Pragmatics and Semantics in Children with Autism Spectrum Disorder

Karece Lopez

Graduate Center, City University of New York

Recommended Citation
http://academicworks.cuny.edu/gc_etds/1026

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Graduate Works by Year: Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
PRAGMATICS AND SEMANTICS IN CHILDREN WITH AUTISM SPECTRUM DISORDER

by

KARECE LOPEZ

A dissertation submitted to the Graduate Faculty in Speech-Language-Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2015
This manuscript has been read and accepted for the Graduate Faculty in Speech-Language-Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Richard G. Schwartz, Ph.D.

__________________________  ________________________________
Date  Chair of Examining Committee

Klara Marton, Ph.D.

__________________________  ________________________________
Date  Executive Officer

Klara Marton, Ph.D.

Sandeep Prasada, Ph.D.

Richard Schwartz, Ph.D.

Supervisory Committee

Joanne Gerenser, Ph.D.

External Reviewer

THE CITY UNIVERSITY OF NEW YORK
Abstract

PRAGMATICS AND SEMANTICS IN CHILDREN WITH AUTISM SPECTRUM DISORDER

by

Karece Lopez

This study examined scalar implicature to investigate semantic bases of pragmatic language impairment in children with autism spectrum disorders (ASD). Scalar implicatures are inferences made by listeners, whereby they strengthen the weaker meaning of a term that can be represented on a scale (Grice, 1975; Horn, 1992). Scalar terms include: some/all; or/and; numerals. These inferences depend on understanding a speaker’s intent and having the cognitive skills necessary to process such information in real time. Informativeness is the value a listener places on having to derive an implicature and assumes that the listener perceives the speakers intentions. Cognitive effort includes the executive functions working memory, interference, and inhibition (Wilson & Sperber, 2004). Children with autism have difficulty with social language and appear to be insensitive to speaker’s intentions, therefore, they may have difficulty understanding scalar implicatures (Happé, 1993). In addition, because it is also believed that inferences place great demands on cognitive resources, children with impaired executive functions, as may be the case in autism, may have difficulty comprehending these sentences (Ozonoff, South, & Provencal, 2005). Method Subjects viewed 1 or 4 pictures while hearing sentences under manipulations of cognitive effect (informativeness level) and cognitive effort (processing costs). Cognitive effect was controlled by experiment (2) and cognitive effort was manipulated by varying stimulus onset asynchrony (3). Three participant groups were tested: seven 7-9 year-old children with autism,
ten 7-9 year-old children with typical language development (TD), and 14 adults with a history of typical language. Performance on three comprehension tasks were compared: a yes/no picture-sentence judgment task, a two-talker picture-sentence judgment task, and a 4AFC picture-sentence matching task using a table-top eye tracker (Tobii TX300). To assess the role of executive functions on scalar comprehension participants also completed a computerized non-verbal task function of attention, interference, and response inhibition. **Results** Children, regardless of clinical presentation, did not perform as well as adults. Within weak scalars (*some*, *or*) children had the greatest difficulty when the target was *or*. Within the strong scalar group, *and* was more difficult than *all*. Children with poorer performance on the executive functions task tended to be younger and also had the poorest accuracy on the scalar implicature tasks. There were no group differences between TD and ASD children in the executive functions task. The three different scalar tasks revealed task effects on inference making in children. TD children were able to accept more implied semantic sentences as correct in the two-talker task compared to the yes/no task. When the stimulus onset asynchrony was set so that the visual preceded the audio stimulus by 750 ms, TD children also performed better. Children with ASD did not demonstrate these differences. The eye tracking data also revealed slower processing of correct responses for children with ASD. Overall children with ASD accepted implied semantic interpretations of the weak scalars *some* and *or*. However, the process in which they did so was different from that of TD children and adults, with more individual variation within the ASD group. **Acknowledgements** This research was supported by two grants from the National Institutes of Health: a pre-doctoral fellowship, 5F31DC013002, to Karece Lopez and a multi-year award, 5R01DC011041, to Richard Schwartz.
Acknowledgments

I would like to express the deepest appreciation to my committee chair Professor Richard Schwartz; committee members, Professors Klara Marton and Sandeep Prasada. I am grateful to all members of the committee including my outside reader, Dr. Joanne Gerenser, for their time and expertise. This dissertation would not have been possible without the mentorship of Dr. Schwartz, and the equipment in his Child Language Laboratory. I am also thankful to Dr. Marton for providing background on executive functions for this project as a committee member and for her support of my completing the degree in her capacity as Executive Officer. Dr. Prasada provided valuable insight into the theoretical considerations of implicature and suggested using numerals.

I am grateful to former Executive Officer, Martin Gitterman, who is an excellent teacher and counselled me on managing my teaching responsibilities while completing my program goals. In this effort, I’d also like to thank former Assistant Program Officer Loretta Walker and current APO Patricia Goodson. Also, Gary Chant, Laboratory Technician, Dr. Simon Henin, graduate of the program and Matlab programmer, and Susan Steinman, Manager of the Language and Hearing Research Lab at the New York Eye and Ear Infirmary and E-Prime wizard, were a tremendous help in getting the hardware and software connected and working to run these studies. I am also thankful to Heidi Udvardy, a former student of mine who volunteered in the lab, for her skills as a graphic designer in creating the images used in my study.

I highly appreciate all their help and support of my peers in the Child Language Laboratory and other current and former students of the Ph.D. Program in Speech-Language Sciences. Christine Rota-Donahue, Iris Fishman, and Zara Waldman DeLuca provided valuable practical advice on completing the dissertation while managing other responsibilities. Dr.
DeLuca and Talita (Dr. Talita) Fortunato-Tavares made valuable comments on a draft of this manuscript; Grace Kim-Lambert edited portions. Dana Battaglia, Emily Root, Erin Reilly, Georgia Drakapolou, and Joanne Gerenser for discussing assessment methods and aspects of language development in autism. Katsiaryna Aharodnik and Ms. Kim-Lambert were instrumental in recruiting potential subjects and helped administer both the language and experimental testing. Experiment 1 also could not have been completed without Erin Reilly and Rejitha Nair who patiently recorded (and in some cases re-recorded) the auditory stimuli. Last, but not least, there were several interns and volunteers over the years who recruited subjects, administered language testing, and cleaned data. Thank you, Irene Karas, Katherine Brown, Kathleen Jedruszczuk, Layla Oubssis, Samia Zahran, Nelly Hernandez, Rebekah Lee, and Shaira Khan. My first intern, Snighda Banerjee, went on to complete a doctorate in psychology. We have kept in touch over the years and I am thankful to her letting me know that I could be a successful mentor.

In addition, I would like to thank former St. John’s University Professor Fredericka (Freddie) Bell-Berti for strongly encouraging me to pursue the doctorate and providing supplemental pep talks; Professor Peggy Jacobson also for knowing what I was going through but telling me I would succeed. I also had the privilege to work with Katherine (Kay) Powell for during the writing of this manuscript. I learned a lot about administration and the institutional review board but more importantly, Kay was one of the best bosses I ever had. It is partly to her that I owe forcing me to write instead of working on other things. Lastly, I would like to thank my mother and sister for providing me the time and resources to pursue what I have always wanted to do. Thank you for not allowing me to quit.
This research was supported by two grants from the National Institutes of Health: a pre-doctoral fellowship, 5F31DC013002, to myself and a multi-year award, 5R01DC011041, to Richard Schwartz.
# Table of Contents

Acknowledgments vi

Table of Contents ix

List of Tables xiii

List of Figures xiv

Introduction 1

Pragmatics-Semantics in Autism Spectrum Disorders 4

Implicature Theory 7

Implicature in Typically Developing Children 13

Scalars 13

Numbers 15

Implicature in Autism Spectrum Disorders 16

Eye Tracking 21

Summary 24

Aim of the present study 25

General Method 28

Participants 28

Additional selection criteria 29

Additional selection criteria for children 29

Group comparisons 31

Stimuli for Experiments 1 and 2 33

Procedure 35
Apparatus .................................................................................................................. 36
Data Reduction .......................................................................................................... 37
  Reaction time and accuracy ................................................................................... 37
  Eye tracking ............................................................................................................. 37

Experiment 1: Truth Value Judgment Task and Felicity Judgment Task 39
  Experiment 1a: Truth Value Judgement Task ...................................................... 40
    Participants .......................................................................................................... 40
    Procedure ............................................................................................................. 40
  Experiment 1b: Felicity Judgment Task ............................................................... 41
    Participants .......................................................................................................... 41
    Procedure ............................................................................................................. 41

Predictions ............................................................................................................... 43
Results ..................................................................................................................... 43
  Adults ..................................................................................................................... 43
  Children overview ................................................................................................. 46

Comparisons between TD and children with ASD ................................................. 47
  Comparisons between individual children .......................................................... 49

Summary ................................................................................................................... 60
Discussion .................................................................................................................. 61

Experiment 2: Four Alternative Forced-Choice Eye Tracking Task 64
  Experiment 2a ........................................................................................................ 64
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scalar Implicature</td>
<td>106</td>
</tr>
<tr>
<td>Accuracy and time course processing.</td>
<td>106</td>
</tr>
<tr>
<td>Executive functions.</td>
<td>108</td>
</tr>
<tr>
<td>Future Directions</td>
<td>110</td>
</tr>
<tr>
<td>Summary</td>
<td>111</td>
</tr>
<tr>
<td>Appendix A. Demographics</td>
<td>113</td>
</tr>
<tr>
<td>Appendix B. The 15 stimuli sets for Experiments 1 and 2.</td>
<td>116</td>
</tr>
<tr>
<td>Appendix C. Trial Timelines for Experiment 1</td>
<td>117</td>
</tr>
<tr>
<td>Appendix D. Individual Results for Scalar Comprehension Tasks</td>
<td>118</td>
</tr>
<tr>
<td>References</td>
<td>121</td>
</tr>
</tbody>
</table>
List of Tables

Table 1. Average Characteristics of the Typically Developing and Autism Spectrum Groups .... 31
Table 2. Adult’s Experiment 1 Average Accuracy and Reaction Time (SD) by SOA ............... 44
Table 3. Children’s Experiment 1 Average Accuracy and Reaction Time (SD) by SOA .......... 46
Table 4. Children Scalar Accuracy (SD) and Reaction Time (SD) on Experiment 1 .............. 49
Table 5. Demographic Data for Children with Low Accuracy Scores on the TVJT and FJT ...... 55
Table 6. Children’s Mean Accuracy (SD) in Experiment 2 ............................................ 69
Table 7. Experiment 2 Children’s Gaze Data during Quantifier Phase and Predicate Phase ..... 78
Table 8. Children’s Executive Function Mean (SD) Percent Correct ................................ 100
Table A1. Detailed Demographics for Typically Developing Children ............................... 113
Table A2. Detailed Demographics for Children with Autism ........................................... 114
Table A3. Detailed Diagnostic Information for Children in the Autism Group .................. 115
Table B. The 15 Stimuli Sets for Experiments 1 and 2 .................................................. 116
Table D1. Typically Developing Children’s Average Accuracy (SD) on Experiment 1a. Truth Value Judgment Task 118
Table D2. Typically Developing Children’s Average Accuracy (SD) on Experiment 1b. Felicity Judgment Task 119
Table D3. Children with Autism Average Accuracy (SD) on Experiment 1a. Truth Value Judgment Task 120
Table D4. Children with Autism Average Accuracy (SD) on Experiment 1b. Felicity Judgment Task 120
List of Figures

Figure 1. Semantic and pragmatic interpretations of two images using sentences with the scalar terms some and all ................................................................. 9
Figure 2. Semantic and pragmatic interpretations of two images using sentences with numerals three and five ................................................................. 10
Figure 3. Schematic design of eye tracker ........................................................................ 22
Figure 4. Communication by socialization domain scaled scores on the Vineland Adaptive Behavior Scales for individual typically developing children and children with autism. ........ 32
Figure 5. Sample trial set for Experiments 1 and 2 .............................................................. 34
Figure 6. Chart of interpretation of responses in yes/no task Experiment 1a: Truth value judgement task ........................................................................... 41
Figure 7. Experiment 1b: Felicity judgment task trial procedure for stimulus onset asynchrony condition 0 ms. ................................................................. 42
Figure 8. Chart of interpretation of responses for two-talker task Experiment 1b: Felicity judgment task ........................................................................... 42
Figure 9. Trial type by reaction time (ms) for Experiment 1 by stimulus onset asynchrony for adults. ......................................................................................... 45
Figure 10. Accuracy scatterplots for typically developing children and children with ASD comparing Experiment 1 tasks for each trial type on felicity and truth value judgment tasks. .... 51
Figure 11. Accuracy of individual typically developing children for each trial type on the truth value judgment task and the felicity judgment task for the three stimulus onset asynchronies. .. 57
Figure 12. Accuracy of individual typically developing children for each trial type on the truth value judgment task and the felicity judgment task for the three stimulus onset asynchronies ... 58
Figure 13. Accuracy of individual children with autism for each trial type on the truth value judgment task and the felicity judgment task for the three stimulus onset asynchronies .......... 59
Figure 14. Accuracy of individual children with autism for each trial type on the truth value judgment task and the felicity judgment task for the three stimulus onset asynchronies .......... 60
Figure 15. Eye tracking Experiment 2 trial procedure for a target sentence in the some/all condition ......................................................................................... 66
Figure 16. Individual children's accuracy on each trial type for the four-alternative forced-choice eye tracking task ........................................................................... 68
Figure 17. Proportions of fixation time to the target and scalar distracter (Alt) images during the four response phases for each group in Experiment 2. ................................................................. 71
Figure 18. Time course of eye movements during visual preview in Experiment 2................. 72
Figure 19. Children’s time time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid correct all trials............... 73
Figure 20. Children’s time course of eye movements to the four image types during the predicate phase of the response period in Experiment 2 for valid correct all trials ............................... 74
Figure 21. Children’s time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid correct and trials.............. 76
Figure 22. Time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid correct some trials................................. 79
Figure 23. Time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid correct or trials ........................................ 81
Figure 24. Time course of eye movements to the four image types during the predicate phase of the response period in Experiment 2 for valid correct or trials .......................................... 82
Figure 25. Time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid correct only some trials................................. 84
Figure 26. Time course of eye movements to the four image types during the predicate phase of the response period in Experiment 2 for valid correct only some trials................................. 85
Figure 27. Time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid incorrect or trials ................................. 87
Figure 28. Time course of eye movements to the four image types during the predicate phase of the response period in Experiment 2 for valid incorrect or trials .................................................. 88
Figure 29. Quantifier phase proportions of fixation times to the target image and response accuracy for a high performing child in each group...................................................... 89
Figure 30. Predicate phase proportions of fixation times to the target image and response accuracy for a high performing child in each group...................................................... 90
Figure 31. Quantifier phase proportions of fixation times to the target image and response accuracy for a low performing child in each group...................................................... 91
Figure 32. Predicate phase proportions of fixation time to the target image and response accuracy for a low performing child in each group...................................................... 92
Figure 33. Sample screens for each condition of the executive function task .......................... 97

Figure 34. Illustration of two trials for the attention condition in the executive function task .... 97

Figure 35. Scatterplots for typically developing children and children with autism accuracy on vigilance compared to distracter interference and compared to response inhibition on the executive function task ................................................................. 101

Figure C. Trial timelines for Experiment 1................................................................. 117
Pragmatics and Semantics in Children with Autism Spectrum Disorder

Autism Spectrum Disorders (ASD) are characterized by deficits in social communication and restricted or repetitive patterns of behaviors or interests (American Psychological Association [APA], 2013). The prevalence of autism in the United States for children under the age of 8 are between 1 in 88 to as high as 1 in 68 (Wingate et al., 2014), an increase of 78% since 2002 (Wingate et al., 2012). Although there is a range of language ability in children with autism, at least 63% of all 8 year olds with ASD have language or learning impairments (Levy et al., 2010). Almost half of children with ASD have average to above average IQ (Wingate et al., 2012). The most persistent and noticeable language deficits in ASD are found within the domain of pragmatics, the appropriate use of language in social situations (Eigsti, de Marchena, Schuh, & Kelley, 2011). It is not surprising that many individuals with these disorders have social communication deficits, which create discrepancies between their IQ and adaptive behavior (Kenworthy et al., 2010). There is a paucity of data on how specific areas of language are affected in the disorder and if any of the language deficits these children have are specific to ASD. The result is that we lack a phenotype for language in autism that can inform and improve diagnostics and treatment for communication impairment for children with ASD.

The heterogeneous nature of ASD also compounds our ability to define a language profile and ascribe an over-arching theory for impairments in ASD. The newly released Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (APA, 2013), DSM – 5, subsumes the autism, childhood disintegrative disorder, Asperger’s syndrome (AsP), and Pervasive Developmental Disorder - Not Otherwise Specified (PDD-NOS) diagnoses into one category: autism disorder (Andrews et al., 2009; APA, 2013; Kupfer & Regier, 2011). Another label often seen in the literature and used by clinicians was high functioning autism (HFA).
These children had typical intelligence and often language within the average range but HFA was a clinical and research label that was never defined in the DSM nor was HFA systematically applied in the field (Ozonoff, South, & Miller, 2000; Sharma, Marks Woolfson, Hunter, 2012). It was unclear if HFA was different from Asperger’s syndrome (AsP) because there was a lot of inter-clinical variation in terms of diagnosing these ASD subgroups (Lord et al., 2012; Williams & Bowler, 2014).

It is still unclear what the clinical effect of the DSM – 5 will have on diagnosis. Review of data from the DSM – IV field trial found that 40% of children who would be diagnosed as ASD under that version of the DSM, would not qualify for the ASD label under the DSM–5 (McPartland, Reichow, & Volkmar, 2012). Other analyses using the CDC database found that 81% of children with ASD based on the DSM – IV would still be labeled with ASD on the DSM – 5 (Maenner et al., 2014). The majority of those children who would not have ASD under the DSM – 5 would fall under the previous high functioning autism umbrella of AsP or PDD (Maenner et al., 2014; Smith, Reichow, & Volkmar, 2015). The revised DSM – 5 also created social (pragmatic) communication disorder (SCD) for children who have social communication impairments but no restrictive or repetitive behaviors (APA, 2013). It did not appear that all the children who were no longer classified with ASD on the DSM – 5 would qualify as SCD (McPartland et al., 2012; Smith et al., 2015). Further complicating the issue of diagnosis is that there are no appropriate assessments for SCD (Norbury, 2014). The lack of a universal theory for autism is also a problem for clinicians and researchers.

Many of the theories of ASD have only been tested in terms of the nonverbal interpersonal behavior of pragmatics; they have not been evaluated in relation to linguistic impairments of children with ASD (Noens & van Berckelaer-Onnes, 2005; Williams & Bowler,
These theories and the plethora of studies on general pragmatic deficits tell clinicians very little about verbal language deficits in ASD. For example, eye contact has been extensively investigated as a pragmatic factor; absence of eye contact and visual joint attention are signs of autism (Gillespie-Lynch, Elias, Escudero, Hutman, & Johnson, 2013; Jones, Carr, & Klin, 2008; Klin, Jones, Schultz, Volkmar, & Cohen, 2002). It is clear however, that intervention focused solely on eye contact does not remediate all language impairment in ASD. Information on the connection between specific pragmatic deficits and other language domains would provide clinicians with suggested diagnostic and treatment tools to facilitate the communicative needs of people with ASD. A more useful method to examining language in ASD is to take a functionalist approach to language instead of the traditional three domain view of language (e.g., Bloom & Lahey, 1978; Traugott, 2011). In a functionalist model, language form (e.g., syntax) and content (e.g., semantics) is embedded within pragmatics (Van Valin, 2003).

The primary purpose of this project was to examine the relation between semantic (vocabulary and literal meanings) and pragmatic (actual use of words and non-literal meanings) language and cognitive skills in comprehension of scalar sentences by 7-10 year-old children with and without autism in two experiments of varying complexity. The words used in this project were familiar to these children. Whether or not semantic-lexical development in ASD is similar to that of typically developing children (TD) remains debatable. The connections among early communication deficits, eye gaze behaviors, deficits in social communication, and later language acquisition also are unclear. Scalar implicatures, a type of conversational inference, are ideal for studying the breakdown between semantics and pragmatics in children with ASD. Scalars have literal semantic and implied (i.e., pragmatic) semantic interpretations (Grice, 1975; Hirschberg, 1985). Scalars, some-all, are terms that can be represented as opposite members on a
SCALAR COMPREHENSION

continuum (Horn, 1992). Numerals can also be considered scalars because number words (e.g., *two, three*) can also be understood with both literal and implied semantic interpretations (Barner, & Bachrach, 2010).

A secondary goal of this research was to use eye tracking to qualify lexical processing in ASD. The current study also addressed conflicting information about how scalar inferences are processed (Grodner, Klein, Carbary, & Tanenhaus 2010; Huang & Snedeker, 2009a). Eye tracking has been used to examine competition across referent pictures within an array following the processing of an auditory or visual prime in TD adults and children (e.g., Huang & Snedeker, 2009a, 2010). This method is underutilized with children and adults with ASD for language-based tasks although its worth as a processing tool is well established in the literature. The current study evaluated the relation between eye tracking and other behavioral measures during the comprehension of scalar sentences to understand inference-making and to evaluate theories of ASD.

