demands, that is, when there was a requirement to use different referring terms for the same referent. This indicated a key role of CF in RC. The effect, however, was not specific to autistic children. Hence, while cognitive flexibility is part of the puzzle of children's ability to

production of referring expressions, it does not appear to fully explain autistic specific difficulties.

Interpretations of these findings will now be considered, along with potential implications for our understanding of pragmatic language in autism.

6.2. Is pragmatic ability universally impaired in autism?

Reduced pragmatic language abilities are accepted as a key feature of autism in most literature to date (Frith, 1989; Martin & McDonald, 2003; Tager-Flusberg, Paul & Lord, 2005; DSM-5, APA, 2013). Indeed, the broad-brush measurement of pragmatic language employed by the study in Chapter 2 (the pragmatic composite of the CCC-2) indicated a pattern of pragmatic language impairment in autistic children; more than 90% were reported by parents to exhibit pragmatic language ability one standard deviation or more below the typical mean. This pattern of impairment held across all five sub-scales comprising the pragmatic composite (coherence, inappropriate initiation, stereotyped language, use of context, non-verbal communication). Across all sub-scales at least 82% of autistic children scored one standard deviation or more below the mean. Closer examination of the experimental evidence from previous studies outlined in the introduction to this thesis, however, indicated some areas of pragmatic functioning in autism may be unimpaired relative to TD controls. This suggested a more nuanced examination of pragmatic language in autism might be valuable. Thus, we applied a finer-grained approach to examine one key element of pragmatic language, namely referential communication, in autism.

The expected pattern of difficulty was observed in the *production* of referring terms relative to TD controls across both the systematic review in Chapter 3 and the empirical studies reported in Chapters 4 and 5. Regarding the *interpretation* of referring terms, however, no such impairment was evident within previous studies or those in this thesis. This

is in line with a recent review of wider pragmatic reasoning ability in autism (Geurts, Kissine & Tiel, 2019), where a pattern of partially intact utterance interpretation is reported across a range of pragmatic abilities. A question remains, however, regarding why reference production is more problematic for autistic individuals than interpretation.

One potential explanation for the discrepancy between autistic ability to produce and interpret reference is task complexity. To illustrate, the narrative and conversation tasks employed to elicit reference in the literature review (Chapter 3) required the formulation of sentences, which burdens working memory (Slevc, 2011). In contrast, listening to an utterance and deducing to which referent it relates, as required in the interpretation version of the director task, might involve lower WM demands. Indeed, speech planning is found the most cognitively demanding aspect of conversation in non-autistic adults (Boiteau, Malone, Peters & Almor, 2014). Notably, all studies of reference comprehension included in the systematic review employed the director task (Volden et al. 1997; Bergeer et al. 2010; Santiesteban et al. 2015), lending support to this explanation. Further, even within the director task, the production of referring terms likely places greater demand on executive function than does comprehension. Producing a referring term requires, for example, identifying visual differences between a target and competitor, accessing the relevant lexical items (the ease of which depends on activation levels of certain lexical items such as ‘black’ versus others, such as ‘stripy’), inhibiting irrelevant lexical items (e.g. suppressing ‘black’ if irrelevant) then formulating and producing an utterance, taking into account another’s visual perspective. That said, EF correlates with comprehension as well as production in director tasks in typical development (Zhao, Wang & Apperly, 2018; Lin, Keysar & Epley, 2010; Brown-Schmidt, 2009; Nilsen & Graham, 2009), hence the effects of EF are not absent in the comprehension version of this task.

Interestingly, autistic children's errors in the production task outlined in Chapter 4 tended to be in the direction of under-informativity. That is, when addressing an unknowledgeable listener, autistic children were more likely than typically developing counterparts to omit a modifier that was required to enable a referent to be identified. During the RC task in Chapter 5, conversely, incorrect responses tended towards over-informativity. That is, a modifier was used when it was not necessary to identify the given referent. When perceptual salience of a referent is increased, speakers are more likely to use a modifier (Davies & Arnold, 2018). Arguably the animal patterns utilised in Chapter 5 had high salience, which may explain the tendency towards over-informativity in this task. As the reference did not have to be uttered as part of a sentence in the RC task in Chapter 5, this may also reflect that more cognitive resources were available to focus on the inclusion of a modifier. Whilst both over and under-informative referring expressions were considered inappropriate in the studies reported in this thesis, over-informative references are arguably less likely to result in communicative breakdown, and may signal an attempt to cater to listener needs (Davies & Katsos, 2016). Identifying the circumstances under which autistic children are more likely to use over or under-informative referring expressions may enhance understanding of the nature of referential communication production difficulties in autism.

The pattern of production difficulties alongside intact reference interpretation summarised in Chapter 4 might also be attributed to differences in task complexity. Here, comprehension involved recalling the previous interaction shared with an interlocutor and selecting one of two potential referents, whilst the production task additionally required the utterance of referring expressions within a sentence e.g. 'Put the <u>big battery</u> into the battery holder'. Hence, a higher cognitive load in this task may again explain discrepancies in performance. The reference interpretation task, however, required children to hold in mind the experiences they had shared with a given adult whilst processing the phrase uttered by

that adult and selecting an activity accordingly, which also likely recruited EF. In sum, it cannot be confidently concluded that the differences between reference production and comprehension in autism are purely a result of task complexity.

Alongside potentially differing EF demand, the requirement to take another's perspective may differ between production and comprehension tasks. Indeed, a debate exists around the extent to which perspective taking is required in utterance interpretation more generally (e.g. Kissine, 2016; Geurts. 2019). It is possible that interpretation of reference might be achieved without complex reasoning about another's mental state, an idea we return to in section 6.3. Difficulty with interpreting why autistic individuals perform differently in reference production versus comprehension tasks partly stems from a lack of clarity regarding the skills required within these tasks. The need to better understand the abilities required to succeed across a range of pragmatic language is returned to in section 6.3.

Finally, before concluding that reference interpretation is intact in autism, a question remains concerning whether this ability would endure in real-life interaction. The findings reported in this thesis suggest that autistic individuals can identify a speaker's intended reference, at the very least, from a limited selection in response to a single utterance. Extraneous demands are clearly higher when interpreting reference in real-life conversation, where a range of referents are tracked during dynamic dialogue and environments, alongside multiple potential distractions. Difficulties with reference interpretation in real-life interaction may therefore occur due to the additional demands placed by the fast-paced nature of the conversational context, rather than inability to identify a speaker's intended reference, per se.

In sum, whilst pragmatic language difficulties appear to be a feature of autism, a deficit compared to TD controls is not evident in all pragmatic language tasks. Whether this reflects some intact pragmatic abilities, or differences in the demands placed by given tasks,

remains unclear. The extent to which the apparent dichotomy between appropriate comprehension and production of referring terms extends into real-life interaction also warrants attention. These findings highlight the potential benefit, firstly, of examining sub-components within areas of pragmatic language to understand the extent to which pragmatic difficulties are universal in autism, and secondly, categorising skills according to the cognitive processes they involve, rather than by pragmatic language 'label'.

6.3. Cognitive underpinnings of referential communication (and broader pragmatic language) in autism

The systematic literature review of verbal reference in autism (chapter 3) highlighted that the cognitive underpinnings of referential communication in autism are not well understood. This is also true for pragmatic abilities in autism more broadly (as discussed in the general introduction). Here, evidence from the studies within this thesis regarding factors that might contribute to autistic difficulties with referential communication and pragmatic language more broadly are considered.

The role of social cognition

Social common ground

The literature review in chapter 3 indicated that autistic individuals can interpret referring terms as well as typically developing controls when *visual* perspective taking is required (e.g. Volden et al. 1997). Given the well-documented difficulties with social cognition in autism (Tager-Flusberg, 2000), it was predicted that interpretation of reference would be problematic for this group when *social* perspective taking (the ability to determine the knowledge one shares with a specific interlocutor; Moll & Kadipasaoglu, 2013) was required. Contrary to this prediction, the study reported in Chapter 4 indicated that autistic

children can take one aspect of social common ground, namely the experiences shared with an interlocutor, into account during both the interpretation *and* production of referring terms. An inability to consider social common ground then does not appear a key factor in autistic children's difficulty with referential communication, at least when it involves tracking the experiences shared with a given interlocutor. Given the difficulties reported in social cognition in autism, this finding is somewhat surprising and warrants further consideration.

A potential explanation for the similar performance of autistic and TD children in the reference interpretation task reported in chapter 4, is the relative simplicity of this task. It is possible that if a greater time delay, number of referents or conversational partners were involved then autistic children may have failed to consider social common ground. That said, as previously discussed, failure might then be attributed to additional factors, such as working memory. The nature of conversational partner interaction in this task might also be relevant. Clark and Marshall (1981) suggest referents are brought into common ground via physical co-presence (when the referent is visually present and attended to by both conversational partners) or linguistic co-presence (when a referent is introduced by a speaker within discourse). In the experimental tasks in chapter 4, referents were visually and linguistically shared. Clark & Marshall (1981) suggest that physical co-presence provides stronger evidence for common ground than linguistic co-presence. To illustrate, in the sentence 'The car is in the garage' identifying which garage is the intended referent is likely easier if there is a garage door clearly visible to both conversational partners, for example, at the car owner's home. Without such physical co-presence, establishing whether the utterance 'the garage' refers to the garage at home, or perhaps a mechanical repair garage, could require the speaker to provide further information, or listener to employ more complex inference. That is, when referents are shared only within speech, identifying the knowledge or intentions of one's interlocutor may be harder. Keeping track of referents may also make higher memory

demands when they are shared only linguistically. In our experimental tasks, potential referents were always physically and linguistically co-present, and limited in number, which may have contributed to the ability of autistic children to utilise common ground knowledge. That said, autistic children are able to correctly answer questions about stories that do not involve consideration of a character's mental state (White, Hill, Happe´ & Frith, 2009), which presumably would involve tracking referents within narrative. Nonetheless, it remains unclear if sharing referents in back and forth conversation (without physical interaction with those referents) would result in the same appropriate consideration of common ground for autistic individuals.

A question also remains regarding which social cognition skills were required to succeed in the interpretation task. As discussed in the general introduction, alongside common ground and ToM, shared intentionality may be a key skill for communication (Tomasello, 2018). Whilst success in the reference interpretation task demonstrated that autistic children are capable of understanding another's goal (in line with Carpenter, Pennington & Rogers, 2002), it did not necessarily require children to engage in *sharing* these intentions with the adult. Rather, as discussed in Chapter 4, it is possible that children simply tracked the adult's actions as they engaged in the shared task, and then connected the adult's reference to the last activity with which the child saw them engage. If autistic individuals do consider social common ground, it remains possible that a difficulty in another area related to shared intentionality, such as the ability to form a joint commitment, could contribute to problems with pragmatic language. A joint commitment is an agreement between two or more people to act together in some way (Gilbert, 2009) and is proposed by Bratman (1992) as a key characteristic of shared cooperative activities. In the sense that conversation requires individuals cooperate to share meaning (Tomasello, 2018), difficulty forming joint commitments might reduce an individual's ability to, for example, make

contributions relevant to the conversation, or recognise how a conversational partners utterance is relevant to the current context.

Whilst our interpretation task was arguably simplistic, performance on the production task provides additional evidence that social common ground was utilised by the autistic children in this study. Specifically, the difference in modifier use with a knowledgeable versus unknowledgeable listener indicates autistic children had awareness of social common ground and attempted to engage in audience design. The interactive and goal driven nature of our reference production task (to provide instructions to enable a listener to make a toy) may have contributed to children's motivation to adjust information to listeners needs (Grigoroglou & Papafragou, 2016; 2019). That said, autistic children adjusted their referring expressions to a lesser extent than their TD counterparts, suggesting difficulties might be related to something other than a failure to consider social common ground. This possibility is retuned to shortly.

Before concluding that autistic children are able to utilise social common ground to inform referential communication, it should be noted that there are multiple types of social common ground. There is some evidence that autistic children may not utilise *all* of these to interpret reference. Abbot-Smith, Williams and Matthews (2020), for example, found autistic children less able than well-matched typically developing peers to interpret an ambiguous reference based upon a speaker's emotional state. In this case, autistic children had more difficulty selecting a referent based on the understanding that a speaker is likely to respond to with greater excitement to a new item than one previously viewed. This task arguably involved a deeper consideration of another's mental state than our social common ground interpretation task, alongside the requirement to track which items are new to the context and understand that new items are generally more comment worthy than old. Nonetheless, it raises the possibility that autistic individuals are not proficient in the use of all aspects of

social common ground to interpret reference. Whether similar discrepancies extend to reference production, or broader areas of pragmatic ability, has yet to be explored.

Finally, despite demonstrating an ability to consider social common ground knowledge (at least the experiences shared with a specific interlocutor) when interpreting and producing referring expressions, autistic children performed significantly worse than TD peers on a battery of ToM tasks. We now consider the role of ToM in autistic pragmatic language difficulties, and its relationship with common ground.

ToM

The consideration of another's mental state intuitively appears important for pragmatic language, including the use and interpretation of referring terms. The nature of the relationship between these abilities, however, is unclear in both typical and atypical populations (Bosco, Tirassa & Gabbatore, 2018). Given the well-documented deficit in ToM reported in autism, we anticipated that this ability might be related to pragmatic language in this population. Unexpectedly, we did not detect a relationship between ToM ability and the pragmatic composite of the CCC-2 (Chapter 2) or the RC production tasks in Chapter 4. Whilst a relationship emerged between ToM and performance on the reference interpretation measure in Chapter 4, three autistic children failed both first-order theory of mind tasks yet succeeded on all interpretation measures. This brings into question the extent to which mentalising was required in this task. A key question then concerns why an association between ToM and pragmatic ability did not emerge across all studies. We consider four possibilities; first, that ToM is not directly related to pragmatic language difficulties in autism; second, that the measures in our studies were not sufficient to detect a relationship; third, that our sample size was insufficient; and fourth, that some, but not all, pragmatic

language tasks require ToM. To address the first possibility, studies that find a relationship between ToM and pragmatic language in autism will now be examined.

Two previous studies reported a relationship between ToM and pragmatic language in autism as measured by the CCC-2. As detailed in Chapter 2, the first did not control for language or non-verbal IQ (Losh et al., 2012), whilst the second measured both ToM and pragmatic language ability using parental questionnaire methods (Baixauli-Fortea Casas et al., 2017). Hence a bandwagon effect may have been responsible for the relationship observed between measures in this study. Regarding direct measurement of reference and ToM, a relationship existed between ToM and reference production within narrative (Kuijper et al., 2016) and a referential communication task (Dahlgren & Sandberg, 2008). Again, structural language ability was not controlled for in either of these studies, rendering the nature of this relationship unclear.

