9-3-2013

Executive and Nonverbal Contributions to Pragmatic Language in Autism Spectrum Disorder

Ashley B. de Marchena Ph.D.
University of Connecticut - Storrs, ashley.demarchena@gmail.com

Executive and Nonverbal Contributions to Pragmatic Language in Autism Spectrum Disorder

Ashley Brooke de Marchena, Ph.D.

University of Connecticut, 2013

Pragmatic language deficits are universal in individuals with autism spectrum disorders. Pragmatic language skills require the integration of multiple communicative and social skills, and as such represent an intersection of two of the three major domains of impairment in ASD: communication and social interaction. Data from typically developing (TD) populations also suggests that pragmatic language is supported by complex skills such as gesture and the executive functions. Here we investigate *common ground*, a pragmatic language skill in which speakers adjust the contents of their speech based on their listener's perceived knowledge, in adolescents with ASD and TD. We designed an experimental narrative paradigm in which participants watched brief cartoons and then described the cartoons to a listener who sometimes shared knowledge about the cartoons and sometimes did not. While the TD sample reliably reduced the number of words in their narrations when the listener shared knowledge about the cartoons, consistent with the common ground literature, this common ground effect was not observed in the ASD sample. The tendency to show a common ground effect was not related to general skills such as IQ, receptive vocabulary, or executive function, in either group. The relationship between common ground and gesture use was difficult to interpret due to an order effect. In the ASD sample only, the common ground effect was positively correlated with age, such that participants 15 and over tended to show the effect while younger participants did not. These results suggest that the tendency to use common ground is relatively stable by adolescence in typical populations; however, it is still undergoing a period of development in adolescents with ASD, pointing to the importance of pragmatic language interventions at this sensitive age. Finally, we present data on gesture use in ASD, suggesting that teens with ASD may be more likely to use gestures when they fulfill a self-serving role rather than a communicative role. We discuss these results in the context of the broader literature on gesture and discourse skills in ASD.

Executive and Nonverbal Contributions to Pragmatic Language in Autism Spectrum Disorder

Ashley Brooke de Marchena

B.A., Tufts University, 2003

M.A., University of Connecticut, 2009

A Dissertation

Submitted in Partial Fulfillment of the

Requirement for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2013

APPROVAL PAGE

Doctor of Philosophy Dissertation

Executive and Nonverbal Contributions to Pragmatic Language in Autism Spectrum Disorder

Presented by

Ashley Brooke de Marchena, B.A., M.A.

Major Advisor: ____

Inge-Marie Eigsti

Associate Advisor: ____

Deborah Fein

Associate Advisor: ____

Marie Coppola

University of Connecticut

2013

Acknowledgements

This work and this degree would have been impossible without the support and encouragement of my many terrific research and clinical mentors. Thanks of course go to my fabulous mentor, Inge-Marie Eigsti. Your trust and encouragement of my abilities as a young scientist has meant so much, and your fascination with autism, language, and the brain is contagious. Thanks to Deborah Fein, for teaching Clinical Neuropsychology with both compassion and zeal. My clinical skills would not be what they are without years of training and heart-to-hearts with Marianne Barton and Sarah Hodgson, who have also supported my research interests all along the way. Finally, thank you to Marie Coppola and James Dixon for your collaboration and support on my committee. I feel truly privileged to have trained with a faculty that is as brilliant as it is humble and approachable.

In addition to my committee, I would also like to thank the many other individuals who I have had the pleasure to work with at UConn. Thanks to my classmates and labmates, and to the many undergraduate research assistants, without whom this research would have been completely impossible. That was a lot of coding.

I am also grateful for the financial support that facilitated my dissertation process, including the *Theodore Millon Dissertation Scholarship,* the *University of Connecticut Doctoral Dissertation Fellowship,* and a fellowship from *the Fulbright U.S. Student Program*. On that note, muchísimas gracias a Rosa García Pérez for listening and encouraging me as went through the writing process from abroad.

I could never have made it this far without the love and support of my amazing family, both biological and legal; each and every one of you has had a bigger impact on my life than you can possibly know. And of course to Matt, for encouraging me every step of the way, and patiently uprooting yourself to help me move forward down this winding career path.

Finally, these acknowledgements would not be complete without thanking the teenagers and families who participated in this study. You have taught me about both love and hard work, and how to get by and stay positive in this world when things are harder than they should be. Your persistence and creativity have truly been an inspiration.

Table of Contents

Page

Abstract……………………………………………………………………….. i
Cover Page……………………………………………………………………. iii
Approval Page…………………………………………………………………. iv
Acknowledgements……………………………………………………………. v
Table of Contents……………………………………………………………... vii
I. Introduction………………………………………………………………… 1
1.1 Discourse and Narratives in Autism Spectrum Disorder…………… 2
1.1a General linguistic features……………………………………… 2
1.1b Register………………………………………………………… 7
1.1c Conversational repair and clarity of reference………………. 10
1.1d Audience design……………………………………………... 14
1.2 The Current Study………………………………………….......... 15
1.3 Areas of Weakness in ASD that May Drive Pragmatic Language Deficits…………………………………….............................. 17
1.3a Executive function…………………………………………… 17
1.3a.a Working memory…………………………………... 18
1.3a.b Fluency…………………………………………….. 18
1.3a.c Inhibition…………………………………………... 19
1.3a.d Set-shifting………………………………………… 20
1.3b Gesture……………………………………………................ 22
1.4 Common Ground in Typical Development…………………………. 24
1.4a What is common ground?.. 24
1.4b Common ground as a dynamic vs. individual process………. 27
1.4c Common ground in children and special populations……….. 29
1.4d A debate: speaker-internal vs. listener-driven motivations for common ground use………………………………………………. 30
1.4e Online processing of common ground……………………… 37
1.4f Executive function and common ground…………………….. 42
1.5 Gesture……………………………………….......................... 45
1.5a Gesture and cognition………………………………………... 45
1.5b Gesture and communication……………………………….… 47
1.5c Gesture restriction….………………………………………... 49
II. Method…………………………………...……………………………… 52
2.1 Participants…………………………………………..................... 52
2.2 Measures…………………………………………....................... 55
2.2a Autism Diagnostic Observation Schedule……………………. 55
2.2b Social Communication Questionnaire……………………… 55
2.2c Social Responsiveness Scale………………………………… 55
2.2d Stanford-Binet Intelligence Scale: Fifth Edition…………….. 56
2.2e Peabody Picture Vocabulary Test: Third Edition…………… 56
2.2f Executive Function Measures………………………………... 56
2.2f.a Working memory…………………………………... 56

2.2f.b Verbal fluency……………………………………... 57
2.2f.c Inhibition…………………………………………… 57
2.2f.d Set Shifting………………………………………… 57
2.3 Design………………………………………........................... 57
2.4 Stimuli……………………………………….......................... 59
2.5 Procedures……………………………………….................... 60
2.6 Behavioral Coding……………………………………….......... 63
2.6a Speech transcription………………………….………… 63
2.6b Gesture coding…………………………………………..... 63
2.6c Ratings of story quality……………………………………… 64
III. Results…………..……………………………...……………….……… 64
3.1 Effects of Order……………………………………….............. 65
3.2 Story Characteristics…………………………………………...... 69
3.2a Baseline word count and gesture frequency…………………. 70
3.2b Enjoyment ratings and quiz performance…………………… 71
3.3 The Common Ground Effect……………………………………... 74
3.3a Common ground: Controlling for quiz performance………... 76
3.4 The Common Ground Effect Variable……………………………… 77
3.5 Age and Common Ground………………………………………….. 78
3.6 Gesture and Common Ground……………………………………… 80
3.7 Executive Function and Common Ground…………………..…… 81
3.8 Social Skills and Common Ground…………………………………. 84
3.9 Qualitative Ratings of Narratives…………………………………… 86
3.10 Gesture in a Non-Social Context………………………………….. 90
IV. Discussion……………………………………….............................. 93
4.1 Order Effect………………………………………..................... 93
4.2 The Common Ground Effect……………………………………... 100
4.2a Common ground and gesture……………………………….. 108
4.2b Common ground and executive function……………………. 110
4.2c Common ground and general skills………………………….. 111
4.2d Common ground and age……………………………………. 112
4.3 Gesture in ASD……………………………………….............. 114
4.3a Gesture in a non-communicative context……………………. 115
4.4 Limitations and Future Directions………………………………….. 117
4.5 Concluding Statements…………………………………………….. 121
References………………………………………................................ 124
Appendix A. Example quiz………………………………………........... 138
Appendix B. Questions given to story raters………………………………… 139
Appendix C. Results split by order…………………………………………… 140

"Adult speech, Edwin used to say, is ridiculously exclusive."

Edwin Mullhouse:
The Life and Death of an American Writer 1943 - 1954 by Jeffrey Cartwright
by Steven Millhauser

Pragmatic language skills are essential to effective communication and social interaction. Pragmatic language includes such diverse skill sets as reciprocal conversational skills, word choice based on specific conversational partners, and the comprehension and use of nonverbal aspects of communication that complement speech. Deficits in pragmatic language are universal in individuals with autism spectrum disorders (ASD; Paul, 2007; Tager-Flusberg, Paul, & Lord, 2005). Because pragmatic language requires the integration of multiple communicative and social skills, it represents an intersection of two of the three major domains of impairment in autism spectrum disorder (ASD): communication and social interaction. Despite the universality of pragmatic impairments in ASD, and the fact that they are often a major source of social anxiety for individuals with ASD (Landa, 2000), the mechanisms underlying pragmatic language impairments in this disorder have yet to be elucidated (Martin & McDonald, 2003).

In this study, we look at a specific pragmatic language skill, *common ground,* and its use in adolescents with and without ASD. Common ground refers to the tendency of interlocutors to modify how they communicate based on shared knowledge. In our study, participants had to describe short cartoon clips to trained confederate listeners; in some conditions, listeners had prior knowledge about the cartoons (i.e., speaker and listener *shared common ground*), and in some conditions, listeners had no knowledge about the cartoons. This allowed us to compare how participants' narratives changed as a result of sharing knowledge with their interlocutor. In addition to assessing common ground use itself, we also examine executive function and gesture, two domains that may underlie pragmatic language and which are likely limited in populations with ASD. Here, we review pragmatic language skills in ASD, with an emphasis on narrative and discourse. We also review executive function and gesture in this population, as potential underlying factors. Finally, we take a close look at common ground in typically developing (TD) populations, including relationships to executive function and gesture.

1.1 Discourse and Narratives in Autism Spectrum Disorder

1.1a General linguistic features. Although there are no published studies to date on common ground in ASD, the literature on discourse skills in this population informed our predictions. Overall, this literature reflects a wide range of difficulties. Storytelling and conversation (discourse domains that have been widely studied in ASD) require a complex set of skills ranging from purely linguistic abilities, such as morphology and syntax, to more basic language skills such as story organization and cohesion, to more purely social skills, such as maintaining listener interest and attention. In general, the literature suggests that verbal individuals with ASD may in fact employ many of the same strategies for conversation and storytelling; however, they employ these strategies less frequently than IQ-matched controls, and are more likely to produce inappropriate responses (Baltaxe & D'Angiola, 1996). Narratives and conversations can be analyzed at many levels, and as such, the conclusions reached by any given study depend heavily on the type and extent of analysis employed. In general, the discourse limitations observed in individuals with ASD relate closely to the core impairments associated with the disorder. Specifically, discourse produced by individuals with ASD reveals general linguistic deficits, limitations in emotion reasoning and theory of mind, and difficulty understanding causal structure, which may be secondary to executive functioning impairments.

In general, studies of discourse and conversation in individuals with ASD show global reductions in communicative output in this population. Narrative studies have shown that children with ASD produce shorter stories in response to the same stimuli (Tager-Flusberg, 1995; Ziatas, Durkin, & Pratt, 2003), although this finding has not been consistently replicated (Diehl, Bennetto, & Young, 2006; Loveland, McEvoy, & Tunali, 1990; Tager-Flusberg & Sullivan, 1995). To engage in a conversation or tell a story, certain linguistic competencies, such as sentence formation, must be in place. In a language-disordered population, such as ASD, we might expect weaknesses in the linguistic skills that underlie discourse. In fact, there is significant debate about how much of the discourse weaknesses observed in the population can be attributed to limitations in language skills. Although some studies have demonstrated that pragmatic deficits during discourse go beyond general impairments in language skills (Eales, 1993), the majority of studies find that general linguistic skills contribute significantly to narrative and discourse abilities in this population (Capps, Kehres, & Sigman, 1998; Capps, Losh, & Thurber, 2000; Liles, Duffy, Merritt, & Purcell, 1995; Norbury & Bishop, 2003).

In addition to linguistic structure, narrative and discourse studies can reveal the specifics of what storytellers may attend to. Often this can provide information as to what aspects of story content hold the most interest or salience for the speaker. In the case of ASD, narrative studies have revealed a lack of attention to the internal states and emotional experiences of story characters, consistent with the extensive literature on limitations in emotion perception in ASD (Prior & Ozonoff, 2007). Individuals with ASD have been shown to use fewer (Ziatas et al., 2003) and a more restricted range (Capps et al., 2000) of references to characters' emotions and internal states. In addition to referring to characters' mental states less frequently, those with ASD are also less likely to use evaluative devices to explain *why* a character might feel the way that he or she does (Capps et al., 2000). For example, a child with ASD might state that a boy was sad or was crying, but would be less likely to explain that this was because he lost his pet frog. In addition to making fewer references to mental states in general, and being less likely to explain the causes underlying these mental states, mental state references by individuals with ASD tend to focus more on desire, and less on belief (Ziatas et al., 2003). Even children who have effectively recovered from ASD (i.e., "optimal outcome" children), continue to demonstrate narrative weaknesses including reduced references to characters' goals and motivations (Kelley, Paul, Fein, & Naigles, 2006).

A consistent weakness that has been observed via the narratives produced by those with ASD is a failure to sequence story events in such a way that the causal structure of narratives is apparent. Related to this phenomenon are findings that individuals with ASD are less likely to use linking statements to connect distinct aspects of stories or conversations (Baltaxe & D'Angiola, 1996). In one of the first studies to examine causal structures produced by individuals with ASD, children with ASD were matched on verbal ability to children with Down Syndrome (Loveland et al., 1990). Participants with ASD were more likely to list the details in the story, rather than to tell the story as a coherent sequence of events. Many other studies have replicated the finding that children with ASD use less causal language than controls when telling stories (Diehl et al., 2006; Kelley et al., 2006; Liles et al., 1995; Losh & Capps, 2003; Tager-Flusberg, 1995).

The problems underlying the lack of inclusion of causal structure in the narratives produced by individuals with ASD are not completely clear. Given the social and pragmatic weaknesses experienced by those with ASD, it could be that when telling stories, individuals with ASD simply do not understand the need to causally connect story events for their interlocutors. That is, they may believe that presenting the details of their story is sufficient for listeners to be able to understand or infer the causal connections between them. However, the evidence suggests that such an explanation is insufficient to account for the observed reductions in causal language. First of all, there is some evidence that the use of causal language is dependent on more basic linguistic skills. For example, children with SLI tell stories with less cohesion than TD children (Liles et al., 1995). Second, children with ASD may have trouble understanding causal structure independently of narrative formation (Gopnik, Capps, & Meltzoff, 2000), which may itself be secondary to the impairments in executive functions that have been observed in this population (Eigsti, 2011; Hill, 2004; S. Ozonoff, B. Pennington, & S. Rogers, 1991a). If the causal structure of a narrative is not understood by the speaker, then it will not be communicated either.

It is no surprise that in domains as complex as discourse, there appear to be multiple dynamic relationships between underlying skills. Because there are so many skills involved in discourse, and only a limited selection of these skills can be measured in any given study, it follows that the relationships between these skills often appear to be different based on the specific research question being asked. The importance of theory of mind skills in narrative competence is a clear example of a possible red herring in narrative weakness. Findings that children with ASD often omit explanations for others' mental states (Capps et al., 2000), that they are generally impaired on false belief tasks and other assessments of theory of mind (Baron-Cohen, 2000), and that performance on theory of mind tasks is correlated with measures of narrative skill (Capps et al., 2000; Tager-Flusberg & Sullivan, 1995), led to the reasonable conclusion that theory of mind deficits underlie the struggles in communicating causal structure demonstrated by many individuals with ASD. However, it may in fact be the case that other factors, such as general linguistic skills or understanding of causal relationships might be driving *both* weaknesses in narrative competence *and* theory of mind performance. For example, Capps and colleagues (1998), like others, found a relationship between conversational abilities and theory of mind performance. However, when they accounted for general linguistic skills, they found the relationship between theory of mind and conversational skills to be attenuated. Similarly, children with SLI have also been shown to use fewer mental state terms in their narratives compared to mental-age matched peers, but use these terms to the same extent as their language-age matched peers, again suggesting that inferences about others' internal states during discourse, including the use of mental state terms, may rely heavily on general linguistic skills rather than simply theory of mind (Johnston, Miller, & Tallal, 2001).

Further, linguistic competence itself may drop as the interpersonal demands of a task increase or the structure of a task decreases, thereby taxing the cognitive and emotional resources available to the child with ASD. For example, Losh and Capps

(2003) compared highly structured storybook narratives to personal narratives in which participants were asked to explain a personal experience such as a favorite vacation. They found that the ASD group used less complex grammar than the TD group in the personal narratives condition only; the ASD group also used a more restricted range of complex syntax in the personal narratives condition. This finding shows that there is likely a dynamic relationship between linguistic skills and more qualitative aspects of narratives that fluctuates as task demands change.

Although no studies to our knowledge have directly investigated the use of common ground by individuals with ASD, several social and linguistic domains that have been studied in this population have implications for hypotheses about common ground usage in ASD. Specifically, studies of register, conversational repair and clarity of reference, and pronoun usage inform our own research questions.

1.1b Register. The concept of register relates closely to common ground. In linguistics, the term ‘register’ refers to changes in language level based on social context. For example, I speak in a different register when explaining my thesis to my doctoral committee than when explaining the project to a new Spanish friend that I meet at a party. Register has both a social component (e.g., speaking to someone with deference vs. authority) and a communicative component (e.g., using simplified language to facilitate the comprehension of someone with limited language skills). Like common ground, register adjustments require speakers both to infer a listener’s needs, and to adjust to the discourse contents based on these perceived needs. When typical adults address listeners with less knowledge of their language (e.g., foreigners or young children), they tend to simplify their syntax and vocabulary (Andersen, 1990). Parallels can be drawn to findings that adult speakers modify their linguistic output when they share knowledge with their interlocutors (Krauss & Weinheimer, 1966). Use of register differs from use of common ground in that it depends less on interlocutors' direct awareness of each others' knowledge and more on a gestalt change in communicative output. Register is similar to common ground, however, because speakers must infer something about listeners' knowledge and use this information in their utterance planning. TD children as young as three years old have been shown to adjust their register down when speaking, for example, to a baby or a doll, relative to a parent or peer (Sachs & Devin, 1976; Shatz & Gelman, 1973). In this respect, common ground and register are likely often co-specified. For example, when discussing the American political system, in English, with a semi-fluent Spaniard, I am likely to simplify the linguistic aspects of my speech, such as my vocabulary and grammar (thus changing my register). I am also likely to modify the content of my speech based on the assumption that a Spaniard will know less about the American political system than I, or my American friends, do (thus modifying my language based on our perceived common ground). The same analogy could be made to discussing the government with a child.

Volden and colleagues (2007) tested 38 high-functioning six- to sixteen-year-old children with ASD. These children were compared to two control groups: one group was matched on nonverbal mental age, and the other was matched on language age. Participants were asked to describe how to "go to a restaurant" to an adult, and to puppets representing a peer, a baby, and a nonnative speaker. Overall, compared to the control groups, the ASD group provided longer explanations with fewer components deemed "core" or "crucial" aspects of the explanation (e.g., ordering your meal), and more components that were deemed inappropriate (e.g., going to the dentist), a finding that fits well with the narrative literature described above. However, the ASD group did succeed in adjusting to listeners' needs, when talking to the baby and nonnative puppets, by reducing the overall length and total number of acts included in their explanations, and including a greater proportion of core explanations. Although successful in a general downward shift in register, the ASD group continued to give longer explanations that included more tangential and irrelevant information. This likely reflects general pragmatic deficits, however, and not specific problems with register adjustment per se.

After the first trial, given to assess spontaneous communication, participants in this study were given more specific prompts (e.g., *she didn't understand that, make sure to explain it so that she can really understand*) to see if they would adjust their language further. Both prompts further reduced the complexity of explanations in both groups. However, although they were attenuated, the same group differences persisted following the prompts, again suggesting that the reduced quality of the ASD group's explanations was due to general pragmatic deficits and not to specific difficulty with adjusting to register. Volden and colleagues (2007) suggest that the use of prompts may have reduced some of the executive demands associated with registral shift (e.g., the child did not have to make a choice about what register to use, as the experimenter specifically instructed him that the listener did not understand, and that he should use simpler language). When demands were decreased, by giving prompts, children were better able to adjust their register. This supports the notion that some pragmatic language deficits in ASD may be secondary to executive weaknesses. Another possible explanation, of course, is that participants with ASD simply did not consider the need to adjust their register, but once prompted to do so, they were able to. In a related study, children and adolescents with ASD were successfully able to modify their register when asked to make “bossy” and “nice” requests of a puppet (Volden & Sorenson, 2009). They were also able to comprehend the same shifts in register, in that they were able to judge which requests (made by the puppet) were bossy and which were nice.

Although register and common ground are certainly distinct skills, the setup of Volden's 2007 study likely represented an example of co-specified registral and common ground needs. Specifically, the “foreigner” puppet was presented as someone “visiting from China” – so shifts in language could also reflect an assumption of less common ground due to cultural differences. The prompt when participants were presented with the baby puppet was similar: “he’s young so he doesn’t understand how everything works,” again implying a lack of general knowledge in addition to a specific lack of linguistic knowledge. Because the setup of the study implies that the participants’ interlocutors lacked both knowledge about the subject matter and linguistic knowledge, participants’ change in *register* could in fact reflect their use of common ground as well. As such, these findings suggest that the use of common ground may actually be an area of relative strength for individuals with ASD.