**Pragmatics-Semantics in Autism Spectrum Disorders**

All children with ASD have deficits in pragmatic language ability; the persistence of pragmatic impairment is well–documented but rarely studied (Booth & Happé, 2010; Eigsti et al., 2011; Gerenser, 2009). Specifically, few studies have examined the relations between pragmatics and other language components, such as semantics, in ASD. There are also few studies within those other language components.

Generally, children with ASD differ pragmatically from TD children because they do not look at a speaker to confirm instructions, attend to objects a speaker is manipulating, nor do they point to desired objects or follow the direction of a speaker’s gesture (McDuffie, Yoder, & Stone, 2006; Yoder, 2006). Absence of these behaviors by early childhood is correlated with
poorer vocabulary later in childhood (Yoder, 2006). Pragmatics is correlated with semantics in TD children (Gillespie-Lynch et al., 2013; Mody & Belliveau, 2013). Paradoxically, many school-age children with ASD have typical vocabulary scores (e.g., McGregor et al., 2012). This dichotomy between semantics and pragmatics is especially common in what some researchers and clinicians refer to as high functioning autism (HFA). These children have average nonverbal IQ; may have a history of language delay but currently have average scores on standardized language tests (Sharma et al., 2012). Yet, they continue to struggle with certain semantic classes such as the mental state verbs *think, guess, and know* (Kelly, Paul, Fein, & Naigles, 2006). This dichotomy may be an artifact of the diagnostic problems concerning the HFA label. It also may be a result of current assessments lacking sensitivity for semantics. The Peabody Picture Vocabulary Test (PPVT - 4: Dunn & Dunn, 2007) is a popular assessment intended to measure single-word receptive vocabulary. In practice, it is limited in its ability to estimate lexical knowledge because recognizing word is not the same as being able to appropriately use that word.

People with ASD do not appear to use social or semantic comprehension strategies to learn, retain, or process novel words and sentences (Gaigg, Gardiner, & Bowler, 2008; Norbury, Griffiths, & Nation, 2010). There may be a specific impairment in ASD at the semantic-pragmatic level, particularly in children with HFA, but not at the semantic-literal level as demonstrated by the development of large vocabularies with persistent impairments in using those words (Kelly et al., 2006). If there were deficits in lexical organization, it could be expected that children with ASD should be unable to list words with similar meanings. However, they do not have problems with standardized synonym tasks (Joseph, McGrath, & Tager-Flusberg, 2005), but remain challenged by contextual word use. There are also cases where
children with ASD use a word repeatedly during conversation and then ask for a definition of the word. This also suggests a separation between word use and word comprehension (Perkins, Dobinson, Boucher, Bol, & Bloom, 2006). Thus, adequate or better semantic skill is not sufficient to comprehend or produce every-day conversation.

Leinonen & Kerbel (1999) examined the spontaneous and structured discourse of three children with pragmatic language impairment. The children could not answer examiner’s wh-questions likely because they required the child to make inferences from a visual scene. Although it is unclear if these children had ASD, social/pragmatic communication disorder (SCD), or other language impairment, the documented difficulties in interpretation may be generalizable to other 6-13 year olds with similar pragmatic deficits. In a later study (Loukusa et al., 2007), children with ASD (7;0 - 12;0) did not correctly answer as many contextually-demanding questions as age-matched controls. This study also provided the insight that children with pragmatic impairment had greater difficulty than their TD peers explaining why their accurate responses were correct without adding irrelevant information (Loukusa et al., 2007). This suggests that children with pragmatic language impairment are unable to anticipate the needs of a listener. These studies are important in showing the linguistic problems related to pragmatic language impairment in ASD. More information is needed to understand the cause of these problems.

Data from language processing tasks suggest that lexical development in ASD is atypical. Children with ASD have persistent impairment in category induction tasks and do not demonstrate robust semantic priming effects (Kamio, Robins, Kelley, Swainson, & Fein, 2007; Naigles, Kelley, Troyb, & Fein, 2013). These results are corroborated by imaging and neurological studies comparing children with and without autism on semantic tasks (Ribeiro,
Valasek, Minati, & Boggio, 2013; Verly et al., 2014). Eye tracking is an underutilized processing tool in studying lexical and pragmatic semantics in ASD. In conjunction with language forms that have both literal and implied meanings, eye tracking can show how children with ASD use their word knowledge.

In summary, children with ASD appear to have impairments in pragmatics and semantics. The nature of these impairments appear to vary across the spectrum but difficulty in social communication remains one of the hallmarks of the disorder. There are signs indicating that several factors play a role in pragmatic language impairment in ASD and that one factor is semantic ability. The evidence suggests that word recognition and retrieval abilities should be separated from conversational requirements on the use of words. Specific linguistic structures, like scalars, that can examine the overlap between semantics and pragmatics are not well studied in these children. Learning more about how the semantic-pragmatic interface works in ASD could eventually lead to possible treatment strategies to improve social functioning through enriched word knowledge.

**Implicature Theory**

Scalar implicatures are inferences made by listeners based on conversational context and the presence of a word that can be represented on an informational scale (e.g., *some* – *all*, Grice, 1975; Hirschberg, 1985; Horn, 1992). On a quantifier scale, *some* is the term at the weak end and *all* is at the opposite, strong, end of the scale (Figure 1). The weak scalar has two construals: The lower-bounded interpretation for *some* is *some and possibly all*. The upper-bound interpretation for *some* is *some though not all*. Conversely, the strong scalar has only one possible interpretation. Scalar implicature occurs when a listener infers that a speaker intends for the upper-bounded interpretation of a weak scalar to be inferred by this audience. This has been
extensively studied in TD children and in adults. It is used to measure the interaction between, and the development of, pragmatics and semantics. There are few studies about scalar implicatures in children with language impairment. By studying scalar implicature, as viewed as a component of pragmatics, we can more directly examine the relation between the impaired social interaction and lexical knowledge rather than the typical and uninformative vocabulary scores in children with high functioning autism (HFA). The ability to make inferences that limit the interpretation of weak lexical items by appealing to stronger scalar alternatives is important in language use.

There are three prominent theories (Sauerland, 2012) on how scalar implicature works in language processing. For the current research, the pragmatic theory (also known as the Gricean approach) (Grice, 1975; Horn, 1979) is considered predominant. The pragmatic theory of scalar implicature is one of the oldest and is based on Grice’s conversational maxims. Under this framework, scalar implicatures occur because in competent communication a listener hears what a speaker says and make inferences about that explicit content based on what the speaker does not say (Sauerland, 2012). In the sentence, *Some students came to the dance*, the scalar term *some* triggers the computation of the implicature, *Not all the students came to the dance*. There are several steps in the process of deriving the implicature. First, the listener recognizes that *some* is the scalar alternative to *all*. The lexical content of the word *some* in the sentence informs the listener that a portion of the students, up to and including all of the students, went to the dance. Second, the listener activates alternative sentences. Assessing what was unsaid, the listener can assume that the speaker did not produce the alternative sentence, *All of the students came to the dance*, because it is false (Grice’s first maxim of quality: Grice, 1975). Deriving the
implicature, *Not all the students came to the dance*, strengthens the meaning of the weaker scalar. Thus, the listener concludes, *Some but not all students came to the dance.*

Figure 1 displays the implied and literal interpretations available from two scalar sentences used in the current studies. If the listener derives the implicature upon hearing the word *some* in the sentence *Some of the carrots are inside of the box*, he will choose the image showing only three carrots inside of the box. Pragmatic interpretation of an utterance requires the listener to derive this implicature. Implicatures are not made for the stronger scalar term (e.g., *all*) because the lexical content of the stronger term includes the weaker. The listener assumes a speaker would not supply more information than is relevant (Grice’s first maxim of quantity: Grice, 1975).

Figure 1. Semantic and pragmatic interpretations of two images using sentences with the scalar terms *some* and *all*

In the case of the disjunction, *or*, listeners derive an exclusivity implicature. For the sentence, *Boys or girls came to the dance*, a listener considers three alternatives: *Boys came to the dance, Girls came to the dance,* and *Boys and girls came to the dance.* The first two alternatives are uncertainty implicatures because the use of *or* implies that the speaker may not be sure that either boys or girls came to the dance (Sauerland, 2012). When it appears that the speaker is sure, then the listener computes the exclusivity implicature, *Both boys and girls did not come to the dance*, based on Grice’s maxims.
Number words also can be understood with literal and implied semantic interpretations (See Figure 2). Under this view, on the lower end of the scale is an *at least* construal whereas on the upper end of the scale is an *exact* construal. Some linguists view number labels as different from other quantifiers because by nature they describe exact sets, whereas terms like *some* are not based on specific quantities (Horn, 1992). Most linguists view numerals as part of the same word class as scalar quantifiers, because numerical expressions can be represented on a number line, which is a scale (Barner, & Bachrach, 2010; Panizza, Chierchia, & Clifton, 2009). A numeral like a quantifier has an upper and lower-bounded interpretation. Although there is some disagreement, most theorists (e.g., Panizza, Chierchia, & Clifton, 2009) think that the basic interpretation of a numeral is the same as for quantifiers; the upper-bound interpretation must be derived through implicature. Interpreting numerals similarly to scalars, listeners use implicature to strengthen the interpretation of a weaker number by contrasting it with a stronger alternative. In the sentence, *Three students went to the dance*, the number 3 can mean that 3 or more (*at least* literal) students went or that exactly 3 (implied) students went.

---

**Figure 2.** Semantic and pragmatic interpretations of two images using sentences with numerals *three* and *five*

Relevance theory (Wilson & Sperber, 2004), a theory of inferential pragmatics, suggests that the process of determining the relevance of input is comparative and based on cognitive
effect and cognitive effort. Although not a model of scalar implicature, relevance theory may explain the general cognitive processing involved in making implicatures under the pragmatic theory (Chemla, & Singh, 2014). Cognitive effect is related to Grice’s first maxim of quality (informativeness). It is based on the listener’s expectations of the speaker’s intention that is determined by previously communicated information or by informativeness level (Wilson & Sperber, 2004). Cognitive effect is the communicative value a listener places on computing an implicature; anticipating enriched meaning leads to the creation of implicatures. Social and grammatical contexts affect cognitive effect. Cognitive effort (Wilson & Sperber, 2004) describes the processing costs or executive functions (e.g., cognitive flexibility, working memory, inhibition; metacognitive skill) associated with deriving an implicature.

Relevance theory applied to scalar implicature suggests that inferences are only derived when cognitive effect is high and the listener has the necessary capacity and management of the required cognitive effort. The components of executive functions most applicable to scalar implicature are not stated in any of that literature. Cognitive effort might include cognitive flexibility, working memory, interference, and inhibition. Our understanding of executive functions in typically developing populations is incomplete (Miyake & Friedman, 2012) so informing the role of cognitive effort in scalar implicature is also lacking. There are data supporting and refuting the Gricean idea that deriving implicatures causes processing costs. Although the model proposes differences in processing time, scalars such as some are not examined in terms of semantic networks and activation of semantic neighbors. It is unclear what happens in terms of semantic organization when implicatures are generated.

There is evidence that deriving implicatures demands greater processing costs for healthy adult participants. These data support the pragmatic view that implicatures are derived. French-
speaking adults took longer to complete a true/false reading task with the scalar *some* when they rejected the un-strengthened basic lexical interpretation of a statement (Bott & Novek, 2004). The assumption is that the additional reading time is caused by making the implicatures. Electrophysiological evidence from ERPs with adults also supports the theory that implicature computation occurs late in the cognitive semantic process (Noveck & Posada, 2003). This suggests that implied semantic interpretations are not defaults within the semantics of weak scalars.

However, relevance theory applied to the lexical theory of scalar implicature suggests that the implied semantic interpretation is not cognitively demanding to derive. According to this approach, competent listeners do not have to compute implicatures every time they hear a scalar; instead, the implied meaning is stored in the lexicon (Chierchia, 2004; Sauerland, 2012). Storing meanings of implicatures, as opposed to deriving them as in the pragmatics theory described above, could reduce cognitive effort. The implicature or strengthened meaning of the scalar is already part of the semantics of the weaker scalar. Thus, in lexical theories of implicature semantics is the highest module of processing. The implicature is theorized to occur in a bottom-up process through the identification of the strongest scalar alternative (Chierchia, 2004). When providing stimuli controlled to isolate semantic and pragmatic contributions Politzer-Ahles et al. (2013) found that some people have difficulty suppressing implicatures when the weaker literal semantic interpretation would be more correct. An eye tracking study also found that when context is clear, there were no delays to deriving implicatures (Grodner et al., 2010). These data suggest that implicatures are automatic; scalar implicature occurs early during sentence processing.
Implicature in Typically Developing Children

**Scalars.** Although typically developing children have many well-developed pragmatic language skills, they also appear to have difficulty deriving scalar implicatures (Papafragou & Musolino, 2003). Classification and task effects play a large role in study outcomes. There also are no studies to date that have examined the relation between children’s cognitive and language ability and their performance on scalar implicature comprehension tasks. These relations may reveal why some children can compute implicatures whereas same-aged and older children cannot.

The problem with using age as a classification variable in research is that age can introduce uncontrolled variables such as language ability. Development plays a strong role on response consistency, or the tendency for some subjects to always choose the literal semantic or implied semantic response to a statement, than any other factors (e.g., Chevallier et al., 2010; Katsos & Bishop, 2011; Papafragou & Musolino, 2003). Only a few studies (e.g., Katsos & Bishop, 2011) have systematically examined the variables of cognitive effort and cognitive effect in typically developing children. Implied semantic responses increased for tasks with lower cognitive effort requirements (Katsos & Bishop, 2011).

Children may perform more poorly than adults because of high metacognitive demands of the task (Guasti et al., 2007). The more pragmatic or felicitous statement does not always require the listener to derive an implicature. Children perform better when alternatives are salient and when distinctions between correctness and felicity are made more apparent within the experiment (Gualmini et al., 2001; Huang & Snedeker, 2009a; Papafragou & Musolino, 2003).

In the typical truth value judgment task (TVJT) of scalar comprehension, children watch a scene (e.g., *All of the children go to a dance*) with a puppet. At the conclusion of the action, the
The puppet makes a statement about the scene. The children are asked whether or not the puppet made an accurate statement (e.g., *Did he answer well?* or *Is the puppet correct?*). The child must consider several things before answering. The child needs to recall what he or she saw and heard. Then the child needs to determine whether the scene matched the puppet’s utterance as remembered. Semantically, two interpretations (e.g., *Some of the students went to the dance* and *All of the students went to the dance*) are correct. But if the child correctly remembers the situation and the puppet’s utterance, she should reject the first description accepting the second as *correct* because it is more felicitous. There are several limitations of the truth value judgment task primarily because it may tap metacognitive ability while measuring metalinguistic skills.

Relevance theory may help explain the problems with the truth value judgment task. According to Relevance theory (Wilson & Sperber, 2004), listeners also weigh cognitive effect (relevance) against cognitive effort (processing cost). Data suggest that TVJT confounds cognitive effect with a possible component of cognitive effort, cognitive flexibility - knowing (or accessing the knowledge) that *some* is a scalar alternative to *all* (Kastos & Bishop, 2011). Barner and Bishop (2011) used the TVJT to demonstrate that 4 – 5 year old TD children could use visual contexts to answer literal questions but could not use those same images to make scalar implicatures to inferential answer questions. The authors concluded that children could not make inferences because they did not know the alternative to the scalar term used in the question.

A second critique of the TVJT is that it tests informativeness of the statement instead of the ability to accept scalar implicatures. Adults may perform well on TVJT because they learn the experimenters’ expectations. An alternative to TVJT, the felicity judgment task (FJT), was more successful in capturing children’s knowledge because it reduced cognitive effort (Processing Limitations hypothesis: Gualmini et al., 2001). By having a different puppet utter
each of the statements, the child is given the two alternatives and avoids the extra processing described above. A critique of both TVJT and FJT are that they are binary and make children choose between two under-informative options (Barner, Brooks, & Bale, 2011; Kastos & Bishop, 2011).

If children are unable to derive scalar implicatures, one way to decrease both cognitive effort and cognitive effect would be to grammatically strengthen statements with the weak scalar term. Using the words only and every with a weak scalar (e.g., or, some) should obviate the process of deriving implicatures (e.g., Barner et al., 2011; Gualmini et al., 2001). Only some dogs and cats are sleeping should be easier (i.e., faster and more accurate) to solve than Some dogs and cats are sleeping. Only in the first statement acts as a grammatical marker, which indicates that it is not necessary to derive the implicature because not all of the dogs and cats are sleeping. Without the grammatical marker present, the listener must derive an implicature. However Barner et al. (2011) found that reducing cognitive effort by clarifying alternatives was more effective than adding the grammatical strengthener in increasing implicatures by 5 year old T children. There are no data examining only some in older children. The current study is the first to directly compare performances on the truth value judgment task and the felicity judgment task in the same group of children and in children with communication disorders. The timing between the speakers and the presentation of the visual stimuli were also manipulated to assess the effect of working memory on comprehension during each task.

**Numbers.** Comparison between scalar implicatures derived for numbers to scalar implicatures derived for other quantifiers can provide insight to the representation of the words in the child’s lexicon. Children can count when they encounter a situation where they need to select items from an array. Eventually children need to learn that the semantic relation between
SCALAR COMPREHENSION

generals, that *three* is one more than *two*, is the basis for counting. Typically developing children start to go through the stages of numerical development after 2 years of age (Brooks, Audet, & Barner, 2013).

If the numbers *two* and *three* were placed on a scale, *two* would be the weaker, lower element whereas *three* would be the stronger upper element of the scale. On this scale, *two* would include *exactly two* up to and including *at most two*. When asked if a display contains *two* items, children readily make the implicature that excludes a display with *three* items although by definition, the three element display includes two items (Papafragou & Musolino, 2003). The implicatures derived by children for numerals but not for other scalars suggests to some researchers that even young children use pragmatics to interpret words (Barner, & Bachrach, 2010; Sullivan, & Barner, 2011). However, for unclear reasons, there are differences between the processing of numerals and other scalars. Children processed implicatures for numbers faster than for quantifiers (Huang, & Snedeker, 2009a). Children also made implicatures for numerals but not for *some* (Papafragou & Musolino, 2003).

**Implicature in Autism Spectrum Disorders**

To date, there are no studies that examine the production or comprehension of scalar implicatures in children with ASD, although there are a few studies with adolescents and adults. More studies examine counting than the lexical-pragmatic function of number words in children with autism. The combined results across these studies are difficult to interpret because of the heterogeneity of age, non-verbal IQ (NVIQ), and verbal IQ (VIQ) across the participants.

When asked to judge the appropriateness of puppet’s responses to questions based on Grice’s Maxims of Conversation (Grice, 1975), 5 out of 8 children with ASD performed below chance (Surian, Baron-Cohen, & van der Lely, 1996). However, the subjects in that study had
very low verbal mental ages (5;7 years, on average) as calculated from the British Picture Vocabulary Test and there was no age-matched control group. In contrast, adults with high functioning autism (HFA), performed as well as age-gender-VIQ- matched controls on a reading-based true/false informativeness judgment task (Pijnacker et al., 2009). There was a significant increase in reaction time when weak scalars (or, some) were interpreted pragmatically compared to semantic interpretations. For example, answering true to the statement, Some sparrows are birds, is a literal semantic interpretation and therefore the sentence would be read faster than if the respondent answered, false.

Children with ASD counted faster than their IQ-matched controls but unlike TD children, their counting speed was not assisted by grouping like elements (Gagnon, Mottron, Bherer, & Joanette, 2004; Jarrold, & Russell, 1997). Children with ASD were also slower to name lower counts suggesting that unlike TD children, they did not subitize, or immediately recognize these small quantities (Gagnon et al., 2004).

Relevance theory, previously mentioned to describe the process of implicature, can also explain pragmatic impairments. By combining new information from the speaker, world knowledge, and context (Leinonen & Kerbel, 1999), meaning is generated by listeners. Children with ASD may have social communication difficulties because they do not have access to or cannot integrate these types of information (Happé, 1993). Based on this reasoning, scalar comprehension would be reduced in ASD compared to chronological and IQ-matched peers. On the other hand, if scalar implicatures are stored in the lexicon as suggested by Chierchia (2014), children with high-functioning autism may comprehend scalars because they have strong receptive vocabularies (Kelly et al., 2006).
There are three often-discussed theories of autism (Noens & van Berckelaer-Onnes, 2005): theory of mind (ToM) (Baron-Cohen, Leslie, & Frith, 1985), weak central coherence (WCC) (Booth & Happé, 2010), and executive function (EF) (Pennington & Ozonoff, 1996). These theories do not explicitly connect to linguistic pragmatics or semantics, but act as a useful scaffold to further develop a theory of autism that can fully explain all the behaviors in the disorder, including language deficits. These theories of autism may also suggest that scalar comprehension would be impaired in autism.