As described in Chapter 2, three previous studies reported a relationship between ToM and discrete areas of pragmatic language in autism when structural language was controlled (Hale & Tager-Flusberg, 2005; Whyte et al. 2014; Schuh, Eigsti & Mirman, 2016). All three studies, however, raised questions regarding the nature of the relationship between these skills. Firstly, the role of working memory in mediating this relationship (Schuh et al., 2016), secondly, the interplay between ToM, pragmatic and structural language (Whyte et al., 2014), and finally, whether ToM should be viewed as a multi-componential construct, whereby not all areas of ToM would exert an equal influence on all areas of pragmatic ability (Hale & Tager-Flusberg, 2005). It is also possible that different pragmatic language tasks place differing demands on ToM, for example, selecting an utterance that follows from one's conversation partner (as found to correlate with ToM by Hale & Tager-Flusberg, 2005) may require more complex consideration of their current mental state than does the production and comprehension of referring expressions. In sum, the three studies to report a relationship

between ToM and pragmatic ability in autism highlight further exploration of the relationship between these two factors is required, particularly the skills that may mediate the relationship between these abilities.

We now consider a second possible explanation for our findings; namely that our measure of ToM (first – and second-order false belief and strange stories) was insufficient. A lack of variance in measures can prevent the detection of relationships. As outlined in Chapter 2, however, there was sufficient variability in ToM scores. Hence, a lack of variance was not a key factor in explaining the lack of relationship between measures. A second issue relates to the extent to which the tasks measured only ToM ability. Whilst groups were matched for structural language ability, a relationship existed between ToM and structural language in the autistic group. This raises the possibility that autistic children with better language ability may have compensated for difficulties with ToM. To obtain a 'purer' measure of mentalising ability, measures of ToM that minimise structural language burden would be a useful addition (e.g. Grueneisen, Wyman & Tomasello, 2015). The development of novel measurements of ToM, including interactive contexts and virtual reality, are gaining interest (see Livingston, Carr & Shah, 2018 for an overview). These too might provide valuable tools for future research. Finally, whilst the tasks we utilised are recognised as suitable for measuring ToM and are frequently utilised in this type of research (Devine & Hughes, 2017; Jones et al. 2018; Beaudoin, Leblanc, Gagner & Beauchamp, 2020), ToM may not be a unitary construct (Hayward & Homer, 2017; Warnell & Redcay, 2019). It is possible that areas of ToM other than those measured in our studies may be more relevant for pragmatic language, or that specific types of ToM are recruited for specific pragmatic language abilities. Beaudoin et al., 2020, for example, separate measures of 'belief' (such as the 'Sally-Anne' task used in our ToM battery) from measures of 'knowledge' (such as the 'Awareness of a reader's knowledge'; Peskin, Prusky & Comay, 2014) and 'emotions' (such

as the ‘Test of emotion comprehension’; Pons and Harris, 2000). Relationships might, therefore, exist between specific areas of ToM and pragmatic ability, rather than global measures of both skill sets.

A third possible reason a relationship was not detected between ToM and pragmatic ability is that our sample size limited the statistical power of analyses. Indeed, achieving a sample size that adequately achieves statistical power within correlational research is a widely recognised challenge (Hedge, Powell & Sumner, 2018). Within both Chapter 2 and 4, however, whilst sample sizes were modest, effect sizes were consistently small. This indicates the sample size was unlikely instrumental in our failure to detect a relationship between pragmatic language and ToM. We, therefore, now turn to a fourth possibility; that ToM was not the key skill required to succeed in our RC tasks.

As previously described, consideration of social common ground was likely required during the interpretation and production RC tasks reported in Chapter 4. Despite a relationship emerging between ToM and the interpretation measure in Chapter 4 for the autistic group, three autistic children achieved perfect performance in the interpretation measure whilst failing both first-order theory of mind tasks. It therefore remains possible that complex consideration of another’s mental state was not required to interpret the interlocutor’s utterance. These findings may be explained by the developmental perspective of common ground proposed by Bohn and Köymen (2017). In this proposal, ToM is an outcome of common ground understanding, rather than a prerequisite for it. Thus, it is possible that the autistic children in our study had at least implicit awareness of social common ground yet diminished ToM. In fact, it has been suggested that some utterances can be understood without the need to infer another’s perspective (Geurts, Kissine & van Tiel, 2019). Some pragmatic interpretation tasks may require only structural language ability, for example, the comprehension of indirect requests (Deliens, Papastamou, Ruytenbeek,

Geelhand and Kissine, 2018) and scalar inference (Katsos, Andrés-Roqueta, Estevan, & Cummins, 2011). Others may additionally require ToM, for example, irony identification requires consideration of speaker intention (Filippova & Astington, 2008; but see also Bosco & Gabbatore, 2017). These task types have been labelled 'linguistic-pragmatics' and 'social-pragmatics' respectively (Andrés-Roqueta and Katsos, 2017).

Our 'co-presence only' control condition demonstrated that children considered social common ground in our RC interpretation task (Chapter 4), but it does not necessarily follow that ToM was required. If this is the case, 'social-pragmatics' might arguably benefit from division into further sub-components, such as tasks requiring common ground, but not fully-fledged ToM, for example. That said, whilst it seems feasible that utterance interpretation could, on some occasions, be achieved without consideration of a speakers mental state, it is harder to envisage how successful conversation could be achieved without some consideration of listener knowledge and needs (Geurts & Rubio-Fernández, 2015; although see Keysar, Barr & Horton, 1998, for suggestion that language may sometimes be produced from an egocentric perspective and Grigoroglu & Papafragou, 2019, for an argument that some adjustments to the information provided by speakers may be based on heuristics concerning what is relevant or salient in a specific situation).

In sum, the skills that mediate the relationship between ToM and pragmatic language in autism, and whether the relationship exists across all areas of pragmatic language ability, remain unclear. Determining how ToM demands vary across pragmatic language tasks might provide a useful basis from which to unravel if, and how, this ability relates to difficulties with pragmatic language in autism. Consideration should also be given to accurate ToM measurement methods. Given the well-documented drawbacks of correlational studies, an approach where ToM demands are experimentally manipulated, or abilities measured over several points in time, would be favourable. Clearly, better understanding the relationship

between ToM and pragmatic language in autism remains a key issue. Heterogeneity, however, also exists in the structural language abilities of autistic individuals. It is therefore possible that 'linguistic pragmatics' tasks (i.e. those requiring structural language ability, but lower ToM requirements, such as scalar inference) may also remain challenging for some autistic individuals. We now consider the role of structural language in autistic difficulties with pragmatic language.

The role of structural language

Contrary to the pattern reported in previous literature, an inconsistent relationship was evident between measures of structural and pragmatic language in the current studies. The only clear relationship to emerge was between the pragmatic and structural language composites within the CCC-2. As discussed in Chapter 2, it is possible this finding was the result of a bandwagon effect, as parents completed all questionnaire items. A relationship between structural and pragmatic composites within the CCC-2 for autistic children was similarly reported in one previous study (Baixauli-Fortea et al., 2017). Other studies utilising the CCC-2 with autistic children (Losh et al.,2012; Kuijper et al. 2017) did not report whether a relationship existed between these composites. The relationship between the CCC-2 composites may also reflect the difficulty discussed in Chapter 1 of separating 'pragmatic' and 'structural' abilities. Whilst the CCC-2 can distinguish between various clinical groups and TD children (Geurts & Embrechts, 2008; Gibson, Adams, Lockton & Green, 2013), classification of some sub-scales as 'pragmatic' or 'structural' (in particular the 'semantics' and 'coherence' sub-scales) has generated debate (see e.g. Volden & Philips, 2010; Ash, Redmond, Timler & Kean, 2017). The relationship between the structural and pragmatic composites of the CCC-2 in Chapter 2 may, therefore, stem from the inevitable crossover between some of the skills comprising these composites. Given the widely reported relationships between pragmatic and structural language (see Matthews et al., 2018 for a

review) we now consider why a broader relationship between structural and pragmatic language was not evident across our studies.

A first consideration is that our direct measures of structural language were used primarily to enable participant matching, and hence consisted of one comprehension and one production task from the CELF in all studies. Had they incorporated a wider range of structural language abilities a correlation may have emerged. That said, previous studies have found pragmatic and structural language to be related even when the latter is measured only by vocabulary measures (Akbar, Loomis & Paul, 2013; De Rosnay et al. 2014; Bacso & Nilsen, 2017), hence our measure of structural language does not seem a compelling explanation for the failure to detect a relationship. Alternatively, as previously discussed, correlational designs require adequate variation in the measures used to detect relationships. Whilst huge heterogeneity exists in structural language ability in autism (e.g. Boucher, 2012; Wittke, Mastergeorge, Ozonoff, Rogers & Naigles, 2017), we reduced this diversity through selecting our autistic participants to be mostly within typical range for structural language. Thus, variance in scores was accordingly limited and may have affected our ability to detect relationships with this measure. That said, some relationships with structural language did emerge, for example, with ToM (Chapter 2). A lack of variance in our structural language measure, therefore, does not appear a completely convincing explanation for our findings.

A second consideration is that pragmatic language difficulties can persist despite intact structural language ability (e.g. Baird & Norbury, 2016). Indeed, intact structural language alongside pragmatic language difficulties have been noted within autism (Lord & Paul, 1997; Tager-Flusberg, 2000). In support of this possibility, despite being matched for structural language, difficulty with pragmatic language relative to TD controls remained to some extent for autistic children across all studies in this thesis. This suggests pragmatic language difficulties stemmed at least partly from areas other than structural language. Even

in TD adults there is some evidence that pragmatic and structural language might be separable (Wilson & Bishop, 2019). If structural and pragmatic language ability are at least partially separable in autism, this may explain why a relationship between these skills did not emerge in our studies. That said, it should be noted that despite the vast majority of children scoring in the typical range on our direct measures of structural language (CELF), a sizable proportion of autistic children were rated as experiencing structural language impairment as rated by their parents on the CCC-2. Structural language difficulties might be more evident for autistic children in less structured situations (King & Palikara, 2018). That said, structural language as measured by the CELF has been shown to strongly correlate with spontaneous speech measures, so should be treated as a robust indication of structural language ability (Condouris, Meyer & Tager-Flusberg, 2003).

In sum, we found only weak evidence for the role of structural language in pragmatic language difficulties in autism. This finding was unexpected given the relationship reported in previous literature (Hale & Tager-Flusberg, 2005; Volden, Coolican, Garon, White, & Bryson, 2009; Whyte & Nelson, 2015). Whilst structural language seems unlikely to be the sole reason for pragmatic language difficulties in autism, clearly reduced structural language ability would be expected to affect pragmatic language skill. For autistic children who have lower than average structural language abilities, improving these skills will contribute towards positive outcomes for pragmatic language (Parsons, Cordier, Munro, Joosten & Speyer, 2017). However, as structural language ability does not appear a key determinant of pragmatic language difficulties for autistic children in the current studies, we now turn our attention to a factor that might potentially limit the ability of children to fully utilise their structural language capabilities, namely executive functioning.

The role of cognitive flexibility

To briefly take stock, Chapters 2 and 4 indicated that social-cognition and structural language did not play a crucial role in the pragmatic language difficulties faced by the autistic children in these studies. In Chapter 4, autistic children produced more modifiers when addressing an unknowledgeable than knowledgeable listener but were less successful at doing so than a well-matched TD group. This suggests listener needs were considered by these autistic children, but another factor limited their ability to produce appropriately informative referring expressions. Given the relationship between executive functioning and pragmatic language reported in TD children, and autistic children's difficulty in this area, this seemed a promising skill set to explore. Cognitive flexibility was identified as a theoretically strong candidate for autistic difficulties with the production of appropriately informative referring expressions. This question was, therefore, the focus of the study reported in Chapter 5.

The study reported in Chapter 5 revealed a clear effect of the demand to switch referring terms for a given referent in a RC task. To our knowledge, this is the first time the role of CF in RC has been demonstrated experimentally. The effect, however, was not greater in autistic than TD children. This indicates that, while cognitive flexibility is important for the appropriate production of referring terms, it does not fully explain autistic specific difficulties with reference production. That said, closer inspection of the referring expressions produced during the RC task suggested perseveration might have contributed to autistic specific difficulties. Specifically, the autistic group were more prone than the TD group to re-use the first referring expression for the target animal in the 'switch' condition (though this did not reach statistical significance). This suggests perseveration was more prevalent in the autistic group. Given the real-life difficulties reported with perseveration in autism (Frith & Happé, 2005) its role in autistic specific pragmatic language difficulties warrants further exploration. Whilst perseveration might have contributed to difficulties in the *switch* condition,

performance was also poorer than the TD group in the *no switch* condition, where switching between referring expressions for the same referent was not required. If difficulties relative to TD controls persist when switching is not required, cognitive flexibility cannot fully explain autistic difficulties with referential communication and future research might explore other cognitive processes.

The role of broader EF

As previously alluded, a glaring difference between the production and comprehension of referring terms is the additional cognitive and linguistic demands of utterance planning and production required during production. One factor that this thesis did not examine, but which may create autistic specific difficulty in the production of referring expressions, is another area of executive function, namely, working memory. The ability to create noun phrases (e.g. 'The green plate') is related to working memory in adults (Sikora, Roelofs, Hermans & Knoors, 2016), which may be compromised in autism (Kercood et al., 2014). Working memory relates to reference comprehension in autistic children (Schuh, Eigsti & Mirman, 2016), hence, a working memory deficit might also contribute to difficulties in producing appropriately informative referring expressions.

Additional cognitive abilities purported to be associated with the ability to produce reference, including attentional resources, perspective taking and integration of visual and verbal information to name but a few (see e.g. De Cat, 2015 for a summary) might also be influenced by EF capacity. For example, allowing the inhibition of one's own perspective to consider divergent social or perceptual viewpoints (Carlson, Claxton, & Moses, 2015) or enabling speakers to bridge the gap between assessing how two referents differ visually and producing an appropriate reference based on this information (Davies & Kreysa, 2018). Unfortunately, as we did not measure ToM and EF abilities in the same children, or indeed

EF ability beyond CF performance, how these factors interact to influence pragmatic language in autism remains a question for future research. Given that cognitive flexibility is considered by some to require a combination of working memory and inhibition (Diamond, 2013), and the clear role it played in the production of referring terms in Chapter 5, better understanding its contribution to pragmatic language difficulty in autism remains an important goal.

Finally, it should be noted that a comprehensive account and model of reference production is lacking even in typical development (Davies & Arnold, 2018). Better understanding referential communication processes more generally will naturally inform the exploration of autism specific difficulties in this area.

Cognitive underpinnings summary

Within this thesis, three key abilities were explored in connection with pragmatic language difficulties in autism, namely, social cognition, structural language, and cognitive flexibility. Social cognitive skills did not appear to relate consistently to referential communication or wider pragmatic ability in autistic children. In fact, we found evidence that autistic children can consider the experiences they shared with a partner during communicative exchanges. That said, further exploration of the sub-components of social cognition is required, for example, elements of shared intentionality such as joint commitment. Secondly, a relationship did not emerge between pragmatic and structural language across studies. However, the vast majority of autistic children in the current studies possessed structural language in the typical range. Difficulties with structural language might be expected to impact negatively on pragmatic language abilities, hence inclusion of autistic children with a wider range of structural language abilities in future research would be a

valuable. Finally, CF was found to play a pivotal role in the production of referring expressions. Whilst this effect was not specific to autism, research that unpicks the contribution of executive functioning to pragmatic language difficulties in autism appears a promising line of inquiry. In sum, whilst it is likely that multiple factors interact to influence referential communication and wider pragmatic ability in autistic children, the role of executive functioning in particular warrants further research attention.