1.1c Conversational repair and clarity of reference. Another area of study that has implications for the use of common ground is the small literature on conversational repair in ASD. Conversational repair refers to what speakers do when it is clear that their listener has missed, or misunderstood, part of the message they were intending to convey. For example, Geller (1998) found a general failure by children with ASD to repair conversational misunderstandings. The children in this study appeared to recognize breakdowns in communication; however, they had trouble using their linguistic skills flexibly enough to repair these breakdowns. Unfortunately, this study only included five children (and no control group), so it is hard to generalize the findings. Volden (2004), in contrast, found that found that children with ASD responded to requests for clarification similarly to TD children matched on language. However, the children in the ASD sample were more likely to give inappropriate responses following requests for clarification. Interestingly, the ASD sample also used gestures in an attempt to repair communicative breakdowns to the same extent as the TD sample. This finding in particular is very consistent with the literature on narratives in ASD. Volden's (2004) sample was also small, limited to only nine children with ASD and nine control participants with TD or minor developmental delays. Due to the small sample size of both of these studies, and the limited amount of research in this interesting area to date, the specifics of conversational repair skills in ASD remain unknown.

Clarity of reference, for example, in the use of determiners and pronouns, is another aspect of narrative performance that is closely linked to the phenomenon of common ground. In a study comparing TD children, children with SLI, and children with ASD using the classic *Frog, Where Are You?* narrative, group differences emerged *only* in terms of the clarity of references produced during the story (Norbury & Bishop, 2003). The authors assessed story organization at global (overall story structure) and local (syntax) levels. They also looked at evaluative comments made by the participants, including comments on characters' states of mind, hedges (which suggest uncertainty about story events), and causal connectives, which link a character's emotion or behavior to a story event. The SLI and ASD groups both made more syntactic errors than the TD group, and the SLI and ASD groups both used more ambiguous pronouns than the TD group, but the ASD group alone tended to confuse determiners ‘a’ and ‘the,’ often using ‘a’ when ‘the’ should have been used, suggesting the presence of an additional character that was not in fact there (three of the 12 participants with ASD made this error, compared to one of 56 control participants). Does this reflect the fact that children with ASD were not taking the listener’s continuing knowledge of the main character into account? Or does it reflect general linguistic weaknesses in understanding the difference between definite and indefinite determiners? The fact that this study included an SLI control group suggests the former. However, these results could also simply reflect poor comprehension of the story itself. In a similar study, children with ASD were found to make the same number of ambiguous statements and unclear references, including using ambiguous pronouns, when compared to children with Down Syndrome (Loveland et al., 1990).

Although several authors who study pragmatic language in ASD have suggested that executive function, specifically set-shifting and flexibility, may be limiting factors for the fluent use of pragmatic language (e.g., Volden, 2004), there is only one study, of which we are aware, that has directly investigated the relationship between processing demands and pragmatic language in ASD. Arnold and colleagues (2009) looked at “referring expressions” in children and adolescents with TD and ASD. Specifically, they investigated the frequency of underspecified referential expressions, that is, pronouns and zeros (e.g., Jane went to the mall, [zero] ate lunch, [zero] and bought a shirt) in children with ASD while narrating a cartoon. They argued that the cognitive load hypothesis of ASD would predict that the ASD sample would use more explicit references and fewer pronouns than the comparison group. This prediction was based on findings that speakers are less efficient in their use of pronouns and zeros under conditions of high cognitive load (Arnold & Griffin, 2007). Overall, they found that the ASD group used fewer references, and produced shorter narratives. However, there were few group differences in the use of underspecified references, with the exception that the youngest subset of their ASD sample (i.e., nine- to twelve-year-olds) was least likely to use pronouns, particularly for characters who had not just been mentioned in the previous clause.

In addition to looking at underspecified pronominal references on their own, Arnold and colleagues also looked at the effect of cognitive load on referential choices. Since individuals with ASD are thought to struggle with complex information processing (Minshew, Goldstein, & Siegel, 1997), it was predicted that they would behave like typical participants under conditions of high cognitive load. Although they did not manipulate cognitive load directly, participants were said to be experiencing high cognitive load when (1) their speech was dysfluent, or (2) when they produced clauses with unusually high word counts (because longer clauses take more planning, the assumption was that more resources were required to produce these utterances). Participants in both groups were more likely to use pronouns in their narratives when they were experiencing less cognitive load, suggesting that adhering to discourse rules about pronoun usage requires significant cognitive resources. In addition, there was a marginally significant effect of clause length on pronoun and zero usage in the predicted direction, such that longer clauses resulted in reduced use of pronouns *and* zeros. The results of this study suggest that cognitive load is important in discourse processing in both TD individuals and those with ASD. In general, the ASD group was quite successful at this task, suggesting that there are factors other than social skills and mentalizing driving the pragmatics of pronoun usage; we discuss several possible contributors in the *Areas of Weakness in ASD that May Drive Pragmatic Language Deficits* section, below.

An early study of pragmatic language in ASD is also fairly relevant to the domain of common ground usage. Baltaxe (1977) found that children with ASD were less likely to foreground and background information that would allow listeners to differentiate between new and old information, for example, in the use of definite articles. Like Arnold and colleagues' (2009) study on pronoun use during storytelling, this early finding also suggests that individuals with ASD may provide overly explicit references in their narrative production, supporting the hypothesis that our ASD sample would fail to use common ground to constrain their speech, perhaps because of a general tendency to use overly explicit references in their language production.

1.1d Audience design. During conversation, TD children and adults take the needs of their interlocutors into account as they are forming utterances; thus, speakers' utterances are tailor-made to specific listeners and specific conversations. This important and ubiquitous pragmatic language process is sometimes called *audience design* (Clark, 1996; Clark & Carlson, 1982). Nadig, Vivanti, and Ozonoff (2009) examined this process in children with ASD. Children in their study completed two tasks, a referential communication task and a hidden object task, in which child participants filled the speaker role. Three increasingly complex levels of audience design were assessed. The authors found group differences at all three levels, with participants with ASD providing less efficient and context-appropriate descriptions. Interestingly, the authors found individual differences within the ASD group, with some participants demonstrating no audience design, some demonstrating audience design at one or two levels, and some demonstrating audience design at all three levels. This suggests that, while audience design may not generally be used as often in ASD samples, that it is incorporated in the discourse of some high-functioning individuals. Nadig and colleagues questioned what underlying factors might contribute to the finding that some participants with ASD to showed audience design while others did not. They found that while symptom severity and social skills did not account for individual differences in audience design, that participants with higher levels of structural language skills were more likely to use audience design.

1.2 The Current Study

The literature on narratives and discourse in ASD leaves questions about common ground in this population wide open. On the one hand, individuals with ASD struggle with multiple aspects of discourse and pragmatic language. On the other hand, certain skills, such as registral shifts, that are closely related to common ground, appear to be relative strengths within the domain of pragmatic language.

The current study utilized a narrative task to explore the use of common ground in high-functioning adolescents with ASD. Narrative studies in TD samples typically find that speakers reduce the amount of their speech when they share knowledge with an interlocutor (this phenomenon will be explored in detail, below). We seek to discover whether individuals with ASD will respond similarly to the presence of common ground. In the TD literature, the debate over what processes underlie and support the use of common ground during discourse is far from settled. On the surface, the use of common ground may appear purely dynamic in nature; that is, establishing shared knowledge and using this knowledge within a discourse seems entirely dependent upon the relationship between interlocutors. However, there is sufficient evidence to believe that other processes, processes that lie within the individual rather than within the relationship, are involved, and in fact may even go beyond social skills. One well-supported line of research suggests that common ground requires complex information processing (Horton & Gerrig, 2005a, 2005b; Horton & Keysar, 1996; Roxβnagel, 2000) and involvement of the executive functions (Brown-Schmidt, 2009; Nilsen & Graham, 2009; Schuh, 2010). A more nascent line of research suggests that certain nonverbal skills, specifically, gesture, may help support the use of common ground during discourse (Holler & Wilkin, 2009). Both of these skill sets (i.e., information processing/executive function and nonverbal communication) are thought to be impaired in ASD. Thus, in addition to testing for the simple presence or absence of the common ground effect in adolescents with ASD, we will also explore what potential limitations might be making common ground less accessible for these individuals.

We predicted that, while a common ground effect will be clearly observable in our typical sample, adolescents with ASD would fail to demonstrate this effect. Given the large amount of processing required by the pragmatics of common ground, and the evidence that individuals with ASD have limited executive function skills, one possibility is that those with ASD may fail to use common ground during discourse because their available resources have been consumed simply by the task of narrating a story, with which they are also known to struggle. To test this possibility, and to explore the role of executive function in typical common ground processing, we compared performance on the experimental common ground task to a series of standardized executive function measures designed to test working memory, verbal fluency, inhibition, and set shifting. Based on the literature, we predicted that, regardless of diagnosis, individuals with better executive function skills would show a greater CG effect.

Stimuli for the common ground task were clips from Looney Tunes cartoons that were selected to be highly visual in nature and to include little to no speech. Highly visualizable cartoon stimuli were specifically selected based on the literature suggesting that narratives of such cartoons should be the most sensitive to the effects of gesture manipulations. To examine the role that gesture plays in the common ground effect, participants were gesture-restricted on half of the experimental trials, allowing us to look at the common ground effect when gesture was not available to our participants. We predicted that gestures would allow participants to maintain representations that might be cut from speech due to the pragmatic demands of common ground. Thus, the common ground effect would be attenuated in our TD sample when participants were restricted from gesturing, because such representations would no longer be maintained by gesture. These data provide insight not only into if and how common ground is used in adolescents with ASD, but also into what processes guide and underlie the use of common ground in typical development as well.

1.3 Areas of Weakness in ASD that May Drive Pragmatic Language Deficits

1.3a Executive function. Executive function is known to be impaired in ASD (Hill, 2004; Hughes, Russell, & Robbins, 1994; Ozonoff et al., 1991a; Robinson, Goddard, Dritschel, Wisley, & Howlin, 2009). Some authors have proposed that executive deficits can account for some of the major symptoms seen in ASD (Russell, 1997a), including impairments in theory of mind (Russell, 1997b), and repetitive behaviors (Turner, 1997), although the centrality of executive function deficits to the core features of ASD has been debated (e.g., Liss, Fein et al., 2001). A complete review of this debate is outside of the scope of this study; however, a brief summary of what is known about how some of the specific executive functions present in individuals with ASD is relevant to our hypotheses and study design. Specifically, we review the four executive functions that were tested in the current study: *working memory, fluency, inhibition, and set shifting.*

1.3a.a Working memory. Working memory consists of an individual's ability to hold information in mind while manipulating or mentally operating on that information. Within the context of pragmatic language, and common ground specifically, working memory is relevant because a speaker must hold multiple sets of information in mind, including the narrative of the story *and* what their listener does and does not know about the narrative, while operating on this information to tell the story with or without reference to any existing common ground. Some studies have shown impaired working memory processes in individuals with ASD relative to typical peers (Bennetto, Pennington, & Rogers, 1996; Joseph, Steele, Meyer, & Tager-Flusberg, 2005; Minshew & Goldstein, 2001); however, other studies have failed to find a difference between ASD samples and control groups in terms of working memory abilities (Ozonoff & Strayer, 2001; Russell, Jarrold, & Henry, 1996). This discrepancy in findings suggests that there are likely subtle, task-dependent differences in how working memory is utilized in individuals with ASD.

1.3a.b Fluency refers to the ability to generate information rapidly and efficiently. Specifically, verbal fluency refers to an individual's ability to generate verbal information, for example, exemplars of a category, or words starting with a certain letter. Fluency is relevant to our experimental task because participants were asked to generate verbal narratives quickly and with relatively little time for planning. Although fluency has not been studied as extensively as working memory in ASD, the findings on fluency in ASD have also been inconsistent. Some studies have shown general impairments in fluency in individuals with ASD as compared to controls (Minshew, Goldstein, Muenz, & Payton, 1992; Rumsey & Hamburger, 1988; Turner, 1999). These fluency impairments have been theorized to be related to the struggle that children with ASD demonstrate when trying to come up with new ideas. However, several other studies have failed to find fluency impairments in ASD (Boucher, 1988; Minshew, Goldstein, & Siegel, 1995; Robinson et al., 2009; Scott & Baron-Cohen, 1996), even in samples where other executive function deficits have been observed (Robinson et al., 2009).

1.3a.c Inhibition is an executive function that requires the ability to attend and respond to the relevant features of a stimulus while simultaneously inhibiting a potentially more salient response in favor of a less salient or automatic response. The classic test of inhibition is the Stroop task, in which a participant is presented, for example, with color words (‘blue’, ‘red’, etc.) printed in different-colored ink. For fluent readers, the written word is more salient, so participants are asked to name the color of the ink while inhibiting the impulse to read the printed word. While most studies of Stroop performance in ASD do not find impaired performance relative to controls (Barnard, Muldoon, Hasan, O'Brien, & Stewart, 2008; Ozonoff & Jensen, 1999; Ozonoff & Strayer, 1997), some early studies (e.g., Rumsey & Hamburger, 1988; Rumsey & Hamburger, 1990) and one recent study (Robinson et al., 2009) have found impaired performance on the Stroop, suggesting poor inhibitory functioning. Other inhibition tasks that do not require fluent reading have also largely failed to consistently demonstrate impairments in ASD, with some studies finding intact performance (Ozonoff, Strayer, McMahon, & Filloux, 1994), some finding impairments (Hill & Bird, 2006; Robinson et al., 2009), and some finding impairments that are not specific to autism (Bishop & Norbury, 2005). Although inhibition does not appear to be specifically impaired in ASD, we have chosen to include a test of inhibitory processing because of the connection that has been shown with the ability to use common ground in typical populations (Brown-Schmidt, 2009). To effectively reduce verbal output in conditions of common ground, speakers must actively inhibit themselves from including details that are not relevant to a knowledgeable interlocutor.

1.3a.d Set-shifting is an executive function that requires the ability to flexibly shift from one cognitive "set" to another, for example, from one type of response to a different, contradictory response. When using common ground to modulate discourse, speakers must shift back and forth between knowledge that is known only to themselves, and knowledge that is shared with a listener. The most well-studied task of set shifting is the Wisconsin Card Sort Test (WCST; Berg, 1948), which requires respondents to match cards based on a single feature (e.g., the shape of items on the card) that changes as the task progresses (e.g., to the number or color of items on the card). Across a wide age and functioning range, many studies have shown impairments on the WCST in ASD samples relative to controls (Kaland, Smith, & Mortensen, 2008; Liss, Fein et al., 2001; Ozonoff & Jensen, 1999; S. Ozonoff, B. F. Pennington, & S. Rogers, 1991b; Shu, 2001), including control samples with other cognitive and learning impairments (Rumsey & Hamburger, 1990; Szatmari & Tuff, 1990). Most studies report that individuals with ASD do not perform as well as controls on the WCST; however, it should be noted that several studies have failed to find group differences (Barnard et al., 2008; Liss, Harel et al., 2001; Minshew et al., 1997). A more recently developed and more specific task of set-shifting, the Intradimensional/Extradimensional Shift task from the Cambridge Neuropsychological Test Automated Battery (Robbins et al., 1994), has also been tested in ASD samples, with some (Hughes et al., 1994; Ozonoff et al., 2004), but not all (Landa & Goldberg, 2005) studies finding impairments in set-shifting in the ASD sample relative to controls.

In a comprehensive attention battery completed with a large sample of children and adolescents with high-functioning ASD and a large sample of controls, sustained attention and information encoding were found to be intact in the ASD sample relative to the controls (Goldstein, Johnson, & Minshew, 2001). However, tests assessing attentional focus and execution, and the ability to shift attention, were impaired in the ASD group relative to controls, suggesting that cognitive flexibility and psychomotor speed were the main areas of attentional impairment in ASD.

Relevant to the current study, Nancy Minshew and colleagues have proposed that ASD is not the result of a single underlying processing deficit (such as a specific executive function), but rather results from impairments in complex, higher-order information processing (Goldstein et al., 2001; Minshew & Goldstein, 1998; Minshew et al., 1997). That is, autistic symptomatology does not emerge from a single, underlying limitation in low-level processing, but rather from difficulties in complex, high-level processing across multiple cognitive domains. This theory has been supported by findings of intact or superior performance on test of low-level memory, language, attention, motor, and visual-spatial domains, alongside impairments, in the same sample, in reasoning, complex language, complex memory, and skilled motor tests (Minshew et al., 1997). Deficits in higher-order processing are of course consistent with the social symptoms that are the hallmark of ASD, as social interaction itself is an almost inconceivably complex set of skills. As described above, common ground requires complex information processing, thus, if individuals with ASD are limited in this domain, it stands to reason that they will struggle with making common ground adjustments.

1.3b Gesture. The presence of communicative gesture deficits in ASD has been widely asserted in the clinical literature. Impairments in gesture are codified on gold-standard ASD diagnostic measures and screeners such as the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 2002), the Autism Diagnostic Interview (ADI; Lord, Rutter, & LeCouteur, 1994), and the Modified Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, Barton, & Green, 2001); on these measures, the absence or infrequency of gesture is rated as symptomatic. Scoring criteria for these diagnostic measures suggest that individuals with ASD use *all* gestures less than their typically developing (TD) peers; however, this assertion has not been demonstrated in the empirical literature. For example, protodeclarative pointing (i.e., pointing to share attention), but not instrumental pointing (i.e., pointing to request), is found to be reliably reduced in frequency in ASD (M. A. Bono, T. Daley, & M. Sigman, 2004; Camaioni, Perucchini, Muratori, & Milone, 1997; Loveland & Landry, 1986; Mundy, Sigman, & Kasari, 1990; Mundy, Sigman, Ungerer, & Sherman, 1986), a finding which highlights that the *social* aspects of these gestures likely contribute to their delayed production (Baron-Cohen, 1989; Klin, Jones, Schultz, Volkmar, & Cohen, 2002).

The literature on non-pointing gestures in ASD is sparser and less conclusive; specifically, reductions in gesture frequency are not well replicated. For example, early studies found reduced rates of gesture in children with ASD (Bartak, Rutter, & Cox, 1975), and more recent studies have observed gesture delays (Charman et al., 2003; Luyster et al., 2007). However, many others have failed to find group differences in gesture frequency after controlling for speech production (Attwood, Frith, & Hermelin, 1988; Capps et al., 1998; de Marchena & Eigsti, 2010; García-Pérez, Lee, & Hobson, 2007; Tantam, Holmes, & Cordess, 1993). In fact, some have found iconic gestures to be a relative communicative strength for children with ASD, perhaps because enacting experiences is more accessible than verbalizing them (Capps et al., 1998). Children with ASD also show a reduced variety of gestures (Colgan et al., 2006; Wetherby & Prutting, 1984), which may strengthen the impression of fewer gestures. As described above, gestures serve many different cognitive and communicative roles. It is likely the case that gesture use by individuals with ASD differs from gesture use in TD in a task-dependent way; that is, individuals with ASD may use fewer gestures in some contexts (e.g., a social context, as in the case of protodeclarative pointing), similar amounts of gesture in other contexts (e.g., during play or conversation), and perhaps even more gestures in other contexts (e.g., during problem solving). Although this possibility has not been investigated to a great extent, preliminary data from our lab, using a range of cognitive and communicative tasks, suggests that this might in fact be the case (Eigsti, de Marchena, & Dixon, 2010).

In addition to the question of gesture frequency in ASD, it certainly remains an open question whether gestures produced by those with ASD benefit both the speaker and listener to the same extent as gestures produced by TD individuals (see below). Here we will examine not only the role that gestures play in common ground processes by TD adolescents, but also the role that gesture plays for those with ASD, in both a communicative and an executive function task.

1.4 Common Ground in Typical Development

According to the principle of audience design, not only do speakers tailor their utterances to their listeners, but listeners in turn make use of this fact when interpreting speakers' utterances (Clark & Murphy, 1982). Thus, audience design is a dynamic process that facilitates communication in terms of both production and comprehension. To facilitate the process of audience design, interlocutors often consider shared knowledge, or *common ground*. That is, speakers consider the knowledge that they share with listeners when planning their utterances (Clark & Bernicot, 2008; Clark & Wilkes-Gibbs, 1986; Fussell & Krauss, 1992; Horton & Keysar, 1996; Isaacs & Clark, 1987; Krauss & Weinheimer, 1966; Lockridge & Brennan, 2002; Nadig & Sedivy, 2002; Nilsen & Graham, 2009; Roxβnagel, 2000).

1.4a What is common ground? According to Grice's maxim of *quantity*, during discourse, speakers should (1) Make their contributions as informative as is required, but also, (2) *Not* provide information that is irrelevant, distracting, or otherwise detracts from the discourse (Grice, 1975). This delicate balancing act requires speakers to provide just the right amount of information to express themselves, while not providing so much information as to make the content of their language irrelevant or inappropriate.