Theory of mind (ToM) is the ability to infer mental states or perspectives of others (Baron-Cohen, 1989; Baron-Cohen et al., 1985). This could cause the impaired social understanding and difficulty participating in conversations experienced by most children with autism. Children with autism also seem to lack the natural preference for social cues that typically developing children have that enables language and social learning (Happé, 1993; Mundy, Gwaltney, & Henderson, & 2010). ToM tasks are often criticized because they seem to measure more cognitive processes than ToM.

Although the ToM view of autism does not make predictions about scalar comprehension, it can be related to the pragmatic theory of scalar implicature. Based on Grice’s Conversational Maxims (Grice, 1975), assuming that the speaker is cooperative and is being informative, the listener must attribute mental states to predict and understand the speaker’s behavior (Surian et al., 1996). This area is predicted to be delayed or deficient in children with ASD (Perkins et al., 2006). If children with autism lack or have impaired ToM, they would have difficulty deriving scalar implicatures. Their general deficit in understanding what a speaker intends to convey through an utterance would cause a decrease in cognitive effect. This lack of
sensitivity to cognitive effect would manifest in more literal lexical responding, particularly in cognitively demanding two-talker truth value judgment tasks.

Weak central coherence (WCC) theory implies that the underlying deficit in autism is a weak understanding of global contexts (Booth & Happé, 2010). WCC theory is based on reports that people with ASD demonstrate relative strengths in local contexts (i.e., detail) and weakness in global contexts (i.e., the gist or big picture) where global contexts include conversational speech (Booth & Happé, 2010; Happé & Frith, 2006). Children with ASD may be pragmatic responders in tasks where cognitive effect does not have a significant impact (Happé, 1993). WCC would predict that the natural local processing in children with ASD would be sufficient to successfully complete these tasks because there is no advantage in holistic processing. In addition, enhanced local bias favors canceling an implicature when a grammatical strengthener (e.g., only) is added to the weaker scalar term.

The executive function (EF) theory of autism attributes impairment in the disorder to atypical cognitive processing (Ozonoff, South, & Provencal, 2005; Pennington & Ozonoff, 1996). Although the relation is unclear, theory of mind and EF may rely on shared processes (Hill, 2004). Like ToM, executive functions are general and largely undefined (Noens & van Berckelaer-Onnes, 2005). The relation between EFs, task performance, and daily behavior is unclear in both ASD and TD (Friedman et al., 2006; Hill, 2004; Miyake & Friedman, 2012; Ozonoff et al., 2004). The type or sub-type of EFs assessed by specific tasks are not always explicit. Proponents of an EF cause of autism do not consistently state which EF components are impaired so the relation between EFs and autism are unclear (Poljac & Bekkering, 2012). Confusion in the ASD EFs literature is also due to the high heterogeneity of the population.
Researchers often do not fully describe their participants’ language, comorbidity, and symptoms related to ASD although these factors may vary drastically between samples.

Inhibition has been one EF documented as impaired in some children with ASD (Robinson, Goddard, Dritschel, Wisley, & Howlin, 2009). Recent data suggests that even children with ASD who improve in aspects of language still have relative deficits in some EFs (Troyb et al., 2014). Children with high functioning autism (HFA) and children who no longer meet the DSM criteria for ASD diagnoses, still had impaired inhibition (Troyb et al., 2014). Findings like these are supported by neurological differences between adolescents with and without ASD on inhibition tasks (Vara et al., 2014).

Inhibition is not the only executive function implicated in ASD. There is also evidence for poorer than average cognitive flexibility, or the ability to adapt to changing situations, in HFA (Semrud-Clikeman, Fine, & Bledsoe, 2014). Lopez, Lincoln, Ozonoff, and Lai (2005) correlated cognitive flexibility with autism symptoms (restricted and repetitive behaviors). However, the task used in their EFs battery were the same ones criticized in the general EF literature (Jurado & Roselli, 2007). There are different types of inhibition performance profiles across subgroups in children with ASD. Distracter inference and proactive interference are two types of inhibition that have been found impaired in some children with ASD (Adams & Jarrold, 2012). Similar issues plague working memory studies with apparent differences between verbal and visuo-spatial working memory in children with ASD can vary with task and individual ability of children on the spectrum.

Some evidence also shows that children with HFA have typical executive functions skills (e.g., automatic response inhibition, cognitive flexibility) (Adams & Jarrold, 2012; Hill, 2004; Robinson et al., 2009). In addition, many children with ASD perform better on non-verbal tasks
compared to verbal tasks, suggesting dissociation between non-linguistic and linguistic processes in those children. Unlike in children without ASD, non-verbal IQ does not correlate with the type and extent of language ability (Joseph et al., 2005; Rapin & Dunn, 2003). However, because EFs are poorly understood many of the standard tasks used to assess them may be invalid (Geurts, Corbett, & Solomon, 2009; Jurado & Rosselli, 2007; Miyake & Friedman, 2012). Some tasks tap more than one EF so it is difficult to determine which component is impaired when a child performs poorly (Poljac & Bekkering, 2012). Components of EF do not develop at the same pace in TD children (Best, Miller, & Naglieri, 2011) which adds to the difficulty in assessing EFs in children with ASD.

The executive function (EF) account of autism would predict that implicatures are not formed because children with ASD cannot allocate the cognitive resources necessary for the extended processing of pragmatic content (e.g., Joseph et al., 2005). Executive function impairment in ASD suggests difficulty processing pragmatic-semantic information. Thus, any cognitive effort (i.e., executive function skills) required to derive implicatures will be too high a demand regardless of the cognitive effect indicated by the social or grammatical context of a statement. Some of the previous research in scalar comprehension in adults has studied processing time between implied and semantic interpretations (e.g., Grodner et al., 2010), but none have examined how task performance relate to general EF ability. The current study examined EFs in children with a nonverbal online task of attention, interference, and response inhibition.

**Eye Tracking**

The current study addressed the lack of information about how scalar implicatures are processed by including an eye tracking task. There are no eye tracking studies in adults or
children with ASD on scalar implicature tasks. The current study used a visual design to examine the time course of scalar implicature in children with and without ASD.

Eye tracking at its most basic, is any method used to measure eye movements in relation to visual or auditory stimuli. It is a language research method that can inform theories of ASD because it captures information about language processing. The principle behind eye tracking is that eye movements are guided by attention so they can provide a measurement of the time course of processing (Altmann & Kamide, 2007; Trueswell & Gleitman, 2004). When a person looks at an array of stimuli and concurrently hears stimuli, his attention, and therefore his eye movements, to the image move based on or in anticipation to auditory stimuli (Altmann & Kamide, 2007). Advanced eye tracking technology, as illustrated in Figure 3, increases synchronicity of the stimuli and movements. The Tobii eye tracker also provides more precise measurements than offline methods of eye tracking, such as video coding. Basic dependent measures from this method include the looking time to each image or word and the number and proportion of fixations to each image or word (Barr, 2008; Mirman, Dixon, & Magnuson, 2008). Fixations are looks of a researcher-determined minimum length of time.

Figure 3. Schematic design of eye tracker

Eye tracking can reveal categorical competition and resolution over time across a picture array following an auditory or visual prime in adults and in TD children (Huang & Snedeker,
2011; Huang & Snedeker, 2009a,b). Eye tracking research in ASD has focused on non-linguistic abilities, such as face processing; children and adolescents with ASD process faces differently than age- and IQ-matched TD peers (Karatekin, 2007; Klin et al., 2002; Jones et al., 2008). In one of the few linguistic eye tracking studies in ASD, there were more looks to a phonological distracter in subjects with and without autism who did not have language impairment than in children with language difficulties (Brock et al., 2008). Diehl, Friedberg, Paul, & Snedeker’s (2014) eye tracking study found no difference between children with ASD and their peers in their ability to use prosody to understand ambiguous sentences.

There are few studies that have examined scalar implicature using eye tracking. Panizza et al. (2009) provided insight as to how scalar implicatures were created in Italian adults during an online reading task. Readers fixated longer on the noun prior to a quantifier or number word in a sentence when they were making an implicature than in contexts where the need to infer was canceled. This longer fixation time in addition to slower reading time support the theory that the implied semantic or pragmatic interpretations of scalars have higher processing costs than the literal semantic interpretation. Another eye tracking study in adults revealed no processing delays associated with the implied versus the literal interpretation of some (Grodner et al., 2010). However, this study did not use a reading design, instead presenting a visual array to adult English speakers. As a result, the overall proportions of fixations were different from Panizza and colleagues’ (2009) study. Grodner et al. (2010) calculated accuracy based on eye gaze and for their proportions of fixation, they did not include the two distracter images. Panizza et al. (2009) did not have distracter images because they used a reading design.

Grodner et al. (2010) directly compared their findings to an earlier paper by Huang and Snedeker (2009a). In that earlier study, the eye gaze of adults did not disambiguate between the
lexical and pragmatic images for 800 ms after the onset of the weak scalar some when they made implied semantic responses. This was much later than when responses were literal (Huang, & Snedeker, 2009a). The designs of these two studies were similar, but the differences in methods may be the reason for the disparate results. It is unclear which of these differences had the larger impact. The later study used a head-mounted eye tracker (Grodner et al., 2010) whereas the earlier study used a camera surrounded by 4 images (Huang, & Snedeker, 2009a). Another issue with comparing these eye tracking studies is that neither research group defined the length of a fixation. In fact, Huang and Snedeker (2009b) did not use fixations as their major dependent variable, and relied on looking time. These are different metrics and may provide different information about the processing method used by the participants in these studies. A third methodological difference, which is expressed by Grodner et al. (2010), is related to theories of scalar inference. Grodner et al. (2010) argued that using the construction quantifier of, opposed to quantifier, as used by Huang and Snedeker (2009b) encouraged delays for weak scalars relative to strong scalars because the word of forces the listener to disambiguate the phrase later.

Huang and Snedeker (2009b) extended their research to 5;0-9;0 year old typically developing children. To date, this is the only eye tracking study of scalar implicature in children. The children, unlike the adults, consistently accepted the upper-bounded interpretation of the weak scalar some. They also took longer than adults to disambiguate between the target and scalar distracter (Huang & Snedeker, 2009b).

Summary

Children with ASD have persistent impairments in pragmatic language (APA, 2013; Eigsti et al., 2011). These deficits extend across the autism spectrum (APA, 2013) yet other language domains have not been well studied. Some children with ASD seem to have impaired
executive function abilities (Ozonoff et al., 2005) and semantic organization (Norbury et al., 2010) even with intact receptive vocabulary (Kelly et al., 2006). However it is unclear how the pragmatics related to semantics in ASD even though we know that pragmatics and semantics are correlated in TD children (Gillespie-Lynch et al., 2013). Scalar implicature (Grice, 1975) has been used to measure inferencing in TD children as a way to link linguistic pragmatics and semantic development. Weak scalars (e.g., *some, or*) have two possible meanings: literal and implied. Implicatures are made when the listener accepts the implied meaning (Grice, 1975). When scalar alternatives are salient, that is, when both scalar terms are provided, and other task demands, such as working memory, are low, TD children are more likely to make implicatures (Barner et al., 2011). However the connection between executive functions and scalar implicature has not been systematically explored in children or adults. There are currently no studies examining scalar comprehension in children with ASD.

**Aim of the present study.** The current study examined the role of cognitive effect and effort in the derivation of scalar implicatures by TD adults, TD children, and children with ASD. As described above, children with ASD, particularly those with high functioning autism (HFA), have persistent errors in using language in conversation even though they do not have impairments in receptive vocabulary. Scalar implicatures are ideal to study in children with ASD because scalars can be interpreted semantically (lexical knowledge) or pragmatically (implied-lexical knowledge). This information can be useful to clinicians in providing a link between semantics and pragmatics in ASD – two areas which seem dissociated in many children with high functioning autism.

This research sought to answer 3 questions:
1. What are the relations between cognitive effort and the derivation of scalar implicatures in children with ASD? What are the effects, if any, of executive functions (especially interference) on scalar comprehension tasks? What are the effects of manipulating temporal information on scalar comprehension? Experiment 1 compared accuracy of implicature derivation when a task demanded high (a. Truth value judgment task) or low (b. Felicity judgment task) cognitive effort.

2. What are the relations between cognitive effect and deriving scalar implicatures in children with ASD? What are the effects of manipulating visual information on scalar comprehension? Experiment 2a was designed, with even lower cognitive effort than Experiment 1b: Felicity judgment task, to determine whether increasing task informativeness and saliency of alternative interpretations facilitated implicature comprehension.

3. How does syntax affect scalar implications in children with ASD? Experiment 2b: Grammatical strengthening assessed if scalar implicatures were incorrectly derived when a weak scalar-term statement contains the grammatical strengthenener, only.

More children than adults were expected to derive scalar implicatures because scalar implicatures are late developing pragmatic skills. According to the executive functions (Ozonoff et al., 2005) or theory of mind (Baron-Cohen et al., 1985) views of ASD, and the relevance theory of pragmatics (Wilson & Sperber, 2004), children with ASD should be primarily lexical responders, choosing mostly under-informative items. Children with ASD would be less sensitive to cognitive effect than TD children because they have difficulty making inferences in other tasks (Leinonen & Kerbel, 1999; Loukusa et al., 2007). Increases to cognitive effort should negatively impact reaction time and accuracy mostly in the children with ASD. Children with ASD would also show greater difficulty applying additional visual information in the eye tracking experiments to make implied semantic responses.
All participants were expected to perform best (more pragmatic responses and faster reaction time) in the eye tracking tasks (Experiment 2). They should perform relatively worse on the two-talker truth value judgment task. The task requires high cognitive effort and should be especially difficult when the auditory stimulus precedes the appearance of the visual display. Performance should correlate with language ability on tests and questionnaires for subjects whose nonverbal IQ is within normal limits.

The eye tracking experiments should reveal within and between group differences in eye movement data, even if accuracy and reaction time do not. As the sentence is heard, eye movements should focus more on the target image and less on distracters. The continuous eye tracking data should show that children with ASD are slower to eliminate distracters in correct responses than the TD groups. Children with ASD will likely fixate toward the target and scalar alternative images than TD children.
General Method

Three experiments are reported. Experiment 1 was a within-subject examination of the yes/no truth value judgment task and the two-talker felicity judgment task. Experiment 2 was a 4 alternative forced-choice eye tracking task of scalar comprehension. During the third experiment subjects completed a nonverbal task of executive functions (vigilance, distracter interference, and response inhibition).

Participants

Speakers of American English were recruited from online advertisements and fliers posted in public areas, online volunteer forums, and schools in New York City. English was required to be the primary language used at home. Due to the lack of previous studies of scalar implicature in ASD, we took the conservative approach of selecting children slightly older than those in the TD literature, but still within the age range consistent with non-adult like performance in implicature tasks (e.g., Katsos et al., 2011).

The first control group consisted of 14 young adults (19;0 – 32;0 years: $M = 23$, $SD = 4.3$) with no self-report of hearing impairment or additional neurological conditions including ADD/ADHD. The adults all completed or were in the process of completing an undergraduate education. The second control group consisted of 10 children (7;1 – 9;11 years: $M = 8;0$, $SD = .77$). The clinical group contained 7 children (6;11 - 9;7 years: $M = 8;3$, $SD = 1.1$) with ASD based on clinical report provided by parents. Five of the children with ASD had ADD or ADHD. Table A3 illustrates the diagnostic categories for each of these children.

Four additional adults were excluded from the study because they reported language impairment, neurological conditions, performed less than -1 $SD$ on nonverbal IQ (NVIQ), or could not be calibrated on the eye tracker. One adult female was excluded from further analysis.
because her reaction times were more than twice the average for the group. One additional child was excluded from the TD group because there were technical difficulties and she did not return for a second session. Two of the children in the ASD group did not complete all the experimental tasks.

**Additional selection criteria for children and adults.** Adults completed a short questionnaire about their medical and language background to determine eligibility for participation. Parents completed a similar questionnaire for their children. All participants reported normal or correct-to-normal visual acuity and passed a hearing screening at 25dB HL (ASHA, 1997). Participants all scored within normal limits on the Test of Nonverbal Intelligence, Fourth Edition (TONI-4: Brown, Sherbenou & Johnsen, 2010). Typically developing children were not excluded if their scores were above average because the children with autism had high TONI scores.

**Additional selection criteria for children.** Demographic data for the children are displayed in Table 1 with additional information in Appendix A.

For the purposes of this study, children with formal diagnoses of high functioning autism, Asperger’s syndrome, and pervasive developmental disorder – not otherwise specified were considered the same. This grouping corresponds to the new DSM – 5 which has eliminated those sub-categories which were difficult to distinguish diagnostically (APA, 2013; Lord et al., 2012). Assignment to the ASD group was determined by clinical report of diagnosis supplied by the parent and Vineland Adaptive Behavior Scales, Second Edition (VABS-II: Sparrow, Cicchetti, & Balla, 2005) scores.

Parents filled out the *Parent/Caregiver Rating Form* of the VABS- II. This survey has strong concurrent validity with the ADOS (Klin et al., 2007; Saulnier & Klin, 2007) making it
ideal for use as an exclusionary tool for children recruited to the TD group who score less than -1 $SD$ on either the Communication or Socialization domain scores. All TD children scored at or above -1 $SD$.

All TD children participating in the research also completed the Comprehensive Evaluation of Language Fundamentals, Fourth Edition (CELF - 4: Semel, Wiig, & Secord, 2003). The purpose of this test was to provide information about the language development of all the children in the study and possibly to use in analysis of the language tasks. It was anticipated that some children with ASD would have language impairment which could affect study outcomes. Children with ASD were not excluded based on CELF scores so that the sample better reflected the actual ASD population. Some children were also given the Peabody Picture Vocabulary Test 4 (PPVT - 4: Dunn & Dunn, 2007) but scores did not differentiate between the TD and ASD group so this test was discontinued to save testing time and reduce participant fatigue. The TD children scored at least -1 $SD$ of the composite language scores on the CELF - 4 and at least -1 $SD$ of the standard score on the PPVT - 4. Children were not excluded if their scores were above 1 $SD$ on the CELF because TD children with high NVIQ are expected to have higher language scores. These boundaries are typical of research in distinguishing TD children from children with language impairment if they have average IQ (e.g., Katsos et al., 2011). Based on these criteria, three of the children in the ASD group had a language disorder. However, the CELF, like many of the omnibus language assessment tools, is not sensitive to pragmatic language impairment. Therefore, the language assessment likely lacked the ability to detect social language impairments at the discourse level (e.g., Levy et al., 2010).

For detailed demographic data about each child see Appendix A, Tables 1 and 2. The IQ score and VABS for one TD child, the composite CELF scores for a second TD child were
missing. One child with ASD did not have the necessary expressive language to complete the CELF.

Table 1.

*Average (SD) Characteristics of the Typically Developing and Autism Spectrum Groups*

<table>
<thead>
<tr>
<th>Group</th>
<th>Agea</th>
<th>Malesb</th>
<th>TONIc</th>
<th>CELFd</th>
<th>VABSf</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agea</td>
<td>Malesb</td>
<td>TONIc</td>
<td>CELFd</td>
<td>VABSf</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Recep.</td>
<td>Express.</td>
<td>Total</td>
</tr>
<tr>
<td>TD</td>
<td>8:0</td>
<td>5</td>
<td>113 (8)</td>
<td>115 (7.7)</td>
<td>114 (6.5)</td>
</tr>
<tr>
<td>ASD</td>
<td>8:3 (1.1)</td>
<td>7</td>
<td>118 (27)</td>
<td>100 (24)</td>
<td>99 (26)</td>
</tr>
</tbody>
</table>

*Note.* aAge in years, bnumber of males; cstandard scores*(standard deviation)* on the Test of Non-verbal Intelligence - 4 (Brown, Sherbenou & Johnsen, 2010); dcomposite standard scores on the Comprehensive Evaluation of Language Fundamentals, Fourth Edition (CELF - 4: Semel, Wiig, & Secord, 2003); eon the Vineland Adaptive Behavior Scales - II (VABS-II: Sparrow, Cicchetti, & Balla, 2005).

**Group comparisons.** Nonparametric statistics were used to compare the standard scores because variances were not similar between the groups and the number of subjects was less than 15 in the child groups. The average scores on the TONI were significantly higher for the children *(M = 115, SD = 18)* than for adults *(M = 99, SD = 22)*, Mann-Whitney-Wilcoxon test *U = 183, p = .002*. However, there was no difference in average IQ scores between the two groups of children, Mann-Whitney-Wilcoxon test *U = 26.5, p = .606*. Nonverbal IQ in the TD group ranged from 103 - 130 whereas the range was 93 - 158 in the ASD group.

One child with ASD did not have adequate expressive language to be tested using the CELF and one child in the TD group did not complete all the subtests necessary for computing the core language scores. There was also no group difference on the CELF for the 9 TD children.
and the 6 ASD children who had scores. The difference on the receptive subtest however approached significance, Mann-Whitney-Wilcoxon test $U = 9, p = .059$. The range of receptive scores in the TD group was 105 - 128 whereas it was 61 - 137 in the ASD group. The relative disadvantage in receptive language for children with ASD is consistent with previous findings (Mody & Belliveau, 2013). Also consistent, the average VABS scores for children with ASD were significantly lower than those of the TD children (Klin et al., 2007; Saulnier & Klin, 2007), most notably for the socialization sub-domain (66 - 97 ASD; 80 - 153 TD), Mann-Whitney-Wilcoxon test $U = 11, p = .003$). Figure 4 below illustrates that there was some overlap between children with and without ASD on the Vineland. One typically developing child, Subject 207, was rated slightly less than one standard deviation below the mean on the communication and socialization sub-domains.