6.4. Conclusions and future directions

Given the prevalence with which pragmatic language difficulties are reported in autism, and the importance of this skill set in social interaction, existing knowledge regarding whether difficulties apply equally to all areas of pragmatic ability was surprisingly limited. The cognitive underpinnings of these difficulties in autism were also unclear. A systematic review of one area of pragmatic ability in autistic individuals, namely referential communication, indicated an inconsistent pattern of ability; reference interpretation was similar to TD individuals, but difficulties existed in production. In our empirical studies, intact interpretation of referring terms was found to extend beyond visual perspective taking to one aspect of social common ground; the ability to consider the experiences shared with an interlocutor. Regarding the production of appropriately informative referring expressions, our empirical studies indicated that autistic difficulty may be more related to executive functioning than ToM or structural language ability, at least for children with structural language and non-verbal IQ in the typical range.

A key difficulty in conducting research in pragmatic language is the unclear delineation and categorisation of this skill set. Traditionally pragmatic language has been defined via a list of loosely related abilities (Levinson, 1983). The findings within this thesis demonstrate that pockets of ability exist within at least one of these abilities for autistic individuals, namely, referential communication. Thus, approaching our understanding of

pragmatic language in autism via this traditional 'list' approach, with the aim of labelling each area of pragmatic ability 'intact' or 'impaired' is likely to be unsuccessful. Instead, an approach such as that suggested by O'Neill (2012) or Andrés-Roqueta and Katsos (2017), where pragmatic language abilities are categorised based on the cognitive skills they recruit, may provide a useful framework within which to understand pragmatic language in autism. This may entail a shift from viewing pragmatic language as a unitary construct, towards identifying separable sub-categories of ability based primarily on the cognitive skills they recruit. Factor analytic work would be useful here, though this approach is currently rare in pragmatics (Matthews et al., 2019). The empirical studies reported in this thesis also demonstrate the important role that experimental manipulation plays in identifying the cognitive underpinnings of pragmatic language skills.

A challenging but vital task then remains to identify which cognitive abilities contribute to specific pragmatic language tasks. As discussed here, some cognitive abilities, such as structural language, might be expected to affect a range of pragmatic language abilities. Others, such as executive functioning, might enhance an individual's ability to succeed across pragmatic language tasks. Finally, some skills, such as mentalising, might only be called into play during certain types of pragmatic language task. As described in this discussion, one approach to categorising pragmatic language abilities is to separate those requiring basic linguistic abilities from those additionally requiring social-cognition (Andrés-Roqueta & Katsos, 2017). 'Social linguistic' activities might further be categorised by the type of social cognition recruited. Alongside ToM, decisions about how to provide or interpret information might be based, for example, on assumptions about knowledge shared with an interlocutor based on their community membership (Clark & Marshall, 1981), personal shared experienced (Chapter 4) or emotional affect (Abbot-Smith, Williams and

Matthews, 2020). This highlights the fact that successful pragmatic language involves consideration of multifaceted and rich contexts.

Organising pragmatic language abilities by the cognitive skills that they recruit might also enable researchers to more easily identify whether difficulties persist across all areas of pragmatic language for autistic individuals, as well as other clinical groups, and to develop increasingly targeted interventions. It should be noted that whilst different pragmatic language tasks may recruit different types of cognitive skill, these cognitive abilities are undoubtedly connected. Identifying interactions between these cognitive abilities presents an additional challenge to understanding pragmatic language development in both typical and atypical populations. A final vital question concerns how pragmatic language abilities, and the skills that underpin them, interact and develop over time. This remains an important aim in both typical and atypical pragmatic language development.

REFERENCES

Abbot-Smith, K., Williams, D., & Matthews, D. (2020). 'What's new for you?': Interlocutor-specific perspective-taking and language interpretation in autistic and neuro-typical children. *Research in Autism Spectrum Disorders, 70,* 101465. https://doi.org/10.1016/j.rasd.2019.101465

Adachi, T., Koeda, T., Hirabayashi, S., Maeoka, Y., Shiota, M., Wright, E. C., & Wada, A. (2004). The metaphor and sarcasm scenario test: A new instrument to help differentiate high functioning pervasive developmental disorder from attention deficit/hyperactivity disorder. *Brain and Development, 26*(5), 301-306. https://doi.org/10.1016/S0387-7604(03)00170-0

Adams, C., Green, J., Gilchrist, A., & Cox, A. (2002). Conversational behaviour of children with Asperger syndrome and conduct disorder. *Journal of Child Psychology and Psychiatry, 43*(5), 679-690. https://doi.org/10.1111/1469-7610.00056

Akbar, M., Loomis, R., & Paul, R. (2013). The interplay of language on executive functions in children with ASD. *Research in Autism Spectrum Disorders*, *7*(3), 494-501. https://doi.org/10.1016/j.rasd.2012.09.001

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology, 106*(1), 20-29. https://doi.org/10.1016/j.jecp.2009.11.003

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Anderson, P., (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, *8*(2), 71-82, doi: 10.1076/chin.8.2.71.8724

Andrés-Roqueta, C., & Katsos, N. (2017). The contribution of grammar, vocabulary and theory of mind in pragmatic language competence in children with autistic spectrum disorders. *Frontiers in Psychology, 8*, 996. https://doi.org/10.3389/fpsyg.2017.00996

Andrés-Roqueta, C., & Katsos, N. (2020). A Distinction Between Linguistic and Social Pragmatics Helps the Precise Characterization of Pragmatic Challenges in Children With Autism Spectrum Disorders and Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research, 63*(5), 1494-1508. doi: 10.1044/2020_JSLHR-19-00263

Antoniazzi, D., Snow, P., & Dickson-Swift, V. (2010). Teacher identification of children at risk for language impairment in the first year of school. *International Journal of Speech-Language Pathology*, 12(3), 244-252. doi:10.3109/17549500903104447

Apperly, I. (2018). Mindreading and psycholinguistic approaches to perspective taking: Establishing common ground. *Topics in Cognitive Science, 10*(1), 133-139. https://doi.org/10.1111/tops.12308

Ariel, M. (2010). *Defining pragmatics*. New York: Cambridge University Press.

Arnold, J. E. (2008). Reference production: Production-internal and addressee-oriented processes. *Language and Cognitive Processes*, 23(4), 495-527. doi:10.1080/01690960801920099

Arnold, J. E., Bennetto, L., & Diehl, J. J. (2009). Reference production in young speakers with and without autism: Effects of discourse status and processing constraints. *Cognition, 110*(2), 131-146. doi:10.1016/j.cognition.2008.10.016

Ash, A. C., Redmond, S. M., Timler, G. R., & Kean, J. (2017). The influence of scale structure and sex on parental reports of children's social (pragmatic) communication symptoms. *Clinical Linguistics & Phonetics, 31*(4), 293-312.

https://doi.org/10.1080/02699206.2016.1257655

Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition & Emotion, 9*(2-3), 151-165. https://doi.org/10.1080/02699939508409006

Bacso, S. A., & Nilsen, E. S. (2017). What's that you're saying? Children with better executive functioning produce and repair communication more effectively. *Journal of Cognition and Development, 18*(4), 441-464. https://doi.org/10.1080/15248372.2017.1336438

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology, 8*(4), 485-493. https://doi.org/10.1037/0894-4105.8.4.485

Baird, G., & Norbury, C. F. (2016). Social (pragmatic) communication disorders and autism spectrum disorder. *Archives of Disease in Childhood*, *101*(8), 745-751. http://dx.doi.org/10.1136/archdischild-2014-306944

Baixauli-Fortea, I., Miranda Casas, A., Berenguer-Forner, C., Colomer-Diago, C., & Roselló-Miranda, B. (2019). Pragmatic competence of children with autism spectrum disorder. Impact of theory of mind, verbal working memory, ADHD symptoms, and structural language. *Applied Neuropsychology: Child, 8*(2), 101-112. https://doi.org/10.1080/21622965.2017.1392861

Baltaxe, C. A. (1977). Pragmatic deficits in the language of autistic adolescents. *Journal of Pediatric Psychology, 2*(4), 176-180. doi:10.1093/jpepsy/2.4.176

Baltaxe, C. A., & D'Angiola, N. (1996). Referencing skills in children with autism and specific language impairment. *International Journal of Language & Communication*

Disorders, 31(3), 245-258. doi:10.3109/13682829609033156

Banney, R. M., Harper-Hill, K., & Arnott, W. L. (2015). The autism diagnostic observation schedule and narrative assessment: Evidence for specific narrative impairments in autism spectrum disorders. *International Journal of Speech-Language Pathology, 17*(2), 159-171. doi:10.3109/17549507.2014.977348

Bates, E. (1976). *Language in context*. New York: Academic Press.

Barceló, F., & Knight, R. T. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia, 40*(3), 349-356. https://doi.org/10.1016/S0028-3932(01)00110-5

Baron-Cohen, S. (1988). Social and pragmatic deficits in autism: Cognitive or affective?. *Journal of Autism and Developmental Disorders, 18*(3), 379-402. https://doi.org/10.1007/BF02212194

Baron-Cohen, S. (1989). The autistic child's theory of mind: a case of specific developmental delay. *Journal of Child Psychology and Psychiatry, 30,* 285-298. doi: 10.1111/j.1469-7610.1989.tb00241.x

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. *Journal of Child psychology and Psychiatry, 38*(7), 813-822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x

Baron-Cohen, S. (1999) The evolution of a theory of mind. In: M. C. Corballis & S. E. G. Lea (Eds). *The descent of mind: Psychological perspectives on hominid evolution*, Oxford University Press.

Baron-Cohen, S. (2001). Theory of mind in normal development and autism. *Prisme, 34*(1), 74-183.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a “theory of

mind"? *Cognition, 21*(1), 37-46. doi:10.1016/0010-0277(85)90022-8

Bauminger-Zviely, N., Karin, E., Kimhi, Y., & Agam-Ben-Artzi, G. (2014). Spontaneous peer conversation in preschoolers with high-functioning autism spectrum disorder versus typical development. *Journal of Child Psychology and Psychiatry, 55*(4), 363-373. https://doi.org/10.1111/jcpp.12158

Beaudoin, C., Leblanc, É., Gagner, C., & Beauchamp, M. H. (2020). Systematic review and inventory of theory of mind measures for young children. *Frontiers in Psychology*, *10*, 2905. https://doi.org/10.3389/fpsyg.2019.02905

Begeer, S., Malle, B. F., Nieuwland, M. S., & Keysar, B. (2010). Using theory of mind to represent and take part in social interactions: Comparing individuals with high-functioning autism and typically developing controls. *European Journal of Developmental Psychology, 7*(1), 104-122. doi:10.1080/17405620903024263

Berger, H. J., Aerts, F. H., Spaendonck, K. P. V., Cools, A. R., & Teunisse, J. P. (2003). Central coherence and cognitive shifting in relation to social improvement in high-functioning young adults with autism. *Journal of Clinical and Experimental Neuropsychology, 25*(4), 502-511. https://doi.org/10.1076/jcen.25.4.502.13870

Berman, R. A., & Slobin, D. I. (1994). *Relating Events in Narrative: A Crosslinguistic Developmental Study*. New Jersey: Lawrence Erlbaum Associates.

Bishop, D. V. (1993). *The children's communication checklist: CCC*. London: Harcourt Assessment.

Bishop, D. V. (2003). *CCC-2. Children's communication checklist* (2nd ed.). London, UK: Pearson.

Bishop, D. V., & Adams, C. (1991). What do referential communication tasks measure? A study of children with specific language impairment. *Applied Psycholinguistics; Applied Psycholinguistics, 12*(2), 199-215. doi:10.1017/S0142716400009140

Bishop, D. V., & McDonald, D. (2009). Identifying language impairment in children: combining language test scores with parental report. *International Journal of Language & Communication Disorders, 44*(5), 600-615.

Bishop, D. V., & Norbury, C. F. (2005). Executive functions in children with communication impairments, in relation to autistic symptomatology: I: Generativity. *Autism, 9*(1), 7-27. https://doi.org/10.1177/1362361305049027

Bishop, D. V., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2016). CATALISE: A multinational and multidisciplinary delphi consensus study of problems with language development. phase 2. terminology. *PeerJ Preprints, 4*, e2484v1.

Blain-Brière, B., Bouchard, C., & Bigras, N. (2014). The role of executive functions in the pragmatic skills of children age 4–5. *Frontiers in Psychology, 5*, 240. https://doi.org/10.3389/fpsyg.2014.00240

Blakey, E., & Carroll, D. J. (2018). Not all distractions are the same: Investigating why preschoolers make distraction errors when switching. *Child Development, 89*(2), 609-619. https://doi.org/10.1111/cdev.12721

Bohn, M., & Köymen, B. (2018). Common ground and development. *Child Development Perspectives, 12*(2), 104-108. https://doi.org/10.1111/cdep.12269

Boiteau, T. W., Malone, P. S., Peters, S. A., & Almor, A. (2014). Interference between conversation and a concurrent visuomotor task. *Journal of Experimental Psychology: General, 143*(1), 295 –311. https://doi.org/10.1037/a0031858

Bosco, F. M., & Gabbatore, I. (2017). Theory of mind in recognizing and recovering communicative failures. *Applied Psycholinguistics, 38*(1), 57-88. https://doi.org/10.1017/S0142716416000047

Bosco, F. M., Tirassa, M., & Gabbatore, I. (2018). Why pragmatics and Theory of Mind do not (completely) overlap. *Frontiers in Psychology, 9*, 1453.

https://doi.org/10.3389/fpsyg.2018.01453

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy, 18*(1), 1-21. doi:10.1191/0265659002ct224oa

Botting, N., Jones, A., Marshall, C., Denmark, T., Atkinson, J., & Morgan, G. (2017). Nonverbal executive function is mediated by language: A study of deaf and hearing children. *Child Development, 88*(5), 1689-1700. https://doi.org/10.1111/cdev.12659

Boucher, J. (2012). Research review: structural language in autistic spectrum disorder–characteristics and causes. *Journal of Child Psychology and Psychiatry*, *53*(3), 219-233. https://doi.org/10.1111/j.1469-7610.2011.02508.x

Branigan, H. P., Tosi, A., & Gillespie-Smith, K. (2016). Spontaneous lexical alignment in children with an autistic spectrum disorder and their typically developing peers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(11), 1821. https://doi.org/10.1037/xlm0000272

Bratman, M. E. (1992). Shared cooperative activity. *The Philosophical Review, 101*(2), 327-341. DOI: 10.2307/2185537

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 22*(6), 1482. https://doi.org/10.1037/0278-7393.22.6.1482

Brink, K. A., Lane, J. D., & Wellman, H. M. (2015). Developmental pathways for social understanding: Linking social cognition to social contexts. *Frontiers in Psychology, 6*, 719. https://doi.org/10.3389/fpsyg.2015.00719

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review, 16*(5), 893-900. doi:10.3758/PBR.16.5.893

Brown-Schmidt, S., & Duff, M. C. (2016). Memory and common ground processes in language use. *Topics in Cognitive Science, 8*(4), 722-736. https://doi.org/10.1111/tops.12224

Cane, J. E., Ferguson, H. J., & Apperly, I. (2016). Using perspective to resolve reference: the impact of cognitive load and motivation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. *43*(4), 591. doi: 10.1037/xlm0000345

Capps, L., Kehres, J., & Sigman, M. (1998). Conversational abilities among children with autism and children with developmental delays. *Autism, 2*(4), 325-344. https://doi.org/10.1177/1362361398024002

Capps, L., Losh, M., & Thurber, C. (2000). "The frog ate the bug and made his mouth sad": Narrative competence in children with autism. *Journal of Abnormal Child Psychology, 28*(2), 193-204. https://doi.org/10.1023/A:1005126915631

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development: An International Journal of Research and Practice, 11*(2), 73-92. https://doi.org/10.1002/icd.298

Carpenter, M., Pennington, B. F., & Rogers, S. J. (2002). Interrelations among social-cognitive skills in young children with autism. *Journal of Autism and Developmental Disorders, 32*(2), 91-106. doi. 10.1023/A:1014836521114

Carroll, D. J., Blakey, E., & FitzGibbon, L. (2016). Cognitive flexibility in young children: Beyond perseveration. *Child Development Perspectives, 10*(4), 211-215. https://doi.org/10.1111/cdep.12192

Carrow-Woolfolk, E. (1999). *CASL: Comprehensive Assessment of Spoken Language*. Circle Pines, MN: American Guidance Services.