Unsurprisingly, children with ASD have been shown to be less sensitive to Grice's conversational maxims when compared to language-matched peers with specific language impairment (SLI; Surian, 1996). Anecdotally, individuals with ASD are known to violate the Gricean maxim of quantity in both directions. Imagine, for example, asking someone, "do you have a favorite movie?" Almost anyone who has worked with individuals with ASD has experienced a violation of the maxim of quantity in response to a similar query; the speaker with ASD may provide too little information, for example, by responding, simply, "yes;" or too much information, for example, "yes, my favorite movie is *The Lion King*, I saw it Tuesday night with my sister Samantha, the lions are Simba, Nala, Mufasa, Sarabi, and Scar, the hyenas are Shenzi, Banzai, and Ed, etc…." The "appropriate" amount of information to provide often falls somewhere in the middle, for example, "yes, my favorite movie is *The Lion King.*" The appropriate amount of information varies based on multiple factors. For example, speakers tend to take a listener's knowledge into account when planning utterances and deciding how much information to include in their speech (Gundel, Hedberg, & Zacharski, 1993). Thus, the presence of shared knowledge shifts the quantity of information that is considered appropriate in a conversation. Specifically, listeners require, and speakers provide, less information when they share knowledge, or *common ground.*

Common ground, or shared knowledge with an interlocutor, is incorporated seamlessly into our conversations and significantly affects the content of our speech. Studies of common ground in TD adults generally find that speakers use fewer words and conversational turns when they share information with an interlocutor (Fussell & Krauss, 1992; Holler & Wilkin, 2009; Isaacs & Clark, 1987). This shortening process typically occurs when conversational partners establish shared referential forms that are used consistently throughout their discourse, replacing lengthier descriptions. For example, in the earliest of these studies, utilizing what is known as a referential communication task, Krauss and Weinheimer (1966) asked pairs of college students to work collaboratively with a set geometric figures (Tangrams). At first, participants used extended descriptions of the figures (e.g., "*the rectangle with two triangles under it*"), but they quickly settled on shorter referential forms (e.g., "*the coffee table*") that were used from then on (also see Clark & Bernicot, 2008; Clark & Wilkes-Gibbs, 1986; Isaacs & Clark, 1987). Referential communication tasks demonstrate that for successful communication, *both* interlocutors must engage in a back-and-forth dynamic process to establish shared references during conversation, consistent with principles of audience design. After going through the process of lexical entrainment – essentially, a process of creating a new vocabulary or shorthand, as is accomplished during referential communication tasks – reduced forms such as "the coffee table" are established, and interlocutors interpret them reliably and consistently (Clark & Wilkes-Gibbs, 1986; Gordon, Grosz, & Gilliom, 1993). Further, overly explicit references can actually slow comprehension (Gordon et al., 1993; Hudson-D'Zmura & Tanenhaus, 1998), providing even more incentive to establish shorter, more efficient forms. In summary, reducing the length of utterances referring to information contained within interlocutors' common ground adheres to Grice's conversational maxim of quantity, facilitating communicative effectiveness.

It should be noted that lexical entrainment does not *universally* lead to reduced referential forms. For example, Brennan and Clark (1996) set up a conversation scenario in which reduced forms did not pick out unique referents, thus participants used longer, subordinate terms (e.g., “the high heel,” instead of, “the shoe,” because multiple shoes were present). Even after the context changed and the reduced forms did in fact pick out unique referents, participants continued to use the longer forms, *because* they had already been established within the discourse. Interestingly, speakers continued to use these same forms even when collaborating with a new interlocutor. This study demonstrates that although lexical entrainment generally leads to reduced forms, there are multiple factors that lead to the establishment of terms within a discourse, which can sometimes result in the use of longer, or simply, different labels.

1.4b Common ground as a dynamic vs. individual process. In referential communication tasks, such as the seminal 1966 Tangrams study by Krauss and Weinheimer, the process of lexical entrainment is highly dynamic in the sense that it depends on a back-and-forth interaction between interlocutors. Other evidence for the process being dynamic comes from findings that when adults overhear a conversation between two interlocutors with established common ground, they are less likely to understand what is being discussed than those who were actively participating in the conversation, even if the overhearer is present from the beginning (Schober & Clark, 1989). This finding suggests that there is something specific to being involved in lexical entrainment, or establishing common ground, over and above just being exposed to the common ground, that allows it to facilitate conversation. There are also cases, however, in which the introduction of terms into common ground can work in a less dynamic way. For example, in a didactic context, experts introduce specific terms into the common ground that are subsequently learned and incorporate by students (Isaacs & Clark, 1987).

The findings of early referential communication tasks described above have clearly demonstrated the dynamic processing involved in establishing common ground. Recent research has further uncovered the fact that common ground use is also driven by processes within the speaker, such as memory and executive function. Thus, the process is not exclusively due to speakers' sensitivity to listeners' communicative needs. In classic referential communication tasks, interlocutors tend to have equivalent roles in the discourse, and work together to establish terms that are incorporated into their common ground. This methodology has provided fascinating insights into the discursive processes that transpire between interlocutors; however, because the process is completely dynamic, these methods do not allow for a close look into the processes that are going on within *individual* interlocutors during utterance production and comprehension. In contrast, most studies that have attempted to specify the internal processes that underlie common ground have used slightly different methodologies, in which interlocutors are given specific "speaker" and "listener" roles, for example, in a narrative or decision-making task. The speaker is then evaluated to investigate processes specific to production based on common ground, while the listener is evaluated to investigate processes involved comprehension based on common ground. The current study focuses on the speaker (as is often the case, the listener is a research assistant) because we are interested in production. In addition, in the current study, there is a parametric (experimental) manipulation of common ground.

In spontaneous conversation, common ground adjustments by the speaker require two distinct but related processes, *commonality assessment* and *message formation* (Horton & Gerrig, 2005a). Commonality assessment involves searching memory to determine what information is shared with an interlocutor; message formation is the process of determining how to modify communicative output based on this shared information. In the current study, participants were explicitly told what information their listener (i.e., the research assistant) shared. Although participants did not have to make a subjective judgment as to which information the listener knew, they did have to hold this information in memory, and use it during message formation.

1.4c Common ground in children and special populations. There is considerable debate as to when the use of common ground emerges during development. In general, children tend to produce speech that is more egocentric in nature, for example, by using pronouns in their speech without giving the relevant antecedents (Warren & Tate, 1992), suggesting that they fail to take listeners' needs into account. Children continue to misinterpret listeners' understanding of ambiguous referents until approximately kindergarten age (Ackerman, Szymanski, & Silver, 1990). Early studies of referential communication suggested that children younger than about six or seven years old did not take listeners' needs into account (Glucksberg & Krauss, 1967; Lloyd, Mann, & Peers, 1998), for example, by failing to resolve ambiguous referents. However, more recent evidence suggests that children as young as five (Nadig & Sedivy, 2002; Nilsen & Graham, 2009) and even as young as two (Clark & Bernicot, 2008; O'Neill & Topolovec, 2001) and three (O'Neill & Holmes, 2002), clarify ambiguous referents based on their listeners' needs, for example, by providing a disambiguating adjective (Nadig & Sedivy, 2002) or gesture (O'Neill & Topolovec, 2001). In addition, evidence suggests that young children may *initially* interpret communicative contexts from an egocentric perspective (i.e., rather than their interlocutor's perspective), but that they are able to monitor and correct these initial interpretations (Epley, Keysar, Van Boven, & Gilovich, 2004), a process that is consistent with the findings from many studies of adults (see the following section). Thus the evidence suggests that children consider and respond to their listeners' perspective and communicative needs, though they likely do so less efficiently than adults do.

1.4d A debate: speaker-internal vs. listener-driven motivations for common ground use. It has been suggested, based on evidence from communication-impaired populations, that pragmatic skills more broadly may be epiphenomena arising from more basic cognitive and linguistic functions (Perkins, 1998). The use of common ground specifically has been shown to be less efficient when additional demands are placed on speakers (e.g., time constraints, Horton & Keysar, 1996, or working memory demands, Schuh, 2010), suggesting that it is dependent on multiple factors. Within the common ground literature, there is much debate about what aspects of common ground processing emerge via unconscious utterance-planning constraints within the speaker, and what aspects are secondary to explicit speaker-listener dynamics, such as deliberately designing utterances based on a listener's known needs.

Several studies have suggested that language modifications based on common ground are largely due to speaker-internal processes. Early evidence that speakers may be planning utterances based on their own internal processes and not necessarily based on listeners' needs comes from a carefully-designed study by Brown and Dell (1987). In this study, speakers read printed stories and then told the stories to confederates from memory. Stories always involved an instrument that was either used for its typical purpose or for an atypical purpose (e.g., a knife was used to stab someone in the typical condition, while an ice pick was used to stab someone in the atypical condition). The authors found that speakers were more likely to explicitly mention the instrument in the atypical condition, suggesting on the surface that they knew listeners would not need to be told that a knife was used for a stabbing, but would need the additional information when the atypical ice pick was used for the crime. However, speakers may have explicitly mentioned the ice pick for multiple reasons, not necessarily because of the listeners' needs. Brown and Dell tested the processing underlying the explicit mention of instruments by manipulating the *listeners'* knowledge in a second experiment: in one condition, the listener was given access to a picture depicting the story, including the typical or atypical instrument (e.g., a sinister-looking man holding an ice pick and following another man), and in another condition, they were not given a visual aid. The authors found that speakers continued to mention the atypical instrument more often than the typical instrument, and that this effect was *not* moderated by whether the listener had knowledge (via the pictures) about what instrument was used. This finding suggests that speakers did not mention the atypical instruments solely to facilitate the listener's comprehension of the story, but rather did so based on a process internal to their own comprehension or processing of the story.

Brennan and Clark (1996) also demonstrated that content selection may be driven by speaker-internal processes. In their task, one participant, the "director," was asked to describe cards so that a second participant, the "matcher," could pick them out of a set. Initially, the set contained many items of the same category, leading the director to refer to the items by subordinate category terms (e.g., 'the penny loafer' instead of 'the shoe'). Later item sets included some of these original cards, but also new cards, such that all cards in a set were then from different categories. After this manipulation, directors continued to use the subordinate terms, as these terms were established within the common ground, even though they were no longer necessary. Critically, even after a new matcher was introduced to the task, the directors often continued to use the subordinate terms, although at this point, they were no longer necessary, *and* were no longer part of common ground. This finding suggests that part of the common ground process may be due to speakers selecting whatever referential expression happens to be most accessible to working memory. Under normal, non experimentally-manipulated circumstances, the most accessible term will likely be one that is shared with the current interlocutor, although, as this clever experiment demonstrates, this does not necessarily have to be the case.

These studies suggest that the accessibility of certain lexical terms in memory contributes to audience design and common ground. A related proposal is that specific conversational partners themselves can serve as cues to speakers' memories. Horton and Gerrig (2005b) tested this theory in a carefully-designed communication task in which speakers were instructed to direct different listeners to arrange different sets of cards. For half of the trials, the listeners' sets of cards were orthogonally distributed (e.g., the speaker directed one listener to match cards depicting dogs and directed another listener to match cards depicting birds). In the other half of the trials, listeners had overlapping sets of cards. It was predicted that audience design (i.e., the tendency to tailor speech to specific listener) would be more apparent on the trials in which card categories were orthogonally distributed, and this is precisely what the authors found. When cards were orthogonally distributed across listeners, speakers readily formed associations between particular listeners and particular cards: they were more likely to use more efficient, "tailor-made" descriptions of the cards, and to initiate these descriptions more quickly, than when there was overlap across different listeners. However, when listeners had cards depicting members of overlapping sets, it was harder for speakers to tailor their utterances to each specific listener, suggesting that tailoring was based largely the speaker's internal processes.

Related to this work is a quickly growing body of evidence suggesting that when speakers are placed under conditions of increased cognitive load, they are less likely to incorporate a listener's perspective into their discourse. In a very interesting study, Roxβnagel (2000) asked participants to give instructions as to how to assemble a model machine either to an adult listener or to a seven-year-old boy. Listeners thus differed in the amount of knowledge they could be presumed to have about machines and mechanics in general, and it was hypothesized that, due to presumed common ground, speakers would be more descriptive with the less-knowledgeable (i.e., child) listener. The author manipulated cognitive load in three conditions. In the *low load* condition, participants gave instructions while assembling the model in front of the listener. The other two conditions increased participants' cognitive load. In the *high load* condition, participants gave instructions with no model available in front of them, thus they had to rely on their memory of the machine when planning their instructions. Finally, in the *dual task* condition, the model was available but participants had to remember a list of seven digits for the entirety of their explanation. Under the low load condition, participants gave more detailed instructions to the boy listener than to the adult, thus adjusting their speech based on the presumed needs of the listener, consistent with hypotheses and what would be predicted by the common ground literature. This finding is interesting in itself, in that it used a somewhat different design than most because the speaker in this study did not actually know what the listener knew, he or she simply inferred this from the listener's age, demonstrating what could be considered a shift in register. Critically, under both the high load and dual load conditions, speakers used less detailed descriptions overall and did *not* adjust their speech for child versus adult listener. This finding suggests that speakers may tailor their utterances to listeners' needs not as a communicative default, but only when they have sufficient cognitive resources available to do so.

In Experiment 2 of the same study, Roxβnagel (2000) tested the moderating effect of increased motivation on the effect of cognitive load on common ground. Motivation was manipulated by telling participants in a high-motivation group that they would be asked to justify their explanations after they had finished. Common ground and cognitive load were manipulated as described above. Roxβnagel found that the detrimental effect of cognitve load on the common ground effect was partially attenuated when participants were highly motivated. That is, when motivated, participants continued to give more detailed instructions to the child than to the adult, even under conditions of high cognitive load. Unfortunately, the manipulation used to increase motivation did not speak directly to the construct of interest – that is, motivation to communicate with a specific listener (rather than just better communication overall). An alternative manipulation would have been to tell participants that their *listeners* would be asked about the explanations, thus increasing their motivation to provide the best communicative input as possible to that *specific listener*. However, the manipulation that was used could actually have made the common ground effect less likely to appear, because participants may have been motivated to provide full, detailed descriptions regardless of who they were instructing. Thus, even conservatively we can conclude that increased motivation likely attenuated the detrimental effects of high cognitive load on common ground. This finding is partiularly relevant to populations with ASD, who may lack the social motivation during discourse to push through the high cogntive load demands that common ground requires.

Speakers' choices about when it is acceptable to use a pronoun rather than its proper or common noun antecedent have also contributed to the common ground debate. While a full discussion of this literature is outside the scope of this paper, the following narrative study is particularly relevant to hypotheses about common ground and cognitive load. Arnold and Griffin (2007) had adult participants complete a simple, two-sentence, story based on two cartoon panels (the experimenter provided the first sentence of the story, based on Panel 1, and the participant was instructed to give a second sentence, based on Panel 2). Cognitive load was increased by adding a second character to the story (i.e., in the picture stimuli and in the experimenter's first sentence). The two characters were always of clearly different genders, ensuring that pronouns used to refer to them would be unambiguous. Regardless, speakers were more likely to use character names over pronouns when telling stories with two characters. This was taken to suggest that speaker's sensitivity to the potential ambiguity of pronoun use during discourse reflects their own internal processes and attention to story characters, and is not exclusively a pragmatic device intended to help the listener's comprehension.

The same study also looked at latency to refer to the main character (by pronoun or name), and found that it was longer in the two-character condition, suggesting that it took additional attentional resources to refer to the character when cognitive load was increased (Arnold & Griffin, 2007). In a second experiment, they demonstrated that this effect persisted even when Panel 2 was identical in the two-character vs. one-character conditions. Because Panel 2 was identical in the two conditions, we can rule out the possibility that participants took longer to refer to the main character in the two-person story due to the added complexity of processing a visual scene with two characters. This suggests that the longer latencies observed in the two-character condition were in fact likely due to the additional processing required to hold multiple characters in mind during storytelling. Reduced pronoun use was also correlated with other measures suggesting increased cognitive load; specifically, speakers who were at times dysfluent tended to use proper names more often than more fluent speakers. Interestingly, on the surface, the increase from one to two story characters (in a *very* brief story) appears as if it should be a relatively subtle increase in cognitive load for typical adults; however, it was enough to show multiple differences in narrative performance, reflecting the strong influence that cognitive load can have on word choice during discourse.

Studies of pronoun usage nicely demonstrate how processes that arise due to constraints within the speaker can contribute to discourse management. For example, some discourse rules that are regularly followed (e.g., pronouns being used more frequently for entities that have recently been the focus of attention, particularly in the subject role) may shift the attention of everyone in the discourse, including the speaker. So it is certainly possible that pronoun usage is simply due to shifts in discursive attention in the speaker's own mind, and not specifically for the sake of the listener. Therefore, consistent with the literature on speaker-internal processes in common ground, we may produce appropriate referring expressions without needing to speculate on the contents of our interlocutors' minds (Arnold & Griffin, 2007; Grosz, Weinstein, & Joshi, 1995).

1.4e Online processing of common ground. Within the common ground literature is an ongoing debate about when in time common ground is incorporated into utterance planning. According to one hypothesis, common ground is considered at the earliest stages of utterance planning. According to an alternative hypothesis, speakers monitor their utterances for comprehensibility and then adjust them as needed; common ground is thus integrated at a later stage of utterance planning, if at all. In a referential communication task testing this hypothesis, speakers and (confederate) listeners were both presented with screens displaying sets of shapes, and the speaker was asked to describe one of the shapes to the listener so that he could pick it out of the display (Horton & Keysar, 1996). In the shared condition, speakers were told that listeners saw all of the same shapes on their display; in the private condition, speakers were told that listeners saw the same target shape, but had different distracter shapes. Distracter shapes were chosen to change the context of the target shape; for example, two circles of different sizes would appear on the speaker's screen, making one (the target) the "big circle" and one the "small circle." In a shared context, if the speaker referred to the target as the "big circle," one could interpret this as a reference to the common ground shared with the listener – the speaker knew that the listener had the same set of reference shapes, and thus selected the adjective 'big' to disambiguate between the two circles. An alternate interpretation, however, is that the speaker used the adjective 'big' simply because it disambiguated the circles *for the speaker herself*, and was thus highly salient and useful for the speaker. Horton and Keysar thus reasoned that, if speakers were truly referring to common ground with the listener, then they should use more of these context-dependent disambiguating terms in the shared context than in the private context. This was exactly what they found. In addition, they found that speakers used more context-independent adjectives (e.g., “the one-inch circle”) to disambiguate in the private condition, suggesting that they relied on other information to help the listener disambiguate. This finding, while demonstrating that speakers did in fact incorporate knowledge that they shared with their listeners into their speech, did not provide any information as to *when* in the process this information was integrated.

To test how early in the process this integration occurred, a sample of participants did the same task under time constraints; specifically, they were instructed to begin their descriptions immediately, within the first 1.5 seconds of stimulus presentation. The authors reasoned that if common ground is incorporated at an early stage of processing, then the same responses described above (the unspeeded condition) should be observed when participants were under time constraints (the speeded condition). However, if common ground is incorporated at a later stage of utterance planning, then it should be seen to a lesser extent in the speeded condition as compared to the unspeeded condition (Horton & Keysar, 1996). The authors found exactly this: speakers in the speeded condition did *not* show the common ground effects that were observed in the unspeeded condition. The results of this study suggest that common ground is likely integrated at later stages of utterance planning, and not at the earliest stages. It further suggests that when processing demands are placed on speakers (such as time constraints, in this case), that they are less likely to incorporate common ground into their speech.

Although our study focuses on the role of the speaker in the common ground dynamic, studies exploring the role of the listener have also shed light on how common ground operates in real time. These comprehension studies, which have primarily tested their hypotheses in online language-processing tasks, largely suggest that listeners begin with an egocentric bias that is subsequently modified by information available within common ground (Hanna, Tanenhaus, & Trueswell, 2003; Horton & Keysar, 1996; Keysar, Barr, Balin, & Paek, 1998). Much of this work has been done by Boaz Keysar and colleagues, who have suggested that common ground it is not driven by audience design; that is, speakers are not tailoring their utterances to listeners based on mutually shared knowledge or experiences. Rather, language production is driven by the information that is most immediately available to the speaker. These processing limitations lead language to have an initial egocentric basis (Keysar, Barr, & Horton, 1998); any listener-specific language output would be due to adjustments and updating of the initial output, rather than being planned from the beginning.

In terms of comprehension, interpretations of speaker output based on common ground would likewise only be incorporated at later stages of processing. Comprehension studies rely on data from *listeners* rather than *speakers.* In one such study, Keysar and colleagues (1998) conducted a visual search task, in which some items (i.e., shared items) were visible to both the confederate speaker and the participant listener, while others (i.e., private items) were visible only to the participant. Speakers provided information about the scene and then listeners were asked questions, some of which were ambiguous based on what the listener could see but not based on what the speaker could see. The authors found that objects to which the speaker could not reasonably have intended to refer (based on what they could see) interfered with the listener's visual search; that is, listeners took longer to respond to the ambiguous prompts, even though these prompts would *not* be ambiguous if listeners were taking the speaker's perspective into account from the earliest stages of processing. Eye tracking confirmed that this effect was due to listeners looking at the conflicting object. Other similar paradigms from this research group have yielded similar findings (Barr & Keysar, 2002; Keysar, Barr, Balin, & Brauner, 2000).

One study from this group found the same pattern of findings in six-year-old children (Epley et al., 2004). Six-year-old participants initially interpreted communicative situations from their own perspective (rather than taking the speaker's perspective into account), and tended not to correct these interpretations – that is, they were much more likely than adults to choose an object that was hidden from the speaker's view. Adult participants, like the children, initially showed an egocentric bias (i.e., they looked at objects that the speaker could not see), but were quick to correct this bias and take the speaker's perspective into account in their final decision. Adults and children did not differ in their latency to look at the egocentric (i.e., hidden) object, but they did differ in their latency to look at the true target object. Children took longer to make this adjustment, again suggesting that incorporating an interlocutor's perspective takes substantial cognitive resources. Experimentally-manipulated increases in cognitive load have also been shown to interfere with listeners' ability to consider common ground, as demonstrated by eye-tracking studies (Kronmuller & Barr, 2007; Schuh, 2010). Interestingly, and relevant to the current study, increased cognitive load may have an even stronger impact on the comprehension of common ground in children and adolescents with ASD (Schuh, 2010).

The use of common ground, like any pragmatic language skills, is clearly driven by multiple interpersonal and intra-individual factors. Of course, along with the accumulating evidence that common ground effects are due in large part to speaker-internal processes, there is also substantial evidence that speakers explicitly tailor their speech to meet their listeners' needs. For example, in a recent study, participants were found to simplify their narratives more for listeners who had already heard a version of the same story than for those who were naïve to the story, even if speakers themselves had already told the story several times in both cases (Galati & Brennan, 2010). In a study with children, using an online language-comprehension paradigm, five- and six-year-old participants were found to incorporate a speaker's knowledge from the earliest stages of processing (Nadig & Sedivy, 2002), and not as a later adjustment to their comprehension, as suggested by Keysar and colleagues. Finally, Lockridge and Brennan (2002) used a similar design to Brown & Dell (1987; see above); however, Lockridge and Brennan used two participants rather than one participant and a confederate. Like Brown and Dell, Lockridge and Brennan found that speakers explicitly mentioned instruments more when they were used for atypical purposes (e.g., when a character used an ice pick for a stabbing). However, in contrast to Brown and Dell, speakers in this study also explicitly mentioned instruments relatively more often when the listener did not have access to pictures of the instrument, suggesting that speakers were primarily adjusting their linguistic output based on their perceptions of listeners' knowledge, and not based on their own response to the atypical instrument. The contrast between these two sets of findings also demonstrates that subtle differences in study design can lead to contradictory findings.