*Figure 4. Communication by socialization domain scaled scores on the Vineland Adaptive Behavior Scales for individual typically developing children and children with autism. Note: Communication domain = x-axis; socialization = y-axis. Typically developing children subject numbers in blue; children with autism subject numbers in red.*
Stimuli for Experiments 1 and 2

The same images and similar sentences were used across experiments and conditions to minimize unfamiliarity effects. Pseudorandomization and the large number of trials was intended to prevent perseveration effects. In creating the sets no distinction was made between noun types (i.e., count vs. mass) because the validity of such distinctions is questionable (e.g., Barner & Snedeker, 2005). The adult version of the tasks consisted of two stimulus trial sets of 15 balanced for noun familiarity and imageability (Set 1, $M = 2.84$, $SD = .79$; Set 2, $M = 3$, $SD = .64$). The child version of the tasks used a total of 15 trial sets to reduce overall testing time. Each set was used to create two trial lists for each experiment. The target of each trial set in one list became the scalar alternative in the other list. The presentation of the lists were counterbalanced by subject. Trial sets were presented in pseudo-randomized order avoiding the same images on adjacent trials.

Each trial set contained 8 images depicting nouns (see Appendix B) although not all images were used in every condition in every experiment. The five component pictures in each image were about 1 x 1 inch in size. The images were black and white line drawings, either from or based on children’s picture database (Cycowicz et al., 1997; Snodgrass & Vanderwart, 1980), which were digitally edited in Photoshop. The images were re-sized to fit inside a shaded box drawn in the upper center of a 2 x 2 inch beige digital canvas (Figure 2).

The targets ($T$) were the pictures that matched the auditory stimuli. The scalar distracter ($SD$) image matched the hypothetical alternative sentence representing the other side of the scale for that condition. $T$ became $SD$ in alternative condition. For example, if $T$ illustrated all then the $SD$ had to show some. In the some/all conditions the images contained one pictured noun. For the other distracter images ($OD1$ and $OD2$) in Experiment 2 some/all condition the same numbers of
pictures were used but they were placed in different configurations. In the or/land conditions T contained two pictured nouns, the target and an unrelated noun. The other distracter images for the or condition in Experiment 2 in each set were OD2 and OD4. The other distracter images for the and condition in Experiment 2 in each set were OD1 and OD3.

Sets were also randomly assigned for control trials. Control trials used number (zero/one, two, three, and four/five) and location (inside of/outside of) as variables instead of scalar terms (e.g., Click on the picture five [Noun] are outside of the box). These trials enabled the distracter images be a correct choice (See Figure 5).

![Figure 5](image)

*Figure 5. Sample trial set for Experiments 1 and 2
Note. [Some/All of] the balls [or/and pineapples] are inside of the box. If the sentence was, Some of the balls are inside of the box, T = some; SD = all; OD1 = distracters."

All auditory stimuli were recorded at a 22 kHz sampling rate and normalized for amplitude. The recordings for Experiment 1 were produced by two female speakers of Mainstream American English in a sound-attenuated booth. These recordings included preambles (Here are [Noun]s, and [Noun]s (~2 s)) and target sentences (e.g., Some of the [Noun]s are inside of the box or Five [Noun]s are inside of the box). The practice trial prompts
[i.e., *Did she describe the picture well?* in the Truth value judgment task (TVJT); *Who said it better?* in the felicity judgment task (FJT)] were recorded by the speaker used for Experiment 2. She was a third native speaker of Mainstream American English. Her recordings also included the carrier phrase (*Click on the picture*) and target sentences. The three speakers were directed not to place any contrastive stress on the quantifiers in each sentence. Independently, two undergraduate research assistants listened to the recordings to verify stress and intelligibility. During the experiments the audio stimuli were presented at 60 dB HL in sound field from two speakers positioned directly below the infrared optics of the eye tracker.

**Procedure**

All data were collected at the Child Language Laboratory in the Department of Speech-Language-Hearing Sciences located at the Graduate Center, City University of New York (CUNY). Subjects completed standardized testing in a quiet clinical room. Experimental tasks (2 two-part scalar implicature experiments and 1 executive function experiment) were completed in a sound-attenuated booth. Children came to the laboratory for two or three sessions each lasting no more than 2 hours whereas many adults completed all activities during one 2-hour session. Participants received $20 (adults) or $30 (children) after each session. Parents also received a written score report with results of the CELF - 4 and PPVT - 4.

**In Session 1:** All subjects and/or their parents completed questionnaires, hearing screenings, NVIQ testing, and language testing, if appropriate. Before the language evaluation began the children were provided with a written and oral assent script that described the tasks in child-friendly terms and age-appropriate language. Subjects also completed the executive function task described below. **Session 2:** Some children needed time during this second session to complete standardized testing from Session 1. All participants, except one TD child who
needed a third session, completed the experimental tasks during Session 2. The following data were collected for all experiments: accuracy (percent correct) and reaction time (ms). Additional data were collected for Experiment 2 which used eye tracking technology. These included: total looking time (ms) to the visual target, proportion of looking time to the target and to each distracter, and sequence and time course of looks across each trial in the continuous gaze data.

**Apparatus**

Subjects were seated inside a sound-attenuated booth approximately 65 cm from the table-top Tobii TX300 eye-tracker (TET). The TET, attached below a widescreen (1920 x 1080 resolution) LCD computer monitor, measured and recorded eye movements during Experiment 2 (also see Figure 3). These movements were observed via binocular infrared bright-pupil corneal-reflection with video-oculography technology. Although the TX300 has a maximum resolution of 300 gaze points per second (Tobii Technology AB, 2011) for the purposes of this study the sampling rate of the eye tracker was set to 60 Hz. The TET software, Tobii Studio (Version 3.2, Tobii, Technology AB, 2012), collected a video recording of Experiment 2 from the computer running E-Prime via a video capture card allowing the experimenter to unobtrusively monitor the experiment from outside the booth.

Experiment 1 and the executive function task were presented on the same LCD monitor. The TET remained in place but the eye tracking capability was disabled. All experiments including the executive function task were presented through E-Prime (Version 2, Psychology Software Tools, Inc., 2011) running on a LAN-connected computer. This experimental control software created gazedata files with the eye tracking data as well as accuracy and response time statistics for each subject. The gazedata files were entered into a custom Matlab program to analyze the eye tracking data.
Data Reduction

Reaction time and accuracy. Correct and incorrect responses were analyzed separately. If the participant failed to make a response, the trial was marked as incorrect but a note was made in the E-Prime file so those responses could be separated from true incorrect responses. Accuracy (percentage correct) in the TD groups were expected to approach ceiling but comparisons across conditions and groups were planned. Non-parametric statistical testing was used for group comparisons because sample sizes were small and variances were sometimes unequal between groups. Testing was set at the .05 level. Data from individual participants were examined in addition to the group analyses because the sample sizes were small and variances were large.

Eye tracking. Data were cleaned in a custom Matlab program. Each data point, a gaze data sample collected by the eye tracker approximately every 18 ms, was analyzed for validity. Invalid points were excluded from further analysis by the program (e.g., McMurray et al., 2010). Invalid data could be one of two levels: sample and trial.

The Matlab program averaged the validity scores from the eye tracker for both eyes for each sample. Averaging gaze across both eyes helped to prevent excluding potentially acceptable data. Validity scores less than 3 were considered for further processing. Validity scores of 4 or higher indicated that the eye tracker collected no data, due to instrumental errors or participant movement. Therefore trials composed only of samples with average validity scores greater than 3 were excluded from analysis. Trials during which there were never a fixation also were excluded from further analysis. Fixation duration was set a priori in the Matlab program as at least 50 ms of continuous looking at an image. This duration was determined based upon previous eye tracking studies in the laboratory, the current paradigm, and the age of the participants.
Analyzes on the valid gaze data included both course and fine-grained summary measurements for each trial by participant and group (Barr, 2008). Course-grain measurements include total fixation time, total looking time, and the total number of fixations. Data also were collected based on eye movements to each of the four images: the proportions of fixation time, proportions of fixation, and proportions of looking time. The proportions of fixation time were computed by the Matlab program as the number of looks to an image type divided by the total number of fixations on a given trial for the total amount of time for which any of the images were fixated. Proportions of fixations without time, the number of fixations to an image type divided by the total number of fixations, can inflate fixation values because they do not account for when subjects may not be looking at the screen. Fine-grain measurements included the sequence of fixations (i.e., changes in fixated image).

The Matlab program also used the sequence of fixations to each image within those periods and proportions of fixation time to create time course graphs. Although 0 ms was the absolute start of the response period for a trial, the start of each phase was re-centered to zero for analysis.
Experiment 1: Truth Value Judgment Task and Felicity Judgment Task

The purpose of this experiment was to study how children with ASD understand scalar implicatures when given high (a. truth value judgment task: TVJT) and low (b. felicity judgment task: FJT) cognitive effort tasks. These binary tasks were modified for computer presentation by removing live-action voices, displays, and puppets. Two different female voices represented the voices of the puppets of traditional TVJT and FJT (see stimuli in general procedures above). The experiments were administered in random order. The scalar quantifiers, *some/all* and *or/and*, tested cognitive effect. Numerals which can also be represented as scalars were included in control trials.

Three stimulus onset asynchrony (SOA) conditions were used to manipulate cognitive effort in the TVJT and in the FJT. The auditory stimulus (i.e., target sentence) played before (-750 ms), during (0 ms), or after (+750 ms) the visual stimulus (i.e., target or scalar distracter) displayed. Appendix C illustrates the three SOA conditions in Experiment 1b (FJT). The figure is the same for Experiment 1a (TVJT) where the voice images, explained below, were replaced by smiley faces. Trial types were randomly assigned a SOA condition. Two lists were created for each experiment. Each list was counterbalanced by subject and the SOA conditions were presented in random order blocks. Within each block the trials were presented in sequential order to avoid the appearance of the same scalar or image set on consecutive trials. The adult version of each task had 90 trials whereas the child version had 45 trials. There were an equal number of trials for each SOA condition and scalar quantifier (i.e., *some/all; and/or*). Breaks were given every 30 (adults) or 15 (children) trials.

Response displays consisted of the target or scalar distracter image from a trial set with decision icons below it. A yellow smiley in the center of the screen appeared between each trial.
for 1000 ms. Participants had up to 5000 ms to respond via mouse click. This response window was determined by piloting the task and averaging the amount of time participants took to move the mouse and click on a choice. Reaction time (ms) was presented onscreen after practice trials. There was no feedback for experimental trials.

**Experiment 1a: Truth Value Judgement Task**

**Participants.** Data were collected from 14 adults, 10 TD children, and 5 children with ASD. All but one of the children with ASD had ADD/ADHD.

**Procedure.** After receiving verbal and onscreen instructions in the sound-attenuated both, subjects saw a sample trial and answered 3 practice trials. Following a preamble, designed to introduce the objects in the upcoming display (e.g., Here are [noun]s), one sentence was played for the subject. Response displays consisted of the target or scalar distracter image and yes/no response icons. The auditory stimulus either matched the image or described a not displayed scalar alternative. The sentence matched the image for half of all trials. Subjects had to click on the yes (green smiley) or no (red frown) icon to indicate that the utterance did or did not describe the image. During the practice, participants heard the prompt, *Did she describe the picture well?* A correct response was when the participant correctly accepted or rejected the utterance (Figure 6). These data were also analyzed by SOA condition and trial type. If the participant failed to make a response, the trial was marked as incorrect but a code in E-Prime file marked those responses so they could be separated from true incorrect responses.
Experiment 1b: Felicity Judgment Task

**Participants.** Data were collected from the same participants who completed Experiment 1a: 14 adults, 10 TD children, and 5 children with ASD.

**Procedure.** After receiving verbal and onscreen instructions in the sound-attenuated booth, subjects saw a sample trial and answered 3 practice trials. After the preamble subjects heard each speaker utter a sentence (~3 s each with a 1 s pause between utterances). Response displays (Figure 7) consisted of the target or scalar distracter image with black and white icons for each speaker below it. One sentence matched the image displayed and the other sentence matched the not displayed scalar alternative. Subjects had to identify who best described the image by clicking on the correct face icon (Figure 7). Each utterance was presented with the appropriate face icon so subjects did not need to be familiar with the speaker’s voices. During the practice trials, a prompt spoken by a third female voice asked, *Who said it better?* The matching voice was counterbalanced between speakers. The first speaker was the correct response for half of all trials. A correct response was one that matched the image.
Data were analyzed by stimulus onset asynchrony (SOA) condition and trial type. If the target was a strong scalar or number (e.g., *all, and, five*) the participant could get the item correct by selecting the speaker who produced the statement with that term (Figure 8). The interpretation of the response was that the participant understands the strong scalar to exclude the weaker. Getting the item incorrect by selecting the other speaker was interpreted to mean that the participant believed that the image could be represented by the weaker term. This was expected to be a rare occurrence. If the target was a weak scalar or number (e.g., *some, or, four*) participants could get the item correct by selecting the speaker who produced the statement with that term. Selecting the other speaker implied that the participant rejected that the strong scalar matched the displayed image.

<table>
<thead>
<tr>
<th>Sentences</th>
<th>Display</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some of the balls are …</td>
<td></td>
<td>Incorrect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Accepts implied SOME</td>
</tr>
<tr>
<td>All of the balls are…</td>
<td></td>
<td>Correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Accepts ALL</td>
</tr>
<tr>
<td>Some of the balls are …</td>
<td></td>
<td>Correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Accepts implied SOME</td>
</tr>
<tr>
<td>All of the balls are…</td>
<td></td>
<td>Incorrect</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Error</td>
</tr>
</tbody>
</table>

*Figure 8.* Chart of interpretation of responses for two-talker task Experiment 1b: Felicity judgment task.
Predictions

Experiment 1 was expected to provide specific evidence of linguistic processing differences—over any basic-level attentional differences found in the nonverbal EF task—between children with and without ASD. If attentional differences play a role in scalar comprehension then performance should be poorest in both groups when the stimulus onset asynchrony was negative (SOA = -750 ms). Given the nature of binary tasks, response bias in some participants with and without ASD could also occur. If children with ASD also have perceptual load and local processing advantages (Happé & Frith, 2006) then there should not be a decrease in accuracy as accuracy increases. It was possible that children with ASD would perform better in the felicity judgment task (FJT) than the truth value judgment task (TVJT), like younger TD children in previous studies (Barner et al., 2011). However, theory of mind (Baron-Cohen et al., 1985) suggested that there would not be any differences in performance between the TVJT and FJT because children with ASD are insensitive to cognitive effort.

Results

Adults. Performance was similar on both tasks of Experiment 1 (Table 2) although adults did slightly poorer and performed more slowly on the more demanding yes/no Experiment 1a: Truth value judgment task than on the two-speaker Experiment 1b: Felicity judgment task (Table 2). Adults performed equally well across trial types. Accuracy was affected by stimulus onset asynchrony (SOA) condition. Participants scored significantly lower when SOA = -750 ms than when SOA = +750 ms, as measured by the Wilcoxon signed rank test on the felicity judgment task ($W = 4.5, p < .01$, one-tailed) and on the truth value judgment task ($W = -91, p = .0008$, one-tailed). There was no difference in accuracy on the two-talker task when the target audio was heard first than when it was heard second. There was also no difference in accuracy in the yes/no
task when the target audio matched the image (yes response) than when it did not (no response).

Individuals performed similar to the group although there was some evidence of individual variability where one or two adults scored less than the average. No single adult subject performed below the group average in both tasks on Experiment 1.

Table 2.

Adult’s Experiment 1 Average Accuracy and Reaction Time (SD) by SOA

<table>
<thead>
<tr>
<th>SOA</th>
<th>Experiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TVJT</td>
</tr>
<tr>
<td>0 ms</td>
<td>85 (36)</td>
</tr>
<tr>
<td>-750 ms</td>
<td>82 (38)</td>
</tr>
<tr>
<td>+ 750 ms</td>
<td>90 (30)</td>
</tr>
</tbody>
</table>

Note: Accuracy [mean percent correct (SD)] and mean reaction time (ms) (SD) on Experiment 1 for stimulus onset asynchrony (SOA) conditions. TVJT: Truth value judgment task; FJT: Felicity judgment task

Overall, reaction times were longer for the -750 ms SOA condition whereas the other two conditions were similar to each other. Reaction time was less affected by trial type than SOA, particularly in the -750 ms condition (Figure 9). There was a large range for reaction time within subjects. This may be an artifact of mouse-click timing in E-Prime. Timing tests for these studies were consistent with the reported average mouse response delay of 70 ms (Psychological Software Tools, Inc., 2007). However the timing tests also showed there was a large inter- and intra-subject variability for mouse click responses. Because this timing delay is inconsistent the reaction time data for these experiments may be inaccurate.
Figure 9. Trial type by reaction time (ms) for Experiment 1 by stimulus onset asynchrony for adults. Left, Experiment 1a: Truth value judgment task; Right, Experiment 1b: Felicity Judgment task

Note: Error bars represent the SEM. Horizontal lines represent the overall mean reaction time for each experiment.
Table 3.

*Children’s Experiment 1 Average Accuracy and Reaction Time (SD) by SOA*

<table>
<thead>
<tr>
<th>SOA</th>
<th>TD (n = 10)</th>
<th>ASD (n = 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TVJT</td>
<td>FJT</td>
</tr>
<tr>
<td>0 ms</td>
<td>80 (40)*</td>
<td>1887</td>
</tr>
<tr>
<td></td>
<td>(1114)</td>
<td>(1143)</td>
</tr>
<tr>
<td>-750 ms</td>
<td>83 (38)</td>
<td>2656</td>
</tr>
<tr>
<td></td>
<td>(1156)</td>
<td>(1015)</td>
</tr>
<tr>
<td>+750 ms</td>
<td>75 (43)</td>
<td>1809</td>
</tr>
<tr>
<td></td>
<td>(1124)</td>
<td>(1235)</td>
</tr>
</tbody>
</table>

*Note:* Accuracy [mean percent correct (SD)] and mean reaction time (ms) (SD) on Experiment 1 for stimulus onset asynchrony (SOA) condition for typically developing children (TD) and children with autism (ASD). TVJT: Truth value judgment task; FJT: Felicity judgment task + n = 6, * n = 3

**Children overview.** Overall children in both groups were less accurate than adults and took longer to respond, as expected (Tables 2 and 3). Reaction time is reported in the tables for reference but as mentioned previously, there is reason to believe that equipment-induced random delay reduces the reliability of those data. Unlike the adults, there were no significant differences in accuracy between stimulus onset asynchronies for typically developing children or for children with ASD on either task (Wilcoxon signed rank, p > .05).

As predicted, accuracy was higher for the strong scalars, including numerals, than the weak ones. The overall accuracy of weak numerals was more similar to the strong quantifiers *all* and *and*. Overall, averages for the children were slightly higher in the two-talker task.
(Experiment 1b: Felicity judgment task) than in the yes/no task (Experiment 1a: Truth value judgment task). However, scores varied substantially within both groups of children. In addition, due to a programming error, the data for the 0 ms stimulus onset asynchrony (simultaneous SOA) condition for 2 TD and 2 children with ASD were not collected for Experiment 1a: Truth value judgment task. Some figures illustrating the group variability are presented after the group comparisons below.

**Comparisons between TD and children with ASD.** The ASD group had lower average accuracy than the TD group when the image preceded the sentence on the FJT or followed the sentence on the TVJT (Table 3). Unlike in the TD children there was no overall average increase in accuracy on the FJT for viewing the display prior to hearing the sentences (SOA = +750 ms, $M = 85\%, SD = 35\%$) compared to the baseline condition (SOA = 0 ms, $M = 85\%, SD = 35\%$). The only between group SOA difference to approach significance was on the truth value judgment task. TD children had higher accuracy than the children with ASD for the SOA = -750 ms condition, Mann-Whitney-Wilcoxon test, $U = 12, p = .063$.

The children with ASD as a group performed similarly to their typically developing peers on all trial types on the yes/no TVJT. The largest difference between the groups was for the weak scalar *or*, Mann-Whitney-Wilcoxon test, $U = 14, p = .0985$. For this trial type accuracy was higher when the correct response was *yes* than when the correct response was *no*, Wilcoxon signed rank $W = -33, p < .05$ for TD; $W = -15, p < .05$ for ASD one-way.

On the two-talker felicity judgment task TD children were more accurate when the target audio was not first ($M = 87\%, SD = 34\%$ vs $M = 84\%, SD = 38\%$) (Wilcoxon signed rank $W = 35, p = .0392$ one-tailed). There was a trend in the opposite direction for the ASD group. They were generally more accurate when the target audio was first ($M = 92\%, SD = 27\%$ vs. $M = 81\%$,
SD = 39%) (Wilcoxon signed rank W = -13, p’s > .05 one-tailed). These differences in accuracy for the order of target audio were not significant, Mann-Whitney-Wilcoxon test, U = 20, p > .05.