Charman, T., Jones, C. R., Pickles, A., Simonoff, E., Baird, G., & Happé, F. (2011). Defining the cognitive phenotype of autism. *Brain Research, 1380*, 10-21. doi:10.1016/j.brainres.2010.10.075

Chevalier, N., & Blaye, A. (2008). Cognitive flexibility in preschoolers: The role of representation activation and maintenance. *Developmental Science, 11*(3), 339-353. https://doi.org/10.1111/j.1467-7687.2008.00679.x

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., & Schultz, R. T. (2012). The social motivation theory of autism. *Trends in Cognitive Sciences, 16*(4), 231-239. https://doi.org/10.1016/j.tics.2012.02.007

Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar inferences in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 40*(9), 1104-1117. doi: 10.1007/s10803-010-0960-8

Clark, E. V. (2014). Pragmatics in acquisition. *Journal of Child Language, 41* (S1), 105-116. doi:10.1017/S0305000914000117

Clark, H.H. (1996). *Using language*. Cambridge, England: Cambridge University Press.

Clark, H.H., Bangerter, A., 2004. Changing ideas about reference. In: I.A., Noveck & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 25–49). London: Palgrave Macmillan.

Clark, H.H., & Brennan S.E. (1991) Grounding in Communication. In L. Resnick, J. Levine & S.Teasley (Eds.), *Perspectives on socially shared cognition* (127-149). Hyattsville, MD: American Psychological Association

Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webher, & I. A. Sag (Eds.), *Elements of discourse understanding*. Cambridge: Cambridge Univ. Press.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J.-F. Leny & W. Kintsch (Eds.), *Language and comprehension* (pp. 287-299). Amsterdam:

North-Holland.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior*, *22*(2), 245-258. doi. 10.1016/S0022-5371(83)90189-5

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155. https://doi.org/10.1037/0033-2909.112.1.155

Colombi, C., Liebal, K., Tomasello, M., Young, G., Warneken, F., & Rogers, S. J. (2009). Examining correlates of cooperation in autism: Imitation, joint attention, and understanding intentions. *Autism, 13*(2), 143-163. https://doi.org/10.1177/1362361308098514

Colle, L., Baron-Cohen, S., Wheelwright, S., & van Der Lely, H.,K.J. (2008). Narrative discourse in adults with high-functioning autism or asperger syndrome. *Journal of Autism and Developmental Disorders, 38*(1), 28. doi:10.1007/s10803-007-0357-5

Colozzo, P., Morris, H., & Mirenda, P. (2015). Narrative production in children with autism spectrum disorder and specific language impairment. *Canadian Journal of Speech-Language Pathology & Audiology*, *39*(4)

Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology. 12*, 349–358. https://doi.org/10.1044/1058-0360(2003/080)

Constantino, J. N., & Gruber, C. P. (2007). *Social Responsiveness Scale (SRS)* Western Psychological Services Los Angeles, CA.

Coull, G. J., Leekam, S. R., & Bennett, M. (2006). Simplifying Second-order Belief Attribution: What Facilitates Children's Performance on Measures of Conceptual Understanding? *Social Development*, *15*(2), 260-275. doi: 10.1046/j.1467-

9507.2006.00340.x

Dahlgren, S., & Sandberg, A. D. (2008). Referential communication in children with autism spectrum disorder. *Autism: The International Journal of Research and Practice, 12*(4), 335-348. doi:10.1177/1362361308091648

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences, 38*(9), 571-578. https://doi.org/10.1016/j.tins.2015.07.003

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental Psychology, 40*(2), 271. https://doi.org/10.1037/0012-1649.40.2.271

Davies, C., & Arnold, J.E. (2018). Reference and informativeness: How context shapes referential choice. In: C. Cummins., & N. Katsos (Eds.), *Handbook of Experimental Semantics and Pragmatics*, Oxford: OUP

Davies, C., and Katsos, N. (2016). Reference and informativeness as cognitive processes in verbal communication. In: A. Rocci and L. de Saussure (Eds.) *Verbal Communication. Handbooks of Communication Science* (pp. 123–140). Berlin: Gruyter Mouton).

Davies, C., & Kreysa, H. (2018). Look before you speak: children's integration of visual information into informative referring expressions. *Journal of Child Language, 45*(5), 1116-1143. https://doi.org/10.1017/S0305000918000120

De Cat, C. (2015). The cognitive underpinnings of referential abilities. In: S.E. Allen, L. Serratrice (Eds.), *The Acquisition of Reference* (pp. 263-283). Amsterdam, Netherland: John Benjamins.

de Marchena, A., & Eigsti., I. (2016). The art of common ground: Emergence of a complex

pragmatic language skill in adolescents with autism spectrum disorders. *Journal of Child Language, 43*(1), 43-80. doi:10.1017/S0305000915000070

De Rosnay, M., Fink, E., Begeer, S., Slaughter, V., & Peterson, C. (2014). Talking theory of mind talk: young school-aged children's everyday conversation and understanding of mind and emotion. *Journal of Child Language, 41*(5), 1179-1193. https://doi.org/10.1017/S0305000913000433

de Villiers, J. G., & de Villiers, P. A. (2014). The role of language in theory of mind development. *Topics in Language Disorders, 34*(4), 313-328. doi: 10.1097/TLD.0000000000000037

Deliens, G., Papastamou, F., Ruytenbeek, N., Geelhand, P., & Kissine, M. (2018). Selective pragmatic impairment in autism spectrum disorder: Indirect requests versus irony. *Journal of Autism and Developmental Disorders, 48*(9), 2938-2952. https://doi.org/10.1007/s10803-018-3561-6

Demetriou, E. A., Lampit, A., Quintana, D. S., Naismith, S. L., Song, Y. J. C., Pye, J. E., Hickie, I., & Guastella, A. J. (2018). Autism spectrum disorders: a meta-analysis of executive function. *Molecular Psychiatry, 23*(5), 1198-1204. https://doi.org/10.1038/mp.2017.75

Dennis, M., Lazenby, A. L., & Lockyer, L. (2001). Inferential language in high-function children with autism. *Journal of Autism and Developmental Disorders*, 31(1), 47-54. https://doi.org/10.1023/A:1005661613288

Devine, R.T. & Hughes, C. (2017). Family correlates of false belief understanding in early childhood: A meta-analysis. *Child Development, 89*(3), 971-987. doi: 10.1111/cdev.12682.

DeVries, R. (1970). The development of role-taking as reflected by behavior of bright,

average, and retarded children in a social guessing game. *Child Development, 41*(3), 759-770. doi: 10.2307/1127222

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64,* 135-168. https://doi.org/10.1146/annurev-psych-113011-143750

Dichter, G. S., Radonovich, K. J., Turner-Brown, L. M., Lam, K. S., Holtzclaw, T. N., & Bodfish, J. W. (2010). Performance of children with autism spectrum disorders on the dimension-change card sort task. *Journal of Autism and Developmental Disorders, 40*(4), 448-456. DOI 10.1007/s10803-009-0886-1

Doenyas, C., Yavuz, H. M., & Selcuk, B. (2018). Not just a sum of its parts: How tasks of the theory of mind scale relate to executive function across time. *Journal of Experimental Child Psychology, 166,* 485-501. https://doi.org/10.1016/j.jecp.2017.09.014

Eigsti, I. M., de Marchena, A. B., Schuh, J. M., & Kelley, E. (2011). Language acquisition in autism spectrum disorders: A developmental review. *Research in Autism Spectrum Disorders, 5*(2), 681-691. https://doi.org/10.1016/j.rasd.2010.09.001

Elliot, C. D., & Smith, P. (2011*). British Ability Scales–Third Edition (BAS-3)*. London: GL Assessment.

Faja, S., & Dawson, G. (2014). Performance on the dimensional change card sort and backward digit span by young children with autism without intellectual disability. *Child Neuropsychology, 20*(6), 692-699. https://doi.org/10.1080/09297049.2013.856395

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development, 79*(1), 126-138. https://doi.org/10.1111/j.1467-8624.2007.01115.x

Fine, J., Bartolucci, G., Szatmari, P., & Ginsberg, G. (1994). Cohesive discourse in pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*(3),

315-329. doi:10.1007/BF02172230

Flavell, J. H., Green, F. L., Flavell, E. R., Watson, M. W., & Campione, J. C. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development, 51*(1-68).

Frith, C. D. (2008). Social cognition. Philosophical Transactions of the Royal Society of London. Series B, *Biological Sciences, 363*(1499), 2033–2039. https://doi.org/10.1098/rstb.2008.0005

Frith, U. (1989). A new look at language and communication in autism. *International Journal of Language & Communication Disorders, 24*(2), 123-150. https://doi.org/10.3109/13682828909011952

Frith, U., & Happé, F. (2005). Autism spectrum disorder. *Current biology, 15*(19), R786-R790.

Frith, U., Morton, J., & Leslie, A. M. (1991). The cognitive basis of a biological disorder: Autism. *Trends in Neuroscience, 14*, 433-438. doi: 10.1016/0166-2236(91)90041-R

Fukumura, K. (2015). Development of audience design in children with and without ASD. *Developmental Psychology,* doi:10.1037/dev0000064

Funabiki, Y., & Shiwa, T. (2018). Weakness of visual working memory in autism. *Autism Research, 11*(9), 1245-1252. https://doi.org/10.1002/aur.1981

Gathercole, S. E., Lamont, E., & Alloway, T. (2006). Working memory in the classroom. In S. J. Pickering (Ed.). *Working Memory and Education*. Amsterdam: Elsevier Press.

Geller, E. (1988). The interplay between linguistic and social-cognitive knowledge in perspective-taking by autistic children. *Communication Disorders Quarterly, 14*(1), 23-44. doi:10.1177/152574018801400103

Geurts, B. (2019). What's wrong with Gricean pragmatics. In: ExLing 2019: *Proceedings of the 10th International Conference of Experimental Linguistics*, Lisbon.

Geurts, B., Kissine, and B. van Tiel (2019). Pragmatic reasoning in autism. In: K. Morsanyi and R. Byrne (eds.), *Thinking, Reasoning and Decision Making in Autism*, (pp. 113-134). Routledge.

Geurts, B., & Rubio-Fernández, P. (2015). Pragmatics and processing. *Ratio, 28*(4), 446-469. https://doi.org/10.1111/rati.12113

Geurts, H. M., Corbett, B., & Solomon, M. (2009). The paradox of cognitive flexibility in autism. *Trends in Cognitive Sciences, 13*(2), 74-82. https://doi.org/10.1016/j.tics.2008.11.006

Geurts, H. M., & Embrechts, M. (2008). Language profiles in ASD, SLI, and ADHD. *Journal of Autism and Developmental Disorders, 38*(10), 1931-1943. DOI 10.1007/s10803-008-0587-1

Geurts, H. M, Sinzig, J., Booth, R., & Happé, F. (2014). Neuropsychological heterogeneity in executive functioning in autism spectrum disorders. *International Journal of Developmental Disabilities, 60*(3), 155-162. https://doi.org/10.1179/2047387714Y.0000000047

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? *Journal of Child Psychology and Psychiatry, 45*(4), 836-854.

Gibson, J., Adams, C., Lockton, E., & Green, J. (2013). Social communication disorder outside autism? A diagnostic classification approach to delineating pragmatic language impairment, high functioning autism and specific language impairment. *Journal of Child Psychology and Psychiatry, 54*(11), 1186-1197. https://doi.org/10.1111/jcpp.12079

Gilbert, M. (2009). Shared intention and personal intentions. *Philosophical Studies, 144*(1), 167-187. https://doi.org/10.1007/s11098-009-9372-z

Gillis, R., & Nilsen, E. S. (2014). Cognitive flexibility supports preschoolers' detection of communicative ambiguity. *First Language*, *34*(1), 58-71. doi:10.1177/0142723714521839

Gioia, G.A., Isquith, P.K., Guy, S.C., & Kenworthy, L. (2000). *Behavior Rating of Executive Function*. Lutz, FL: Psychological Assessment Resources.

Glenwright, M., & Pexman, P. M. (2010). Development of children's ability to distinguish sarcasm and verbal irony. *Journal of Child Language, 37*(2), 429-451. https://doi.org/10.1017/S0305000909009520

Glucksberg, S., & Krauss, R. M. (1967). What do people say after they have learned how to talk? studies of the development of referential communication. *Merrill-Palmer Quarterly of Behavior and Development, 13*(4), 309-316.

Gordon, A. C., & Olson, D. R. (1998). The relation between acquisition of a theory of mind and the capacity to hold in mind. *Journal of Experimental Child Psychology, 68*(1), 70-83. https://doi.org/10.1006/jecp.1997.2423

Gottmann, J., Gonso, J. & Rassmussen, B. (1975). Social interaction, social competence and friendship in children. *Child Development, 46*: 709-718. doi: 10.2307/1128569

Graf, E., & Davies, C. (2014). The production and comprehension of referring expressions. In Matthews, D. (Ed.) *Pragmatic Development in First Language Acquisition*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Granader, Y., Wallace, G. L., Hardy, K. K., Yerys, B. E., Lawson, R. A., Rosenthal, M., ... & Schultz, R. T. (2014). Characterizing the factor structure of parent reported executive function in autism spectrum disorders: The impact of cognitive inflexibility. *Journal of Autism and Developmental Disorders, 44*(12), 3056-3062. doi: 10.1007/s10803-014-2169-8

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. J. Morgan. (Eds.) *Syntax and*

*Semantics, Volume 3: Speech Acts (*pp. 225-242). New York: Academic Press.