1.4f Executive function and common ground. The importance of processing limitations in articulating and interpreting speech points to the potential involvement of executive functions. Regardless of whether executive deficits *cause* autistic symptoms, executive function likely impacts on many of the complex social behaviors in which we engage, including the pragmatics of discourse. Although the role of executive function in common ground has not been fully explored in either the TD or ASD literature, recent findings in TD samples have begun to demonstrate a relationship between these skill sets. Although the literature is small, two recent studies have directly examined executive functioning in common ground paradigms, and have found relationships between traditional tasks of executive function and common ground use. In a comprehension task conducted with adults (Brown-Schmidt, 2009), participants were presented with a visual array of items and were then asked by the speaker to answer questions about the array (e.g., "what's above the cow that's wearing shoes?"). The questions were ambiguous, in that, for example, there were two cows wearing shoes; however, the item above one of the cows had been mentioned previously in the experiment (and thus, established in common ground), while the item above the other cow had never been mentioned. Participants' eye movements were monitored and demonstrated that, during these ambiguous trials, participants shifted their attention away from the item that had already been mentioned, and fixated on the item that had not been mentioned previously. This effect was not seen on control trials in which neither item had been mentioned. The author compared this eye tracking data to the results of two inhibitory control tasks, a go-no go task, and a Stroop task, completed by the participants. Participants who performed better on the Stroop task were more efficient in their shifts away from the shared item (and thus more efficient in their use of common ground during this task), suggesting that an element of inhibitory control is involved in common ground use. Performance on the go-no go task did not correlate with eye tracking data; however, this may have been due to the near ceiling performance of participants on this measure.

In the other study directly examining the relationship between common ground and executive function, five-year-old children acted as both speakers and listeners (Nilsen & Graham, 2009). Interlocutors were adult confederates. Conversational partners were presented with a visual array consisting of multiple objects, and the speaker was cued to instruct the listener to pick a specific object out of the array. In the shared condition, all four objects could be seen by both speaker and listener; in the private condition, one of the four objects could only be seen by the speaker. The obscured object was designed to have redundant features with another object in the array (e.g., the obscured object was a small duck when another duck was present in the display). As speakers, children were more likely to use adjectives to disambiguate between objects (e.g., "pick up the *big* duck" vs. "pick up the duck") in the shared condition (i.e., in the face of ambiguity), suggesting that they were sensitive to the listener's need for additional information. When children were listeners, adult confederates gave prompts such as "pick up the duck" which would have been ambiguous in the shared, but not the private, condition. Children were more likely to select the object that was visible to the speaker than the hidden object, even if the hidden object was also an appropriate referential match, suggesting that they used knowledge shared with the speaker to guide their selection.

Children had the option of selecting more than one object following the speaker's prompt; they did so more often in the shared condition, suggesting that they understood that when two plausible referential matches were visible to the speaker, but not when only one match was visible, that the prompt could refer to more than one object. In this task, children's eye movements were also monitored. Children were found to look less at the referential alternative object when it was not visible to the speaker, further suggesting that they were taking the speaker's perspective into account when interpreting their utterances.

In the same study, children were tested on executive function measures assessing working memory, inhibitory control, and cognitive flexibility. Performance on these tasks did not correlate with performance on the production task (i.e., the number of disambiguating adjectives produced). However, performance on two tasks measuring inhibitory control, but not tasks of working memory or flexibility, correlated with performance on the comprehension task. The comprehension task was repeated with three- and four-year-olds in Experiment 2. Participants were also tested on two different inhibitory control tasks. Three- and four-year-old children showed the same pattern of findings as five-year-olds, although the degree of difference between the private and shared conditions was less, suggesting that although the younger children were in fact sensitive to the speakers' perspective when making their choice, they were perhaps less attuned to it, or less able to inhibit an egocentric response, than the older five-year-olds. They also found that preschoolers who performed better on the inhibitory control tasks looked less at the referential alternative, suggesting a link between executive function and common ground usage, even at this early age.

In summary, studies of TD children and adults suggest that, although audience design and language adjustments based on common ground are driven in part by interlocutors' sensitivity to their conversational partners' perspective and needs, these adjustments require substantial amounts of information processing. Cognitive load studies suggest both that speech is more effortful when speakers are adjusting to their listeners' needs, and that audience design is less apparent when speakers struggle with heavier processing demands. Studies directly correlating the use of common ground to standardized tasks of executive function have found that the integration of common ground into language comprehension is associated with executive function skill, particularly inhibitory control. In the current study, we compare the use of common ground to multiple executive functions, including inhibitory control, set shifting, working memory, and fluency.

1.5 Gesture

1.5a Gesture and cognition. In addition to executive function, we explore the role of gesture as a facilitator of common ground. Gesture, which is thought to be reduced throughout the lifespan in ASD, may impact on both the cognitive and communicative resources needed to modify language in response to common ground. Through our everyday experiences, we can observe that gesture is related to cognition, through, for a example, a finger tap on a table or forehead during deep thought, or a hand tracing the steps of a mental arithmetic problem in the air. Experimentally, gesture has been shown to "lighten the cognitive load" (Church & Goldin-Meadow, 1986; Garber & Goldin-Meadow, 2006; Goldin-Meadow, 2005; Goldin-Meadow, Nusbaum, Garber, & Church, 1993; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001; Iverson & Goldin-

Meadow, 1998; Perry, Church, & Goldin-Meadow, 1988). In general, these studies have found that children and adults use their gestures to help themselves reason through new problem solving situations. Specifically, gestures appear to facilitate reasoning and problem solving by allowing speakers (and thinkers) to entertain more than one hypothesis about the world simultaneously (one in speech and one in gesture), thus providing more available cognitive resources at the same time. Interestingly, when problem solvers do not have access to gesture, their ability to learn new concepts is slowed, and when they are encouraged to gesture, learning is facilitated, suggesting a potential causal role of gesture in the learning process (Broaders, Cook, Mitchell, & Goldin-Meadow, 2007).

The *information packing hypothesis* of gesture, developed by David McNeill (1992), is based on the idea that gestures themselves help constitute thought (consistent with the studies presented above), and that they are involved in the conceptual planning of to-be-verbalized information (Kita, 2000). That is, gestures do not only present communicatively useful information themselves, but they are critically involved in both thought and speech production. Although the information packing hypothesis was initially based on the performance of adults, evidence supporting this hypothesis also comes from studies with children (Alibali, Kita, & Young, 2000). For these reasons, gesture may be an additional resource available for information processing that could be harnessed for common ground, which, as described above, requires substantial cognitive resources. If the use of common ground requires significant processing resources, and if gesture itself serves as one of these resources, then it stands to reason that if gesture use is limited in some way (as it is proposed to be in ASD), that common ground would be detrimentally affected. This hypothesis has not been tested directly, and we propose to test it here.

Gesture may aid cognition specifically by supporting working memory processes, particularly for spatial and visual aspects of thought and language. For example, Wesp and colleagues (2001) proposed that, through repetitive motor patterns, gestures may help maintain visual representations in working memory, on something like a visuospatial sketchpad (Baddeley, 1992). They asked participants to describe paintings, and, consistent with their hypothesis, they found that participants gestured more when they could not look at the painting as they spoke (Wesp et al., 2001), suggesting that gesture served to draw up a visual representation and maintain it in working memory, to which participants could then refer as they gave their descriptions. Morsella and Krauss (2004) had participants describe objects that varied in how verbally codable they were (e.g., a clock vs. an abstract line drawing). They further manipulated whether or not the participants could see the stimuli as they described them, or whether they had to base their descriptions on memory. Consistent with Wesp and colleagues, Morsella and Krauss found that their participants gestured more when the stimuli were not visible, further supporting a role for gesture in memory processes. Further, participants gestured more when the stimuli were not easily verbally codable, suggesting, consistent with many previous studies, that gesture plays a role in communicating visual information.

1.5b Gesture and communication. To date, very little research has been done on the relationship between common ground and gesture. One interesting study points to the role that gesture can play in relationship to audience design, even in very young children. Two-year-old children were more likely to use a word *in addition* to a pointing gesture to disambiguate between two objects when the objects were close together (and thus, when a gesture may have provided insufficient information; O'Neill & Topolovec, 2001). Although all child participants used a pointing gesture, only about half added a disambiguating word, suggesting that sensitivity to listeners’ needs is emerging at this time. Although this study did not contrast shared versus private knowledge, it suggests that the ability to tailor communication toward a listener’s needs develops early, and further, shows that, from an early age, speech and gesture work together to serve the communicative purpose of clarifying meaning.

Several studies (primarily conducted with TD adult participants) have demonstrated the communicative power of gestures during interaction. With respect to comprehension, information from gesture is integrated into information provided by speech (Cassell, McNeill, & McCullough, 1998). Gesture tends to be particularly useful for conveying spatial information, such as size and relative position of multiple referents (Beattie & Shovelton, 1999; Cook & Tanenhaus, 2009; Holler, Shovelton, & Beattie, 2009), and, particularly in these domains, communicates information that goes beyond the information included in speech (Beattie & Shovelton, 1999). Additionally, there seems to be a dynamic quality to how listeners glean information from gesture. Common ground studies have found that outside observers, in contrast to those actively participating in a conversation, do not benefit from established common ground as much as the interlocutors themselves (Schober & Clark, 1989). Similarly, listeners glean more information from gesture in face-to-face contexts than when watching the same information presented on video (Holler, Shovelton, & Beattie, 2009), suggesting that there is an interactive component to how gesture communicates.

From the speaker's perspective, which is more relevant to the current study, gesture appears to benefit pragmatic and cognitive aspects of communication in several ways. Pragmatically, gesture can act as a source of information available to listeners that remains present, even after the auditory content is over, or when the content of speech is markedly reduced (Holler & Wilkin, 2009). Speakers also use gestures to clarify ambiguous words, such as homonyms (Holler & Beattie, 2003). As suggested above, gesture may free up cognitive resources needed to consider and account for shared knowledge with an interlocutor, thereby facilitating common ground effects by allowing speakers to tell shorter, more efficient stories. In a recent study, typical adults were found to decrease the amount of semantic information included in their speech when they shared common ground with an interlocutor, consistent with the literature (Holler & Wilkin, 2009). However, when gestures produced during the same narratives were analyzed, they were found *not* to show a decrease in semantic content during conditions of common ground. This finding suggests that gesture may be an important cognitive or communicative resource that that may facilitate the common ground effect. Specifically, if speakers reduce the information contained in their speech, but maintain semantic information in their gestures, this may facilitate their ability to follow the discourse ideal of audience design and the Gricean maxim of quantity. This study provides compelling evidence for the idea that gestures are involved in processing and communicating common ground. However, the specific contribution of gestures to this process was not tested experimentally.

1.5c Gesture restriction. In general, there are two major methodologies for testing the role that gesture plays in cognitive and communicative processes. The more common method, employed by all of the studies described above, is to allow participants to gesture freely and naturally during the task at hand, and then to analyze what cognitive and communicative processes are associated with spontaneous gesture production. The other method is to use an experimental design in which participants' gestures are restricted, and then to examine how gesture restriction impacts on the processes in question. In the current study, we tested the influence of gesture on common ground during discourse in both ways: 1) by examining gestures during unrestricted spontaneous speech, and 2) by having our participants narrate stories under conditions of private and shared ground with their gestures restricted (i.e., by placing them in gloves Velcroed to a lap desk) on half of the trials.

Although, to our knowledge, gesture restriction has not been used during a common ground task, other studies have investigated the effects of gesture restriction on different communicative processes. Children and adults who are prevented from gesturing retrieve fewer words from their mental lexicons (Beattie & Coughlan, 1999; Butterworth & Hadar, 1989; Dunn & Dunn, 1997; Frick-Horbury & Guttentag, 1998; Pine, Bird, & Kirk, 2007), suggesting a role for gesture in word finding. Although very few studies have employed gesture restriction with narrative stimuli, one study that did so found that participants who were prevented from gesturing showed higher rates of dysfluencies in their descriptions of cartoons (Rauscher, Krauss, & Chen, 1996), particularly, their descriptions of the cartoons' spatial content. The authors concluded that gesture facilitates speech production and lexical access. Specifically, when gesture is prevented, speech becomes more dysfluent, especially when using language with spatial content, a finding that is consistent with other studies (Graham & Heywood, 1975;

Morsella & Krauss, 2004). An alternative interpretation (acknowledged in the Discussion) is that gesture facilitates conceptualization of the stories; when conceptualization is hindered by restricting gestures, dysfluencies appear. Other studies using narrative or dialogue formats have consistently found that when gesture is restricted, speech is less visualizable in nature, and includes less spatial content, pointing to the role that gesture plays in verbal imagery (Graham & Heywood, 1975; Rauscher et al., 1996; Rime, Schiaratura, Hupet, & Ghysselinckx, 1984).

Gesture studies in TD populations tentatively suggest that co-speech gestures may be a resource available to speakers when processing language, including common ground. This could be because gesture facilitates speech production, thereby freeing up other resources to consider common ground during utterance planning, consistent with findings from the gesture restriction literature. Alternatively, gestures may help speakers maintain multiple representations simultaneously, from which they can then select the most relevant information to include in their speech, consistent with the information packing hypothesis of gesture (Kita, 2000). In addition to aiding the speaker in multiple ways, gestures also provide valuable information to the listener; thus they serve to benefit both interlocutors in distinct ways. In the current study, we tested the role that gesture plays in the common ground process by physically restricting gestures under conditions of shared and private ground. We also compared participants' spontaneous production of gestures (in a non-restricted condition) to their use of common ground.

In summary, children, adolescents, and adults with ASD struggle with pragmatic language at all levels. Studies of TD populations suggest that executive function and gesture, two domains that are known to be impaired in ASD, may underlie the use of *common ground*, an important discourse skill. Here we seek to study the use of common ground by adolescents with ASD on a narrative task. To test the relationship between the executive functions and common ground, we compare task performance to standardized assessments of executive function. Finally, we examine the role of gesture during narrative and cognition and its relationship with common ground.

Method

2.1 Participants

Participants were 19 adolescents with ASD and 19 adolescents with TD matched on chronological age, gender, receptive vocabulary (Peabody Picture Vocabulary Test, Third Edition; Dunn & Dunn, 1997), and verbal, nonverbal, and full scale IQ (Stanford-Binet, Fifth edition, Roid, 2003).

Diagnoses were confirmed in the ASD group using the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2002), and were confirmed in the ASD group and ruled out in the TD group using the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003), and the Social Responsiveness Scale (SRS; Constantino & Gruber, 2005). Participants with ASD were not excluded for comorbid learning or psychiatric disorders. Likewise, participants with TD were not excluded for non-ASD learning or psychiatric disorders, to avoid obtaining a "hypernormal" comparison group. Participants with TD, however, were excluded if they had any first-degree relatives with ASD, or if they had any history of neurological problems.

One participant with ASD was excluded because he left his residential school before completing the study procedures. One participant with TD was excluded because there were concerns about his social development, including an elevated score (*t*-score of

62) on the SRS. These exclusions resulted in a final sample of 18 adolescents with ASD and 18 adolescents with TD. All participants in the final sample had IQ and PPVT scores in the average range (for details on matching variables, see Table 1).[1]

This study was approved by the University of Connecticut Institutional Review Board. Before beginning testing, written consent was obtained from parents, and written assent was attained from the participants themselves. Participants received financial remuneration (ten dollars per hour) for their efforts. Participation involved two sessions of two hours apiece.

Table 1

Demographic information for participants with autism spectrum disorders (ASD) and typically developing (TD) control participants

	ASD *M* (*SD*) Range	TD *M* (*SD*) Range	χ^2 or *F*	*p*	η^2_p
N	18	18			
Gender (M:F)	17 : 1	15 : 3	1.13	.29	
Chronological Age (years)	14.9 (1.4)	15.3 (1.4)	0.87	.36	.03
	13 – 17	12 – 17			
PPVT	110 (13.6)	118 (15.5)	3.01	.09	.08
	86 – 135	104 – 170			

[1] Of the final sample, four TD participants, and five participants with ASD, also participated in a study of speech-gesture integration (de Marchena & Eigsti, 2010) conducted in our laboratory two years prior.

Stanford-Binet					
Nonverbal	10 (2.2)	9 (1.9)	1.89	.18	.05
	7 – 14	6 – 13			
Verbal	10 (1.9)	10 (1.5)	1.42	.24	.04
	6 – 14	8 – 14			
Full Scale IQ	100 (10.6)	99 (6.6)	0.08	.78	.002
	82 – 118	85 – 106			
SCQ (total score)[a]	20 (6.2)	2 (2.2)	125.81	<.001	.80
	7 – 30	0 – 8			
SRS (total t-score)[b]	82 (11.1)	44 (8.7)	126.67	<.001	.80
	56 – 90	34 – 69			
ADOS (ASD group only)					
Communication (C)	3 (1.2)				
	1 – 5				
Social Reciprocity (SR)	7 (2.0)				
	4 – 10				
C + SR[c]	10 (2.7)				
	6 – 15				

[a] When used as a screening instrument, a cutoff score of 15 is recommended as an indication of a possible ASD (Rutter et al., 2003).
[b] When used as a screener, *t*-scores less than or equal to 59 are in the moderate range, *t*-scores from 60 to 75 suggest mild to moderate ASD symptoms, and *t*-scores of 76 or higher suggest severe symptoms.
[c] On the ADOS, 7 is the cutoff for a diagnosis on the autism spectrum, 10 is the cutoff for autism. All ASD participants in the final sample, except one, were above the cutoff for an ASD diagnosis on the ADOS; this participant had elevated SCQ (24) and SRS (73) scores, and was judged to carry an ASD diagnosis by clinicians on the study.

2.2 Measures

2.2a Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2002). The ADOS is a semistructured assessment for the diagnosis of pervasive developmental disorders, which provides multiple opportunities for social and communicative engagement. Only participants with ASD were administered the ADOS, which was used to confirm ASD diagnoses. All participants given the ADOS were administered Module 3 by author, who has established reliability within the laboratory.

2.2b Social Communication Questionnaire (SCQ; Rutter et al., 2003). The SCQ is a 40-item parent questionnaire for the screening of ASD symptoms in children. Items on this measure were derived from the Autism Diagnostic Interview – Revised (ADI-R; (Lord et al., 1994), a major tool used for diagnosing ASD. Concurrent validity between the SCQ and the ADI-R was reported as follows: for the Reciprocal Social Interaction domain, the Pearson intercorrelation was .92, for the Communication domain, the Pearson intercorrelation was .73, and for the Restricted, Repetitive, and Stereotyped Patterns of Behavior domain, the Pearson intercorrelation was .89. When used as a screening instrument, a cutoff score of 15 is recommended as an indication of a possible ASD (Rutter et al., 2003).

2.2c Social Responsiveness Scale (SRS: Constantino & Gruber, 2005). Parents were also asked to fill out the SRS, a 65-item questionnaire on which the frequency of social and communicative behaviors are rated on a four-point scale of (1) not true, (2) sometimes true, (3) often true, and (4) almost always true. The SRS provides distinct *t*-scores for males and females, and provides cutoffs for mild to moderate ASD symptoms (*t*-score of 60 to 75) and severe ASD symptoms (*t*-scores above 75). SRS scores are stable over time (Constantino et al., 2009) and differentiate well between TD, at-risk, and ASD samples (Constantino, Przybeck, Friesen, & Todd, 2000). The SRS was included, in addition to the SCQ, because it assesses a wide range of social skills and thus provides a sufficient range of scores in both ASD and TD samples for correlational analyses (Constantino & Todd, 2003).

2.2d Stanford-Binet Intelligence Scale: Fifth Edition (SB5; Roid, 2003). The SB5 is a factor-analytic measure of intellectual functioning from preschool age to adulthood. Participants were administered the abbreviated battery. The battery includes a vocabulary test, in which participants must explain the meanings of words, and a matrix reasoning test, in which participants must select the missing piece of a visual puzzle. This battery provides a reliable estimate of current cognitive functioning.

2.2e Peabody Picture Vocabulary Test: Third Edition (PPVT; Dunn & Dunn, 1997). The PPVT is a measure of receptive vocabulary from preschool age to adulthood. The PPVT presents participants with pictures of objects, actions, and events; the participants must then select the appropriate referent of a word stated by the experimenter. The reliability and validity of this measure are well established.

2.2f Executive function measures.

2.2f.a Working memory was assessed using the letter-number sequencing subtest from the Wechsler Intelligence Scale for Children, 4th edition (WISC-IV; Wechsler, 2003). The letter-number sequencing task requires participants to sequence increasingly long strings of letters and numbers into alphanumeric order while holding them in working memory.

2.2f.b Verbal fluency was assessed with the word generation subtest of the NEPSY-II (Korkman, Kirk, & Kemp, 2007). This task assesses letter fluency by asking participants to generate as many F- and S-initial words as they can in one minute (letter probe), and assesses category fluency by asking participants to generate as many animals and things to eat or drink as they can in one minute (category probe).

2.2f.c Inhibition was assessed using the inhibition task of the inhibition subtest from the NEPSY-II (Korkman et al., 2007), which requires participants to inhibit prepotent responses when naming different geometric forms (e.g., by saying 'square' when they see a circle and vice versa).

2.2f.d Set Shifting was assessed with the shifting task of the inhibition subtest from the NEPSY-II (Korkman et al., 2007). For this task, participants must alternate between one verbal rule and another, based on the specific stimuli (e.g., "if you see a black circle, say, 'circle,' but if you see a white circle, say, 'square'). This task also requires inhibition.

2.3 Design

Two independent variables were manipulated in this study: (A) the ability to gesture (*unconstrained condition,* in which participants were able to gesture vs. *constrained condition*, in which participants' gestures were restricted), and (B) common ground (*private condition*, in which information was known only to the participant, thus the experimenter and participant had no common ground *vs. shared condition,* in which information was shared with the experimenter and thus the experimenter and

Table 2

Counterbalancing.