There were several differences in performance between the yes/no truth value judgment task and the two-talker felicity judgment task. The overall averages were greater for both groups on the two-talker task. This difference was significant only in the ASD group, Wilcoxon signed rank, W = 15, p < .05. The average differences were made apparent in both groups by looking at the data by trial type (Table 4) but these differences were not statistically significant due to the small sample sizes and high variability within the groups. When the target was the weak scalar some, TD children scored on average 88% (SD = 33%) on Experiment 1b (FJT) and 87% (SD = 33%) on Experiment 1a (TVJT). Children with ASD demonstrated a larger average difference but a higher standard deviation (M = 81%, SD = 40% on FJT; 72%, SD = 45% on TVJT).

The advantage for the two-talker felicity judgment task was most apparent when the word or was the target. Children were more accurate when the alternative to the disjunction was played (FJT: two-talker task) than when they had to determine if or described an image (TVJT: yes/no task), Wilcoxon signed rank, W = 31, p < .05 for TD; W = 15, p < .05 for ASD. Or was the only weak scalar for which children appeared more likely to accept a literal semantic interpretation but only on the TVJT. The ASD group (M = 23%, SD = 50%) had greater average difficulty with the or trial target type than the TD group (M = 53%, SD = 50%) for that condition although the scores overlap between groups (Table 4).
Table 4.

Children Scalar Accuracy (SD) and Reaction Time (SD) on Experiment 1

<table>
<thead>
<tr>
<th>Group</th>
<th>Scalar Accuracy</th>
<th>Range</th>
<th>Time</th>
<th>Accuracy</th>
<th>Range</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>TD</td>
<td>Some</td>
<td>87 (33)</td>
<td>54 – 100</td>
<td>2197 (463)</td>
<td>88 (33)</td>
<td>25 – 100</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>91 (28)</td>
<td>0 – 100</td>
<td>2153 (1077)</td>
<td>85 (36)</td>
<td>63 – 100</td>
</tr>
<tr>
<td></td>
<td>Or</td>
<td>53 (50)</td>
<td>0 – 100</td>
<td>2203 (459)</td>
<td>82 (40)</td>
<td>56 – 100</td>
</tr>
<tr>
<td></td>
<td>And</td>
<td>77 (42)</td>
<td>43 – 100</td>
<td>2052 (461)</td>
<td>81 (40)</td>
<td>33 – 100</td>
</tr>
<tr>
<td></td>
<td>WNu</td>
<td>85 (37)</td>
<td>60 – 100</td>
<td>1925 (1192)</td>
<td>89 (32)</td>
<td>67 – 100</td>
</tr>
<tr>
<td></td>
<td>StrNu</td>
<td>86 (35)</td>
<td>70 – 100</td>
<td>2158 (1250)</td>
<td>88 (33)</td>
<td>50 – 100</td>
</tr>
<tr>
<td>ASD</td>
<td>Some</td>
<td>72 (45)</td>
<td>34 – 100</td>
<td>1705 (531)</td>
<td>81 (40)</td>
<td>50 – 100</td>
</tr>
<tr>
<td></td>
<td>All</td>
<td>95 (23)</td>
<td>80 – 100</td>
<td>1785 (1296)</td>
<td>90 (30)</td>
<td>78 – 100</td>
</tr>
<tr>
<td></td>
<td>Or</td>
<td>23 (52)</td>
<td>0 – 67</td>
<td>1923 (291)</td>
<td>83 (38)</td>
<td>67 – 100</td>
</tr>
<tr>
<td></td>
<td>And</td>
<td>89 (32)</td>
<td>73 – 100</td>
<td>1522 (369)</td>
<td>86 (35)</td>
<td>60 – 100</td>
</tr>
<tr>
<td></td>
<td>WNu</td>
<td>71 (46)</td>
<td>50 – 100</td>
<td>1513 (1084)</td>
<td>87 (35)</td>
<td>44 – 100</td>
</tr>
<tr>
<td></td>
<td>StrNu</td>
<td>93 (26)</td>
<td>75 – 100</td>
<td>1841 (1060)</td>
<td>83 (38)</td>
<td>75 – 100</td>
</tr>
</tbody>
</table>

Note: Mean accuracy (SD) and range as percent correct and reaction time in ms (SD). WNu = weak numerals; StrNu = strong numerals

Comparisons between individual children. As shown in Figures 10 - 14 and Tables 3 and 4, accuracy varied within groups on Experiment 1. Figure 10 shows the percent correct for each child as a comparison between the two-talker felicity judgment task (FJT) and the yes/no truth value judgment task (TVJT). For the most part, children with and without ASD did not form individual clusters on either task and there were overlaps in performance between the two
groups. The overlapping appeared greatest in the strong scalar conditions (i.e., *all, and, strong numerals*) especially on the TVJT. The number of children within each group who scored at or near ceiling varied by trial type. For the overall score on the TVJT, 40% of the TD children and 0% of the children with ASD scored at least 85%. On the FJT, the percentage of high scorers were 40% in both groups. In all three of those conditions there was a least one TD child, usually TD Subject 208, who scored lower than children with ASD. The distributions for weak scalar (i.e., *some, or, weak numerals*) accuracy were more diffuse. There were also fewer typically developing children who scored less than children with ASD. The children with the lowest scores on Experiment 1 did not differ from the high performers in age or standardized scores (Tables 4 and 5). Individual data tables for each task are in Appendix D.

Figure 10 also demonstrates that words in the same class of scalars had unequal rates of comprehension by the same child; performance could vary with task. For example, TD Subject 204 was an overall high scorer but she had higher accuracy on the FJT for *and* and strong numerals targets than when the target was *all*. On the TVJT she performed best when the target was *all*. 
Figure 10. Accuracy scatterplots for typically developing children and children with ASD comparing Experiment 1 tasks for each trial type on Felicity and Truth Value Judgment Tasks. Note: Individual subject numbers for typically developing children = blue; for children with ASD = red. x-axis = FJT (felicity judgment task); y-axis = TVJT (truth value judgment task). Trial types from top left to bottom right: all, some, and, or, strong numerals, weak numerals. Data missing from two children in each group for the truth value judgment task.

**Responder types and task differences.** One of the goals of the research was to examine if children with ASD would be literal responders. All children tended to choose the implied semantic interpretation more than the literal interpretation for weak scalars. There were differences in performance based on task and trial type though.
Truth value judgment task. See Appendix D for individual accuracy scores. In the yes/no task (Experiment 1a: Truth value judgment task), one TD girl, Subject 212, was a literal responder (Figure 10 and Table 5). She accepted the literal semantic interpretation of the images 62% of the time. There were also two literal responders in the ASD group. One boy, Subject 301, accepted the literal semantic interpretations of the image 65% whereas the other, Subject 304, accepted them 60% of the time if the target was a weak scalar or number. Literal responses included clicking the smiley face to indicate a correct statement if the image showed all five apples were inside of the box and the speaker said some of the apples are inside of the box. Another literal response included clicking the smiley face to indicate a correct statement when an image that showed three apples inside of the box was paired with the speaker statement, All of the apples are inside of the box. When no response trials were separated from these children’s incorrect items, only Subject 304 remained a literal semantic responder (60% literal).

On Experiment 1a: Truth value judgment task, Subjects TD 212 (M = 47%) and ASD 304 (M = 62%) had the lowest overall accuracy scores for scalar trials (Figure 10). Accuracy for these children depended on trial type. TD Subject 212 scores ranged from 0% (or) to 70% (strong numerals) whereas ASD Subject 304 scored from 0% (or) to 100% (all) (Table 5). The number of children in the TD group who scored at or near ceiling ranged from 3 (or) to 8 (some and all). In the ASD group, there were 0 (or) to 4 (all, and, strong numerals) children who scored at least 85%.

Mean scores on the TVJT of the TD group may have been skewed by 3 participants: Subjects 208, 209, and 212. Unexpectedly, Subjects 208 and 212 scored very low in the all condition (M = 0%; M = 50%, respectively). Further inspection of those data revealed their scores were calculated from misses rather than incorrect responses. Other TD children also
scored below the group mean on scalars. Except for the and and or trial types these low average scores for TD children were above chance. Mean scores on the TVJT in the ASD group were also heavily influenced by individual performances, mostly by Subjects 302 (M = 69%) and 304 (M = 62%). Both subjects scored less than the group average in the and, or, and some conditions. Subject 304 also performed poorer than the other children in the ASD group in the two numeral conditions (Table 5 and Appendix D).

*Felicity judgment task.* Only one child, a male in the TD group, was a literal responder in Experiment 1b: Felicity judgment task (FJT). Subject 209 chose the speaker who produced the literal semantic interpretation of the image 54% of the time a weak scalar or numeral was the target. If the image showed that there were three apples inside and two outside the box, a literal responder would click on the speaker who said, *All of the apples are inside of the box.* Further examination of Subject 209’s data revealed that he did not respond at all to 5 of those trials. The rest of the TD participants were pragmatic responders that chose the implied semantic sentence 83% (range = 68% - 100%) of the time. If the image showed that there were three apples inside and two outside of the box, they were more likely to click on the speaker who said, *Some of the apples are inside of the box* than on the one who said, *All of the apples are inside of the box.* The ASD group chose the implied semantic sentence 76% (range = 54% - 96%) of weak target trials.

Three of the poor performing children (TD Subjects 208 and 212; ASD Subject 304) on the yes/no TVJT had higher mean accuracies on the two-talker FJT (Table 5). The two other low TVJT performers (TD Subject 209 and ASD Subject 302) had similar mean scores on the FJT (Table 5). Overall, two TD children (Subjects 208 and 209) and 1 ASD child (302) depressed group averages on the felicity judgment task but again, performance varied with trial type. These children’s overall scores averaged above chance except the some (TD Subject 208 and ASD
Subject 302) and weak numerals condition (ASD Subject 302). More children in both groups scored at or near ceiling for certain trial types on this task than on the yes/no task. In the TD group 4 (and, or) to 7 (some, strong numerals) out of the 10 children had at least 85% accuracy. The respective range in the ASD group was 2 (some, and) to 4 (all) out of the 5 children.
Table 5.

Demographic data for children with low accuracy scores on the TVJT and FJT

<table>
<thead>
<tr>
<th>Subject</th>
<th>TD 208</th>
<th>TD 209</th>
<th>TD 212</th>
<th>ASD 302</th>
<th>ASD 304</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age&lt;sup&gt;a&lt;/sup&gt;</td>
<td>7;1</td>
<td>7;5</td>
<td>7;9</td>
<td>8;0</td>
<td>9;4</td>
</tr>
<tr>
<td>TONI&lt;sup&gt;b&lt;/sup&gt;</td>
<td>113</td>
<td>108</td>
<td>112</td>
<td>106</td>
<td>93</td>
</tr>
<tr>
<td>CELF&lt;sup&gt;c&lt;/sup&gt;</td>
<td>106</td>
<td>102</td>
<td>120</td>
<td>66</td>
<td>108</td>
</tr>
<tr>
<td>Soc&lt;sup&gt;d&lt;/sup&gt;</td>
<td>122</td>
<td>98</td>
<td>91</td>
<td>66</td>
<td>97</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>TVJT</th>
<th>FJT</th>
<th>TVJT</th>
<th>FJT</th>
<th>TVJT</th>
<th>FJT</th>
<th>TVJT</th>
<th>FJT</th>
<th>TVJT</th>
<th>FJT</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0</td>
<td>63</td>
<td>100</td>
<td>88</td>
<td>50</td>
<td>100</td>
<td>88</td>
<td>100</td>
<td>88</td>
<td>100</td>
</tr>
<tr>
<td>Some</td>
<td>85</td>
<td>86</td>
<td>78</td>
<td>25</td>
<td>54</td>
<td>100</td>
<td>69</td>
<td>50</td>
<td>33</td>
<td>57</td>
</tr>
<tr>
<td>And</td>
<td>57</td>
<td>80</td>
<td>71</td>
<td>60</td>
<td>43</td>
<td>100</td>
<td>86</td>
<td>60</td>
<td>73</td>
<td>100</td>
</tr>
<tr>
<td>Or</td>
<td>38</td>
<td>63</td>
<td>50</td>
<td>57</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>71</td>
<td>25</td>
<td>88</td>
</tr>
<tr>
<td>Strong numeral</td>
<td>80</td>
<td>67</td>
<td>70</td>
<td>100</td>
<td>70</td>
<td>100</td>
<td>90</td>
<td>100</td>
<td>75</td>
<td>100</td>
</tr>
<tr>
<td>Weak numeral</td>
<td>80</td>
<td>67</td>
<td>60</td>
<td>80</td>
<td>60</td>
<td>91</td>
<td>100</td>
<td>44</td>
<td>57</td>
<td>73</td>
</tr>
</tbody>
</table>

*Note:* <sup>a</sup>Age in years, <sup>b</sup>standard scores on Test of Non-verbal Intelligence - 4 (TONI-IV: Brown, Sherbenou & Johnsen, 2010); <sup>c</sup>composite standard score on Comprehensive Evaluation of Language Fundamentals, Fourth Edition (CELF-4: Semel, Wiig, & Secord, 2003); <sup>d</sup>standard score on Socialization domain of the Vineland Adaptive Behavior Scales - II (VABS-II: Sparrow, Cicchetti, & Balla, 2005); TVJT (truth value judgment task) and FJT (felicity judgment task) accuracy as percent correct

**Stimulus onset asynchrony task differences and scalar types.** There was a lot of variation within individual children in the two groups when the trial type data is further explored.
by SOA condition. Figures 11 – 14 illustrate individual accuracy in Experiment 1 for each SOA condition.

Overall, most TD children were more accurate when the auditory and visual stimuli were presented simultaneously (SOA = 0 ms) (Figures 11 – 12). In addition, more scores were below the group averages on the more difficult truth value judgment task (TVJT) than on the two-talker felicity judgment task (FJT). Mean accuracies appeared to improve the most for the or condition. The figures also show that there were individual differences among the TD children. These children did not always perform their best when SOA = 0 ms. Of the poorest performers on the tasks, half were more accurate when the visual appeared before the auditory stimulus (SOA = +750 ms) (e.g., Subject 208). The other half was better when the auditory preceded the visual stimuli (SOA = -750 ms) (e.g., Subject 209). However, this relation varied with task. For example, Subject 208 had the highest average for some on the TVJT when SOA = +750 ms but on the FJT, her highest average for some was when SOA = -750 ms.

Similar reversals can be seen in Figure 14 for ASD Subject 304. The figures also indicate some different trends for children with ASD compared to their typically developing peers. ASD Subject 302 was consistent in performing better when SOA = +750 ms than when SOA = -750 ms. Overall, the simultaneous (0 ms) SOA yielded higher average accuracies on both tasks only for the strong scalars.
Figure 11. Accuracy of individual typically developing children for each trial type on the truth value judgment task and the felicity judgment task for the three stimulus onset asynchronies SOA: 0 ms, -750 ms, and +750 ms (SOA = 0 ms was missing for two subjects (202 and 203) on the truth value judgment task).
Figure 12. Accuracy of individual typically developing children for each trial type on the truth value judgment task and the felicity judgment task for the three stimulus onset asynchronies. Note. From top to bottom: all, some, and, or; SOA: 0 ms, -750 ms, and + 750 ms (SOA = 0 ms was missing for two subjects (202 and 203) on the truth value judgment task).
Figure 13. Accuracy of individual children with autism for each trial type on the truth value judgment task and the felicity judgment task for the three stimulus onset asynchronies. Note. From top to bottom: all, some, and, or; SOA: 0 ms, -750 ms, and + 750 ms (SOA = 0 ms was missing for two subjects (202 and 203) on the truth value judgment task).
Truth Value Judgment Task

Felicity Judgment Task

Strong Numbers

Weak Numbers

Figure 14. Accuracy of individual children with autism for each trial type on the truth value judgment task and the felicity judgment task for the three stimulus onset asynchronies. Note. From top to bottom: all, some, and, or; SOA: 0 ms, -750 ms, and + 750 ms (SOA = 0 ms was missing for two subjects (202 and 203) on the truth value judgment task).

Summary. Children with and without autism accepted implied semantic interpretations of weak scalars some and or and numerals in both the yes/no truth value judgment task and the two-talker felicity judgment task. In general accuracy was higher for some and the weak numerals than for or. Accuracy improved on the TVJT when the correct response was yes. On the FJT, the TD children performed better when the target sentence was played after the first whereas there was a trend in the opposite direction for children with ASD. More implicatures were also accepted in the two-talker task compared to the yes/no task. This improvement was significant only the ASD group. There were few differences between children with and without
autism but there was much individual variation within each group. The demographic data
collected on each participant did not relate to performance on these tasks but given the individual
variation and small groups, it is unclear which variables can contribute to high or low
performances.

**Discussion**

These data support the critiques of the yes/no truth value judgment task. According to
relevancy theory (Wilson & Sperber, 2004), implicatures should be derived when cognitive
effect is high and cognitive effort is low. The pragmatic view of scalar implicature (Grice, 1975)
also suggests that making implied semantic interpretations is more cognitively demanding than
making literal semantic interpretations.

One of the critiques of the yes/no truth value judgment task is that it is metacognitive and
is more a test of cognitive effort than comprehension (Barner et al., 2011; Katsos & Bishop,
2011). The two-talker task (Experiment 1b: Felicity judgment task) was thought to reduce
cognitive effort because the alternative interpretations were salient. The results from this study
suggest that TD and children with ASD benefit from hearing the alternative interpretations of the
images and use that information in determining whether or not a statement is truthful. Even
adults performed best in the two-talker task. These differences in performance suggest that weak
and strong scalars are processed differently. The data support the pragmatic theory that
inferences are inherently more difficult to make than literal interpretations (Grice, 1975)

Adults also performed better on the two-talker task than on the yes/no task. There would
be no reason to assume that adults would need alternatives made salient for them the way
children might require (Katsos & Bishop, 2011). Another explanation for this task difference
might be that the computerized version of the tasks used in the current study did not adequately
replicate the cognitive demands of the traditional live-speech tasks (Papafragou & Musolino, 2003). Live-speech may add cognitive demands (e.g., attention, natural variation in speech) to a task. It is also worth noting that participants were more likely to agree to an uninformative statement (yes) than to disagree (no). Katsos & Bishop (2011) used multidimensional scaling to show that both adults and children hesitate to reject under-informative statements because they are somewhat correct.

Surprisingly, there were no strong differences between the TD and ASD children. The results in this study are similar to one with adults with high functioning autism who selected implied semantic choices as often as their TD peers (Pijnacker et al., 2009). In fact, only one child with ASD was one of the overall bottom three performers. This child also had language impairment. Adult scalar implicature studies found a correlation between implied semantic responses and VIQ only in the ASD groups (Chevalier et al., 2010; Pijnacker et al., 2009).

Children with impaired theory of mind, the ability to understand a speaker’s intention, and the pragmatic skill could be assisted by the presentation of both the target and alternative statement. When the target was and, TD and ASD children agreed that a statement using or was correct. That is, if the image displayed in the yes/no task was two balls and two apples inside of a box and the statement was Balls or apples are inside of the box, the children chose yes. In the two-talker task, the children were more likely to click on the icon of the person who said, Balls and apples are inside of the box than on the person who said, Balls or apples are inside of the box. It is also possible that the tasks themselves were not sensitive enough to show weakened pragmatic skills in ASD children. Computerized tasks reduce communicative demands in children with ASD (Ozonoff, 1995). Relatedly, if children with autism have weak central
coherence (Booth & Happe, 2010), they would perform well on computerized tasks that lack a strong need to understand more than the trial at hand.

The current study does illustrate processing differences between TD and children with ASD. Although the group differences were not significant, there was a trend for children with ASD being less able to capitalize in conditions which were easier for their typically developing peers. Children with ASD also did not improve accuracy in the positive stimulus onset asynchrony period. This suggests ASD participants did not use previewed visual information to make judgments later. This may be another example of weak central coherence (Booth & Happe, 2010). The children with ASD only responded to local information so that knowledge about the task gained from previous trials may not have transferred to future trials.
Experiment 2: Four Alternative Forced-Choice Eye Tracking Task

A criticism of traditional binary tasks is that children may prefer one option for unspecified reasons (Katsos & Bishop, 2011). Therefore, the purpose of this experiment was to study how children understand the principle of informativeness and how children are affected by syntax in a 4 alternative, forced-choice task. Informative trials used the stronger scalar value and under-informative trials used the weaker value. Critical trials were further divided into some/all trials and or/and trials; all - and are the stronger terms.

Prior to each eye tracking experiment, participants were individually calibrated using a package call or script that communicated between E-Prime and the eye tracker. This calibration performed at the start of testing affirmed that the eye tracker could identify the location of both eyes when the subject looked at the screen. The calibration program presented a series of 5 moving circles subjects visually followed while minimizing head movement. The software alerted the examiner if there was not enough data to create an accurate calibration profile for the participant. When calibration profiles were invalid, participants were re-calibrated to ensure the highest level of success in obtaining valid gaze data. Following this 1 - 2 minute procedure, the participants were introduced to the experiment with verbal and onscreen instructions.