Grigoroglou, M., & Papafragou, A. (2016). Are children flexible speakers? Effects of typicality and listener needs in children's event descriptions. In *Proceedings from the 38th Annual Meeting of the Cognitive Science Society* (pp. 782–787). Austin, TX: Cognitive Science Society.

Grigoroglou, M., & Papafragou, A. (2019). Interactive contexts increase informativeness in children's referential communication. *Developmental Psychology, 55*(5), 951 –966. https://doi.org/10.1037/dev0000693

Grueneisen, S., Wyman, E., & Tomasello, M. (2015). "I know you don't know I know…" Children use second-order false-belief reasoning for peer coordination. *Child Development, 86*(1), 287-293. https://doi.org/10.1111/cdev.12264

Gundel, J. K., Hedberg, N. & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language, 69*(2): 274-307. http://www.jstor.org/stable/416535

Hale, C. M., & Tager-Flusberg, H. (2005). Social communication in children with autism: The relationship between theory of mind and discourse development. *Autism, 9*(2), 157-178. https://doi.org/10.1177/1362361305051395

Hamilton, A. F., (2009). Research review: Goals, intentions and mental states: Challenges for theories of autism. *Journal of Child Psychology and Psychiatry, 50*(8), 881-892. https://doi.org/10.1111/j.1469-7610.2009.02098.x

Hamilton, A. F., Brindley, R., & Frith, U. (2009). Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition, 113*(1), 37-44. https://doi.org/10.1016/j.cognition.2009.07.007

Happé, F. G. (1993). Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition, 48*(2), 101-119. https://doi.org/10.1016/0010-

0277(93)90026-R

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129-154. doi:10.1007/BF02172093

Happé, F. G. (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. https://doi.org/10.1111/j.1467-8624.1995.tb00909.x

Happé, F., Cook, J. L., & Bird, G. (2017). The structure of social cognition: In (ter) dependence of sociocognitive processes. *Annual Review of Psychology, 68,* 243-267. https://doi.org/10.1146/annurev-psych-010416-044046

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. doi:10.1007/s10803-005-0039-0

Happé, F., & Frith, U. (2014). Annual research review: Towards a developmental neuroscience of atypical social cognition. *Journal of Child Psychology and Psychiatry, 55*(6), 553-577. https://doi.org/10.1111/jcpp.12162

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166-1186. https://doi.org/10.3758/s13428-017-0935-1

Hatton, C. (1998). Pragmatic language skills in people with intellectual disabilities: A review. *Journal of Intellectual and Developmental Disability*, 23(1), 79-100. https://doi.org/10.1080/13668259800033601

Hayward, E. O., & Homer, B. D. (2017). Reliability and validity of advanced theory-of-mind measures in middle childhood and adolescence. *British Journal of Developmental*

Psychology, 35(3), 454-462. https://doi.org/10.1111/bjdp.12186

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology*, 57(2), 243-259. doi: 10.2307/1416950

Helland, W. A., Lundervold, A. J., Heimann, M., & Posserud, M. B. (2014). Stable associations between behavioral problems and language impairments across childhood–The importance of pragmatic language problems. *Research in Developmental Disabilities*, *35*(5), 943-951. doi:10.1016/j.ridd.2014.02.016

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences, 8*(1), 26-32. doi:10.1016/j.tics.2003.11.003

Hobson, R. P. (1984). Early childhood autism and the question of egocentrism. *Journal of Autism and Developmental Disorders, 14*(1), 85-104. https://doi.org/10.1007/BF02408558

Hollebrandse, B., Van Hout, A. & Hendriks, P. (2014). Children's first and second order false belief reasoning in a verbal and a low-verbal task. *Synthese*, *191*(3): 321-33.

Horn, L. (1988): Pragmatic theory. In F, Newmeyer (ed.). Linguistics: *The Cambridge Survey*, *Vol. I* (pp. 113-145). Cambridge, UK: Cambridge University Press.

Horn, L. R. (2006). Implicature. In R. Horn & G. Ward (Eds), *The Handbook of Pragmatics*. (pp. 3 – 28). Oxford, UK: Blackwell.

Huang, Y. (2014). *Pragmatics*. Oxford, UK: Oxford University Press.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*(6), 1447. https://doi.org/10.1037/0012-1649.43.6.1447

Hughes, C., & Leekam, S. (2004). What are the links between theory of mind and social relations? Review, reflections and new directions for studies of typical and atypical development. *Social Development*, *13*(4), 590-619. https://doi.org/10.1111/j.1467-

9507.2004.00285.x

Hutchins, T. L., Prelock, P. A., & Bonazinga, L. (2012). Psychometric evaluation of the Theory of Mind Inventory (ToMI): A study of typically developing children and children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 42*(3), 327-341. https://doi.org/10.1007/s10803-011-1244-7

Hutchison, S. M., Müller, U., & Iarocci, G. (2019). Parent reports of executive function associated with functional communication and conversational skills among school age children with and without autism spectrum disorder. *Journal of Autism and Developmental Disorders, 50* (2019-2029). https://doi.org/10.1007/s10803-019-03958-6

IBM. (2016). *SPSS Version 25.0.* Chicago, IL: IBM Corp.

Jaswal, V. K., & Akhtar, N. (2018). Being versus appearing socially uninterested: Challenging assumptions about social motivation in autism. *Behavioral and Brain Sciences, 42* (1-84). https://doi.org/10.1017/S0140525X18001826

Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press, Clarendon Press.

Jones, C. D., & Schwartz, I. S. (2009). When asking questions is not enough: An observational study of social communication differences in high functioning children with autism. *Journal of Autism and Developmental Disorders,* 39(3), 432-443. https://doi.org/10.1007/s10803-008-0642-y

Jones, C. R., Simonoff, E., Baird, G., Pickles, A., Marsden, A. J., Tregay, J., Happé, F., & Charman, T. (2018). The association between theory of mind, executive function, and the symptoms of autism spectrum disorder. *Autism Research, 11*(1), 95-109. https://doi.org/10.1002/aur.1873

Jones, C. R., Swettenham, J., Charman, T., Marsden, A. J., Tregay, J., Baird, G., ... & Happé,

F. (2011). No evidence for a fundamental visual motion processing deficit in adolescents with autism spectrum disorders. *Autism Research, 4*(5), 347-357. https://doi.org/10.1002/aur.209

Joseph, R. M. (2011). The significance of IQ and differential cognitive abilities for understanding ASD. *In D.* Fein *(Ed.), The Neuropsychology of Autism (pp.* 281-294). Oxford, United Kingdom*:* Oxford University Press.

Joseph, R. M., & Tager-Flusberg, H. (2004). The relationship of theory of mind and executive functions to symptom type and severity in children with autism. *Development and Psychopathology*, *16*(1), 137-155. https://doi.org/10.1017/S095457940404444X

Kalandadze, T., Norbury, C., Nærland, T., & Næss, K. A. B. (2018). Figurative language comprehension in individuals with autism spectrum disorder: A meta-analytic review. *Autism*, *22*(2), 99-117. https://doi.org/10.1177/1362361316668652

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*, 217–250

Katsos, N., Roqueta, C. A., Estevan, R. A. C., & Cummins, C. (2011). Are children with Specific Language Impairment competent with the pragmatics and logic of quantification?. *Cognition*, *119*(1), 43-57. https://doi.org/10.1016/j.cognition.2010.12.004

Keehn, B., Müller, R. A., & Townsend, J. (2013). Atypical attentional networks and the emergence of autism. *Neuroscience & Biobehavioral Reviews*, *37*(2), 164-183. https://doi.org/10.1016/j.neubiorev.2012.11.014

Kemner, C., Verbaten, M. N., Cuperus, J. M., Camfferman, G., & Van Engeland, H. (1998). Abnormal saccadic eye movements in autistic children. *Journal of Autism and Developmental Disorders*, *28*(1), 61-67. https://doi.org/10.1023/A:1026015120128

Kenny, L., Cribb, S. J., & Pellicano, E. (2019). Childhood executive function predicts later autistic features and adaptive behavior in young autistic people: A 12-year

prospective study. *Journal of Abnormal Child Psychology, 47*(6), 1089-1099. https://doi.org/10.1007/s10802-018-0493-8

Kenworthy, L., Black, D. O., Harrison, B., Della Rosa, A., & Wallace, G. L. (2009). Are executive control functions related to autism symptoms in high-functioning children?. *Child Neuropsychology, 15*(5), 425-440. https://doi.org/10.1080/09297040802646983

Kerbel, D., & Grunwell, P. (1998). A study of idiom comprehension in children with semantic-pragmatic difficulties. Part II. Between groups results and discussion. *International Journal of Language and Communication Disorders, 33* (1), 23–44. doi: 10.1080/136828298247910

Kercood, S., Grskovic, J. A., Banda, D., & Begeske, J. (2014). Working memory and autism: A review of literature. *Research in Autism Spectrum Disorders, 8*(10), 1316-1332. https://doi.org/10.1016/j.rasd.2014.06.011

Kern, A. & Moll, H. (2017). On the transformative character of collective intentionality and the uniqueness of the human. *Philosophical Psychology, 30*(3): 315-333. doi: 10.1080/09515089.2017.1295648.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science, 11*(1), 32-38. doi:10.1111/1467-9280.00211

Keysar, B., Barr, D. J., & Horton, W. S. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science, 7*(2), 46-49. https://doi.org/10.1111/1467-8721.ep13175613

King, D., & Palikara, O. (2018). Assessing language skills in adolescents with autism spectrum disorder. *Child Language Teaching and Therapy, 34*(2), 101-113. https://doi.org/10.1177/0265659018780968

Kissine, M. (2012). Pragmatics, cognitive flexibility and autism spectrum disorders. *Mind &*

Language, 27(1), 1-28. https://doi.org/10.1111/j.1468-0017.2011.01433.x

Kissine, M. (2016). Pragmatics as metacognitive control. *Frontiers in Psychology, 6*, 2057. https://doi.org/10.3389/fpsyg.2015.02057

Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger syndrome: The social attribution task. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 41*(7), 831-846. https://doi.org/10.1111/1469-7610.00671

Kloo, D., & Perner, J. (2005). Disentangling dimensions in the dimensional change card-sorting task. *Developmental Science, 8*(1), 44-56. https://doi.org/10.1111/j.1467-7687.2005.00392.x

Korkman, M., Kirk, U., & Kemp, S. L. (1998). *NEPSY - A Developmental Neuropsychological Assessment*. San Antonio, TX: Psychological Corporation

Krauss, R. M., & Glucksberg, S. (1977). Social and nonsocial speech. *Scientific American, 236*(2), 100-105.

Kuijper, S. J., Hartman, C. A., Bogaerds-Hazenberg, S., & Hendriks, P. (2017). Narrative production in children with autism spectrum disorder (ASD) and children with attention-deficit/hyperactivity disorder (ADHD): Similarities and differences. *Journal of Abnormal Psychology, 126*(1), 63. https://doi.org/10.1037/abn0000231

Kuijper, S. J., Hartman, C. A., & Hendriks, P. (2015). Who is he? Children with ASD and ADHD take the listener into account in their production of ambiguous pronouns. *PloSOne, 10* (7). 10:e0132408. doi. 10.1371/journal.pone.0132408

Lai, C. L. E., Lau, Z., Lui, S. S., Lok, E., Tam, V., Chan, Q., Cheng., K.M., Lam, S.M., & Cheung, E. F.C. (2016). Meta-analysis of neuropsychological measures of executive functioning in children and adolescents with high-functioning autism spectrum disorder. *Autism Research*. *10*(5), 911-939. doi:10.1002/aur.1723

Landa, R. J. (2005). Assessment of social communication skills in preschoolers. *Mental Retardation and Developmental Disabilities Research Reviews*, *11*(3), 247-252. https://doi.org/10.1002/mrdd.20079

Landa, R. (2000). Social language use in Asperger's syndrome and high-functioning autism. In A. Klin, F. R. Volkmar, & S. S. Sparrow (Eds.), *Asperger's Syndrome* (pp. 125-155). New York: Guilford Press.

Le Sourn-Bissaoui, S., Caillies, S., Gierski, F., & Motte, J. (2009). Inference processing in adolescents with Asperger syndrome: Relationship with theory of mind abilities. *Research in Autism Spectrum Disorders, 3*(3), 797-808. https://doi.org/10.1016/j.rasd.2009.03.003

Lecce, S., Bianco, F., Devine, R. T., & Hughes, C. (2017). Relations between theory of mind and executive function in middle childhood: A short-term longitudinal study. *Journal of Experimental Child Psychology, 163*, 69-86. https://doi.org/10.1016/j.jecp.2017.06.011

Lee, B. P. (2001). Mutual knowledge, background knowledge and shared beliefs: Their roles in establishing common ground. *Journal of Pragmatics, 33*(1), 21-44. https://doi.org/10.1016/S0378-2166(99)00128-9

Leech, G. N. 1983. *Principles of Pragmatics*. London & New York: Longman.

Leitão, S., & Allan, L. (2003). *Peter and the Cat: Narrative Assessment* Black Sheep Press.

Levinson, S. C. (1983). *Pragmatics.* Cambridge University Press. Cambridge UK.

Liebal, K., Behne, T., Carpenter, M., & Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Developmental Science*, 12(2), 264-271. doi: 10.1111/j.1467-7687.2008.00758.x

Liebal, K., Colombi, C., Rogers, S. J., Warneken, F., & Tomasello, M. (2008). Helping and cooperation in children with autism. *Journal of Autism and Developmental Disorders,*

38(2), 224-238. doi. 10.1007/s10803-007-0381

Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551-556. doi:10.1016/j.jesp.2009.12.019

Lind, S. E., Williams, D. M., Bowler, D. M., & Peel, A. (2014). Episodic memory and episodic future thinking impairments in high-functioning autism spectrum disorder: An underlying difficulty with scene construction or self-projection? *Neuropsychology*, *28*(1), 55-67. doi:10.1037/neu0000005

Livingston, L. A., Carr, B., & Shah, P. (2019). Recent advances and new directions in measuring theory of mind in autistic adults. *Journal of Autism and Developmental Disorders*, *49*(4), 1738-1744. https://doi.org/10.1007/s10803-018-3823-3

Loucas, T., Charman, T., Pickles, A., Simonoff, E., Chandler, S., Meldrum, D., & Baird, G. (2008). Autistic symptomatology and language ability in autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, *49*(11), 1184-1192. https://doi.org/10.1111/j.1469-7610.2008.01951.x

Lord, C., & Bishop, S. L. (2015). Recent advances in autism research as reflected in DSM-5 criteria for autism spectrum disorder. *Annual Review of Clinical Psychology, 11*, 53-70. doi: 10.1146/annurev-clinpsy-032814-112745

Lord, C. & Paul, R. (1997). Language and communication in autism. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of Autism and Pervasive Developmental Disorders* (2nd Edition). New York: John Wiley & Sons.