	Order 1:		**Order 2:**	
Cartoon	**Constraint**	**Ground**	**Constraint**	**Ground**
Tom & Jerry (1)	Unconstrained [a]	Shared [c]	Constrained [b]	Private [d]
Dog & Kitten (1)*	Unconstrained	Shared	Constrained	Private
Daffy Duck (1)	Unconstrained	Private	Constrained	Shared
Roadrunner (1)	Constrained	Private	Unconstrained	Shared
Tweety Bird (1)*	Constrained	Shared	Unconstrained	Private
Pink Panther (1)	Constrained	Shared	Unconstrained	Private
Tom & Jerry (2)*	Unconstrained	Private	Constrained	Shared
Dog & Kitten (2)	Unconstrained	Private	Constrained	Shared
Daffy Duck (2)	Unconstrained	Shared	Constrained	Private
Roadrunner (2)	Constrained	Shared	Unconstrained	Private
Tweety Bird (2)	Constrained	Private	Unconstrained	Shared
Pink Panther (2)*	Constrained	Private	Unconstrained	Shared

[a] *Unconstrained* Condition: Participant was seated naturally in chair
[b] *Constrained* Condition: Participant was seated in chair with hands in Velcro gloves
[c] *Shared* Condition: Participant watched preview with listener
[d] *Private* Condition: Participant watched preview alone
* Participants received a quiz after narrating the story
Note: Constrained vs. unconstrained conditions were blocked in three due to time constraints of putting on and taking off the Velcro gloves.

participant had common ground). This resulted in a total of four conditions: (1) unconstrained-private, (2) unconstrained-shared, (3) constrained-private, and (4) constrained-shared. Three narratives were given for each of the four conditions, for a total of 12 narratives per participant, as shown in Table 2.

The 12 cartoons used as stimuli were presented in a fixed order to every participant. There were, however, two different counterbalanced orders for the above four conditions; for full counterbalancing, see Table 2. Nine (of 18) participants from the TD group received Order 1 and nine received Order 2; eight (of 18) participants from the ASD group received Order 1 and ten received Order 2.

2.4 Stimuli

Stimuli for the narrative task consisted of 12 approximately 60-second cartoons clips (e.g., *Tom & Jerry, The Pink Panther),* selected from six children's cartoons (two clips were selected from each cartoon). Cartoons were primarily wordless, although some cartoon characters made occasional brief statements (e.g., "you're never leaving!"). Cartoons depicted a range of possible and impossible events, for example: a cat chasing a mouse across a kitchen; a cat swallowing a bowling ball, rolling down a hill into a bowling alley and making a strike; a kitten playing with a ball; the Pink Panther lighting a light bulb with a match.

Each cartoon clip was preceded by an approximately 30-second "preview" of the clip that participants watched with a research assistant (i.e., during the shared condition) or alone (i.e., during the private condition). The research assistant did not watch the entire cartoon under any condition. Previews contained three approximately eight-second events from the cartoon clip, presented in a pseudo-random order, and separated by three seconds of a black screen. Cartoons were presented on a portable DVD player.

2.5 Procedures

All participants were tested at our laboratory at the University of Connecticut or in their homes or schools in Connecticut and Massachusetts. In all cases, testing was conducted in a private room with a table.

There were two experimenters present during data collection. One experimenter (the first author) assumed the role of the "researcher." The researcher administered standardized measures, explained all study instructions, and presented cartoon stimuli. The second experimenter (trained research assistants and graduate students) assumed the role of the "listener," and was present only to listen to the participants narrate the cartoons. The presence of a second listener who was not the primary experimenter was necessary because it was important that study participants believe that the person to whom they were telling the stories was unfamiliar with the cartoons that he or she was explaining. This was, in fact, the case; research assistants and graduate students serving as the listener were never shown the full cartoons, although they were familiar with the cartoon previews. Listeners were trained to respond to participant narratives by nodding attentively, smiling and laughing, and providing non-specific verbal responses (e.g., "oh," "ok," "mh hm") when appropriate. Listeners were specifically instructed *not* to indicate (either verbally or nonverbally) any confusion or difficulty that they may have had in following the participants' narratives.

The researcher explained the study procedures by telling the participant that this was a study about communication, in which the participant would be asked to communicate about 12 cartoons with the listener, who had never seen the cartoons before. The participant was told that communication would be assessed by recording their narratives, and by the results on “quizzes” taken by the participant and the listener. Quizzes were included in the study procedures because during pilot testing, participants given general instructions (e.g., “tell the story”) tended to give a thematic or plot summary (e.g., “the cat wants to catch the mouse but can never succeed”) that was insufficient to demonstrate a common ground effect. After brief quizzes were added, pilot participants included more detail in their narrations, and the anticipated effect emerged. Quizzes may also have increased participant motivation, thus enhancing the common ground effect (Roxβnagel, 2000). For a sample quiz, see Appendix A.

Participants were told that they and the listener would receive the same quiz about some of the cartoons; even though the listener would not be able to watch the cartoons, he or she should be able to respond correctly to some of the questions on the quizzes based on what the participant had communicated about the cartoon. The researcher also explained that participants would see a brief preview of each cartoon before seeing the whole thing, so they would know something about the cartoon before it started (“like when you see a preview for a movie, you know something about the movie before you see it but not everything”). Participants were given a chance to ask questions about study procedures and were then given a practice trial (unconstrained-private) in which they watched a preview, watched a cartoon, and narrated the cartoon to the listener. The listener and the participant then took a practice quiz. Although listeners were often familiar with the plot of the practice cartoon, they made a genuine effort to respond to the quiz based exclusively on the participant’s actual narration. The researcher then reviewed the quizzes and gave constructive feedback. Occasionally this feedback involved instructing participants that they would need to tell more detailed stories so that the listener would be able to answer the quiz.

After completing and reviewing the practice trial, the experiment began. Half of the participants (evenly divided between ASD and TD participants) began with the unconstrained-private condition, in which study procedures proceeded as in the practice trial. The other half of the study participants began with the constrained-shared condition. During the private conditions, the listener left the room so that it was apparent that he or she could not see the preview and cartoon while the participant was watching them; participants were also asked to wear headphones so that the listener could not hear. Headphones were worn during all cartoons as well as during the previews in the private conditions. During the shared conditions, participants were asked to remove the headphones for the preview, and the participant and listener sat next to each other and watched the preview together. Prior to beginning the first trial of the shared condition, the researcher explained that the participant and listener would watch the preview together, but the participant would then watch the cartoon alone. As such, the listener would know something about the cartoon, but not everything ("like if you watched a preview for a movie together, but then you saw the movie alone, without [the listener]"). During the private condition, participants watched the preview alone; this served as a control for the possibility that watching certain cartoons events twice might affect narrations. The 12 trials were presented in four blocks of three cartoons each; blocks were separated by breaks and executive functioning measures.

Conversational gestures were constrained by asking participants to place their hands in a pair of gardening gloves that were Velcroed to a lap desk during storytelling.

One participant with ASD found the gloves to be uncomfortable and distracting; he was asked instead to grip the bottom of the lap desk while he narrated the cartoons in the constrained conditions. Participants were not told the purpose of the gloves and desk until after all study procedures were completed.

2.6 Behavioral Coding

2.6a Speech transcription. Narratives were fully transcribed by trained undergraduate research assistants. All complete words were included; non-words (e.g., ‘um’) and partial words (e.g., ‘st-’) were excluded. Narrations from eight participants (22% of the sample), including four with ASD (for a total of 96 narrations) were independently transcribed by two separate coders for the purposes of obtaining reliability data. Agreement was very high; the intraclass correlation coefficient (ICC) for word count per narrative was .99.

2.6b *Gesture coding.* Gesture count was coded by the first author. One narrative was selected from each of the two unconstrained conditions (unconstrained-private and unconstrained-shared).[2] The narrative was watched in real time and the number of gestures used in each narrative was counted. Only hand movements were included as gestures; self-regulating movements, such as scratching and hair touching, were excluded. Head motions and movements of other body parts (e.g., feet) were also excluded. Gestures were not categorized according to gesture type. In a previous narrative study from our lab, also conducted with adolescents with ASD and TD, reliability for the analysis was conducted with nine participants (seven with ASD). Percent agreement was .90 and ICC was .98 for the number of gestures per narration.

[2] All six narratives from the unconstrained conditions will be gesture coded prior to submission for publication.

2.6c *Ratings of Story Quality.* College students (n = 49) read transcriptions of the narratives (24 transcriptions per rater) and rated them on three dimensions. (1) To assess whether naïve raters would have an impression about whether or not interlocutors shared common ground, raters were asked to judge whether or not the speaker had watched the cartoon alone or with the listener (simple forced-choice shared vs. alone rating); (2) to assess overall quality of the narratives, raters were asked to judge how easy the narratives were to follow, on a 1 (*very difficult, confusing; the plot didn't make sense*) to 7 (*totally coherent; a very clear plot*) scale; (3) based on the literature suggesting that gesture facilitates speech high in visuospatial content, we were interested in whether gesture restriction would lead to differences in the imageability of speech, so raters were asked to judge how well they were able to visualize each story as the read it, on a 1 (*poorly – hardly pictured anything*) to 7 (*very well – pictured every detail*) scale. For a sample rating sheet, see Appendix B. Order and group were counterbalanced across raters.

Results

Dependent variables were examined for deviations from the assumptions of normality and sphericity and were found to be normally distributed. Effect size was calculated using partial η^2 and Cohen's d. Effects of order were analyzed based on the two counterbalanced order included in the study. Order effects were found to be present, as described below. Some analyses were thus excluded (as noted when appropriate).

Data presented below address the influence of common ground on narrative characteristics in adolescents with and without ASD. Further, we shed light on how gesture, executive function, and social skills interact with the use of common ground, as well as how naïve raters responded to stories told during the experiment. Finally, we present data on gesture use in a non-communicative task, as a possible contrast to its use during storytelling.

3.1 Effects of Order

Order effects were examined independently and in relationship to the other variables of interest, that is, group, constraint, and ground. A mixed-model ANOVA was conducted with order and group as between-subjects variables, constraint and ground as within-subjects variables, and word count as the dependent variable. Word counts for all conditions, grouped by order and diagnosis are presented in Table 3. Overall, there was a significant main effect of order, $F(1,32) = 7.46$, $p = .01$, partial $\eta^2 = .19$, with participants in Order 2 telling longer stories than participants in Order 1. Orders 1 and 2 presented the same cartoon stimuli in the same order, and alternated between constraint and ground in the same way (i.e., three consecutive trials of a single constraint in a row before switching, and two consecutive trials of either private or shared ground before switching to the other; see Table 2); the orders differed in that participants in Order 1 started with the unconstrained-shared condition while participants in Order 2 started with the constrained-private condition. The observed order effect did not interact with diagnostic group, $F(1,32) = 0.85$, $p = .36$, partial $\eta^2 = .03$. There was a significant interaction of

Table 3

Word count by condition, order, and diagnosis

	Unconstrained		Constrained	
	Private	Shared	Private	Shared
Order 1				
ASD	122.2 (52.4)	137.3 (61.6)	104.4 (38.4)	107.5 (48.1)
	35 – 176	46 – 146	45 – 146	39 – 178
TD	179.2 (50.6)	164.9 (49.7)	165.7 (45.6)	138.4 (25.8)
	95 – 248	98 – 236	86 – 239	93 – 166
Order 2				
ASD	159.8 (66.0)	159.7 (59.8)	216.3 (81.3)	182.6 (72.8)
	94 – 283	95 – 274	117 – 343	100 – 291
TD	195.1 (57.1)	164.6 (32.6)	222.4 (60.3)	188.5 (39.0)
	136 – 319	115 – 216	161 – 349	148 – 253

Note: All data presented as Mean (*SD*); Range

order and constraint, $F(1,32) = 36.26$, $p < .001$, partial $\eta^2 = .53$, see Figure 1, below. There was a trend for an interaction between order and ground, $F(1,32) = 4.06$, $p = .052$, partial $\eta^2 = .011$. Because our main study hypotheses revolved around the interaction of group and ground, as well as ground and constraint, and because order effects were found to be present, three-way interactions were examined as well. Neither the order by group by ground interaction, $F(1,32) = 0.62$, $p = .44$, partial $\eta^2 = .02$, nor the order by ground by constraint interaction, $F(1,32) = 0.18$, $p = .68$, partial $\eta^2 = .01$, was significant. Because neither three-way interaction was significant and both yielded a small effect size, they were not further examined.

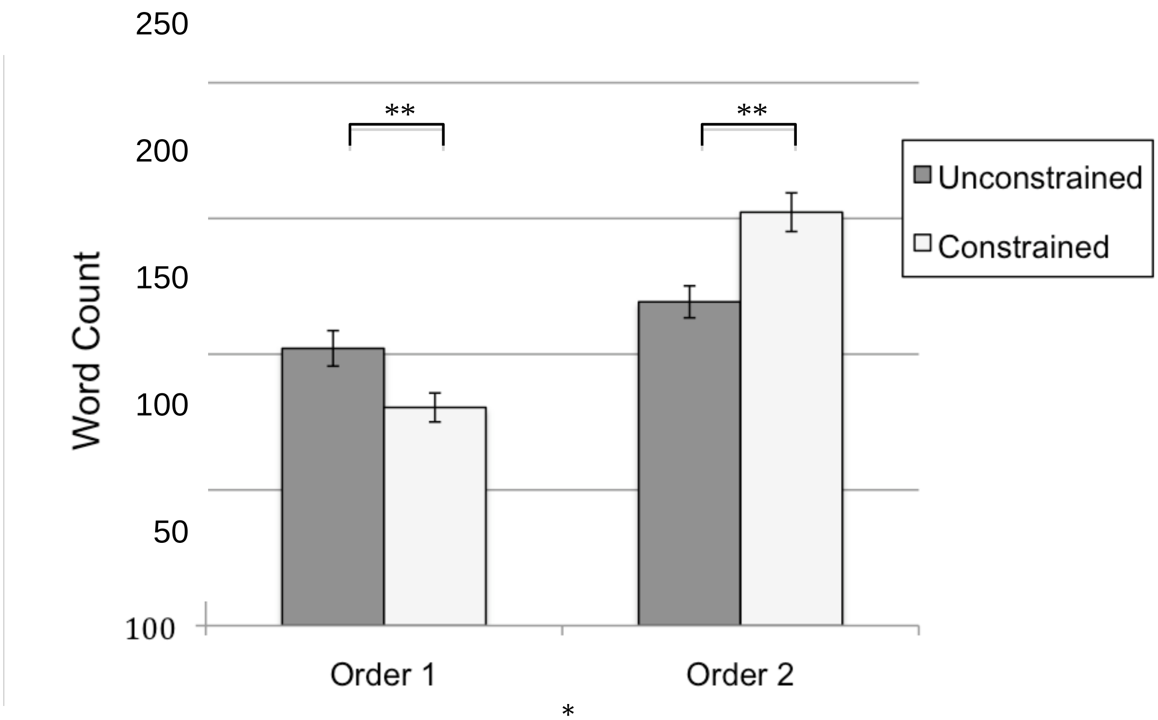

Figure 1. Mean word count in unconstrained vs. constrained conditions, by order.
* $p < .05$ ** $p < .01$
Note: Order 1: unconstrained condition first; Order 2: constrained condition first

The significant effects of order, and the order X ground and order X constraint interactions, were examined further. Post-hoc independent-samples *t*-tests revealed that the order effect was driven primarily by the effect of *constrained* stories (during which participants were prevented from gesturing). That is, participants in both orders told stories of approximately the same length in the *unconstrained* condition (i.e., in the condition that approximates a more typical story-telling scenario), t (34) = 0.98, p = .34, Cohen's d = 0.35. However, during the constrained condition, participants in Order 2 (who began with the constrained condition) told significantly *longer* stories than participants in Order 1, t (34) = 3.97, p < .001, Cohen's d = 1.36, with a large effect size, see Figure 1. Separate repeated-measures t-tests also revealed that the change in word count between unconstrained and constrained conditions was significant in both orders. For Order 1, stories were significantly *longer* in the unconstrained condition relative to the constrained condition, t (16) = 3.67, p = .002, Cohen's d = 0.44, whereas for Order 2, stories were significantly *shorter* in the unconstrained condition, t (18) = -5.00, p < .001, Cohen's d = -0.58, see Figure 1. Thus, for both orders, participant stories were longest for the constraint condition given first, keeping in mind that constrained and unconstrained conditions then alternated every three trials.

Demographic variables were compared for the two orders via a one-way ANOVA. Although participants with and without ASD were assigned in equal numbers to each order, the orders were not specifically matched for demographic variables as participants were enrolled. The results of a one-way ANOVA indicated that participants in Order 2 were significantly older than participants in Order 1, $F(1,34) = 8.42$, $p = .01$, partial $\eta^2 = 0.20$, with Order 2 participants having a mean (SD) age of 15.7 (1.1) years, and Order 1 participants having a mean (SD) age of 14.4 (1.4) years. No other demographic variables, including receptive vocabulary, verbal, nonverbal, and full scale IQ, or autism symptom severity, differed by order, all p's > .22. Because age was significantly different between orders, the mixed-model order by group by constraint by ground ANOVA, described above, was re-run with age as a covariate. The main effect of order and the order by constraint interaction remained significant, suggesting that, although age did differ by order, age differences did not account for the distinctions observed between Order 1 and Order 2. In addition, relationships among age, word count and gesture rate, separated into all four conditions and collapsed across conditions, was examined for correlations. Age did not correlate with word count (all r's $< .27$, all p's $> .12$) or with gesture rate (all r's $< .16$, all p's $> .34$) in any of these comparisons, again suggesting that the age difference observed between orders did not account for the order effect.

Potential explanations for the effect of order and the order by constraint interaction will be fully addressed in the Discussion section; for now, we present the following analyses collapsed across order. We believe this is appropriate since there was no interaction with diagnostic group, which was the primary contrast of interest, nor any three-way interaction with our independent variables of interest, and because dividing the sample on this between-subjects factor drastically reduces power and increases the risk of Type II errors. For completeness, separate data for each order is presented in Appendix C.

3.2 Story Characteristics

Prior to looking at the common ground effect and the factors that may underlie it, we first sought to compare broad narrative characteristics between the two groups, including word count, gesture count, and participant responses to story quizzes. We predicted that narrative characteristics would be roughly comparable between the two groups, with the possibility that participants with TD may communicate more, resulting in somewhat longer stories and more gestures.

3.2a Baseline word count and gesture frequency. For common ground analyses, word count was our primary dependent variable; thus, it was important to establish the baseline number of words used to narrate the cartoons by the ASD and TD groups. To compare baseline word count between the two samples, an independent-samples *t*-test was conducted to compare mean word count in the two groups, averaged across the 12 stories included in the study. This analysis revealed no significant difference between groups, t (34) = 1.35, p = .19, Cohen's d = 0.45, with the ASD mean (*SD*) word count at 153 (66) words per story and the TD mean (*SD*) word count at 178 (43) words per story.

In addition to assessing word count for each story, we coded the number of gestures used during storytelling. As gestures were intentionally prevented during the two constrained conditions, they were only coded for the two unconstrained conditions (i.e., unconstrained-private and unconstrained-shared). In addition, because gesture coding is more complex and requires a greater degree of training than speech transcription, only one story per condition was coded for gesture count (rather than all three, as in the word count analyses). Stories for the gesture count coding were selected so that the private and shared conditions would be roughly matched for word count, as story length is a major factor in how many gestures will be used by the speaker. The word count for stories selected for gesture coding did not differ significantly for the shared versus private condition, t (35) = 0.205, p = .84, Cohen's d = 0.02; stories in the private condition had a mean (*SD*) of 154 (49) words, and stories in the shared condition had a mean (SD) of 155 (52) words.

Although stories subjected to gesture analysis were matched overall for word count, individual differences in story length likely affect each participant's amount of gesturing. To control for individual differences in story length, a gesture *rate* variable was created by dividing the total number of gestures used in each story by the total number of words in that story, a practice that is common in the TD gesture literature (e.g., Beattie & Aboudan, 1994; Nobe, 2000). This variable was used in all subsequent gesture analyses. An independent samples *t*-test compared gesture rates between the two groups. The TD sample was found to have a higher gesture rate; thus, participants from the TD sample gestured more even after controlling for the number of words in each participant's individual stories, t (34) = 2.33, p = .03, Cohen's d = 0.78. Participants from the TD sample gestured approximately eight times for every 100 words; the ASD sample, in contrast, gestured approximately four times for every 100 words. Of the 36 participants in the study, four of them never gestured on either of the two stories subjected to gesture analysis; all four of these participants were from the ASD sample. These findings suggest that adolescents with ASD do not gesture as often as adolescents with TD during storytelling.

3.2b Enjoyment ratings and quiz performance. Prior to analyzing common ground effects, it was also important to establish that the TD and ASD samples were responding to the cartoons and the task itself in roughly the same manner. We sought to establish this in two ways, (1) by assessing participants' enjoyment of the cartoons, and (2) by measuring the amount of information participants were able to retain from the cartoons. Participants' enjoyment of the cartoons, which also serves as a proxy for task engagement, was assessed by asking participants to rate how much they enjoyed each cartoon on a one ("*hated it*") to five ("*loved it*") scale. An independent-samples *t*-test revealed that participants from the two groups found the cartoons similarly enjoyable, t (34) = 0.04, p = .97, Cohen's d = .01, with the ASD group providing a mean (SD) rating of 3.7 (0.9), and the TD group providing a mean (SD) rating of 3.7 (0.5), demonstrating that participants from both groups found the cartoon stimuli moderately enjoyable.