Experiment 2a

This study examined whether children with ASD respond to informativeness (cognitive effect) when presented with logical scales (e.g., some/all; or/and).

Participants. Data were collected from 14 adults, 10 TD children, and 6 children with ASD. The number of participants included for the eye tracking portion of the report, however, depended on eye tracking data validity.
Procedure. The adult version of the task had 90 trials whereas the child version had 40 trials. There were similar numbers of trials for each trial type (i.e., some/all, and/or, and numeral control). Breaks were given after every 30 trials for adults and after 15 trials for children. After calibrating to the eye tracker and receiving directions, subjects answered 3 practice trials. In the practice trials, the correct response was highlighted in red. Following a response (or the end of the practice trial if there was no response), participants saw a gaze replay demonstrating the eye tracking technology.

A trial began with a 1000 ms fixation item (i.e., yellow smiley face) followed by an auditory preamble. The pictures in the horizontal display (see Figure 15) appeared in one of four random positions. The critical images were the target ($T$) and the scalar distracter ($SD$). The target sentences (e.g., *Click on the picture [some] of the [$T$]s are inside of the box*) were played 1000 ms after the images appeared - timing consistent with scalar experiments and linguistic eye tracking studies in ASD (Brock et al., 2008; Chevallier et al., 2010). This permitted an examination of gaze patterns before and after the delivery of the auditory stimulus. Subjects responded by clicking on the picture that matched the sentence. Accurate responses were images that matched the sentence. Participants had up to 5000 ms after the stimulus sentence, determined through piloting, to respond. This response time provided enough time for multiple saccades and fixations and the motor movement for the mouse (Huang & Snedeker, 2009b).
Each response period was divided into four phases based on the timing of the target sentence: Prompt phase, quantifier phase, predicate phase, and trial end. The prompt phase contained the carrier phrase, *Click on the picture*, and lasted from the beginning of the response period to the onset of the quantifier phase (~1000 ms). The next phase containing, for example, *some/all of the balls*, continued until approximately 2050 ms of the period. The third phase (predicate phase) began at the end of the quantifier phase and ended around 3400 ms into the response period. Most mouse click responses occurred towards the end of the predicate phase. Some delay was introduced to the reaction time because of the use of a mouse for responses (Psychological Software Tools, Inc., 2007). Trial end represents the time between the end of the predicate phase and the end of the trial. This final phase was not present in all files. It was designed to catch any post-response eye movements and varies with response time and the 60 Hz sampling rate of the eye tracker. Because eye tracking is a dynamic process, trials could be rejected for analysis in Matlab for one phase but not for another across the response period. Therefore, it was possible for a participant not to have data for every phase, as each period was analyzed individually.

**Experiment 2b**

The purpose of Experiment 2b was to study how children with ASD respond to syntactic cues in sentences containing scalars.
Participants. Data were collected from the same participants from Experiment 2a.

Procedure. Experiment 2b directly followed administration of Experiment 2a and used the same trial procedure and displays with two differences: The grammatical strengthener only was added to or and some target sentences to decrease the cognitive effect of deriving an implicature. The adult version of the task had 90 trials whereas the child version had 27 trials. There were approximately an equal number of trials for each weak quantifier (i.e., some, or, and numeral control). Breaks were given approximately every 30 or 16 trials, depending on the age of the participant.

Predictions

Looks and fixations to the target as the sentence unfolds should increase while looks and fixations to the other distracters markedly decrease. Children with ASD should take longer to eliminate the scalar distracter from their decision-making process than either comparison group, evidenced by fewer fixations on the target. Alternatively, the ASD group would be quicker to eliminate distracters because they generally neglect related competitors, consistent with weak central coherence theory. The ASD group may also have faster reaction times to correct responses than TD children, as TD children will likely look at all images before responding. In this case, executive functions (particularly interference) scores would correlate to reaction time.

Reaction times will likely be faster in Experiment 2b than Experiment 2a for only some statements if children respond to the grammatical strengthener. Reaction time will correlate with language scores and nonverbal IQ. Children with ASD who have typical language scores should have more similar accuracy to TD children if standardized language scores correspond to the ability to complete scalar comprehension tasks. All children with ASD may perform poorer than TD children on Experiment 2b (grammatical strengthening) because adding the syntactic cue
SCALAR COMPREHENSION

makes the task more complex linguistically. Children with ASD may not have knowledge of the
cue or have poor executive functions (high cognitive effort) to process and apply *only*.

Experiment 2 may be more difficult for the group because there are more images to look at, with
interference and inhibition potentially playing a role in making correct responses. Theory of
mind deficits in autism would also suggest an inability to relate *only* to the images or the task.

**Results**

**Reaction time and accuracy.** Fourteen adults, 10 TD children, and 6 ASD children
completed Experiment 2: Four alternative forced-choice eye tracking task. Compared to
Experiment 1 (truth value judgment and felicity judgment tasks), responses in Experiment 2 were
faster overall. There was no advantage in reaction time or accuracy for Experiment 2b:
Grammatical strengthening compared to 2a for adults. Adults performed at ceiling on these tasks.
Incorrect responses for adults were mostly attributed to a failure to respond.

![Figure 16](image)

*Figure 16. Individual children's (typically developing, left; with autism, right) accuracy on each trial type for the four-alternative forced-choice eye tracking task*

In children, incorrect responses were due to non-response when the child failed to click
any image (9% of all trials for TD and 6% for ASD) or because he selected the scalar alternative.
When the target was a numeral, accuracy was high, ranging from 70 - 100% correct across
groups. The targets *four* and *five* accounted for the number trials with poorer accuracy. As the
case in Experiment 1b: Felicity judgment task, accuracy was the poorest when the target was *or* (Figure 16 and Table 6). Forty-four percent of the TD children and 80% of the ASD children incorrectly matched *or* with the target image. ASD Subject 601 scored the poorest on the task overall. His averages were also the lowest in the ASD group, with one exception. ASD Subject 304 averaged more incorrect items where *and* was the target. Overall accuracy was increased in Experiment 2b: Grammatical strengthening for TD children but not for children with ASD. Accuracy for *only some* was less than that for *some* in the ASD group.

### Table 6.

*Children’s Mean Accuracy (SD) and Ranges in Experiment 2*

<table>
<thead>
<tr>
<th>Target</th>
<th>TD (n = 10)</th>
<th>ASD (n = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Range</td>
</tr>
<tr>
<td>Some</td>
<td>93 (13)</td>
<td>67 – 100</td>
</tr>
<tr>
<td>All</td>
<td>96 (8)</td>
<td>78 – 100</td>
</tr>
<tr>
<td>Only Some</td>
<td>88 (23)</td>
<td>33 – 100</td>
</tr>
<tr>
<td>Or</td>
<td>35 (41)</td>
<td>0 – 100</td>
</tr>
<tr>
<td>And</td>
<td>96 (7)</td>
<td>83 – 100</td>
</tr>
</tbody>
</table>

On a given trial, the mouse click response occurred from any point after the onset of the target sentence. Participants were asked to delay their response until the sentence was completed, though some people (mostly adults and children with autism) did not wait. The average reaction time was 4647 ms (SD = 852 ms) for TD children and 4278 ms (SD = 469 ms) for children with ASD. All children responded as quickly when the strong scalars were the target as when the weak scalars were the target. However, reaction time was recorded from mouse clicks, resulting in some random instrumentation delay.
Eye tracking. More details regarding the Matlab program used to analyze eye tracking data can be found in the data reduction section under the general procedure of this paper. A custom Matlab program filtered the eye tracking data by rejecting trials where there were no fixations (defined as at least 50 ms of looking at one image), or samples that had poor validity. The number of trials used to compile the average eye tracking data varied within each subject due to variability of response accuracy, sample validity, number of fixations, and period duration.

Matlab rejected many of the children’s trials in the critical response period for analysis, mainly because of poor validity in eye tracking (Table 7). In Experiment 2a, 13% of the ASD trials and 41% of the trials from the TD group were rejected. In Experiment 2b: Grammatical strengthening, 25% of ASD trials and 68% of TD trials were eliminated. Data for Experiment 2a were corrupted for one TD and one ASD subject. Most of the adult data were retained. Trials that were answered correctly were then processed separately from incorrect trials.

The proportions of fixation times to the target image were larger for adults than for children (Figure 17). When subjects clicked on the target image, they significantly fixated more on the target than the other images. Children took longer to reject the alternative image (scalar distracter) than adults, continuing to fixate relatively equally on both the target and alternative image during the quantifier phase of the trial. However, among the children, those with ASD fixated more on the alternative than the TD group during that part of the response period (Figure 17).
Figure 17. Proportions of fixation time to the target and scalar distracter (Alt) images across the four response phases for each group in Experiment 2.

Note: Error bars indicate the standard error of the mean. Horizontal line indicates chance accuracy (.25)

Matlab created line graphs showing the proportions of fixation times across each analysis period. During the preview phase (the 1000 ms visual preview period where participants saw the answer choices prior to the start of the auditory stimuli), all participants looked relatively equally at each of the four areas of interest (Figure 18). The average proportions of looking time and fixation time to each picture type were equivalent across groups. Participants appeared to look at one image before moving on to another. In figure 18, this appears as roughly stable horizontal proportion lines in the TD groups and proportion lines with changing slopes in the ASD group. Children with ASD was the only group to exhibit greater than .30 proportions of fixation time on an image. The difference between the ASD and TD groups, however, appears to be in the sequence of looks. Adults and TD children appear to stop fixating on the other distracter image.
control 2 by 400 ms into the period. Children with ASD do not show this behavior until 700 ms into this quantifier phase (Figure 19).

Figure 18. Time course of eye movements during visual preview in Experiment 2
Note: Top = adults; middle = TD children; bottom = ASD children
Figure 19. Children’s time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid correct all trials
Note: Top = TD children; bottom = ASD children
Eye tracking for strong scalars. There were processing differences between the strong scalars and between groups. For correct responses to *and* and *all* during the quantifier phase, the TD children continued the same pattern of fixation as during the preview period (Figures 18 and 19, top). They initially looked back and forth between the images. Figure 19 suggests that the children with ASD were quicker to respond to the target word than TD children for strong scalars *all* and *and*. The ASD children fixated more on the target image without any noticeable
competition in looking behavior between the target and the alternative. This seems to be atypical, not only because it is different from the TD children, but also because it is different from the adults, who share similar looking behaviors with the TD children. However, it is imperative to note the data for the ASD children is composed of fewer valid correct trials than for the TD children (Table 7).

The following sections describe the time course of processing for the various trial types. Estimates are provided in the text based on the figures indicated.

*Eye tracking for all.* Figure 19 also indicates that the ASD group fixated on the target and alternative approximately 100 ms later than the TD children. When the target was *all*, the TD children’s looking behavior demonstrated visual competition from about 200 ms to 1000 ms in the quantifier phase. At around 400 ms (as the children started to make eye movements in response to the auditory stimulus), there was competition among all four images. Between about 500 ms and 800 ms after the onset of the target word, there was a visual preference for the alternative. Competition between the target and alternative appeared approximately 800 – 1000 ms after the onset of the target *all*. Fixations increased to the alternative image while the proportions of fixation time to the other image types decreased near 1000 ms (Figure 19). In the ASD group, there was competition between the target and alternative for a shorter period than in the TD group – about 300 ms – 500 ms. Children with ASD fixated on the scalar distracter for longer proportions of fixation time than for the two control images, although all three were fixated on significantly less than the target after approximately 900 ms.
Figure 21. Children’s time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid correct *and* trials. 

*Note:* Top = TD children; bottom = ASD children

In the TD group, the average proportions of fixation time to the target image *all* peaked at .70 during the predicate phase near 200 ms (Figure 20). The ASD group proportions of fixation time peaked around 1200 ms in the predicate phase and was maintained until the end of the trial (Figure 20). This peak occurred much later in the ASD group than among their TD peers and adults.
Eye tracking for the target and. The offset in the auditory stimulus for the scalar and occurred at approximately 300 ms into the quantifier phase. Between 0 ms and about 400 ms, all images competed for attention in both groups (Figure 21). At around 500 ms into the quantifier phase (200 ms after the offset of the target and), the TD children fixated the most on a distracter (control 1) and the target images. After approximately 900 ms into the phase, the proportions of fixation time to the target were much greater than to the other images. The children with ASD demonstrated a longer period of fixating on all image types than the TD children (Figure 21). The ASD group fixated relatively equally on the target, alternative, and a distracter (control 1) between about 500 ms and 1000 ms into the quantifier phase. There was no competition between the target and alternative images. The target image did not become the most fixated image until nearly 1400 ms into the period (1100 ms after the offset of the target word and). In terms of looking behavior, the predicate phase for the target and was similar to that for the target all for both groups (Figure 19), with the exception the ASD children exhibited late competition between the target and a distracter (control 1) for approximately the first 200 ms of the phase.
Table 7.

Experiment 2 Children's Overall Gaze Data during Quantifier Phase and Predicate Phase

<table>
<thead>
<tr>
<th>Measures</th>
<th>Quantifier Phase</th>
<th>Predicate Phase</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TD</td>
<td>ASD</td>
</tr>
<tr>
<td>Average Percent Accuracy</td>
<td>80 (40)</td>
<td>65 (49)</td>
</tr>
<tr>
<td>Number of Trials</td>
<td>151</td>
<td>194</td>
</tr>
<tr>
<td>Number of Fixations</td>
<td>387</td>
<td>524</td>
</tr>
<tr>
<td>Proportions of Fixations (Target)</td>
<td>.30</td>
<td>.28</td>
</tr>
<tr>
<td>Proportions of Fixations (Alternative)</td>
<td>.25</td>
<td>.28</td>
</tr>
<tr>
<td>Proportions of Fixation Time (Target) a</td>
<td>.32 (.34)</td>
<td>.29 (.33)</td>
</tr>
<tr>
<td>Proportions of Fixation Time (Alternative) a</td>
<td>.26 (.32)</td>
<td>.31 (.33)</td>
</tr>
</tbody>
</table>

*Note.* a Group average (SD). Accuracy dependent on the number of trials retained by Matlab.

Gaze data includes valid correct and incorrect trials.

Eye tracking for weak scalars. Overall, the proportions of fixation times to the weak scalars showed that participants gave more attention to the alternative image than to the target image (Figures 22 - 24). When the target was *some*, the average proportion of fixation times was similar to that of the strong scalars. Closer examination of the weak scalars revealed processing differences between *or* and *some*. The time course for sentences with the word *or* was different from all the other scalar types.
Figure 22. Time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid correct *some* trials

*Note:* TD on top; ASD on bottom

The time course for *some* compared to the one for *all* suggests that the TD children started their fixations to the target image faster for the weak scalar (Figures 19 and 22, top).

However, differentiation of the proportions of fixation time to the target *some* did not begin until around 1200 ms into the quantifier period. This was approximately 200 ms later than the disambiguation point for *all* trials. The ASD group’s time course for *some* shows there was
competition between the target and alternative from around 650 ms to 1200 ms after the onset of the target word (Figure 22, bottom). This differed from the relatively faster disambiguation between target and scalar distracter for all trials (Figures 19 and 22, bottom). In both groups, the proportions of fixation time to the target some during the predicate phase were lower than during the same period for the strong scalar all.

Figures 23 and 24 show the time course of fixations for correct or trials. There were few correct trials for this weak scalar; as a result, the graphs were composed of fewer data points than for the other scalars (see Table 7). The offset of the target or occurred at approximately 400 ms into the quantifier phase. At around 200 ms, the TD children increased their fixations to one distracter (control 1) while fixating relatively equally to the target and alternative images (Figure 23). The ASD group’s fixations showed competition between the target and alternative at around 200 ms. TD children switched their attention to the target near 1000 ms, but also continued to fixate on the alternative and distracter control 1. In the ASD group, proportions of fixation time to the target increased at approximately 1100 ms in the quantifier phase, with competition between the target, alternative, and distracter control 1 resuming at close to 1500 ms.

Figure 24 shows by the time TD children heard the predicate of the sentence, they were overwhelmingly fixated on the target image, while ASD children continued to fixate on the same three images as during the end of the quantifier phase. ASD children fixated primarily on the target image about 1200 ms into the predicate phase. And although occurring later than the TD children, the proportions of fixation time to the target for ASD children were higher than those for TD children by the end of the trial.
Figure 23. Time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid correct or trials

Note: TD on top; ASD on bottom
Figure 24. Time course of eye movements to the four image types during the predicate phase of the response period in Experiment 2 for valid correct or trials

Note: TD on top; ASD on bottom

Eye tracking for only some. When the target was only some (Figures 25 - 26), there was competition between the target and alternative images for the TD group, occurring from the offset of the phrase to around 900 ms into the quantifier phase. The target image received the most proportions of fixations after about 1000 ms. This time course shares similarities to that of the some trials, when fixations to the target separated from the other images. During the
predicate phase, TD children primarily fixated on the *only some* target image. There were fewer correct trials in the ASD group for *only some* than for their TD peers (see Table 6). In the ASD group, the alternative image received the greatest proportion of fixations up to approximately 800 ms after the onset of *only some*, at which point the proportions of fixation time to the target increased and visually competed with the alternative image for the rest of the phase. Unlike in the *some* trials, the target did not receive the majority of fixations at the end of the quantifier phase for the ASD group. The disambiguation between target and other image types did not occur until around 200 ms into the predicate phase, approximately 1800 ms after the onset of the phrase *only some* (Figure 26).
Figure 25. Time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid correct only some trials.

Note: TD on top; ASD on bottom.
Figure 26. Time course of eye movements to the four image types during the predicate phase of the response period in Experiment 2 for valid correct *only some* trials

*Note:* TD on top; ASD on bottom

Eye tracking for incorrect responses. The eye tracking data from the few error trials show that when participants incorrectly gazed at an alternative after hearing the scalar *or*, they fixated later on the target image. When participants did look at the target in incorrect trials, fixations were relatively shorter than fixations to the target in correct trials. Once eye movements responded to the end of the target word, the greatest proportions of fixation time were associated with a distracter (*control I*) (TD group) or the alternative image (ASD group). The TD group initially exhibited greater competition between the alternative image and a distracter.
Competition for visual attention then occurred between those two image types and the target (Figure 27). TD children did not look primarily at the alternative until about 1400 ms after the onset of the scalar or. When TD children were incorrect, their average longest look was to the alternative image in the predicate phase. TD children continued to fixate on the alternative until about 200 ms in the predicate phase, at which time they began to fixate on distracter control 1. The group then did not return to fixating primarily on the alternative image until at the end of the trial (Figure 28).

As seen in Figure 27, the ASD group had greater proportions of fixation time to the alternative image than on the other images. Participants continued to focus on the alternative image until the end of the trial. However, like their TD peers, the ASD children demonstrated competition between the alternative and the distracter control 1 during the predicate phase (Figure 28). This competition did not resolve until the alternative became the most fixated image at around 1600 ms.
Figure 27. Time course of eye movements to the four image types during the quantifier phase of the response period in Experiment 2 for valid incorrect or trials

Note: TD on top; ASD on bottom
**Figure 28.** Time course of eye movements to the four image types during the predicate phase of the response period in Experiment 2 for valid incorrect *or* trials

*Note:* TD on top; ASD on bottom

*Individual eye tracking results in children.* The Matlab program was able to provide some insight into individual processing differences. Although time course plots cannot be created with individual data, the overall gaze data from each phase of the response period can be examined. Figures 29 and 30 illustrate the accuracy and visual fixation results for a high-performing child in each group. Both Subject 204 and Subject 304 were very accurate on the eye tracking task. The gaze data, shown as proportions of fixation time to the target image for each trial type, were
very variable within subject. As a result, some of the accuracy data appear to be at odds with the
gaze data. Subject 204 was 100% accurate in the scalar conditions yet her proportions of fixation
time to the target were not always greater than chance (e.g., target = *or*). The gaze data for these
children during the predicate phase show less variability as well as the expected larger
proportions of fixation time to the target image (Figure 30). However, in some cases, a data point
was created from as few as one trial. Although Subject 204 was correct in the single *only some*
trial retained by the Matlab program for analysis, the eye tracker did not collect any fixations to
the target during the predicate phase.

![Figure 29](image)

*Figure 29.* Quantifier phase proportions of fixation times to the target image and response accuracy
for a high performing child in each group.
*Note:* Subject 204 is TD and Subject 304 is ASD. Error bars indicate the standard error of the
mean. Horizontal line indicates chance accuracy (.25)
Figure 30. Predicate phase proportions of fixation times to the target image and response accuracy for a high performing child in each group.

Note: Subject 204 is TD and Subject 304 is ASD. Error bars indicate the standard error of the mean. Horizontal line indicates chance accuracy (.25)

The data for two low performing children were also examined (Figures 31 – 32). In the eye tracking task, both low performers were more variable in their eye gaze and in their mouse click responses than the high performers. Overall, the low performers were less accurate and fixated less on the target image than their high performing peers. Neither low performer did well when or was the target word (quantifier phase: \( M = 0\%, SD = 0\% \)), while barely fixating on the target image (Figures 31 - 32). The non-eye tracking data, however, showed better accuracy for the TD child. When the target was or in Experiment 2, the TD child correctly clicked on the target image (\( M = 83\%, SD = 33\% \)). The eye tracking data during the predicate phase showed that by the end of the trial, the low performing TD child was fixating on the target image (Figure 32). These fixations were greater than those of the high performing TD child for the same scalar
(Figure 30). However, this relation for these two TD children was reversed for all the other trial types. The only eye tracking differences between the high and low performing ASD children (besides the already mentioned or condition during the quantifier phase) were in the only some condition during the quantifier and predicate phases. Subject 302 had lower accuracy and fewer proportions of fixation time to the only some target than the higher performing Subject 304.