Lord, C., DiLavore, P. C., & Gotham, K. (2012). *Autism diagnostic observation schedule* Western Psychological Services Torrance, CA.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders*,*24*, 659-686.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (1999). ADOS. *Autism Diagnostic Observation Schedule Manual*. Los Angeles: WPS.

Losh, M., & Capps, L. (2003). Narrative ability in high-functioning children with autism or asperger's syndrome. *Journal of Autism and Developmental Disorders*, *33*(3), 239-251. doi:10.1023/A:1024446215446

Losh, M., Martin, G. E., Klusek, J., Hogan-Brown, A. L., & Sideris, J. (2012). Social communication and theory of mind in boys with autism and fragile X syndrome. *Frontiers in Psychology, 3*, 266. https://doi.org/10.3389/fpsyg.2012.00266

Loukusa, S., Leinonen, E., Kuusikko, S., Jussila, K., Mattila, M. L., Ryder, N., Ebeling, H., & Moilanen, I. (2007). Use of context in pragmatic language comprehension by children with Asperger syndrome or high-functioning autism. *Journal of Autism and Developmental Disorders, 37*(6), 1049-1059. https://doi.org/10.1007/s10803-006-0247-2

Loukusa, S., & Moilanen, I. (2009). Pragmatic inference abilities in individuals with Asperger syndrome or high-functioning autism. A review. *Research in Autism Spectrum Disorders*, *3*(4), 890-904. https://doi.org/10.1016/j.rasd.2009.05.002

Loveland, K. A., McEvoy, R. E., Tunali, B., & Kelley, M. L. (1990). Narrative story telling in Autism and Down's Syndrome. *British Journal of Developmental Psychology, 8*(1), 9-23. doi:10.1111/j.2044-835X.1990.tb00818.x

Loveland, K. A., Tunalia, B., Mcevoy, R. E., & Kelley, M. L. (1989). Referential communication and response adequacy in autism and down's syndrome. *Applied Psycholinguistics; Applied Psycholinguistics, 10*(3), 301-313. doi:10.1017/S014271640000864X

Mäkinen, L., Loukusa, S., Leinonen, E., Moilanen, I., Ebeling, H., & Kunnari, S. (2014). Characteristics of narrative language in autism spectrum disorder: Evidence from the

Finnish. *Research in Autism Spectrum Disorders, 8*(8), 987-996. doi:10.1016/j.rasd.2014.05.001

Malkin, L., Abbot-Smith, K., & Williams, D. (2018). Is verbal reference impaired in autism spectrum disorder? A systematic review. *Autism and Developmental Language Impairments, 3*, 1–24. https://doi.org/10.1177/2396941518763166

Manolitsi, M., & Botting, N. (2011). Language abilities in children with autism and language impairment: Using narrative as a additional source of clinical information. *Child Language Teaching and Therapy, 27*(1), 39-55. doi:10.1177/0265659010369991

Manor, R. (2001). On the overlap of pragmatics and semantics. *Synthese, 128* (1/2), 63-73.

Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and language*, *85*(3), 451-466. https://doi.org/10.1016/S0093-934X(03)00070-1

Matthews, D. (2014). Introduction: An overview of research on pragmatic development. In Matthews, D. (Ed.) *Pragmatic Development in First Language Acquisition* (pp. 1-12). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Matthews, D., Biney, H., & Abbot-Smith, K. (2018). Individual differences in children's pragmatic ability: a review of associations with formal language, social cognition, and executive functions. *Language Learning and Development, 14*(3), 186-223. https://doi.org/10.1080/15475441.2018.1455584

Matthews, D., Lieven E & Tomasello, M. (2007). How toddlers and preschoolers learn to uniquely identify referents for others: A training study. *Child Development*. *78*(6), 1744-1759. doi: 10.1111/j.1467-8624.2007.01098.x

Mayer, M. (1969). *Frog, where are you?* Dial Press New York.

McTear, M. and Conti-Ramsden, G. (1992). *Pragmatic Disability in Children*. London:Whurr

Melis, A. P., Hare, B., & Tomasello, M. (2006). Engineering cooperation in chimpanzees: tolerance constraints on cooperation. *Animal Behaviour, 72*(2), 275-286. doi. 10.1016/j.anbehav.2005.09.018

Michael, J., Sebanz, N., & Knoblich, G. (2016). The sense of commitment: A minimal approach. *Frontiers in Psychology, 6*, 1968. https://doi.org/10.3389/fpsyg.2015.01968

Michael, J., & Székely, M. (2018). The developmental origins of commitment. *Journal of Social Philosophy, 49*(1), 106-123.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child Development, 78*(2), 622-646. https://doi.org/10.1111/j.1467-8624.2007.01018.x

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49-100. doi:10.1006/cogp.1999.0734

Moll, H., & Kadipasaoglu, D. (2013). The primacy of social over visual perspective-taking. *Frontiers in Human Neuroscience, 7*, 558. doi:10.3389/fnhum.2013.00558

Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unsolved mysteries of the mind. Tutorial essays in cognition*. Hove, UK: Lawrence Erlbaum Associates Ltd.

Moore, R. (2014). Ontogenetic constraints on Paul Grice's theory of communication. In: D., Matthews (ed.) *Pragmatic Development in First Language Acquisition*. (pp 87–104). Amsterdam: John Benjamins.

Morris, C. W. (1938). Foundations of the Theory of Signs. In *International Encyclopaedia of Unified Science* (pp. 1-59). Chicago University Press.

Morris, N., & Jones, D. M. (1990). Memory updating in working memory: The role of the

central executive. *British Journal of Psychology, 81*(2), 111-121. https://doi.org/10.1111/j.2044-8295.1990.tb02349.x

Mottron, L. (2004). Matching strategies in cognitive research with individuals with high-functioning autism: Current practices, instrument biases, and recommendations. *Journal of Autism and Developmental Disorders, 34*(1), 19-27. https://doi.org/10.1023/B:JADD.0000018070.88380.83

Murphy, S., Faulkner, D., & Farley, L. (2014). The behaviour of young children with social communication disorders during dyadic interaction with peers. *Journal of Abnormal Child Psychology, 42*(2), 277-289. doi:10.1007/s10802-013-9772-6

Nadig, A., Seth, S., & Sasson, M. (2015). Global similarities and multifaceted differences in the production of partner-specific referential pacts by adults with autism spectrum disorders. *Frontiers in Psychology, 6.* doi:10.3389/fpsyg.2015.01888

Nadig, A., Vivanti, G., & Ozonoff, S. (2009). Adaptation of object descriptions to a partner under increasing communicative demands: A comparison of children with and without autism. *Autism Research,* (2), 334-347. doi:10.1002/aur.102

Nilsen, E. S., & Fecica, A. M. (2011). A model of communicative perspective-taking for typical and atypical populations of children. *Developmental Review, 31*(1), 55-78. doi:10.1016/j.dr.2011.07.001

Nilsen, E. S., & Graham, S. A. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58*(2), 220-249. doi: 10.1016/j.cogpsych.2008.07.002

Nilsen, E. S., Varghese, A., Xu, Z., & Fecica, A. (2015). Children with stronger executive functioning and fewer ADHD traits produce more effective referential statements. *Cognitive Development, 36,* 68-82. doi: 10.1016/j.cogdev.2015.09.001

Norbury, C. F. (2005). The relationship between theory of mind and metaphor: Evidence

from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology, 23*(3), 383-399. https://doi.org/10.1348/026151005X26732

Norbury, C. F. (2014). Practitioner review: Social (pragmatic) communication disorder conceptualization, evidence and clinical implications. *Journal of Child Psychology and Psychiatry, 55*(3), 204-216. https://doi.org/10.1111/jcpp.12154

Norbury, C. F., & Bishop, D. V. (2002). Inferential processing and story recall in children with communication problems: a comparison of specific language impairment, pragmatic language impairment and high-functioning autism. *International Journal of Language & Communication Disorders,* 37(3), 227-251. https://doi.org/10.1080/13682820210136269

Norbury, C. F., & Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders, 2003, 38; Vol.38*(3; 3), 287; 287-313; 313. doi:10.1080/136820310000108133

Norbury, C. F., Gemmell, T., & Paul, R. (2014). Pragmatics abilities in narrative production: A cross-disorder comparison. *Journal of Child Language*, *41*(3), 485-510. doi:10.1017/S030500091300007X

Norbury, C. F., Nash, M., Baird, G., & Bishop, D. V. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: a validation of the Children's Communication Checklist—2. *International Journal of Language & Communication Disorders, 39*(3), 345-364. doi: 10.1080/13682820410001654883

Noveck, I. A., & Sperber, D. (Eds.). (2004). *Experimental Pragmatics*. Basingstoke: Palgrave Macmillan.

Novogrodsky, R. (2013). Subject pronoun use by children with autism spectrum disorders (ASD). *Clinical Linguistics & Phonetics, 27*(2), 85-93.

doi:10.3109/02699206.2012.742567

Novogrodsky, R., & Edelson, L. R. (2016). Ambiguous pronoun use in narratives of children with autism spectrum disorders. *Child Language Teaching and Therapy, 32*(2), 241-252. doi:10.1177/0265659015602935

O'Madagain, C., & Tomasello, M. (2019). Joint attention to mental content and the social origin of reasoning. *Synthese,* 1-22. https://doi.org/10.1007/s11229-019-02327-1

O'Neill, D. K. (2014). Assessing pragmatic language functioning in young children: Its importance and challenges. In Matthews, D. (Ed.) *Pragmatic Development in First Language Acquisition*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

O'Neill, D. K. (2012). Components of pragmatic ability and children's pragmatic language development. In H. Schmid (Ed.), *Cognitive Pragmatics*, (Vol. 4, pp. 261–287). Berlin: Walter de Gruyter.

O'Neill, D. K. (2005). Talking about "new" information: The given/new distinction in children's developing theory of mind. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 84–105). Oxford, United Kingdom: Oxford University Press

O'Neill, D.K. & Happé, F. G. (2000). Noticing and commenting on what's new: differences and similarities among 22-month-old typically developing children, children with Down syndrome and children with autism. *Developmental Science, 3*(4), 457-478. doi:10.1111/1467-7687.00139

Ostashchenko, E., Geelhand, P., Deliens, G., & Kissine, M. (2019). Struggling with alternative descriptions: Impaired referential processing in children with Autism Spectrum Disorder. *Research in Autism Spectrum Disorders, 66*, 101414. https://doi.org/10.1016/j.rasd.2019.101414

Osterhaus, C., Koerber, S., & Sodian, B. (2016). Scaling of advanced theory-of-mind tasks. *Child Development, 87*(6), 1971-1991. https://doi.org/10.1111/cdev.12566

Ozonoff, S., & McEvoy, R. E. (1994). A longitudinal study of executive function and theory of mind development in autism. *Development and Psychopathology*, *6*(03), 415-431. doi:10.1017/S0954579400006027

Parsons, L., Cordier, R., Munro, N., Joosten, A., & Speyer, R. (2017). A systematic review of pragmatic language interventions for children with autism spectrum disorder. *PLoS One, 12*(4), e0172242. https://doi.org/10.1371/journal.pone.0172242

Paul, R., Orlovski, S. M., Marcinko, H. C., & Volkmar, F. (2009). Conversational behaviors in youth with high-functioning ASD and Asperger syndrome. *Journal of Autism and Developmental Disorders, 39*(1), 115-125. https://doi.org/10.1007/s10803-008-0607-1

Pearson, A., Marsh, L., Ropar, D., & Hamilton, A. (2016). Cognitive Mechanisms underlying visual perspective taking in typical and ASC children. *Autism Research, 9*(1), 121-130. https://doi.org/10.1002/aur.1501

Pellicano, E. (2007). Links between theory of mind and executive function in young children with autism: Clues to developmental primacy. *Developmental Psychology, 43*(4), 974. https://doi.org/10.1037/0012-1649.43.4.974

Pellicano, E. (2010). The development of core cognitive skills in autism: A 3-year prospective study. *Child Development, 81*(5), 1400-1416. https://doi.org/10.1111/j.1467-8624.2010.01481.x

Pellicano, E. (2013), Testing the Predictive Power of Cognitive Atypicalities in Autistic Children: Evidence from a 3-Year Follow-Up Study. *Autism Research, 6*: 258–267. doi:10.1002/aur.1286

Pennington, B. & Ozonoff, S. (1996). Executive functions and developmental

psychopathology. *Journal of Child Psychology and Psychiatry, 37*(1): 51-87. 10.1111/j.1469-7610.1996.tb01380.x

Perkins, M. R. (2005). Pragmatic ability and disability as emergent phenomena. *Clinical Linguistics & Phonetics*, 19(5), 367-377. https://doi.org/10.1080/02699200400027155

Perkins, M. R. (2010). Pragmatic impairment. In J. S. Damico, N. Muller & M. J. Ball (Eds). *The Handbook of language and speech disorders* (pp. 227-246). Chichester: Wiley-Blackwell.

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*, 60, 689-700. doi:10.2307/1130734

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief. *British Journal of Developmental Psychology*, *5*, 125 – 137. doi: 10.1111/j.2044-835X.1987.tb01048.x

Peskin, J., Prusky, C., & Comay, J. (2014). Keeping the reader's mind in mind: development of perspective-taking in children's dictations. *Journal of Applied Developmental Psychology, 35*(1), 35-43. https://doi.org/10.1016/j.appdev.2013.11.001

Phelps-Terasaki, D., & Phelps-Gunn, T. (2004). *Test of Pragmatic Language.* Austin, TX: Pro-Ed.

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J. P., & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with autism and Asperger syndrome. *Journal of Autism and Developmental Disorders, 39*(4), 607. doi: 10.1007/s10803-008-0661-8

Pons, F., & Harris, P. (2000). *Test of Emotion Comprehension: TEC*. University of Oxford.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences, 1*(4), 515-526. https://doi.org/10.1017/S0140525X00076512

Prutting, C., & Kirchner, D. (1983). Pragmatic Protocol. *Pragmatic Assessment and Intervention Issues in Language.* New York: College-Hill.

Pugliese, C. E., Anthony, L., Strang, J. F., Dudley, K., Wallace, G. L., & Kenworthy, L. (2015). Increasing adaptive behavior skill deficits from childhood to adolescence in autism spectrum disorder: Role of executive function. *Journal of Autism and Developmental Disorders, 45*(6), 1579-1587. https://doi.org/10.1007/s10803-014-2309-1

Rakoczy, H. (2017). The development of collective intentionality. In K. Ludwig & M. Jankovic (Eds.), *The Routledge Handbook of Collective Intentionality*. London: Routledge.

Renfrew, C. E. (1991). *The Bus Story: A test of narrative speech*. Oxford, UK: Winslow Press.