We were also interested in assessing how much information participants were able to retain from the action-packed cartoons, as group differences in retention would have confounded the common ground effect. To assess participants' attention to and recall of the cartoons, participants were given short quizzes on four of the cartoons (one per condition) that consisted of two short answer and three multiple choice questions; see Appendix A for a sample quiz. A mixed model ANOVA was conducted with group as the between subjects factor and condition as the repeated measure to test for group differences in cartoon recall as well as differences in recall for the shared vs. private conditions. Mean performance on the quizzes was higher in the TD group than in the ASD group; however, this difference did not reach significance, $F(1,34) = 3.10$, p = .09, partial η^2 = 0.08, see Table 4. No main effect of condition, $F(1,34) = 0.03$, p = .86, partial η^2 = 0.001, or group by condition interaction, $F(1,34) = 0.03$, p = .86, partial η^2 = 0.001, was observed for quiz performance, suggesting that retention was similar regardless of condition. We were also interested in attention to and recall of information specifically presented in the cartoon *previews*, as retention of this information had the most potential to influence the common ground effect. Each quiz included one multiple-choice question that asked specifically about what information was contained in the preview. An independent-samples *t*-test revealed that the TD and ASD groups did not differ significantly in their ability to differentiate the information contained in the previews, *t* (34) = 1.78, *p* = .08, Cohen's *d* = 0.59, although again there was a trend for the TD participants to perform better on this task than the participants with ASD; see Table 4 for details on quiz performance.

Table 4

Percent correct on cartoon quizzes

	ASD *M* (*SD*) Range	TD *M* (*SD*) Range	*t*	*p*	*Cohen's d*
Private condition	92 (16) 50 – 100	97 (7) 75 – 100	1.173	.249	0.39
Shared condition	92 (11) 62 – 100	97 (5) 88 – 100	2.000	.056	0.67
Preview questions	78 (27) 25 – 100	90 (13) 75 – 100	1.183	.083	0.59

In summary, results from analyses of story characteristics suggest that in general participants with ASD and participants with TD told stories that were comparable in length, although participants with TD gestured more than participants with ASD.

Participants from both groups appeared to enjoy the cartoon stimuli to a similar extent, and to glean the same amount of information from both the cartoons themselves, and the previews, although there was a trend for participants with TD to recall more details about the cartoons. Overall, the pattern of findings above was comparable for participants in Orders 1 and 2.

3.3 The Common Ground Effect

Our primary study goal was to examine the common ground effect, its presence or absence in adolescents with ASD, and what factors may underlie its use in TD and ASD samples. Although common ground effects have repeatedly been demonstrated in TD adults (Brown-Schmidt, 2009; Fussell & Krauss, 1992; Holler & Wilkin, 2009; Horton & Gerrig, 2005b; Horton & Keysar, 1996; Isaacs & Clark, 1987; Krauss & Weinheimer, 1966; Roxβnagel, 2000) and children (Epley et al., 2004; Nilsen & Graham, 2009), to our knowledge the effect has not specifically been demonstrated in an adolescent population. Since adolescents have been shown to respond qualitatively differently to social stimuli than both children and adults (for reviews, including proposed neurodevelopmental explanations for this phenomenon, see Nelson, Leibenluft, McClure, & Pine, 2005; Steinberg, 2005), there is a possibility that the common ground effect may not hold for this sensitive age group, even in our TD sample. Overall, we predicted that the common ground effect would be present in our TD sample and absent in our ASD sample. Further, although our ASD sample may not be as sensitive to speaker-listener pragmatics, and thus may fail to show a common ground effect at the group level, we predicted that common ground, if used by individual members of the sample, would be driven by similar underlying processes. Specifically, we predicted that participants with stronger executive function skills would be more likely to demonstrate a common ground effect.

To investigate the "core" common ground effect, a 2 (group) by 2 (ground) mixed-model ANOVA was conducted with word count as the dependent variable. Only the unconstrained conditions were used in this analysis, as we were interested in the common ground effect *independent* of the possible effects of gesture constraint. This analysis revealed no main effect of group, $F(1,34) = 3.02$, $p = .09$, partial $\eta^2 = 0.08$, suggesting that, in the two unconstrained conditions (e.g., shared and private), the TD and ASD groups used roughly the same number of words in their stories. This analysis also revealed no main effect of ground, $F(1,34) = 1.54$, $p = .22$, partial $\eta^2 = 0.04$, suggesting that, collapsed across groups, word count was similar for private and shared conditions. There was, however, a significant interaction of group and ground, $F(1,34) = 5.31$, $p = .03$, partial $\eta^2 = 0.14$; see Figure 2. Post hoc paired-sample *t*-tests revealed that while the TD group showed a significant drop in word count in the shared condition relative to the private condition, $t\ (17) = 2.78$, $p = .01$, Cohen's $d = 0.47$, word count in the two conditions was equivalent within the ASD group, $t\ (17) = 0.69$, $p = .50$, Cohen's $d = 0.11$. This analysis demonstrates that the TD and ASD samples responded differently to the pragmatics of sharing common ground with an interlocutor, as assessed through their word counts; specifically, while the TD group reduced the number of words in their stories during the shared condition, consistent with studies of common ground in TD samples, the ASD group failed to do so.

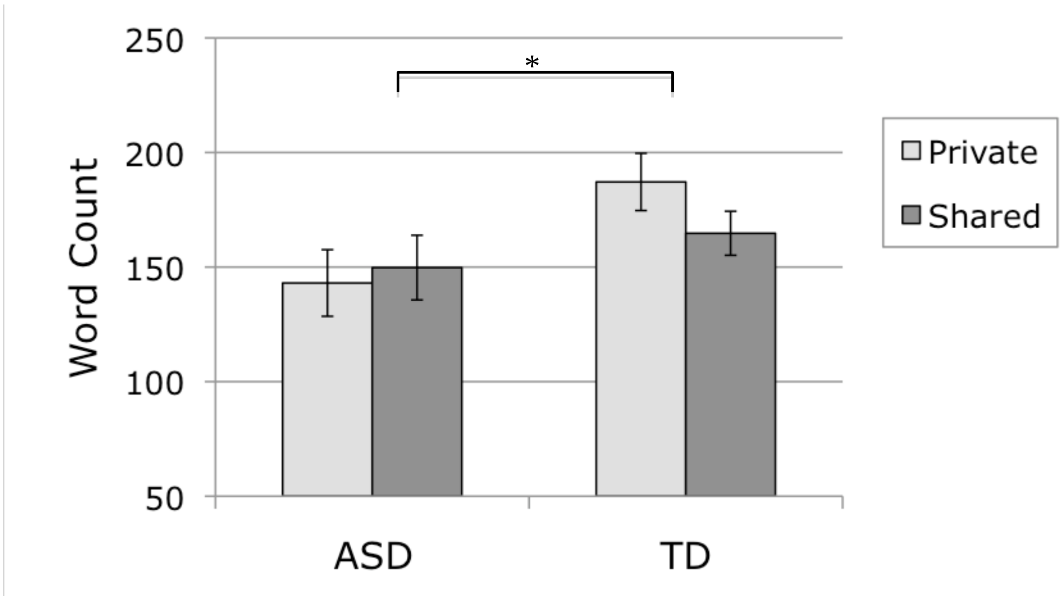

Figure 2. Word count in private vs. shared stories, by diagnostic group.

* *P* < .05

It is worth noting that when only the unconstrained conditions were included, as in the this analysis, order did not appear to impact the results; when order was added to the mixed-model ANOVA, there was no main effect of order, $F(1,32) = 0.42$, $p = .52$, partial $\eta^2 = 0.01$, nor any interaction effects between order and the other variables of interest. This finding, along with the fact that the same pattern of findings was observed when the sample was split by Order (see Appendix C), suggests that order effects did not account for common ground effects observed in this study.

3.3a Common ground: Controlling for quiz performance. Although the difference did not reach significance, there was a trend for the ASD sample to perform worse on the preview component of the post-trial quizzes, suggesting that they were perhaps not as able to recall whether they had seen certain parts of the cartoon with the listener (see *Story Characteristics,* above). Failure to encode which parts of the cartoon were in the preview would significantly impact the common ground effect, as participants would have a weaker understanding of the specific details altogether, and thus have more difficulty recalling which details were known to their interlocutor. Although the ASD group performed well on the quiz questions addressing the preview (with a 78% accuracy rate), the possibility remains that the group difference observed in the common ground effect may have been driven by the ASD group's relatively poorer encoding or retention of preview content (relative to the TD sample's 90% accuracy rate). To test this possibility, we ran the same mixed-model ANOVA, and added accuracy on the preview questions as a covariate. The same pattern of results described above held: there was no main effect of group, $F(1,33) = 1.60$, $p = .22$, partial $\eta^2 = 0.05$, or ground, $F(1,33) = 0.23$, $p = .63$, partial $\eta^2 = 0.01$, but there was again a significant interaction between group and ground, $F(1,33) = 4.97$, $p = .03$, partial $\eta^2 = 0.13$. This analysis suggests that the difference between the common ground effect in the two groups was not a simple reflection of poorer retention of the information presented in cartoon or preview by the ASD group.[3]

3.4 The Common Ground Effect Variable

To look at the relationship between individual characteristics, such as age, gesture use, and executive function skill, with common ground use, we created a common ground effect variable by dividing the total number of words used in the unconstrained-private condition by the total number of words used in the unconstrained-shared condition. Thus,

[3] This finding is also supported by the fact that there was no correlation between the common ground effect variable (see below) and either total quiz performance, $r(36) = .17$, $p = .33$, or performance on preview questions, $r(36) = .07$, $p = .68$.

participants with scores above one on this variable used more words in the private condition, showing a common ground effect, while participants with scores below one used fewer words in the private condition, failing to show a common ground effect. The TD group had a mean (SD) of 1.15 (0.22) on this variable, meaning that, on average, their stories were 15% longer when the listener had not seen the preview (i.e., when they did not share common ground). The ASD group, in contrast, had a mean (SD) score of 0.97 (0.24), suggesting that they used somewhat *fewer* words when they did not share common ground, and consistent with the above finding that adolescents with TD demonstrated a common ground effect while adolescents with ASD did not. Importantly, the common ground effect variable did not differ between Orders 1 and 2 (those who started with the constrained versus the unconstrained condition), either for the entire sample, t (34) = 0.95, p = .35, Cohen's d = 0.32, or for the TD, t (16) = 0.67, p = .51, Cohen's d = 0.34, or ASD, t (16) = -0.38, p = .38, Cohen's d = -0.19, samples independently.

3.5 Age and Common Ground

Collapsed across diagnostic group, we found that age was strongly and positively correlated with the common ground effect, such that older participants showed a greater common ground effect, r(36) = .47, p = .004. Interestingly, after the sample was split by diagnosis, the TD group did not show a relationship between age and common ground, r(18) = .29, p = .24, but the ASD group showed a strong positive correlation, r(18) = .59, p = .01, with age accounting for approximately 34% of the variance in common ground, as shown in Figure 3. Visual inspection of the data suggested that most participants in our ASD sample below the age of 15 had common ground effect variable scores below one,

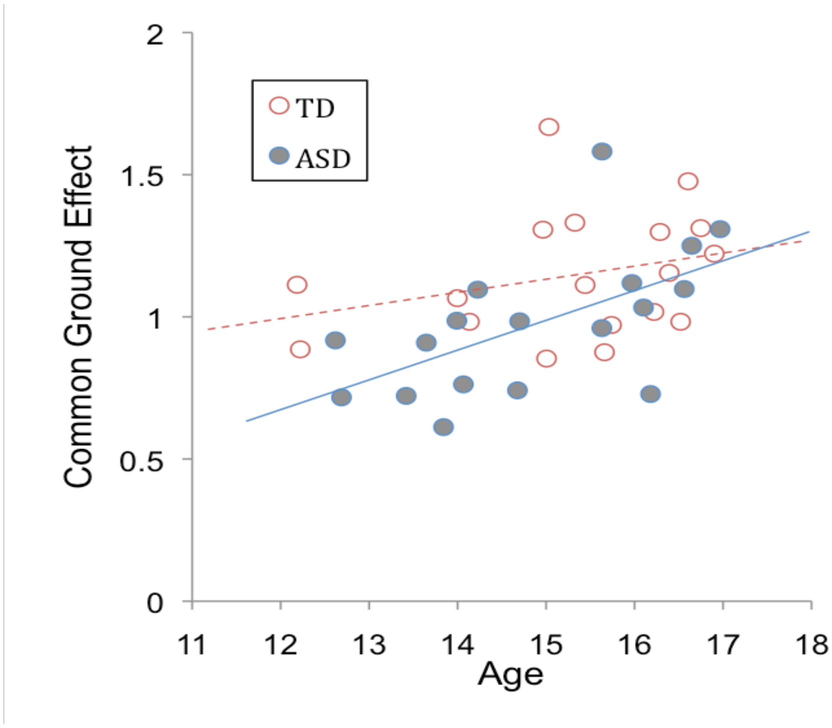

Figure 3. Correlation between age and the common ground effect, by group.

while participants above 15 tended to have common ground effect variable scores above one. This cut point was tested with an independent samples *t*-test, and participants with ASD over 15 (N = 8) were indeed found to have higher common ground effect variable scores than participants below this age (N = 10), t (16) = 3.00, p = .01, Cohen's d = 1.5, a finding which did not hold in the TD sample, t (16) = 0.89, p = .39, Cohen's d = 0.45. A mixed-model ANOVA was conducted for the ASD sample with ground as the repeated measure, age group (under 15 vs. over 15) as the between-subjects measure, and word count as the dependent variable. This analysis revealed no main effect of ground, $F(1,16)$ = 0.25, p = .62, partial η^2 = .02, again reflecting that, as a whole, the ASD sample did not reduce their speech in response to shared common ground. The main effect of age group was also found to be not significant, $F(1,16)$ = 1.00, p = .33, partial η^2 = 0.06, demonstrating that both age groups told stories of approximately the same length.

However, the ground by age group interaction was significant, $F(1,16) = 6.59$, $p = .02$, partial $\eta^2 = 0.29$, with older participants showing a decrease in word count from the private to the shared condition, and younger participants showing an increase. The results of these analyses suggest that, while TD individuals may have mastered common ground by early adolescence, teens with ASD may be on the path to developing this important pragmatic language skill.

3.6 Gesture and Common Ground

Our second line of investigation was to determine what factors might underlie the common ground effect in our TD and ASD samples. We planned to investigate the relationship between common ground and the use of gesture, in two ways: (1) by exploring the relationship between spontaneous gestures and common ground, and (2) by examining the common ground effect when spontaneous gestures were constrained.

Gesture rate (i.e., the number of gestures per number of words) was subjected to order analysis (previous order effect analyses included word count as the dependent variable), and a significant main effect of order was again observed, with participants in Order 2 (who began with the constrained condition) gesturing at a higher rate than participants in Order 1 (controlling for word count), $F(1,32) = 5.86$, $p = .02$, partial $\eta^2 = 0.16$. This effect did not interact with group, $F(1,32) = 0.39$, $p = .54$, partial $\eta^2 = 0.01$, or ground, $F(1,32) = 2.46$, $p = .13$, partial $\eta^2 = 0.07$, or contribute to a three-way interaction between order, group, and ground, $F(1,32) = 2.36$, $p = .14$, partial $\eta^2 = 0.07$. However, we have chosen to exclude analyses examining the relationship between gesture rate and common ground, because data from Order 1 and Order 2 yielded contradictory findings.

In addition to analyses of gesture rate, analyses of the influence of gesture constraint on common ground are excluded for the same reason.

3.7 Executive Function and Common Ground

Studies of executive function in ASD in general tend to find impairments, although many studies report contradictory results (for a review, see Eigsti, 2011). Our battery included measures of four different executive functions: (1) *working memory*, as assessed by the WISC-IV letter-number sequencing task; (2) *set-shifting*, as assessed by the switching component of the NEPSY-2 inhibition task; (3) *inhibition*, as assessed by the inhibition component of the NEPSY-2 inhibition task; and (4) *fluency*, as assessed by the NEPSY-2 word generation task.[4] A multivariate analysis of variance was conducted with group as the between-subjects factor and the above four executive functions as repeated-measure variables. Group averages suggested performance in the average range for both the ASD and TD samples;[5] for group means on all executive function tasks, as shown in Figure 4. This analysis revealed a main effect of group, $F(4,31) = 3.30$, $p = .02$, partial $\eta^2 = 0.30$, with the TD group showing significantly higher scores overall on the executive function tasks than the ASD group. For individual tasks, there was a significant group difference on working memory, $F(1,34) = 7.50$, $p = .01$, partial $\eta^2 = 0.18$ and a trend toward a difference on set shifting, $F(1,34) = 3.78$, $p = .06$, partial $\eta^2 = 0.10$. There

[4] The NEPSY-2 word generation task consists of two subtasks, one of which assesses semantic (category) fluency, and one of which assesses phonemic (letter) fluency. Age-based standard scores for these two subtasks were averaged to provide a composite fluency standard score, which was used for all analyses.

[5] In the ASD group, five participants (out of 18) received scores in the clinically impaired range for the working memory task, five for the set-shifting task, five for the inhibition task, and two for the fluency task. In the TD group, one participant (out of 18) received a score in the clinically impaired range for the working memory task, none for the switching task, four for the inhibition task, and three for the fluency task.

were no group differences on inhibition, $F(1,34) = 0.01$, $p = .93$, partial $\eta^2 < .001$, or fluency, $F(1,34) = 0.78$, $p = .39$, partial $\eta^2 = 0.02$. Overall, this pattern suggests general weaknesses in executive function in our ASD sample, consistent with the executive function literature, with a specific weakness in working memory.

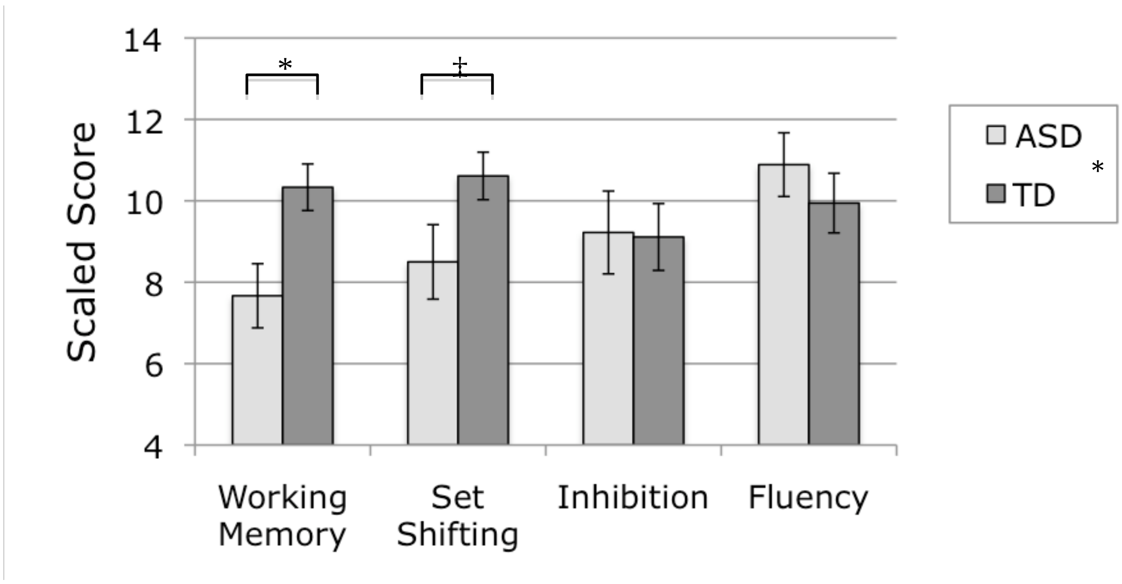

Figure 4. Scaled scores for executive function measures, by diagnostic group.
‡ $p = .06$
* $p < .05$

One of the major goals of this study was to identify cognitive processes that could be contributing to pragmatic language (and thus, to pragmatic language impairments in ASD). We hypothesized that executive function skills may contribute to the common ground effect based on previous findings that certain executive functions, in particular, inhibitory control, were related to common ground in children (Nilsen & Graham, 2009; Schuh, Mirman, & Eisgti, under review) and adults (Brown-Schmidt, 2009). Order effects were not apparent for this set of analyses.

To analyze the relationship between executive function and common ground, we examined the correlation between the common ground effect variable (i.e., the ratio of words in the private condition to words in the shared condition) and a composite executive function variable. This variable was created by averaging standard scores for the four components of executive function described above. A regression analysis was conducted to see if executive function might be a contributor to common ground *after* controlling for individual differences in participant demographics, specifically, age and full scale IQ. A multiple linear regression was conducted with the common ground effect variable as the dependent variable. Age and IQ were entered into the first step. The first model contributed 24.1% of the variance in the common ground effect, $F(2,33) = 5.25$, $p = .01$. The individual beta weight for age was .51, $p = .003$, while beta for IQ was .16, $p = .32$, suggesting that, while age was a significant contributor to the common ground effect, IQ was not. In the second step of the regression, executive function omnibus variable was added as an additional predictor. The second step did not add predictive power to the model. R-squared change for the second step was .003, suggesting a negligible (and not significant) change in the model, $F(1,32) = 0.13$, $p = .72$. This analysis did not support the hypothesis that executive functioning contributes to common ground, based on our task and sample.[6]

[6] In addition to the executive function composite variable, the common ground effect variable did not correlate with scores on any individual executive function tasks: working memory, r(36) = -.01, p = .97, set-shifting, r(36) = .11, p = .51, inhibition, r(36) = .11, p = .51, or fluency, r(36) = .17, p = .31. No significant correlations were found between the common ground effect variable and any individual executive function tasks when split by diagnostic group.

After splitting the sample by diagnostic group, we found that this pattern of findings only held up in the ASD sample. For the TD group, neither step one, $F(2,15) = 1.53$, $p = .25$, nor step two, $F(1,14) = 0.41$, $p = .53$, contributed to the variance in the common ground effect, suggesting that age was a non-contributor to common ground, as were IQ and executive function, at least in a relatively homogenous group. In the ASD group, however, the first step of the model was significant, as in the collapsed analysis above, $F(2,15) = 4.11$, $p = .04$. Individual beta-weights demonstrated that age was a significant contributor to common ground, $p = .01$, while IQ was not, $p = .60$, again consistent with the collapsed analysis. The second step of the model, which added general executive function skills as a predictor, was not significant for the ASD group either, $F(1,14) = 0.88$, $p = .37$. The results of the above regression analyses demonstrate that age contributes to common ground use in ASD but general IQ does not (as described in the *Participant demographics and common ground* section, above), nor does executive function, even after controlling for age and IQ.