Figure 31. Quantifier phase proportions of fixation times to the target image and response accuracy for a low performing child in each group.
Note: Subject 212 is TD and Subject 302 is ASD. Error bars indicate the standard error of the mean. Horizontal line indicates chance accuracy (.25)
Figure 32. Predicate phase proportions of fixation times to the target image and response accuracy for a low performing child in each group. 

Note: Subject 212 is TD and Subject 302 is ASD. Error bars indicate the standard error of the mean. Horizontal line indicates chance accuracy (.25)

Discussion

Accuracy. Overall, children and adults had greater accuracy on the strong scalars in Experiment 2 compared to Experiment 1. In the case of the children, accuracy also improved for the weak scalar *some* in Experiment 2 compared to Experiment 1. Individual children chose the implied semantic interpretations of the weak scalar *or*, though overall, group accuracies were lower for the eye tracking task than for the two judgment tasks. The presence of visual alternatives in these eye tracking experiments, then, may be more interfering than facilitative. The fact that most incorrect responses were for the scalar distracter showed that children between
the ages of 7 and 10 years knew that the scalar distracter was semantically consistent with the auditory instruction.

Experiment 2b used a lexical-syntactic cue (by adding the word *only* to a weak scalar) to decrease cognitive effect. The presence of *only* should have strengthened the implicature, resulting an increase in implied semantic response to *only some* compared to *some*. Adults were already at ceiling on 2a, and there was no gain in accuracy for Experiment 2b. Children did not have greater accuracy for *some* than for *only some*. Although language impairment may have affected the comprehension of scalars in this lexical-syntactic cue context, the trend for lower accuracy in the *only* condition was noticed even among children with autism who had typical language scores. Previous research supports the lack of improvement in accuracy when *only* is added to *some* in younger children (Katsos & Bishop, 2011). The accuracy in ASD children for *only some* trials was poorer than the accuracy for *some* and *all* trials.

**Eye tracking.** Overall, the gaze data support the pragmatic view of scalar implicature (Grice, 1975). These data also agree with the accuracy data and corroborate previous research in children’s processing of *some/all* (Huang & Snedekar, 2009b). When participants fixated more on the target, they were more likely to select the target image. The gaze data also supplied information about the speed of processing of scalar implicature. In children, implied interpretations (Figures 23 – 24) exhibited slower resolution between competing target and scalar alternative images as compared to when children made literal semantic interpretations (Figures 27 – 28).

Children with ASD demonstrated different processing than TD children even when they correctly made the implicature. In the TD group, incorrect responses in the predicate phase to weak scalars revealed competition between the alternative and a control. (Competition indicates
the children were deciding between the two images as choices). In contrast, when the children with ASD were incorrect, they had significantly longer looks on average to the alternative than to the target. This suggests that by the onset of the predicate phase, children with ASD had already made an implicature decision and were not going to revise their response.

Inspection of the few incorrect some trials revealed that distracter control 1 could have been interpreted as an option for some in both groups; specifically, in the case where the image showed one item inside of a box and four items outside. Semantically, some includes one. This can explain the visual competition between the scalar distracter and the other distracter images. The children overwhelmingly selected the alternative as their answer, which may indicate a preference for not accepting one as the best option for some.

During the visual preview period, attention was driven by perceptual factors. The most visually contrastive image (control 2 in Figure 19) initially received the most fixations. For example, if the target was All of the balls are inside of the box, the control with the most fixations would be the image of an empty box with 5 balls outside. This was expected because no information was provided during this time to direct the participants to fixate on any particular image. Participants also tend to first look at images with the largest quantities (Grodner et al., 2010).

There were differences in processing for trial type between the TD and ASD groups. Children with ASD were slower to fixate on both the target and scalar distracter than TD children. The fact that the TD group had shorter looks during the predicate phase for all trials than the ASD group corroborates the unexpectedly higher proportion of fixations in the ASD group; however, this could reflect the lower number of participants and valid trials in the ASD group (Figure 20). The differences between looking time and proportion of fixation time more
likely suggests that the children with autism may have spent more time fixating on images overall, while the TD children spent more of their fixation time on the correct image.

The individual analyses also highlight variability in both TD and ASD children. Both of the high performing subjects selected for Figures 29 and 30 also performed with high accuracy in the executive functions task. In contrast, Subject 212 (TD) and Subject 302 (ASD), the low performing pair in Figures 31 and 32, scored low overall on the executive functions task. Based on her CELF-4 scores, Subject 302 also has language impairment (see Appendix A for demographic data). More data are needed, but it appears there may be a relation between scalar comprehension on a 4 alternative forced-choice task and executive functions.

A caveat to these eye tracking results is that the eye tracker’s validity code (the reliability measurement of the eye gaze determined by the eye tracker) was poor for many trials. The probability of poorer validity was higher for the younger subjects. A possible limitation of setting an overly strict validity criterion is the elimination of data that may be useful. Eye tracking in TD children from Experiment 2b had fewer total trials but an even higher rate of rejection than Experiment 2a.

It is also unclear how age, language ability, and other intrinsic factors may relate to eye tracking tasks. During testing, the children with autism appeared more focused and eager to complete the tasks whereas the TD children indicated boredom and fatigue. Only the TD children agreed to take their assigned breaks. ASD participants were less likely to move during the eye tracking Experiment 2 tasks. More participants are required to see if any of these differences are truly related to the ASD diagnosis or whether other factors are involved.
Executive Function Task

Because executive functions is believed to be impaired in children with autism and involved in deriving implicatures a nonverbal baseline of executive functions was obtained. This computerized task was administered with 20 trials each of three conditions: Vigilance, distracter interference, and response inhibition (Figure 33). This task is part of a larger executive functions battery developed by Marton, Campanelli, Eichorn, Scheuer, and Yoon (2014).

Participants

Data were obtained from the 14 adults mentioned in the general methods section. All 10 of the TD children completed the full version of the executive function task. Six children with ASD also completed the task. Five of these children also had diagnoses of ADD or ADHD. None of them were medicated for that condition at the time of this study.

Procedure

The conditions were presented in fixed sequential order and each condition was preceded by written instructions and five practice trials. Responses were made on a keyboard with a red sticker on the “2” key and a yellow sticker with arrows on the “1” and “3” keys on the numerical keypad.
For all conditions the participant first held down the *start* button (red “2”). To continue to the actual trial, he had to release the key when a red dot (bold Times New Roman 180 font *period*) appeared on the screen. The duration of the button press ranged from 1100 to 2000 ms, randomly assigned in 100 ms increments. The trial ended if the participant prematurely released or did not release the *Start* button (Figure 34).

*Figure 34.* Illustration of two trials for the attention condition in the executive function task

During the first condition, vigilance, which served as a baseline for the other conditions, subjects had to press a key to respond to where (yellow right “3” or left “1”) a green circle appeared (Figure 34). In the second condition, participants still saw a green circle but there was also a blue circle. Participants had to ignore the blue circles and respond only to the position of the green circles (yellow right “3” or left “1”). For the final condition, participants had to ignore the green circle and only respond (red “2”) to the presence of a blue circle.
Predictions

Task difficulty was expected to increase from the first to last condition. Overall children would be slower than adults and make more errors (i.e., inaccurate and premature release). Performance on the nonverbal vigilance task was expected to be comparable between the groups of children because the task is easy, with low cognitive demands. On the other tasks, performance was thought to also be comparable for similar reasons and because executive functions are still developing at the participants’ ages. If children with autism have deficits in response inhibition, as theorized in the EF theory of autism, then they should perform selectively worse in that condition compared to the other groups.

Results

The results of the nonverbal executive function task suggest that basic vigilance, distracter interference, and response inhibition were preserved in all subjects. Overall there was a difference in accuracy and reaction time by condition for the executive function task, Friedman Test $X^2(2, N = 30) = 17.882, p < .001$; $X^2(2, N = 30) = 34.07, p < .001$. Pairwise comparisons with the Boneferroni correction demonstrated that accuracy was poorer, although not significantly worse, in the response inhibition condition than in the other two conditions, Wilcoxon signed-rank test $z = .683, p < .024$; $z = .867, p < .002$. Performance in the vigilance condition was significantly faster than in the other two conditions, Wilcoxon signed-rank test $z = -1.233, p < .000; z = -1.367, p < .001$. The difference in accuracy and reaction time for the congruent and incongruent trial types in the distracter inference condition did not reach significance although the trend was that congruency increased accuracy while decreasing speed.

Adults. There was significant difference in reaction time for adults on the three conditions, Friedman Test $X^2(2, N = 14) = 14.92, p < .0001$. Due to experimenter error, five of
the adult participants completed the full version of the executive function task as described in the methods section. The other eight subjects did a test version, which had only five test trials per condition. All adult data were analyzed together because there was no difference for any measure based on the number of trials administered.

As expected, the adults performed fastest and with the highest accuracy on the first (vigilance) condition [percent correct $M = 94\%$, $SD = 23\%$; reaction time $M = 323$ ms, $SD = 121$ ms]. This reaction time was significantly faster than either of the other two conditions, Wilcoxon signed-rank test $z = -3.233$, $p = .0006$. Adults also did well in the most difficult condition (response inhibition) with mean of $81\%$, $SD = 40\%$ under mean 513 ms ($SD = 235$ ms) reaction time. Performance in the second (distracter interference) condition was between the other two [$M = 88\%$, $SD = 30\%$; $M = 447$ ms, $SD = 207$ ms]. Across the experiment, the rate of skipped trials due to premature release of the Start button was only 3\%. Adults were never late to release the start button during the critical trials. There was only one case where a subject did not release the start button but it occurred in a practice condition.

**Children.** Overall the ASD group was less accurate but faster than the TD children (see Table 8) but these differences in accuracy were not statistically different between the TD and ASD groups (student $t$, $p > .05$). However, there was also greater variability within ASD participants and across that group than for TD children.

Like the adults, the vigilance condition yielded the best performance. As expected, even with higher accuracy, children were slower than adults. The rate of skipped trials because of premature release was $4\% – 18\%$, similar to adults. Unlike the adults, a few trials were ended because of late releases of the start button. Also, unlike the adults, there were cases in which the trial continued past the start phase but the child incorrectly answered the test item.
Children were slower and less accurate than adults on the distracter interference part of the executive function task. Children were more accurate when the target and distracter were on the same side of the screen (congruent trial types) than when they were on opposite sides of the screens (incongruent trial types). This difference in accuracy was not significant in the TD group (congruent $M = 79\%$, $SD = 40\%$; incongruent $M = 88\%$, $SD = 30\%$), Wilcoxon signed-rank test $W = 4.5$, $p > .05$. The corresponding mean accuracy scores in the ASD group were $75\%$ ($SD = 40\%$) and $82\%$ ($SD = 40\%$), respectively. Because there were only 6 children in the ASD group a Sign test was performed instead of a Wilcoxon signed-rank test. The difference in accuracy between the congruent and incongruent conditions was significant (one-tailed $p < .05$). Children with ASD were more negatively distracted by the incongruency than TD children. As a group though, the children with ASD had similar executive functioning abilities as the TD group. In the final condition of the executive function task, response inhibition, children were slower and less accurate than they were in the baseline attention condition. All children also performed less well and slower than the adults on the same condition.
Figure 35. Scatterplots for typically developing children and children with autism accuracy on vigilance compared to distracter interference and compared to response inhibition on the executive function task

Note: Individual subject numbers for typically developing children = blue; for children with autism = red. x-axis = vigilance; y-axis = distracter interference (top) and response inhibition (bottom).

There were individuals within each group who performed differently from the group average (Figure 35). Accuracy scores and reaction time on the executive function task did not correlate with age or nonverbal IQ although two-thirds of the 7 year olds had the lowest accuracy of all children. The range of accuracy within the TD group on the baseline condition was 65% – 100%. Three children (TD Subjects 203, 209, and 211) scored less than the group average (88%) on the task. These children also responded faster than the other TD children ($M = 525$ ms, $SD =$
92 ms compared to $M = 774$ ms, $SD = 253$ ms). For the interference task, 3 children (TD Subjects 208, 209, and 212) scored less than the group average (84%). These children were slower overall than the rest of the group ($M = 1138$ ms, $SD = 488$ ms vs. $M = 913$ ms, $SD = 157$ ms). Dividing the interference by congruency revealed TD Subject 209 performed significantly poorer when the target and distracter were on opposite sides of the screen (40%) than when they were on the same side (90%). TD Subject 212 was similarly impacted but had low scores in both sub-conditions (30% incongruent; 60% congruent). TD Subject 208 was the only child who trended in the opposite direction: 80% in the incongruent condition and 70% in the congruent condition. This child was also a common outlier for the TD group on Experiment 1. For response inhibition, 5 children performed above the group average (76%) and they included TD Subject 212 who did not perform well on the previous condition or on the yes/no truth value judgment task. Again, poor performers were more likely to respond faster ($M = 691$ ms, $SD = 189$ ms) than high scorers ($M = 1094$ ms, $SD = 238$ ms).

Within the ASD group, two children consistently scored lower than the group average on the executive function task: ASD Subjects 302 and 306. The first subject scored 35% on the vigilance task, 55% on the interference task, and 40% on the response inhibition task. ASD Subject 306 scored 50%, 30%, and 8% on each respective task. ASD Subject 302 reacted faster than the other children in the ASD group. Although ASD Subject 306 had poorer scores, he did not take more or less time responding. Overall, when looking at the individual data, it appears that only children with ASD who also had language impairment showed signs of impaired executive functions. However, ASD Subject 300 also stood out from the group because his accuracy in vigilance and interference was at 100% but only 55% on the inhibition task. The
other 3 children with ASD scores consistently obtained over 80% on all portions of the executive functions task.

**Discussion**

Although the adults performed well on the executive function task, it was expected that more of their individual scores would be above 90%, especially for the baseline attention condition. However, there was no relationship between accuracy on the executive function task and the scalar implicature experiments for adults. There was also no relation between scalar comprehension and executive functions in children but there were a few individual children who tended to have poorer accuracy on EFs and the scalar experiments (Figure 35 and also Appendix D).

There appeared to be a speed-accuracy trade-off for both groups of children on the executive functions task. Unlike some previous findings (Robinson et al. 2009; Troyb et al., 2014) inhibition was not a deficit in the average means for the ASD compared to their TD peers. This is consistent with Adams and Jarrold (2012). These authors also reported intact response inhibition in children with ASD. The task used for the current study was nonverbal and not like tasks (Geurts et al., 2009) criticized for their poor construct validity. It should be noted however that there were individual differences and some children with ASD did not perform near the TD group average.

The only difference found between the ASD and TD group data was the clinical group performed significantly poorer in the incongruent condition than the congruent condition on the interference task. Distractor interference is one of the executive functions found to be impaired in some children with ASD (Adams & Jarrold, 2012) though not in others (Hill, 2004). It is unclear from much of the literature how participant factors relate to these findings. Another
executive function implicated but not directed assessed in the current study is impulse control. Children with ASD responded quicker than TD children but they tended to make greater errors at faster reaction times.

The disparate results across conditions for individual children highlight the importance of testing distinct aspects executive functions in children and the distinctiveness of individual EFs (Best et al., 2011; Hill, 2004). Children from both groups with the lowest scalar implicature comprehension scores on Experiment 1 also had poor scores on the distracter interference task. However, there was a trend for younger children to perform worse on the executive function task, probably due to development of executive functions (Best et al., 2011). Although the difference between the median age of TD and ASD children was not significant, there were more young children in the TD group, which may have affected overall task performance averages.

Two of the 3 ASD children who scored the worse on executive function tasks had impaired language based on their scores on the Comprehensive Evaluation of Language Fundamentals - 4 (Table A2). Both children also had ADD or ADHD but so did several other children in the ASD group who performed well on the executive function task (Table A3). The third child was one of the youngest in the group (7;9).

The executive function task suggests a developmental difference in interference and inhibition as opposed to a difference based on ASD diagnosis. Although some children with ASD had difficulty with the interference and inhibition portions of the task, they performed similarly to the TD group. But more children across a wider age range with and without ASD are needed to verify this finding. Based on these simple tasks and few participants, executive dysfunction (Ozonoff et al., 2005) may not be a global underlying cause of deficits in children with autism.
General Discussion

Overview

The goal of this study was to understand the relation between language and cognitive skills for the comprehension of scalars by children with and without autism (ASD) under various manipulations of cognitive effect and effort. Participants completed three computerized experiments: 1) non-verbal executive function task measuring speed and accuracy during three conditions (vigilance, interference, and response inhibition); 2) a - truth value judgment task (TVJT) measuring scalar comprehension under high verbal memory requirements (high cognitive effort), b - felicity judgment task (FJT) measuring scalar comprehension under low verbal memory demands (low cognitive effort); 3) Eye tracking tasks a – 4 Alternative forced-choice task, b – 4 Alternative forced-choice task with grammatical strengthener only with the weak scalar some.

The comprehension of scalar terms was used to examine the relation between literal and implied semantics in children with ASD for the first time. Based on the often-described weakness in pragmatic aspects of language (e.g., Eisti et al., 2011), it was hypothesized that children with autism would not derive implicatures, whereas typically developing children and adults would comprehend implied semantic interpretations with little to no difficulty. The significance of scalar implicature for children with ASD is to connect semantics and pragmatics. The results in these two experiments however indicated that overall, children with autism were able to choose implied semantic interpretations. Much variation was seen in the TD and ASD children but the small sample size allowed for a more detailed examination of individual performance on these tasks. There is evidence that some children with ASD have difficulty making implicatures which may be related to deficits in executive functions with or without
language impairments. These data suggest that semantic knowledge can be interrupted by nonlinguistic factors in some children with ASD.

An important issue in ASD research is the method(s) used to identify and describe research participants. Much of the current research in younger children with ASD report Autism Diagnostic Observation Schedule testing (ADOS and ADOS 2: Lord et al., 2000, 2012) scores. Yet, the scores themselves do not provide a diagnosis nor do they independently represent differences in behavior (Lord et al., 2012). Exclusive use of the ADOS also ignores formal language ability. Two children may have the same autism score but present with very different symptoms thus requiring different treatments. There must be better ways to describe participants in research than using ADOS scores. A single test cannot be used to make a diagnosis. In fact, the gold standard for ASD diagnosis is experienced clinical assessment used with three standardized tools: the Vineland Adaptive Behavior Scales, the Autism Diagnostic Inventory, and the ADOS (Charman & Baird, 2002; Filipek et al., 2000).

Scalar Implicature

Accuracy and time course processing. This study provided experimental confirmation that weak scalars have two interpretations. Given four options, participants rarely selected anything but the target and the scalar alternative in the four-alternative forced-choice task. The data also points to implicatures being derived in real time as suggested by the pragmatic account of scalar implicature (Grice, 1975). The scalar and the complexity of the task appeared to be the biggest factors in encouraging implied semantic interpretations.

The participants in this study had greater difficulty with the or/and scale than with the some/all scale. The disparity in performance between the two scales was most prominent for the children with autism. Among the weak scalars, or was more difficult than some for children
SCALAR COMPREHENSION

across tasks. This is surprising, considering that the *and/or* words are considered to be learned earlier than *some/all*. However, this is in line with previous research (Chevallier et al., 2010; Pijnacker et al., 2009). It may be that another variable, like speech perception, is responsible for this difference because *or* and *and* are acoustically more similar than *some* and *all*. Previous studies about scalar implicature used live instead of recorded speech (Papafragou & Mousolinno, 2001). In online computer tasks the relatively short duration of the quantifier *or* and its scalar alternative may have negatively affected performance in all groups. Other computerized studies (Grodner et al., 2010; Huang & Snedeker, 2009b) used recorded speech but did not test the scalars *and/or*. It is possible that the development of scalar implicature for *or* occurs at later ages than those tested in the current study.

Accuracy was generally better, that is children chose the implied semantic interpretation for *some*, in the two-talker felicity judgment task and in the eye tracking tasks because these two tasks provided the scalar alternatives. This is consistent with results in younger children (Katsos & Bishop, 2011). The results in Experiment 2 also support these findings but the increased presence of visual alternatives did not increase accuracy in the ASD group. Children with ASD did not optimize their use of the visual array and were slower to selectively fixate to the target image. This deficit in attention would appear to be specific to the linguistic task as performance on the nonverbal executive function (EF) task was similar across groups. However, the executive functions task may have been too easy for these children. One area of overlap between language and EFs might be inhibition; this is discussed further below.