Robertson, C. E., & Baron-Cohen, S. (2017). Sensory perception in autism. *Nature Reviews Neuroscience*, *18*(11), 671. DOI: 10.1038/nrn.2017.112

Rollins, P. R. (2014). Narrative skills in young adults with high-functioning autism spectrum disorders. *Communication Disorders Quarterly*, *36*(1), 21-28. doi:10.1177/1525740114520962

Rosenthal, M., Wallace, G. L., Lawson, R., Wills, M. C., Dixon, E., Yerys, B. E., & Kenworthy, L. (2013). Impairments in real-world executive function increase from childhood to adolescence in autism spectrum disorders. *Neuropsychology*, *27*(1), 13. https://doi.org/10.1037/a0031299

Rubin, K. H. (1972). Relationship between egocentric communication and popularity among peers. *Developmental Psychology*, *7*(3), 364. doi. 10.1037/h0033348

Rubio-Fernández, P. (2017). The director task: A test of Theory-of-Mind use or selective attention?. *Psychonomic Bulletin & Review*, *24*(4), 1121-1128.

https://doi.org/10.3758/s13423-016-1190-7

Russell, J., Jarrold, C., & Hood, B. (1999). Two intact executive capacities in children with autism: Implications for the core executive dysfunctions in the disorder. *Journal of Autism and Developmental Disorders, 29*(2), 103-112. https://doi.org/10.1023/A:1023084425406

Russo, N., Flanagan, T., Iarocci, G., Berringer, D., Zelazo, P. & Burack, J. (2007). Deconstructing executive deficits among persons with autism: implications for cognitive neuroscience. *Brain and Cognition, 65*(1): 77-86.

Sacrey, L. A. R., Armstrong, V. L., Bryson, S. E., & Zwaigenbaum, L. (2014). Impairments to visual disengagement in autism spectrum disorder: a review of experimental studies from infancy to adulthood. *Neuroscience & Biobehavioral Reviews, 47*, 559-577. https://doi.org/10.1016/j.neubiorev.2014.10.011

Salatas, H. & Flavell, H. (1976). Perspective taking: the development of two components of knowledge. *Child Development, 47*: 103-109. http://www.jstor.org/stable/1128288

Santiesteban, I., Shah, P., White, S., Bird, G., & Heyes, C. (2015). Mentalizing or submentalizing in a communication task? evidence from autism and a camera control. *Psychonomic Bulletin & Review, 22*(3), 844-849. doi:10.3758/s13423-014-0716-0

Schmerse, D., Lieven, E., & Tomasello, M. (2015). Young children use shared experience to interpret definite reference. *Journal of Child Language. 42*(5): 1146-1157.

Schuh, J. M., Eigsti, I. M., & Mirman, D. (2016). Discourse comprehension in autism spectrum disorder: Effects of working memory load and common ground. *Autism Research*, *9*(12), 1340-1352. doi. 10.1002/aur.1632

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge, UK. Cambridge university press.

Semel, E. M., Wiig, E. H., & Secord, W. (2006). *CELF 4: Clinical evaluation of language*

fundamentals. Pearson: 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), 883-885. doi: 10.1126/science.1176170

Sikora, K., Roelofs, A., Hermans, D., & Knoors, H. (2016). Executive control in spoken noun-phrase production: Contributions of updating, inhibiting, and shifting. *The Quarterly Journal of Experimental Psychology, 69*(9), 1719-1740. doi:10.1080/17470218.2015.1093007

Simmons, E. S., Paul, R., & Volkmar, F. (2014). Assessing pragmatic language in autism spectrum disorder: the Yale in vivo Pragmatic Protocol. *Journal of Speech, Language, and Hearing Research, 57*(6), 2162-2173. https://doi.org/10.1044/2014_JSLHR-L-14-0040

Slevc, L. R. (2011). Saying what's on your mind: Working memory effects on sentence production. *Journal of experimental psychology: Learning, Memory, and Cognition, 37*(6), 1503 –1514. https://doi.org/10.1037/a0024350

Sodian, B., & Kristen-Antonow, S. (2015). Declarative joint attention as a foundation of theory of mind. *Developmental Psychology, 51*(9), 1190. doi: 10.1037/dev0000039

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales*: *(VABS)*. NCS Pearson.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition* (Vol. 142). Cambridge, MA: Harvard University Press.

Sperber, D. and Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind and Language,*17(1-2), 3-23. https://doi.org/10.1111/1468-0017.00186

Stalnaker, R. C. Pragmatic Presuppositions. In A. Rogers, B. Wall, & J. P. Murphy, (Eds.) *Proceedings of the Texas Conference on Performatives, Presuppositions, and*

Implicatures. (pp. 135-148). Arlington, Va.: Center for Applied Linguistics.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology, 18*(6), 643.

Suh, J., Eigsti, I., Naigles, L., Barton, M., Kelley, E., & Fein, D. (2014). Narrative performance of optimal outcome children and adolescents with a history of an autism spectrum disorder (ASD). *Journal of Autism and Developmental Disorders, 44*(7), 1681-1694. doi:10.1007/s10803-014-2042-9

Shu, B. C., Lung, F. W., Tien, A. Y., & Chen, B. C. (2001). Executive function deficits in non-retarded autistic children. *Autism, 5*(2), 165-174. https://doi.org/10.1177/1362361301005002006

Tager-Flusberg, H. (1981). On the nature of linguistic functioning in early infantile autism. *Journal of Autism and Developmental Disorders, 11*(1), 45-56. https://doi.org/10.1007/BF01531340

Tager-Flusberg, H. (1995). 'Once upon a ribbit': Stories narrated by autistic children. *British Journal of Developmental Psychology, 13*(1), 45-59. doi:10.1111/j.2044-835X.1995.tb00663.x

Tager-Flusberg, H. (1996). Current theory and research on language and communication in autism. *Journal of Autism and Developmental Disorders, 26*, 169-172.

Tager-Flusberg, H. (2000). Understanding the language and communicative impairments in autism. In L. M. Glidden (Ed.), *International Review of Research on Mental Retardation* (Vol. 20, pp. 185–205). San Diego, CA: Academic Press

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of Autism and Pervasive Developmental Disorders* (3rd ed.) (pp. 335–364). Hoboken, NJ: Wiley and Sons.

Tager-Flusberg, H., & Sullivan, K. (1994). A second look at second-order belief attribution in autism. *Journal of Autism and Developmental Disorders, 24*(5), 577-586. https://doi.org/10.1007/BF02172139

Tager-Flusberg, H., & Sullivan, K. (1995). Attributing mental states to story characters: A comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics, 16*(3), 241-256. https://doi.org/10.1017/S0142716400007281

Tomasello, M. (2018). How children come to understand false beliefs: A shared intentionality account. *Proceedings of the National Academy of Sciences, 115*(34), 8491-8498. https://doi.org/10.1073/pnas.1804761115

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences, 28*: 675-735. doi: 10.1017/S0140525X05440121

Tomasello, M., & Hamann, K. (2012). The 37th Sir Frederick Bartlett Lecture: Collaboration in young children. *Quarterly Journal of Experimental Psychology, 65*(1), 1-12. doi. 10.1080/17470218.2011.608853

Tan, J., & Harris, P. L. (1991). Autistic children understand seeing and wanting. *Development and Psychopathology, 3*(2), 163-174. https://doi.org/10.1017/S0954579400000055

van den Wildenberg, W.P. & Christoffels, I.K. (2010). STOP TALKING! Inhibition of speech is affected by word frequency and dysfunctional impulsivity. *Frontiers in Psychology, 1*:145 doi: 10.3389/fpsyg.2010.00145

Verschueren, J. (1999). *Understanding Pragmatics*. London: Edward Arnold.

Volden, J. (2002). Features leading to judgements of inappropriacy in the language of speakers with autism: A preliminary study. *Journal of Speech Language Pathology and Audiology, 26*(3), 138-146.

Volden, J. (2004). Conversational repair in speakers with autism spectrum disorder. *International Journal of Language & Communication Disorders, 39*(2), 171-189. doi: 10.1080/13682820410001663252

Volden, J., Mulcahy, R. F., & Holdgrafer, G. (1997). Pragmatic language disorder and perspective taking in autistic speakers. *Applied Psycholinguistics, 18*(2), 181-98. doi:10.1017/S0142716400009966.

Volden, J., Coolican, J., Garon, N., White, J., & Bryson, S. (2009). Brief report: Pragmatic language in autism spectrum disorder: Relationships to measures of ability and disability. *Journal of Autism and Developmental Disorders, 39*(2), 388. https://doi.org/10.1007/s10803-008-0618-y

Volden, J., & Phillips, L. (2010). Measuring pragmatic language in speakers with autism spectrum disorders: Comparing the Children's Communication Checklist—2 and the Test of Pragmatic Language. *American Journal of Speech-Language Pathology*. *19*, 204–212. https://doi.org/10.1044/1058-0360(2010/09-0011)

Wallace, G. L., Kenworthy, L., Pugliese, C. E., Popal, H. S., White, E. I., Brodsky, E., & Martin, A. (2016). Real-world executive functions in adults with autism spectrum disorder: Profiles of impairment and associations with adaptive functioning and co-morbid anxiety and depression. *Journal of Autism and Developmental Disorders, 46*(3), 1071-1083. doi: 10.1007/s10803-015-2655-7

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: the role of prosody and context. *Brain, 129*(4), 932-943. https://doi.org/10.1093/brain/awl032

Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition, 191*, 103997. https://doi.org/10.1016/j.cognition.2019.06.009

Wellman, H. & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development, 75*(2), 523-541. 10.1111/j.1467-8624.2004.00691

Wellman, H.M. & Woolley, J.D. (1990) 'From Simple Desires to Ordinary Beliefs: The Early Development of Everyday Psychology', *Cognition* 35 (3), 245–75. https://doi.org/10.1016/0010-0277(90)90024-E

White, S., Hill, E., Happé, F., & Frith, U. (2009). Revisiting the Strange Stories: Revealing Mentalizing Impairments in Autism. *Child Development*, *80*(4), 1097-1117. doi:10.1111/j.1467-8624.2009.01319.x

Whitehouse, A. J., Watt, H. J., Line, E. A., & Bishop, D. V. (2009). Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and autism. *International Journal of Language & Communication Disorders*, *44*(4), 511-528. https://doi.org/10.1080/13682820802708098

Whyte, E. M., & Nelson, K. E. (2015). Trajectories of pragmatic and nonliteral language development in children with autism spectrum disorders. *Journal of Communication Disorders, 54*, 2-14. https://doi.org/10.1016/j.jcomdis.2015.01.001

Whyte, E. M., Nelson, K. E., & Scherf, K. S. (2014). Idiom, syntax, and advanced theory of mind abilities in children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*. 57, 120–130. doi: 10.1044/1092-4388(2013/12-0308)

Wiesner, D. (1991). *Tuesday*. New York: Clarion Books.

Wiig, E. H., Secord, W. A., & Semel, E. (2013). *Clinical Evaluation of Language Fundamentals: CELF-5*. Pearson.

Wilson, A. C., & Bishop, D. V. M. (2019). "If you catch my drift...": ability to infer implied meaning is distinct from vocabulary and grammar skills [version 3; peer review: 3 approved]. *Wellcome Open Research, 4*, 68. doi:10.12688/wellcomeopenres.15210.3

Wittke, K., Mastergeorge, A. M., Ozonoff, S., Rogers, S. J., & Naigles, L. R. (2017). Grammatical language impairment in autism spectrum disorder: Exploring language

phenotypes beyond standardized testing. *Frontiers in Psychology, 8*, 532. https://doi.org/10.3389/fpsyg.2017.00532

Yerys, B. E., Wallace, G. L., Harrison, B., Celano, M. J., Giedd, J.N & Kenworthy, L. (2009) Set-shifting in children with autism spectrum disorders: reversal shifting deficits on the Intradimensional / Extradimensional Shift Test correlate with repetitive behaviors. Autism: *The International Journal of Research and Practice*. *13*(5), 523–538. doi:10.1177/1362361309335716.

Ying Sng C, Carter M, Stephenson J. A systematic review of the comparative pragmatic differences in conversational skills of individuals with autism. *Autism Dev Lang Impair*. 2018;3(2):1–24. doi:10.1177/2396941518803806

Yuill, N. (1992). Children's production and comprehension of trait terms. *British Journal of Developmental Psychology*, *10*(2), 131-142. https://doi.org/10.1111/j.2044-835X.1992.tb00567.x

Yule, G. (1997). *Referential Communication Tasks*, New York: Erlbaum.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature protocols, 1*(1), 297-301. https://doi.org/10.1038/nprot.2006.46

Zelazo, P., Müller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*(3), 1–137. https://doi.org/10.1111/j.0037-976X.2003.00261.x

Zhao, L., Wang, J. J., & Apperly, I. A. (2018). The cognitive demands of remembering a speaker's perspective and managing common ground size modulate 8-and 10-year-olds' perspective-taking abilities. *Journal of Experimental Child Psychology, 174*, 130-149. https://doi.org/10.1016/j.jecp.2018.05.013

Zufferey, S. (2016). Pragmatic acquisition. In Östman, J.-O. & Verschueren J. (Eds.)

Handbook of Pragmatics. Amsterdam: John Benjamins

APPENDICES

APPENDIX 1: Theory of mind battery (Maximum score on ToM composite = 8 points).

The Theory of Mind battery was applied in a stepwise fashion (see e.g. Charman et al., 2011). Our testing procedure was follows. The first author started with 8.1 'Second-order false belief', which all children were administered. If the child failed both questions on the Robot Story, then the two first-order Theory of Mind tasks were administered (and the two Strange Stories items were not administered). Thus eight children with ASD and two TD children were not administered Strange Stories.

If the child passed one or both questions on the Robot Story, then it was assumed that he or she would pass the first order Theory of Mind and thus the child was accredited two points for first order Theory of Mind without administration and the child progressed to the administration of the two adapted Strange Stories items. All responses to Theory of Mind tests were audio-recorded, transcribed and scored by the first author. The fourth author also scored 18% of Theory of Mind tasks (8 participants - 4 ASD/ 4 TD) whilst blind to diagnostic status, with excellent agreement (Cohen's k = .921).

1.1 Second order false belief: The Robot Story (Coull, Leekam, & Bennett, 2006) accompanied by five pictures drawn by the fourth author (Maximum score = 2)

It is Paul's birthday. Paul and Sally are in his play room. He is showing Sally his favourite new present—a robot. Paul puts the robot back in the box with the lid on and then has to go outside. While Paul is away, Sally decides to play a trick on Paul and move the robot from its box and hide it away in the cupboard. While Sally is hiding the robot in the cupboard, Paul passes by the window and sees Sally hiding the robot in the cupboard. But Sally doesn't see Paul watching her hide the robot in the cupboard. She doesn't see him! Paul then returns to the toy room.