3.8 Social Skills and Common Ground

The above analyses suggested that executive function skills did not significantly contribute to the common ground effect in our task. To investigate other individual differences that may be related to the ability to modulate one's discourse based on shared knowledge, we compared the common ground effect variable to scores on a measure assessing social skills, the SRS. We predicted that participants across groups with better social skills would be more likely to consider and incorporate common ground.

Specifically, we probed for a relationship between SRS scores (measuring social skills) and the common ground effect variable. When collapsed across groups, this hypothesis was supported; SRS score was significantly negatively correlated with common ground effect, $r(34) = -.34$, $p = .048$, suggesting that social skills were related to the common ground effect that we observed. However, visual inspection of the data suggested that the correlation between SRS and common ground scores was driven primarily by the fact that the ASD sample as a whole had both higher SRS scores and lower common ground effect scores, see Figure 5. A follow-up analysis was therefore conducted with the same bivariate correlation (between SRS scores and common ground effect scores) split by diagnostic group. Although the significant negative correlation was maintained in the ASD group, $r(16) = -.56$, $p = .03$,[7] there was a significant *positive* correlation in the TD group, $r(18) = .50$, $p = .04$. The same pattern of findings was observed within Order 1 and Order 2. Overall, the common ground effect as measured here did not appear to be associated with either executive functions or general social skills.

[7] Note: SRS forms for two participants in the ASD group were not returned.

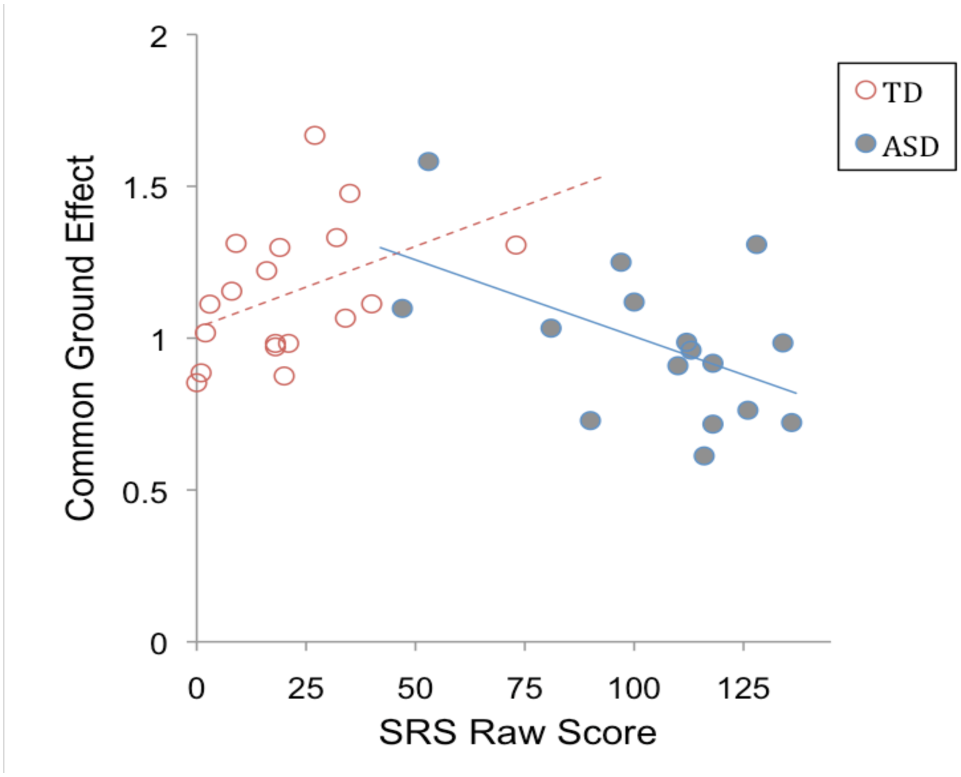

Figure 5. Correlation between SRS score and the common ground effect, by diagnostic group.

3.9 Qualitative ratings of narratives

In addition to looking at the common ground effect within individual participants, we were also interested in how adult observers would respond to the communicative exchange between participant and listener. After all stories were transcribed, students enrolled in the University of Connecticut participant pool were invited into the lab to read transcriptions and make subjective ratings about each story, as shown in Appendix B. Raters were naïve to study hypotheses, and did not know until after completing the study that any of the narratives were produced by individuals with ASD (for more details, see Method section). In general, the pattern of findings between Order 1 and Order 2 was similar in these analyses.

We asked whether naïve raters would have a gestalt impression about any common ground that might be shared between speaker and listener. To address this question, raters responded to forced-choice questions about whether the story they had just read was based on a cartoon that the speaker had watched with the listener, or based on a cartoon that the speaker had watched alone. Because we were interested in the effect of common ground, and not gesture, only stories from the unconstrained conditions were included in this analysis, to avoid any effects that gesture constraint might have had on raters' judgments. Ground and group were entered as within-subjects factors (in this case, group was a within-subjects factor rather than a between-subjects factor because raters read stories produced both by participants with and without ASD) into a 2 by 2 repeated-measures ANOVA. As described above, there was a main effect of ground, $F(1,48) = 5.74$, $p = .02$, partial $\eta^2 = 0.11$; the direction of this effect was opposite of what was predicted, such that raters were more likely to say that participants had watched the cartoon alone in response to narratives from the *shared* ground condition, suggesting that listeners were either unable to determine whether speakers and listeners shared common ground, or that they reliably misinterpreted common ground cues that were apparent in the narratives.

Importantly, there was no main effect of group, $F(1,48) = 0.66$, $p = .42$, partial $\eta^2 = 0.01$, and no interaction between ground and group, $F(1,48) = 0.11$, $p = .74$, partial $\eta^2 = 0.002$, suggesting that raters responded to the effect of common ground similarly for stories produced by those with and without ASD.

We were further interested in overall communicative performance by the two groups, and how communicative quality may or may not have been affected by the influence of common ground or gesture constraint. For this item, naive raters were asked to judge how easy the narratives, from all four conditions, were to follow. A 2 (group) by 2 (constraint) by 2 (ground) repeated-measures ANOVA was conducted with rating as the dependent variable. There was a significant main effect of group, $F(1,39) = 21.80$, $p < .001$, partial $\eta^2 = 0.36$, with a large effect size, with the TD participants telling stories that were rated as easier to follow. There were no main effects of condition (i.e., constraint or ground); however a significant group by ground interaction was observed, with the difference between groups being more pronounced in the shared condition than in the private condition. A post-hoc, paired-samples *t*-test demonstrated that, in the ASD group, stories were significantly harder to follow in the shared condition as compared to the private condition, $t(40) = 2.1$, $p = .04$, Cohen's $d = 0.66$; for descriptive statistics, see Table 5. This interaction suggests that participants with ASD may in fact have been responding in some way (that was detectable by our raters) to the common ground they shared with listeners; however, the way in which they responded served to make their narratives even more difficult for outside readers to follow. No other interactions were observed.

Finally, we were interested in the imageability of narratives and whether imageability would be affected by constraining gesture in either sample. The literature on gesture and speech production has consistently shown that gesture facilitates the production of highly visualizable speech (Graham & Heywood, 1975; Morsella & Krauss, 2004; Rauscher et al., 1996; Rime et al., 1984; Wesp et al., 2001). To test this hypothesis, we asked our raters to judge how well they

Table 5

Qualitative Ratings of Narratives

	ASD *M* (*SD*) Range	TD *M* (*SD*) Range
How easy was story to follow?***	3.9 (0.9)	4.5 (0.8)
	2.4 – 5.8	2.6 – 6.4
Private condition*	3.9 (1.0)	4.3 (1.0)
	2.1 – 6.3	1.8 – 6.5
Shared condition***	3.8 (1.0)	4.6 (0.8)
	1.8 – 6.0	3.2 – 6.3
How well could you visualize it?***	4.1 (1.1)	4.8 (0.9)
	1.9 – 6.3	2.6 – 6.7
Unconstrained condition***	4.1 (1.0)	4.8 (0.9)
	2.0 – 6.0	2.8 – 6.8
Constrained condition ***	4.1 (1.1)	4.7 (1.0)
	1.8 – 6.5	2.3 – 6.7

Note: Mean ratings from a 1 – 7 likert scale
*p < .05
***p < .001

were able to visualize each story as they read it. Since this question did not specifically include effects of common ground, only data from private stories was included. Group and constraint were entered as within-subjects factors in a 2 by 2 repeated-measures ANOVA with rating as the dependent variable. The main effect of constraint was not significant, $F(1,48) = 2.13$, $p = .15$, partial $\eta^2 = 0.04$, suggesting that, at this gross level of analysis, constraining gestures did not have a notable effect on the imageability of speech. Further, this effect did not interact with group, $F(1,48) = 0.42$, $p = .52$, partial $\eta^2 = 0.01$, suggesting that the effect of gesture on the imageablity of speech was comparable for both groups. Finally, there was a significant main effect of group, $F(1,48) = 9.53$, $p = .003$, partial $\eta^2 = 0.17$, see Table 5 for descriptives. This effect suggests that, in addition to finding the stories told by adolescents with ASD harder to follow, naïve raters were also less able to visualize these stories as they read them.

3.10 Gesture in a non-social context

Although the primary goal of this study was to investigate gesture use in a social communicative context, one measure that was collected, the NESPY-2 Inhibition task, allowed us to look at gesture use in a non-social, non-communicative context. In this task, participants were presented with a page of up/down arrows (or circles and squares), and instructed to name the direction of the arrow based on different instructions, for a total of six trials (see Method for more details).[8] The experimenter modeled the task before each trial with a practice set of stimuli by pointing to each stimulus and giving the

[8] The Inhibition task is comprised of three subtasks: Naming, Inhibition, and Switching. When comparing performance between groups, there was a no significant difference on the Naming task, t (34) = 0.88, $p = .39$, Cohen's $d = 0.30$, the Inhibition task (see above), or the Switching task, although there was a trend for better performance in the TD group (see above).

correct response. Participants were not instructed to point; however, the experimenter recorded whether or not the participant pointed on each of the six trials, for a total score ranging from zero (never pointing) to six (pointing on every trial). Pointing presumably helps participants keep track of where they are on the page and thus increases their speed. In this sense, the pointing gesture benefits the participant him or herself, and does not serve a social or communicative role. The same pattern of findings was observed for Order 1 and Order 2 participants, reflecting the reliability of this data.

An independent-samples *t*-test was used to compare the number of trials on which participants with ASD and TD pointed during this task. Participants in the ASD group pointed on more than twice as many trials as participants in the TD group, t (34) = 2.45, p = .02, Cohen's d = 0.84; on average, participants from the ASD group pointed on 3.6 out of 6 trials (SD = 2.7), while participants from the TD group sample pointed on 1.6 (2.3) trials. This finding suggests that when gesture serves a cognitive or organizational role, rather than a communicative role, individuals with ASD may gesture just as much, if not more, than their TD counterparts. This finding is in stark contrast to our above finding that, within a communicative context, participants with ASD gestured significantly *less* than participants with TD. To compare gesture use for the two tasks, we converted the gesture rate variable (in the narrative task) and the gesture count variable (in the executive function task) to z-scores and performed a mixed-model ANOVA with group as the between-subjects variable, task as the within-subjects variable, and gesture z-score as the dependent variable. There was no main effect of group, $F(1,34) = 0.004$, $p = .95$, partial $\eta^2 < .001$, or task, $F(1,34) < .001$, $p = 1.0$, partial $\eta^2 < .01$ (due entirely to the fact that both tasks had the same mean and standard deviation as they were both normalized to z-scores). However, there was a significant group by task interaction with a large effect size, $F(1,34) = 15.26$, $p < .001$, partial $\eta^2 = .31$. This interaction reflects the fact that participants with TD gestured more than those with ASD on the narrative task, while participants with ASD gestured more than their TD counterparts on the executive function task, as shown in Figure 6.

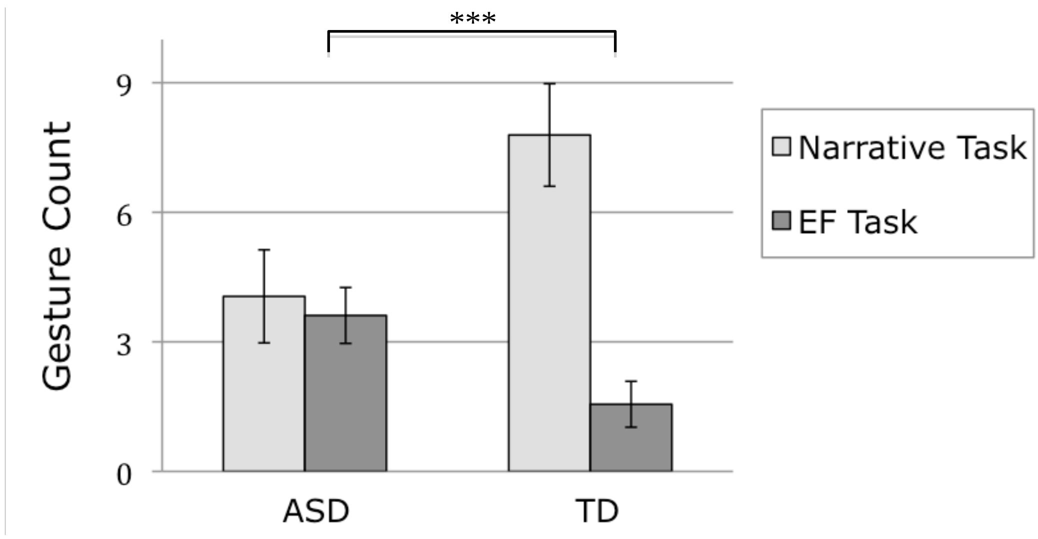

Figure 6. Interaction between group and task type for gesture count. For the narrative task, gesture count represents the number of gestures produced per 100 words. For the executive function task, gesture count represents the total number of gestures (out of six opportunities).
* $p < .001$

In addition to the above noted group differences in gesture use, the amount of pointing on the executive function task was positively correlated with task performance in the TD group, r(18) = .50, p = .04; participants in this group who pointed more were also able to name the direction of the arrows more quickly. However, this effect was not found in the ASD group, $r(18) = -.23$, $p = .37$.

Discussion

The current study was designed to test the common ground effect, specifically, speakers' tendency to reduce the content of their speech when information is shared with a listener, in both TD adolescents and adolescents with ASD. Using an experimental narrative design, we investigated whether or not this effect was present in both of our samples. Further, we were interested in factors that may underlie the common ground effect, both in typical development and in ASD. Specifically, we looked at the effect of gesture, executive function, and general social skills as factors potentially supporting the use of common ground during storytelling. Overall, our results indicated that TD adolescents showed a common ground effect; that is, their stories were shorter when they shared knowledge about the cartoon stimuli with a listener. This effect, however, was not present in our ASD sample: stories were just as long regardless of whether or not information was shared between interlocutors. We did not find strong evidence for gesture, executive function, or general social skills as factors underlying the common ground effect; here we present the relationship between these skills and the common ground effect, as well as potential explanations for why they do not appear to be related in our dataset. We also include a section on age, which emerged as a factor that could potentially account for differences in common ground processing between diagnostic groups. Finally, we take a close look at gesture use in autism, across two different types of tasks: the narrative task, and one of the executive function tasks.

4.1 Order Effect

Our study was designed so that participants would be taken through the task in one of two orders of presentation. The two orders presented the same cartoon stimuli in the same order, but applied the four conditions of the study in the reverse order, see Table 2. This design was implemented so that, should differences between the conditions emerge, we could be confident that these differences were due to differences in the conditions themselves and not, for example, due to differences in the cartoon stimuli or other factors, such as participant fatigue. Although we anticipated that performance between the two orders would be roughly the same, we actually found a significant order effect, such that participants in Order 2 told longer stories than participants in Order 1. Moreover, order interacted with gesture constraint, such that participants in Order 1 told longer stories in the unconstrained condition than in the constrained condition, while the opposite was true of Order 2. In fact, participant word count did not differ between orders for the unconstrained condition, but greatly differed for the constrained condition, suggesting that gesture constraint was the driving force behind the order effect that we observed.

Because the presentation order of the actual cartoons was consistent between Order 1 and Order 2, these findings suggest that there was something about the order of the conditions, rather than about specific cartoons, that affected how participants responded to the task. All participants received one unconstrained-private practice trial before beginning the actual task. For the remaining 12 coded trials, participants in Order 1 began the task with three unconstrained trials in a row, the first two of which were shared. Conditions were reversed for participants in Order 2: the first three trials were constrained, the first two of which were private; for full counterbalancing, see Table 2. Unfortunately, beginning the task with private and constrained conditions was confounded in the study design, so it cannot be determined conclusively whether it was starting off with the private condition or the constrained condition that led to the longer stories observed in Order 2. It should be noted that the order by constraint interaction did not interact with group, suggesting that participants with TD and with ASD responded to the pragmatics of the task in the same way. Furthermore, visual inspection of the data suggested that, with few exceptions, all participants in the task responded in this way, so this effect was not driven simply by a few outliers in either order, see Figure 7.

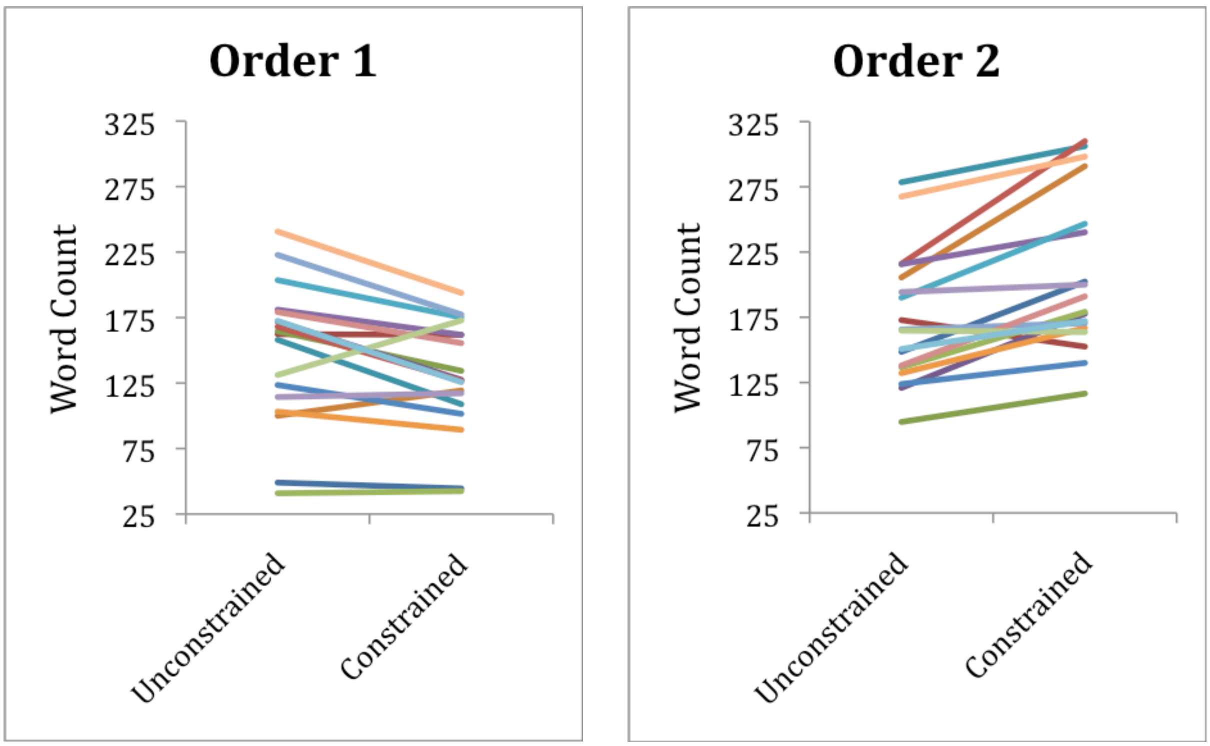

Figure 7. Individual differences in word count between the unconstrained and constrained conditions, by order, collapsed across diagnostic group.

Data were examined visually to establish hypotheses as to what may have differed between condition blocks. Since gesture constraint was alternated every three trials, there were four blocks of gesture constraint in the study, and since ground was alternated every two trials, there were six blocks of ground. When the data were separated into blocks according to gesture constraint, it was apparent that participants from both orders told longer stories in Blocks 1 and 3, and shorter stories in Blocks 2 and 4, see Figure 8a. Thus participants told longer stories consistently within the same type of constraint; however, the constraint with the longer stories differed by condition: for Order 1, participants consistently told longer stories in the unconstrained condition, whereas for Order 2, participants consistently told longer stories in the constrained condition. When the data were separated into the six ground blocks, however, no pattern emerged, see Figure 8b, suggesting that constraint, and not ground, was responsible for the observed order effects. The lack of a pattern across Orders 1 and 2 for the ground blocks also suggests that order effects were not due simply to some cartoon stimuli resulting in longer narratives than others.