The eye tracking data also provided evidence for processing differences of scalars between children with and without autism. Although the accuracy and reaction time for TD and ASD children were similar, differences became apparent when examining the time course of
processing. The time courses from these data show that there is a cognitive difference in processing implied compared to literal interpretations of weak scalars. As suggested in the pragmatic theory of scalar implicature (Grice, 1975), and supported by other research in children and adults (Huang & Snedeker, 2009a,b), scalar implicature increases processing. The children in the current study showed a greater processing gap between implied and literal choices than previous eye tracking studies (Grodner et al., 2010; Huang & Snedeker, 200b) but that may be the result of equipment differences. The gap was wider for children with ASD than TD children.

**Executive functions.** Previous research on scalar comprehension had not examined EF demands within task or measured executive functions in participants (Papafragou & Musolino, 2003). Within Experiment 1, presenting the audio before the visual was expected to increase working memory demands, thus increasing the difficulty of making implied semantic interpretations. It was also predicted that those with poorer interference and inhibition abilities, would accept more literal semantic than implied semantic interpretations. This was verified by a few participants from each group of children. Recall that TD Subjects 208 and 209 and ASD Subject 302 performed poorly on the executive functions task and on Experiment 1. The poorest performers on Experiment 2 also had low scores on the EFs task. There were also two findings in relation to EFs: 1) age and 2) language ability. There also seemed to be an effect of development, possibly related to executive functions ability, on the scalar comprehension tasks. Children under the age of 8 and children with language impairment were typically poor performers. These finding would suggest developmental status should be considered in scalar comprehension.

Performance in the current study would indicate two possible strategies used to complete the two-talker task. The first requires a constant high level of attention. There is evidence that this was a method used in this study. When the stimulus onset asynchrony (SOA) was negative,
accuracy was low for all groups (Tables 2 and 3). This suggests that participants had difficulty listening to at least one of the sentences and maintaining it in memory until the image appeared. In this case, only the negative SOA challenged working memory enough to make the two-talker FJT difficult for language typical participants.

The second strategy used to complete the two-talker task is to turn it into a yes/no task by focusing on just one speaker’s utterance. This strategy would facilitate a quick response once the target had been identified. Typically developing children had greater accuracy when the target audio was the second sentence presented. This suggests the memory method of completing the task because the last item placed into memory is typically easier to recall than an earlier one. The children with ASD were more accurate at the two-talker task when the target audio was presented before the alternative sentence. This processing method would be expected if those subjects listened to the first sentence and only used that one to inform their answers. Although this finding was not significant this was a trend in individual subjects in the ASD group. This could suggest working memory capacity limitations or interference effects in children with ASD.

If children with ASD completed the two-talker task as a yes/no task, then it was also probable they would have performed better with an actual yes/no task (Experiment 1a). In this case, the cognitive effect may have played a role. It is possible that cognitive effect was too low in this task, thus there was less reason for accepting only upper-bounded interpretations of a weak scalar. If children with ASD have deficits in theory of mind (Booth & Happé, 2010) it would be more difficult for them than for TD children to figure out what a test is asking. The connection between ToM and Grice’s Maxims has been found in more impaired children with ASD (Surian et al., 1996). This could explain why ASD children did poorer on the weak scalars for the yes/no task than TD children. When tasks are essentially local and require detailed
processing (Happé, 1993), as can be said of computerized tasks like this one, children with autism perform better because that is the processing style in which they excel.

Displaying the alternative image during the yes/no truth value judgment task could have increased cognitive effect because the relevance of alternatives is made explicit. However, in ASD children there was an effect of congruency on the distracter interference executive functions task. This would suggest that displaying the alternative in the yes/no task would increase cognitive effort by increasing distracter interference. Performance on the four-alternative forced-choice task (Experiment 2) appeared to support this statement. In that task the target and the alternative images were visible to the participants. Children with ASD made the implied semantic choice less often than they did in the two-talker task, which had only one image. Based on anecdotal clinical evidence it is unlike that children with ASD have a preference for auditory as opposed to visual information. A more likely explanation for the poorer accuracy on the eye tracking task is that children with ASD were affected by increased demands on executive functions, primarily interference. Previous eye tracking studies in adults have also found that auditory-visual tasks have an additional cognitive component requiring the participant to integrate those two sources of information (Ferreira, Foucart, & Engelhardt, 2013). These findings imply that visual aids should be used cautiously in children with ASD if they are meant to facilitate complex linguistic tasks.

**Future Directions**

Future research should explore the scalars *and* and *or* to uncover how and why disjunctions are processed differently from the quantifiers *some* and *all* and numerals. Eye tracking methods have the added value of capturing how language is understood in real time. Future research should also study the relationship between the amount of visual information
presented and processing speed and responder (i.e., literal and pragmatic) type. Time course data can also be further explored by finding the specific timing for when the implicature appears to occur. Suggestions for improving the reliability of eye gaze data include using a child-sized office chair with removal boosters and cushions (for children of different sizes) and a wired intercom (for redirecting attention without interrupting the testing).

Other scalars should also be studied as a tool to examine relation between semantics and pragmatics but there is also a need to vary the cognitive effect of tasks to improve their ecological validity. In order to ultimate affect clinical practice, research findings must be generalizable outside of the laboratory. DSM – 5 standards are moving the professions away from subtyping and towards a more dimensional approach to ASD diagnosis (APA, 2013). In addition, although ADOS scores are frequently provided in research articles, ADOS testing is not required for diagnosis of autism and financial coverage for the test widely varies (NYS, 1999). Although it may be more difficult to generalize research results, the sample in this study follows the DSM’s direction in that regard.

Summary

The results on responder type in Experiment 1 suggest that regardless of clinical group, participants were able to choose scalar implicatures in both the two-talker felicity judgment task and the yes/no truth value judgment tasks. But it was easier to derive implicatures for some than all. Children with ASD were less able to benefit from visual information than TD children. It is possible that distracter inference was playing a large role in those effects. The results of the executive function task suggest that the greater within and between subject variability on the language tasks in the ASD group is not cognitive-general. However, aspects of atypical working memory and interference handling may impact their ability to use linguistic contextual cues.
Experiment 2 (forced-choice with eye tracking) was lower in cognitive effort and/or higher in cognitive effect than Experiment 1 because accuracy was higher and reaction times were faster. Using the time course of eye movements and sequences of fixations was valuable in examining comprehension. Although reaction time and accuracy were similar across these groups of children, eye gaze told a different story. Children with ASD took longer to make pragmatic interpretations than TD children. If trial validity is maintained eye tracking can provide insights to behavior in a range of abilities that standardized testing is unable to target.
Appendix A. Demographics

Table A1.

_Detailed Demographics for Typically Developing Children_

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender</th>
<th>Age&lt;sup&gt;a&lt;/sup&gt;</th>
<th>TONI-4&lt;sup&gt;b&lt;/sup&gt;</th>
<th>CELF-4</th>
<th>VABS-II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rec</td>
<td>Exp</td>
<td>Com</td>
</tr>
<tr>
<td>202</td>
<td>F</td>
<td>7;9</td>
<td>122</td>
<td>119</td>
<td>120</td>
</tr>
<tr>
<td>203</td>
<td>F</td>
<td>7;11</td>
<td>110</td>
<td>108</td>
<td>108</td>
</tr>
<tr>
<td>204</td>
<td>F</td>
<td>8;0</td>
<td></td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>206</td>
<td>M</td>
<td>7.1</td>
<td>130</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>207</td>
<td>M</td>
<td>9;11</td>
<td>109</td>
<td>128</td>
<td>120</td>
</tr>
<tr>
<td>208</td>
<td>F</td>
<td>7;1</td>
<td>113</td>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>209</td>
<td>M</td>
<td>7;5</td>
<td>108</td>
<td>108</td>
<td>103</td>
</tr>
<tr>
<td>210</td>
<td>M</td>
<td>8;1</td>
<td>103</td>
<td>115</td>
<td>113</td>
</tr>
<tr>
<td>211</td>
<td>M</td>
<td>8;7</td>
<td>113</td>
<td>121</td>
<td>118</td>
</tr>
<tr>
<td>212</td>
<td>F</td>
<td>7;9</td>
<td>112</td>
<td>115</td>
<td>120</td>
</tr>
</tbody>
</table>

*Note.* <sup>a</sup>Age in years;months, standard scores on the <sup>b</sup>Test of Non-verbal Intelligence - 4 (Brown, Sherbenou & Johnsen, 2010); composite standard scores on the <sup>c</sup>CELF - 4 (Comprehensive Evaluation of Language Fundamentals, Fourth Edition: Semel, Wiig, & Secord, 2003) Receptive, Expressive, and Composite; on the <sup>d</sup>VABS-II (Sparrow, Cicchetti, & Balla, 2005) Communication and Socialization Scales and Composite. Blank = missing data; * = composite could not be generated with subtests completed.
Table A2.

**Detailed Demographics for Children with Autism**

<table>
<thead>
<tr>
<th>ID</th>
<th>Gender</th>
<th>Age</th>
<th>TONI-4</th>
<th>CELF-4</th>
<th>VABS-II</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Rec</td>
<td>Exp</td>
<td>Com</td>
</tr>
<tr>
<td>300</td>
<td>M</td>
<td>7;9</td>
<td>158</td>
<td>101</td>
<td>108</td>
</tr>
<tr>
<td>301</td>
<td>M</td>
<td>9;7</td>
<td>108</td>
<td>137</td>
<td>136</td>
</tr>
<tr>
<td>302</td>
<td>M</td>
<td>8</td>
<td>106</td>
<td>92</td>
<td>65</td>
</tr>
<tr>
<td>303</td>
<td>M</td>
<td>7</td>
<td>152</td>
<td>105</td>
<td>105</td>
</tr>
<tr>
<td>304</td>
<td>M</td>
<td>9;4</td>
<td>93</td>
<td>105</td>
<td>105</td>
</tr>
<tr>
<td>306</td>
<td>M</td>
<td>9;4</td>
<td>113</td>
<td>61</td>
<td>74</td>
</tr>
<tr>
<td>601</td>
<td>M</td>
<td>6;11</td>
<td>93</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

*Note.*  

*a* Age in years;months, standard scores on the *b*Test of Non-verbal Intelligence - 4 (Brown, Sherbenou & Johnsen, 2010); composite standard scores on the *c*CELF - 4 (Comprehensive Evaluation of Language Fundamentals, Fourth Edition: Semel, Wiig, & Secord, 2003) Receptive, Expressive, and Composite; on the *d*VABS-II (Sparrow, Cicchetti, & Balla, 2005) Communication and Socialization Scales and Composite. * = could not be computed because subtests were missing.
Table A3.

*Detailed Diagnostic Information for Children in the Autism Group*

<table>
<thead>
<tr>
<th>ID</th>
<th>Diagnosis</th>
<th>Age of Diagnosis</th>
<th>Method of Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>Autism*</td>
<td>1;6</td>
<td>ADOS; Formal interview/survey</td>
</tr>
<tr>
<td>301</td>
<td>Asperger’s*</td>
<td>3;10</td>
<td>ADOS; Formal interview/survey; Informal interview/survey; Cognitive tests; Language tests</td>
</tr>
<tr>
<td>302</td>
<td>PDD*</td>
<td>5;0</td>
<td>Not stated</td>
</tr>
<tr>
<td>303</td>
<td>Asperger’s*</td>
<td>6;11</td>
<td>Formal interview/survey; Clinical observation &amp; parental interview</td>
</tr>
<tr>
<td>304</td>
<td>PDD</td>
<td>7;0</td>
<td>Formal interview/survey; Clinical observation &amp; parental interview</td>
</tr>
<tr>
<td>306</td>
<td>Autism*</td>
<td>2;0</td>
<td>Not stated</td>
</tr>
<tr>
<td>601</td>
<td>Autism</td>
<td>2;0</td>
<td>Formal interview/survey; Speech tests; Language tests</td>
</tr>
</tbody>
</table>

*Note.* * indicates child also diagnosed with ADD or ADHD. Age in years. Formal = standardized assessment.
Appendix B. The 15 stimuli sets for Experiments 1 and 2.

Table B.

*Stimuli Sets for Experiment 1 and 2*

<table>
<thead>
<tr>
<th>Target</th>
<th>Distracter</th>
<th>Target</th>
<th>Distracter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bikes</td>
<td>Pears</td>
<td>Boats</td>
<td>Vests</td>
</tr>
<tr>
<td>Cars</td>
<td>Strawberries</td>
<td>Planes</td>
<td>Ties</td>
</tr>
<tr>
<td><em>Balls</em></td>
<td>Pineapples</td>
<td><em>Arrows</em></td>
<td>Scissors</td>
</tr>
<tr>
<td><em>Mouse</em></td>
<td>Peanuts</td>
<td><em>Monkey</em></td>
<td>Whistles</td>
</tr>
<tr>
<td><em>Pants</em></td>
<td>Light bulbs</td>
<td><em>Hats</em></td>
<td>Harps</td>
</tr>
<tr>
<td><em>Shoes</em></td>
<td>Telephones</td>
<td><em>Belts</em></td>
<td>Corns</td>
</tr>
<tr>
<td>Pepper</td>
<td>Ferns</td>
<td>Grapes</td>
<td>Kites</td>
</tr>
<tr>
<td><em>Bread</em></td>
<td>Pipes</td>
<td>Oranges</td>
<td>Paintbrushes</td>
</tr>
<tr>
<td>Carrots</td>
<td>Lamps</td>
<td>Apples</td>
<td>Irons</td>
</tr>
<tr>
<td>Lemons</td>
<td>Umbrellas</td>
<td><em>Tomatoes</em></td>
<td>Palms</td>
</tr>
<tr>
<td>Pens</td>
<td>Fish</td>
<td>Hammers</td>
<td>Sleds</td>
</tr>
<tr>
<td>Computers</td>
<td>Locks</td>
<td><em>Keyboards</em></td>
<td>Beds</td>
</tr>
<tr>
<td><em>Flowers</em></td>
<td>Balloons</td>
<td>Trees</td>
<td>Keys</td>
</tr>
<tr>
<td>Forks</td>
<td>Brushes</td>
<td><em>Books</em></td>
<td>Mushrooms</td>
</tr>
<tr>
<td>Stars</td>
<td>Bells</td>
<td><em>Rings</em></td>
<td>Clocks</td>
</tr>
</tbody>
</table>

*Note: * Set is not in Child version of tasks*
Appendix C. Trial Timelines for Experiment 1

SOA = 0 ms

SOA = -750 ms

SOA = +750 ms

Fixation Preamble Voice 1 Image + Visual Stimuli Voice 2 Image + Visual Stimuli Response

Fixation Preamble Voice 1 Image (No Image) Voice 2 Image (No image) Response

Appendix D. Individual Results for Scalar Comprehension Tasks

Typically Developing Children’s Average Accuracy (SD) on Experiment 1a. Truth Value

Judgment Task

<table>
<thead>
<tr>
<th>Subject</th>
<th>All</th>
<th>And</th>
<th>Or</th>
<th>Some</th>
<th>StrNo</th>
<th>WNo</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>202</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>67</td>
<td>52</td>
<td>100</td>
</tr>
<tr>
<td>203</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>204</td>
<td>100</td>
<td>0</td>
<td>86</td>
<td>38</td>
<td>88</td>
<td>35</td>
<td>85</td>
</tr>
<tr>
<td>206</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>63</td>
<td>52</td>
<td>100</td>
</tr>
<tr>
<td>207</td>
<td>100</td>
<td>0</td>
<td>86</td>
<td>38</td>
<td>25</td>
<td>46</td>
<td>100</td>
</tr>
<tr>
<td>208</td>
<td>0</td>
<td>0</td>
<td>57</td>
<td>53</td>
<td>38</td>
<td>52</td>
<td>85</td>
</tr>
<tr>
<td>209</td>
<td>100</td>
<td>0</td>
<td>71</td>
<td>49</td>
<td>50</td>
<td>53</td>
<td>78</td>
</tr>
<tr>
<td>210</td>
<td>100</td>
<td>0</td>
<td>43</td>
<td>53</td>
<td>88</td>
<td>35</td>
<td>92</td>
</tr>
<tr>
<td>211</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>90</td>
</tr>
<tr>
<td>212</td>
<td>50</td>
<td>71</td>
<td>43</td>
<td>53</td>
<td>0</td>
<td>0</td>
<td>54</td>
</tr>
<tr>
<td>Mean</td>
<td>91</td>
<td>29</td>
<td>77</td>
<td>43</td>
<td>53</td>
<td>50</td>
<td>88</td>
</tr>
</tbody>
</table>

Note. Accuracy as mean percent correct (SD). StrNo = strong numerals; WNo = weak numerals.
Table D2.

_Typically Developing Children's Average Accuracy (SD) on Experiment 1b. Felicity Judgment Task_

<table>
<thead>
<tr>
<th>Subject</th>
<th>All</th>
<th>And</th>
<th>Or</th>
<th>Some</th>
<th>StrNo</th>
<th>WNo</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>202</td>
<td>100</td>
<td>0</td>
<td>83</td>
<td>41</td>
<td>56</td>
<td>53</td>
<td>71</td>
<td>49</td>
</tr>
<tr>
<td>203</td>
<td>71</td>
<td>49</td>
<td>67</td>
<td>52</td>
<td>67</td>
<td>50</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>204</td>
<td>88</td>
<td>35</td>
<td>100</td>
<td>0</td>
<td>89</td>
<td>33</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>206</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>207</td>
<td>78</td>
<td>44</td>
<td>33</td>
<td>52</td>
<td>100</td>
<td>0</td>
<td>50</td>
<td>58</td>
</tr>
<tr>
<td>208</td>
<td>63</td>
<td>52</td>
<td>80</td>
<td>45</td>
<td>63</td>
<td>52</td>
<td>86</td>
<td>38</td>
</tr>
<tr>
<td>209</td>
<td>88</td>
<td>35</td>
<td>60</td>
<td>55</td>
<td>57</td>
<td>53</td>
<td>25</td>
<td>50</td>
</tr>
<tr>
<td>210</td>
<td>75</td>
<td>46</td>
<td>83</td>
<td>41</td>
<td>78</td>
<td>44</td>
<td>71</td>
<td>49</td>
</tr>
<tr>
<td>211</td>
<td>89</td>
<td>33</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>212</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
<tr>
<td>Mean</td>
<td>85</td>
<td>36</td>
<td>81</td>
<td>40</td>
<td>82</td>
<td>39</td>
<td>88</td>
<td>33</td>
</tr>
</tbody>
</table>

_Note._ Accuracy as mean percent correct (SD). StrNo = strong numerals; WNo = weak numerals
Table D3.

Children with autism’s Average Accuracy (SD) on Experiment 1a. Truth Value Judgment Task

<table>
<thead>
<tr>
<th>Subject</th>
<th>All</th>
<th>And</th>
<th>Or</th>
<th>Some</th>
<th>StrNo</th>
<th>WNo</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>33</td>
<td>52</td>
<td>89</td>
</tr>
<tr>
<td>301</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>302</td>
<td>100</td>
<td>0</td>
<td>86</td>
<td>38</td>
<td>0</td>
<td>0</td>
<td>69</td>
</tr>
<tr>
<td>303</td>
<td>80</td>
<td>45</td>
<td>100</td>
<td>0</td>
<td>67</td>
<td>52</td>
<td>80</td>
</tr>
<tr>
<td>304</td>
<td>100</td>
<td>0</td>
<td>73</td>
<td>47</td>
<td>25</td>
<td>50</td>
<td>71</td>
</tr>
<tr>
<td>Mean</td>
<td>95</td>
<td>23</td>
<td>89</td>
<td>32</td>
<td>23</td>
<td>43</td>
<td>72</td>
</tr>
</tbody>
</table>

Note. Accuracy as mean percent correct (SD). StrNo = strong numerals; WNo = weak numerals

Table D4.

Children with autism’s Average Accuracy (SD) on Experiment 1b. Felicity Judgment Task

<table>
<thead>
<tr>
<th>Subject</th>
<th>All</th>
<th>And</th>
<th>Or</th>
<th>Some</th>
<th>StrNo</th>
<th>WNo</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>300</td>
<td>100</td>
<td>0</td>
<td>83</td>
<td>41</td>
<td>89</td>
<td>33</td>
<td>100</td>
</tr>
<tr>
<td>301</td>
<td>100</td>
<td>0</td>
<td>83</td>
<td>41</td>
<td>67</td>
<td>50</td>
<td>83</td>
</tr>
<tr>
<td>302</td>
<td>88</td>
<td>35</td>
<td>60</td>
<td>55</td>
<td>71</td>
<td>49</td>
<td>50</td>
</tr>
<tr>
<td>303</td>
<td>78</td>
<td>44</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>75</td>
</tr>
<tr>
<td>304</td>
<td>88</td>
<td>35</td>
<td>100</td>
<td>0</td>
<td>88</td>
<td>35</td>
<td>57</td>
</tr>
<tr>
<td>Mean</td>
<td>90</td>
<td>30</td>
<td>86</td>
<td>35</td>
<td>83</td>
<td>38</td>
<td>81</td>
</tr>
</tbody>
</table>

Note. Accuracy as mean percent correct (SD). StrNo = strong numerals; WNo = weak numerals
References


B Figure C. Trial Timelines for Experiment 1

*Note: SOA = stimulus onset asynchrony. Experiment 1b: Felicity judgment task is illustrated above. The SOA timelines are the same for Experiment 1a: Truth value judgment task.*


PMID: 7373248