Probe question 1: 'Does Paul know that the robot is in the cupboard'? *Yes* = Correct, *No* = Incorrect

Probe question 2: 'Does Sally know that Paul saw her hide the robot'? *No* = Correct, Yes = Incorrect

Probe question 3: 'Where will Paul look for the robot'? *In the cupboard* = Correct

1. Test question 1: 'Where does Sally think Paul will look for the robot'? *Box* = correct.

2. Test question 2: 'Why does Sally think Paul will look for the robot in the _____'? (Box/ Cupboard)

Correct = *Because she doesn't know that Paul knows the robot is in the cupboard*.

1.2 First order false belief (Maximum combined score = 2)

1.2.1 *Contents false belief* : We used playmobil figures to act out Wellman and Liu's (2004) adaptation of Perner, Leekam and Wimmer's (1987) 'Smarties' task. E shows the child a 'plasters' box and asks what the child thinks is inside the box. The child invariably replies 'plasters'. Then E shows the child that there is in fact a toy rabbit inside the box. E closes the box.

a) *[1st CONTROL Q] 'What's in here*?'

b) *Test Question*: Your teacher / brother has not seen inside this box. When s/he comes in later, I'll show her / him this box just like this and ask her / him what's in here. What will s/he say?

c) *[2nd CONTROL Q] 'Is that what's really in the box*?

(1 item; Score either 1 or 0).

1.2.2 'Sally-Anne' change-of-location task (Baron-Cohen, Leslie & Frith, 1985).
We used playmobil figures to act this out. (1 item; Score either 1 or 0).
Scott has a favourite toy; his crayon. One day when he has finished playing with his crayon, he puts it under this blanket and goes out to play in the garden. While Scott is in the garden, his mother comes in to tidy up. She takes the crayon from under the blanket and puts it in this box. Then she goes out the room to the kitchen. Now Scott has come back.
Test Question ''So, where will Scott look for his crayon? In this box or under the blanket?''

a) ''Where is Scott's crayon really? this box or under the blanket? (The Reality Control Question).
b) ''Where did Scott put his crayon? this box or under the blanket? (The Memory Control Question).

1.3. Happe's (1994) 'Strange Stories' items adapted for younger children
1.3.1 Kittens (persuasion) (acccompanied by pictures) (Maximum score = 2)
Tom wanted to buy a kitten, so he went to see Mrs Smith. Mrs Smith had lots of kittens she didn't want. Mrs Smith loved the kittens, and would never hurt them, but she couldn't keep them all herself. When Tom visited he wasn't sure if he wanted one of Mrs Smith's kittens. All of Mrs Smith's kittens were white and Tom wanted a black kitten. Then Mrs Smith said, "If no one buys the kittens I'll have to throw them away with the rubbish"
Mental State Question: Why did Mrs Smith say that to Tom?
Scoring for mental state question:
2 points—reference to persuasion, manipulating feelings, trying to induce guilt/pity (e.g. "*she was trying to make her feel sorry for the kittens so she would buy one*")

1 point—reference to outcome (to sell them or get rid of them in a way which implies not hurting) or simple motivation (e.g. "*to make Tom sad*")

0 points—reference to general knowledge or dilemma without realization that the statement was not true (e.g. "*she's a horrible woman*")

Check Question: Was it true, what Mrs Smith said? Would Mrs Smith really have thrown the kittens away with the rubbish?
Scoring for check question: If child says 'no' to the check question, despite scoring 0 for the mental state question, then the child will score 0.5 overall for this item.

1.3.2 Hidden Biscuits (Double Bluff), adapted from Happe's original 'Prisoner's Double-Bluff' Strange Story item (acccompanied by pictures). (Maximum score = 2)
Lucy has a big brother who is called David. David is always eating Lucy's biscuits so she has hidden them from him. David knows Lucy put the biscuits either in the bedroom or in the garden. David thinks that Lucy will not want to tell him where the biscuits are because she will want to eat them herself. David thinks she will lie about where the biscuits are. Lucy is very clever, she will not let David find her biscuits. The biscuits are in her bedroom. When David asks Lucy where her biscuits are, she says, "They are in the bedroom".

Mental State Question: Why did Lucy say that?
Scoring for mental state question:

2 points—reference to fact that David will not believe and hence look in other place, reference to Lucy's realization that that's what he'll do, or reference to double bluff.

1 point—reference to outcome (to stop David eating the biscuits) or to mislead them.

0 points—reference to motivation that misses the point of double bluff (e.g. "*she was angry*").

Memory check questions: Is it true what Lucy said?

YES – score as correct
NO/ DON'T KNOW – score as incorrect

1.4 Practice and Control Items for 'Strange Stories

1.4.1 Practice item "Ghost" story

This was presented, accompanied by pictures, prior to the aforementioned two 'Strange Stories' items. Feedback was given for this story, using mental state terms (e.g. "Chris is pretending. He wants Mr Brown to think that….") to ensure children understood the task.

It is Halloween and Chris is going to a fancy-dress party. He is going as a ghost. He wears a big white sheet with eyes cut out to see though. As he walks to the party in his costume, he bumps into Mr. Brown. It is dark and Mr. Brown says, "Oh! Who is it?". Chris answers, "I'm a ghost, Mr. Brown".

Mental State Question: Why did Chris say that?

1.4.2 Physical state inference control story "Mountains"

This item was taken from White, Hill, Happe and Frith's (2009) selection of 'natural physical state stories' control stories. This item was found by White et al. (2009) to be well-matched to the mental state stories in difficulty level. This was presented accompanied by pictures.

When there is a storm, rocks can fall from the top of mountains. One day on a mountain a very large rock becomes loose and starts rolling down the mountain. It rolls and rolls and rolls, going faster and spinning and bouncing off the mountain side. Suddenly, there is a very noisy splash.

KEY INFERENCING QUESTION: "Why is there a loud splash?"

PASS —reference to the boulder falling into water to make the splash (e.g. "the rock must have fallen into a lake") or reference to water without reference to the boulder (e.g. "there was a pool at the bottom of the mountain").

FAIL—reference to irrelevant or incorrect factors (e.g. "It's very big so it's very noisy".)

APPENDIX 2. Overall procedure for Common Ground study (Chapter 4)

While the order of presentation of the activities within the 'Reference Interpretation' and 'Elicited Reference Production' tasks was counterbalanced, the order of presentation of these experimental tasks in relation to the assessment of Language Ability, Non-verbal IQ and Theory of Mind was fixed. That is, the experimental tasks were presented as a 'break' from the assessment of Formal Language Ability, Non-verbal IQ and Theory of Mind. This set order is illustrated in Table S1 below.

Table S1: Illustration of overall testing procedure

Task Type	Activity
Language ability Test 1	Clinical Evaluation of Language Fundamentals (CELF) 'Sentence Structures' (comprehension) sub-test
1st Interpretation measure experimental trial	Example (counterbalanced order A) a) One of the experimenters (e.g. E1 counterbalanced) suggests a construction activity disguised as 'having a break' from the CELF task. b) E1constructs **electric circuit** with Participant (P). c) E1covers electric circuit & leaves to find a *battery*. d) E2 arrives carrying materials for **remote-controlled car** and constructs this with P. e) E2 covers remote-controlled & leaves to find a *battery*. f) E (counterbalanced whether E1 or E2) returns with *battery* and says 'Here you go. Now you can do it'. (Both activities are covered at this point). g) P takes *battery* and goes to either electric circuit or car.
1st Production trial	P explains to an experimenter (counterbalanced whether Naïve or Knowledgeable Experimenter) via web-cam how to construct toy (in this case electric circuit).
Theory of Mind	2nd Order Theory of Mind "Robot Story" (Coull, Leekam, & Bennett, 2006) – see section 8.1 below for details.
	If P fails "Robot Story", then "Contents 1st order False Belief" is administered. If P passes "Robot Story", then the practice story for the 'Strange Stories' is administered, followed by "Kittens" Strange Stories item. (See Introduction to Section 8).
2nd Interpretation measure experimental trial	Example (counterbalanced order A) a) One of the experimenters (e.g. E2 counterbalanced) suggests a construction activity disguised as 'having a break' from the Theory of Mind task. b) E2 constructs **slingshot** with P c) E2 covers slingshot & leaves to find *elastic band* d) E1 arrives carrying materials for **woodpecker** and constructs this with P e) E1 covers woodpeckter & leaves to find an *elastic band*

	h) E (counterbalanced whether E1 or E2) returns with *elastic band* and says 'Here you go. Now you can do it'. (Both activities are covered at this point). P takes *elastic band* and goes to either slingshot or woodpecker.
2nd Production trial production	P explains to an experimenter (counterbalanced whether Naïve or Knowledgeable) via web-cam how to construct toy (in this case woodpecker).
Theory of Mind	If P had failed "Robot Story", then "Location change 1st order False Belief" is administered. If P had passed "Robot Story", then the "Biscuits" Strange Stories item is administered.
Theory of Mind	"Mountains" physical state inference control task administered.
Language ability Test 2	Clinical Evaluation of Language Fundamentals (CELF) 'Formulated Sentences' (expressive) sub-test.
3rd Interpretation measure experimental trial	Example (counterbalanced order A) a) One of the experimenters (e.g. E1 counterbalanced) suggests construction activity disguised as 'having a break' from the CELF task. b) E1constructs **blowfish** with P. c) E1covers blowfish & leaves to find a *straw*. d) E2 arrives carrying materials for **spinning bird** and constructs this with P. e) E2 covers spinning bird & leaves to find a *straw*. i) E (counterbalanced whether E1 or E2) returns with straw and says 'Here you go. Now you can do it'. (Both activities are covered at this point). P takes *straw* and goes to either blowfish or spinning bird.
PIQ Test 1	British Ability Scales (BAS) "Pattern Construction" sub-test.
4th Interpretation measure experimental trial	Example (counterbalanced order A) a) One of the experimenters (e.g. E2 counterbalanced) suggests construction activity disguised as 'having a break' from the Theory of Mind task. b) E2 constructs **aeroplane** with P. c) E2 covers aeroplane & leaves to find *glue*. d) E1 arrives carrying materials for **certificate** and constructs this with P. e) E1 covers certificate & leaves to find *glue*. j) E (counterbalanced whether E1 or E2) returns with *glue* and says 'Here you go. Now you can do it'. (Both activities are covered at this point). P takes *glue* and goes to either aeroplane or certificate.
Experimental – Interpretation Co-presentence CONTROL	Example (counterbalanced order A) a) One of the experimenters (e.g. counterbalanced which one) suggests 'snack / puzzle break'

	b) E2 sits next to uncovered **tumbling bunny** activity, but has snack or completes puzzle with P. (E2 does not see marble run activity). c) E2 leaves. d) E1 arrives, constructs **tumbling bunny** with P e) E1 can not find *marble* so leaves tumbling bunny uncovered and leaves to find marble. f) E1 returns carrying materials for with **marble run** activity with P. g) E1 leaves to find *marble* (leaving marble run uncovered) h) E2 returns with *marble* and says 'Here you go. Now you can do it'. (Both activities uncovered at this point). P takes *marble* and goes to either marble run or tumbling bunny.
3rd Production trial	P explains to an experimenter (counterbalanced whether Naïve or Knowledgeable) via web-cam how to construct toy (in this case telephone).
Non-verbal IQ	BAS Matrices **E2**
Experimental – Comprehension CONTROL	Example (counterbalanced order A) i) One of the experimenters (e.g. counterbalanced which one) suggests 'snack / puzzle break' j) E2 arrives, constructs **telephone** with P. k) E2 covers up telephone and leaves to find string. l) E2 returns carrying materials for **flying bird** and constructs this with P. m) E2 leaves again to find *string.* n) E1 arrives sits next to uncovered **flying bird** activity (uncovered), but has snack or completes puzzle with P. (E1 does not see telephone activity). o) E1 leaves. (E2 returns, ensures both activities uncovered, then leaves). p) E1 returns with *string* and says 'Here you go. Now you can do it'. (Both activities uncovered at this point). P takes *string* and goes to either telephone or flying bird.
4th Production trial	P explains to an experimenter (counterbalanced whether Naïve or Knowledgeable) via web-cam how to construct toy (in this case tumbling bunny).
Vocabulary check	The first author presented the child with a series of picture pairs and asked the child to tell her how each pair was different. This was designed to elicit the target adjectives: short / long; big / small; black / white / yellow / pink / green. All child spontaneously named this adjectives.
Episodic memory check	At the end of the experiment, each child was asked to tell either a parent or teacher which activity they had completed with each experimenter. All children successfully remembered this.

Preference check	If a child had scored zero for any of the experimental interpretation trials, at this point the child was asked which activities he or she liked best. (The purpose of this was to determine whether the child chose the foil because it was his / her favourite option).

APPENDIX 3: Animal pairings and Appropriately Informative Responses

Demonstration items

Trial Type	Target	Distractor	Correct Response
Experimenter modelled	Fish	Butterfly	Fish
Experimenter modelled	White Mouse	Brown Mouse	White Mouse
Experimenter modelled	Small stripy cat	Big stripy cat	Small cat

= Fixed trial order

Experimental Trials

Switch Condition

Experimental Trials

Pairing	Target	Distractor	Appropriately informative Response	Simple or Complex
1	**Spotty dog**	Snail	**Dog**	Simple
2	**Spotty dog**	Spider	**Dog**	Simple
3	**Spotty dog**	Horse	**Dog**	Simple
4	**Spotty dog**	Cricket	**Dog**	Simple
5	**Spotty dog**	Cat	**Dog**	Simple
6	**Spotty dog**	Spotty dog (pink)	**Black dog**	Complex
7	**Spotty dog**	Spotty dog (small)	**Big dog**	Complex
8	**Spotty dog**	Striped dog	**Spotty dog**	Complex
9	**Spotty dog**	Spotty dog (small)	**Big dog**	Complex
10	**Spotty dog**	Plain black dog	**Spotty dog**	Complex

No Switch Condition

Experimental Trials

Pairing	Target	Distractor	Appropriately informative Response	Simple or Complex
1	**Spotty cat (grey)**	Butterfly	**Cat**	Simple
2	**Horse (black/white)**	Sheep	**Horse**	Simple
3	**Horse (spotty)**	Fish	**Horse**	Simple
4	**Spotty cat (yellow)**	Snail	**Cat**	Simple
5	**Spotty rabbit**	Butterfly	**Rabbit**	Simple
6	**Spotty snake**	Stripy snake	**Spotty snake**	Complex
7	**Spotty frog**	Stripy frog	**Spotty frog**	Complex
8	**Cow (spots)**	Small cow (spots)	**Big cow**	Complex
9	**Brown pig (white spots)**	Pink pig (white spots)	**Brown Pig**	Complex
10	**Spotty cat (brown)**	Stripy cat	**Spotty cat**	Complex

NB:

All targets in *no switch* condition have a pattern (so equal opportunity to be over-informative)

Order of trials - randomised by PsychoPy

Left / Right presentation of target – counterbalanced by PsychoPy

APPENDIX 4: List of animals included in filler trials for switch and no-switch condition as target and distractors

1. Sheep
2. Butterfly
3. Snail
4. Chicken
5. Duck
6. Butterfly
7. Spider
8. Hamster
9. Fish
10. Mouse
11. Donkey
12. Bird
13. Deer