It should be noted that, for the constraint blocks, participants from both orders told longer stories in the condition with which they began. This suggests that they may have established a particular response set as a function of gesture constraint or non-constraint, and that the change in set elicited shorter stories. However, it remains unclear why participants would tell shorter stories when switching to a new constraint, regardless of what that constraint was. One possibility is that the material presented between blocks drove participants' differential performance. This possibility was especially intriguing because each constraint block was separated by a different executive function task; it is certainly possible that some executive function tasks may have been more demanding than others, resulting in weakened performance on the storytelling task once participants returned it. This could explain why participants in the two orders performed similarly on

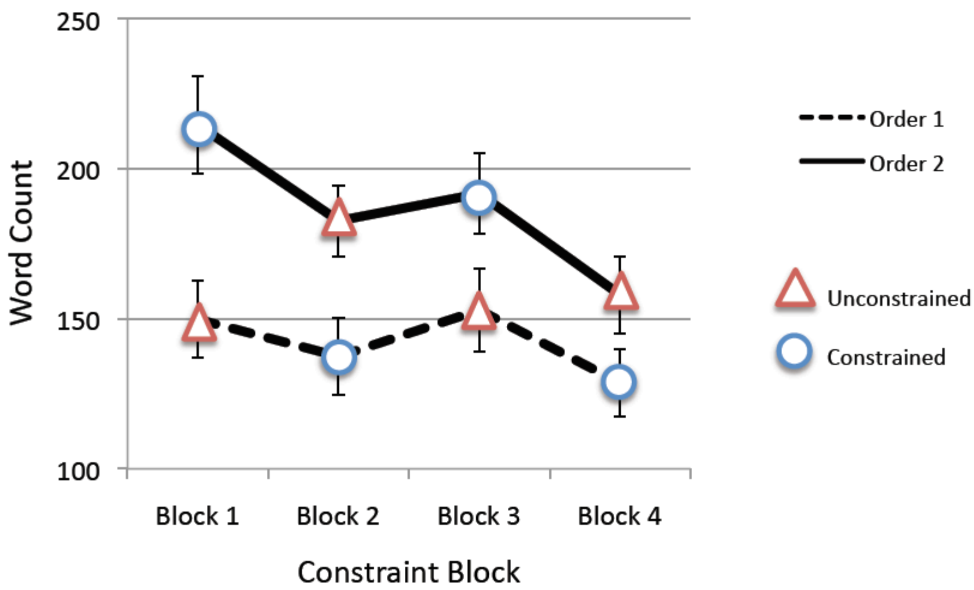

Figure 8a. Performance on the narrative task divided into four blocks, by order. Each block represents three stories told with the same gesture constraint. Constraints differ between orders. Order 1 begins with the unconstrained condition while Order 2 begins with the constrained condition, and blocks then alternate between constrained and unconstrained.

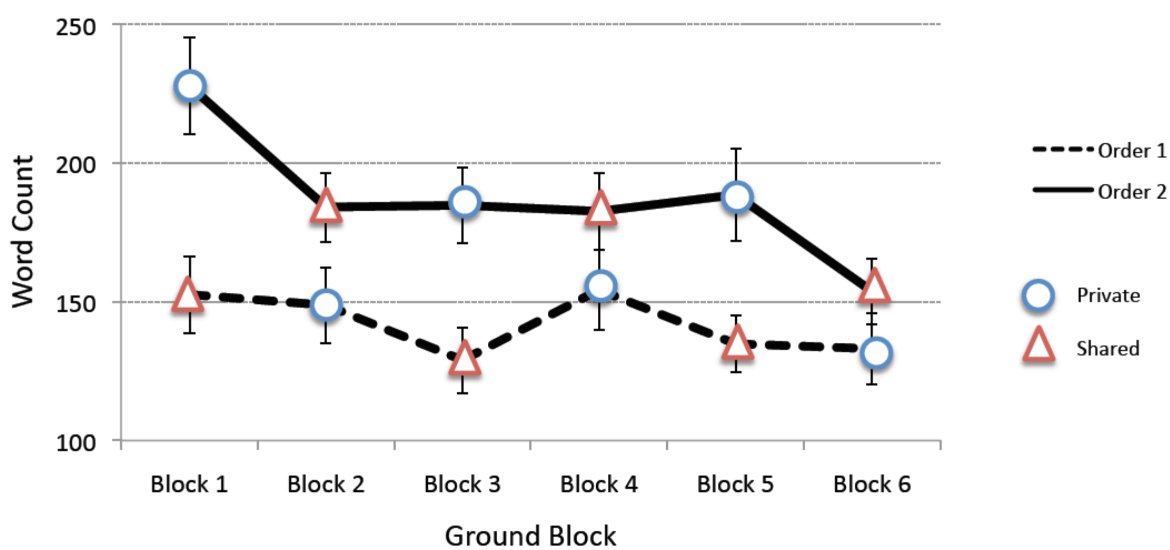

Figure 8b. Performance on the narrative task divided into six blocks, by order. Each block represents two stories told within the same type of ground. Ground differs between orders. Order 1 begins with the shared condition while Order 2 begins with the private condition, and blocks then alternate between shared and private.

the same constraint block, regardless of what the actual condition was. The ground blocks, in contrast, were not consistently separated by non-narrative tasks, which could explain why no pattern emerged. Although intriguing, we ruled out this possibility based on the pattern of data between trials. If, for example, a challenging executive function task between Blocks 1 and 2 were responsible for a drop in story length in Block 2 relative to Block 1, we would expect to see the shortest stories in Block 2 for the trial immediately following the task (i.e., Trial 4). This was not the case, suggesting that intervening executive function trials could not account for the interaction between order and constraint. Further, there is no a priori reason to believe that stories would consistently be longer or shorter following a demanding executive function task; stories may have been shorter for participants who felt overwhelmed and under-motivated, or longer for participants who were mentally tired and thus struggling more with the complex pragmatics of narration.

The overall pattern of findings of the order effect suggests that differences between Order 1 and Order 2 were driven primarily by how participants responded to the gesture constraint, rather than ground, fatigue, or some other factor. We come to this conclusion for the following reasons, described above: (1) order interacted with constraint, but not with ground, (2) when only the unconstrained conditions were included in order analyses, no order effect was observed, and (3) visual inspection of the data suggested that word count changed predictably between constraint blocks, but not between ground blocks, in both orders. It is possible that order may have had a subtle influence on ground as well; however, the effect of ground in our study was strong enough that order did not outweigh it. For the effect of constraint, however, which turned out not to be a major factor in how participants told stories, the effect of order was strong enough to show clearly through the data.

Our data do not allow for a conclusive interpretation of why the order effect existed, or why it had such a strong relationship with gesture constraint. However, the presence of such a relationship has several implications for how we can interpret the rest of our data. First, since participants in both orders told stories of the same length during the unconstrained conditions, and since the order effect itself vanished when only data from the unconstrained conditions were included, analyses that include only data from these two conditions (unconstrained-private and unconstrained-shared), should be fully interpretable collapsed across order. This includes analyses of the common ground effect itself, and analyses comparing the common ground effect to participant demographic variables and measures of executive function. However, analyses focusing on gesture are more limited in their ability to be interpreted, due to the significant effect of order.

To our surprise, we were unable to find any data describing the effect of task order on gesture constraint in the literature. Although studies using gesture constraints have been very important in developing theories about the functions that gestures serve, they represent a small section of the gesture literature. We did not find any studies addressing the potential effect that the order of gesture constraint could have on study findings. Of the seven studies we found that used a constraint manipulation, one (Rime et al., 1984) used the same order (constrained-unconstrained-constrained) for every participant, and one (Pine et al., 2007) did not mention condition order in the method section. Two further studies mentioned that counterbalancing of constraint conditions had been used, but did not present any analysis or discussion of order effects (Graham & Heywood, 1975; Rauscher et al., 1996). In the remaining three studies, gesture constraint was a between-subjects variable, so counterbalancing was not necessary (Beattie & Coughlan, 1999; Frick-Horbury & Guttentag, 1998; Morsella & Krauss, 2004). The current body of literature on gesture restriction thus provides no insight into how gesture constraint may affect unconstrained conditions administered later in the study. Future studies using gesture restriction should either employ a between-subjects design, or should carefully counterbalance conditions of gesture constraint *and* analyze the data for effects of condition order.

4.2 The Common Ground Effect

To get at our principle questions of interest, this study compared narrative data in two crossed conditions: ground (private vs. shared) and gesture constraint (unconstrained vs. constrained). However, to look at the common ground effect on its own, we put aside stories that were told during the constrained conditions and looked only at stories told during the unconstrained conditions (which resembles a more naturalistic storytelling context).

With respect to the TD sample, we found that these adolescents produced significantly shorter stories when they shared information with the person listening to them narrate their stories. We should note that information shared between interlocutors was relatively subtle. Preview stimuli consisted of three brief clips (approximately 9 seconds each) from 60-second cartoons. In addition, the clips were presented in a pseudo-random order (i.e., they were not necessarily presented in the order in which they appeared in the cartoon), and were chosen such that the listener would be able to follow the three isolated events, but would not be able to infer much about the actual plot of the cartoon. Thus, the 12% reduction in word count observed in the TD sample is impressive given the relative subtlety of the actual shared content.

This finding is completely consistent with the existing literature on common ground in typical development. Our study looked specifically at the role of the speaker in the common ground dynamic. Other studies like ours have consistently found that the content of speech produced by speakers is reduced when speakers and listeners share knowledge (Fussell & Krauss, 1992; Holler & Wilkin, 2009; Isaacs & Clark, 1987). We believe that this is the first study to specifically look at the common ground effect in TD *adolescents.* As is often the case, extensive research in this area has been conducted with both children and adults, leaving a gap in the developmental timeline during the adolescent period. Given that this age group is characterized by both heightened social awareness and sensitivity, as well relatively weak social reasoning (Steinberg, 2005), and given that frontal brain circuitry, including those networks involved in executive function, continue to develop throughout adolescence (Best & Miller, 2010) it stands to reason that complex pragmatic language skills, which likely require both social skills and executive function, may be under a period of change during adolescence. Our data did not specifically support this hypothesis, as adolescents in our TD sample showed a reliable common ground effect. However, it remains possible that adolescents may respond subtly differently in response to common ground than both children and adults. Studies addressing this question in the future could use the same task across a wider age range, including child, adolescent, and adult samples.

This is the first study to investigate the use of common ground in ASD, with participants in the *speaker* (as opposed to the *listener*) role. Other research in our lab has explored how children and adolescents with ASD interpret common ground as listeners in a problem-solving task (Schuh, 2010). This project used a referential communication task to examine on-line processing of shared and private information about a visual puzzle. Using eye tracking, the author found that children and adolescents with ASD were sensitive to their partner's perspective on this task. Specifically, like TD participants, they showed longer latency in their decisions, and looked more toward competing puzzle pieces when these competitors were private than when they were shared, suggesting that they needed time to rule out the private pieces as factors potentially influencing their partners' communicative intent. Increased latency and looks to the competing objects under conditions of common ground also suggests that additional processing resources were required. In fact, this study also demonstrated that when the cognitive load required by the task was increased, participants from both groups were less accurate in their decisions, suggesting a role for executive function in the processing of common ground. Interestingly, the drop in accuracy was associated with concurrent increases in latency and looks to competing pieces in the TD group, but not in the ASD group, suggesting that increased load for children and adolescents with ASD might leave them overwhelmed and unable to perform. Although participants with TD also failed to show a clear common ground effect under conditions of high load, their increased latency and looks to competitors suggests that they were taking the steps needed to accurately solve the puzzle, but that the amount of information was too great for them; not so for the participants with ASD who did not change their behavior under increased cognitive load.

In the current study, we found that, while participants with TD showed a clear common ground effect, participants with ASD failed to demonstrate the expected decrease in word count from the private to the shared condition. For adolescents with ASD, story length was comparable regardless of whether or not they shared common ground with the listener, suggesting that they did not adapt their speech based on their interlocutor's knowledge. This difference in performance between the TD and ASD sample existed despite the fact that both samples told stories that were approximately the same length on average, enjoyed the cartoon stimuli to the same degree, and performed similarly on quizzes tapping their comprehension and retention of story events.

Although the common ground effect specifically has yet to be examined in the ASD literature, our findings fit into the broader literature on pragmatic language skills in ASD. Several research groups have studied clarity of reference and pronoun use by children with ASD in a narrative format. Clarity of reference is relevant to common ground because speakers who use language that is less ambiguous to their listeners are more likely to be taking listeners' needs into account rather than simply responding to their own mental representation of story events and characters. Overall, the results of these studies have been inconclusive. While one study found that children with ASD used more ambiguous referential terms than control groups (Norbury & Bishop, 2003), two studies have found clarity of reference by individuals with ASD to be comparable to Down Syndrome (Loveland et al., 1990) and TD (Arnold et al., 2009) comparison groups. We did not specifically study clarity of reference in the current study; however, our finding that adolescents with ASD failed to demonstrate a common ground effect is consistent with the theory that they are not adapting their speech to meet listeners' needs, as in those studies that have demonstrated the use of more ambiguous referential terms in this population.

The choice of referential terms, including pronouns, may be driven in part by the relative salience of distinct referents to the speaker (Arnold & Griffin, 2007) rather than solely being based on consideration of the listener's knowledge. According to this theory we might predict that individuals with ASD would produce referential terms that are as clear as terms produced by other groups (assuming that similar aspects of the stimuli in question were considered salient), consistent with the findings of Loveland and colleagues (1990) and Arnold and colleagues (2009). If it is in fact the case that clarity of reference was driven primarily by salience to the speaker, as seems likely, then these studies would be less relevant to the common ground effect, which critically depends on speakers' ability to incorporate information about listeners' knowledge into referential choices. Because we hypothesized that the common ground effect would be driven by some speaker-internal processes, including potentially the salience of certain stimuli, we deliberately designed our task such that the salience of the critical preview events would be balanced across the private and shared conditions. Specifically, participants watched cartoon previews regardless of whether they were in the private or shared condition, such that events included in the previews would not be relatively more salient in the shared condition simply because participants had seen them before. Thus, even if salience is a factor in referential clarity, it is unlikely to explain the findings from the current study.

The linguistic construct of register is even more closely related to common ground than clarity of reference, as it specifically requires adjusting speech based on the needs of the listener or the social context of an interaction. To our knowledge, two studies (by the same research group) have studied register in ASD, and both have found that children with ASD do adjust their speech based on the social context or communicative needs of their interlocutor (Volden et al., 2007; Volden & Sorenson, 2009). These studies are inconsistent with our finding that adolescents with ASD did *not* adjust their speech based on their interlocutors' knowledge. However, there are differences between these studies and ours that can account for the discrepancy in findings; interestingly, these differences have to do with the processing demands placed on listeners. First, in our task, participants were not instructed that they should perform differently in the shared and private conditions, thus, it was dependent on them to infer that they should alter their communicative style between these two conditions. In Joanne Volden's studies, speakers were instructed (A) that their listener did not understand well, or (B) that they should make "bossy" or "nice" requests. Thus, the demands placed on speakers to moderate the communicative interaction were reduced by the presence of specific instructions. Second, in our task, participants had to keep track of the specific information that their interlocutor had access to (i.e., the information contained in the previews), and use this specific information to modify their speech. In the Volden studies, participants responded with a gestalt change in their communicative output, either by producing speech that was in general more easily understandable, or by producing speech that was bossy or nice. Again, their design required a lesser degree of processing than ours. In fact, within Volden and colleagues' (2007) study, when the task was simplified, participants were more likely to adjust their register, further pointing to the fact that these pragmatic skills require substantial cognitive resources. The contrast between their set of findings and ours suggests that, although individuals with ASD likely have some of the pragmatic language skills underlying common ground, that impairments in other domains may be preventing the use of common ground, a complex, multi-faceted skill. The most likely contributor, executive function, is discussed below.

Finally, we were curious as to whether naïve raters would be able to detect the presence of common ground between speaker and listener by reading transcripts of the narratives produced by our participants. We tested this hypothesis by having college students read the narratives and then make subjective ratings of them. Overall, they rated the stories told by adolescents with ASD as being more difficult to follow than the stories told by adolescents with TD, replicating a previous finding from our lab (de Marchena & Eigsti, 2010), and suggesting that raters were sensitive to qualitative differences in the stories at a group level. Raters were not, however, able to correctly determine whether or not interlocutors shared common ground. In fact, raters were actually *more* likely to say that speaker and listener shared common ground in response to stories produced during the private condition, when there was in fact no common ground. This finding suggests that the common ground effects observed in our study were likely due to subtle reductions in linguistic output that were too small to be detected by naïve raters, especially in a written format. Interestingly, the common ground effects present in the current study may even have been too subtle for our trained listeners to detect. Although we did not explore this idea quantitatively, several listeners reported that they thought the study might not be "working" since participants appeared to be telling stories in the same way regardless of whether or not they shared common ground with the listener. Again, this observation highlights the subtlety of the common ground effect. Neither the data provided by naïve raters nor the observations of our listeners differed by diagnostic group.

We also evaluated naïve raters' impressions of common ground by comparing ratings of how easy the stories were to follow across the shared and private conditions. When collapsed across groups, stories produced in both conditions were equally easy to follow. However, for the ASD sample only, stories were harder to follow when produced under conditions of shared ground. This finding suggests that raters may in fact have been sensitive to something that our participants with ASD were doing communicatively when they shared knowledge about the cartoons with the listener. Participants with ASD may have been attempting to tailor their narratives to what the listener knew during the shared condition, resulting in narratives that were harder for raters, who had no knowledge of the cartoons, to follow. For example, TD participants most likely responded to common ground by subtly reducing the amount of information they included about the shared previews. Participants with ASD, in contrast, may have responded to the shared condition by deliberately omitting whole sections of the narrative that were seen in the shared previews, and then overcompensating for the resulting fractured narrative by including more detail about the non-shared sections of the cartoon. Following such a response, stories from the shared condition would be roughly the same length as stories from the private condition; however, stories from the shared condition would be significantly harder to follow. These data, unfortunately, do not provide any specific information as to what participants with ASD may have been doing that made their narratives harder to follow under conditions of shared ground. Future analyses could explore specific features of the narratives in an attempt to explain why stories were rated as relatively harder to follow in the shared condition. One possibility is that narrative weaknesses typical of ASD populations may have been exacerbated in the shared condition if our participants assumed some knowledge on the part of the listener. For example, children with ASD have been shown to include fewer statements linking story events during narration (Diehl et al., 2006; Loveland et al., 1990), something that they may be even less inclined to do when their interlocutor is already knowledgeable about the story they are telling. This type of analysis could provide insight into what aspects of common ground are being used by adolescents with ASD, and what aspects are not being incorporated into their discourse.

4.2a Common ground and gesture. Our results did not provide a clear conclusion as to what forms of processing may be underlying common ground, either in TD or ASD samples. We explored the relationship between common ground and cognitive load in this study with respect to the use of gesture and in relation to standardized tasks of executive function. Due to strong order effects, the results of our gesture analyses are difficult to interpret, and any conclusions drawn from these data can be thought of as tenuous at best. Previous research has demonstrated that gesture use is held constant during conditions of shared ground, despite concurrent decreases in the semantic content of co-gesture speech, resulting in a relative increase in gesture rate when common ground exists (Holler & Wilkin, 2009). These findings suggest that gesture may be a cognitive resource available the speakers to help them maintain information in working memory while simultaneously decreasing their verbal output. Based on this research, on studies that suggest more generally that gestures help manage high cognitive load (Church & Goldin-Meadow, 1986; Garber & Goldin-Meadow, 2002;

Goldin-Meadow, 2005; Goldin-Meadow et al., 1993; Kita, 2000; McNeill, 1992; Morsella & Krauss, 2004; Wesp et al., 2001), and on findings that common ground requires a relatively high cognitive load (Brown-Schmidt, 2009; Horton & Keysar, 1996; Nilsen & Graham, 2009; Roxβnagel, 2000; Schuh, 2010), we predicted that gesture would moderate the common ground effect in our study.

Specifically, we predicted that participants who either (A) had access to gestures (i.e., in the unconstrained conditions as compared to the constrained conditions) or (B) who spontaneously used more gestures (i.e., during the unconstrained conditions) would show the strongest common ground effect, since they would have gestures available to them as an additional cognitive resource while processing common ground. We predicted that participants with ASD, who are thought to gesture less than TD samples, would show less of a common ground effect in part because they would not be able to depend as much on gesture. Our results did not support these predictions. Participants with ASD did in fact show a reduced gesture count and gesture rate, consistent with some (Bartak et al., 1975; Luyster, Lopez, & Lord, 2007), but not all (de Marchena & Eigsti, 2010; García Pérez, Lee, & Hobson, 2007) studies of gesture in ASD (gesture itself will be discussed in the *Gesture in ASD* section, below). However, participants who spontaneously used more gestures did not show an enhanced common ground effect in either group. Our results did not replicate Holler and Wilkin's (2009) finding of an increased gesture rate during conditions of shared ground. Further, when participants' gestures were constrained, they actually showed an *enhanced* common ground effect, going against the direction of our original hypothesis. However, because the effects related to gesture and common ground were compromised by significant influences of order, these findings cannot be interpreted conclusively.

4.2b Common ground and executive function. We also explored processing in relation to common ground by comparing participants' performance on four standardized measures of executive function to their tendency to reduce their word count from the private to the shared condition. Common ground use has been linked to processing capacity more broadly (Horton & Keysar, 1996; Kronmuller & Barr, 2007; Roxβnagel, 2000), and to executive function specifically (Brown-Schmidt, 2009; Nilsen & Graham, 2009; Schuh, 2010), though not all studies have found a link between common ground and executive functioning (Hupet, Berrewaerts, & Feyereisen, 2007). We hypothesized that common ground use in our study would be associated with executive function, as measured by standardized assessments. Further, we predicted that limitations in executive functioning within the ASD sample may contribute to their reduced use of common ground. This set of findings did not show an effect of order, thus they are more easily interpretable than the findings on gesture and common ground.

Overall, participants in the ASD sample showed weaker executive function skills than participants in the TD sample, consistent with the literature on executive function in ASD (Eigsti, 2011; Hill, 2004; Hughes et al., 1994; Ozonoff et al., 1991a). Our results did not suggest, however, that general executive function skills impacted on participants' tendency to adjust their speech to common ground, as regression analyses indicated that executive function contributed no significant variance to the common ground effect. Further, when we looked at the four executive function tasks separately (tapping working memory, inhibition, set shifting, and fluency), none of the tasks individually correlated with the common ground effect, in either sample. While consistent with findings that common ground does not correlate to working memory or flexibility (Nilsen & Graham, 2009), our results are inconsistent with findings demonstrating a relationship between common ground and working memory (Schuh, 2010) and between common ground and inhibition (Brown-Schmidt, 2009; Nilsen & Graham, 2009). Interestingly, all three studies that have demonstrated a link between common ground and specific executive functions have used eye-tracking designs focused on common ground *comprehension*. This kind of design allows for the detection of very subtle differences in the processing of common ground. Production studies, in contrast, must rely on more substantial differences in language output as their dependent variables, such as overall word count (the current study) or adjective count (Nilsen & Graham, 2009), potentially obscuring important relationships between common ground production and the factors that underlie it. In fact, one of the above studies included a production task as well as a comprehension task; the production task yielded no relationship between common ground and any executive function (Nilsen & Graham, 2009), consistent with our findings.

4.2c Common ground and general skills. In addition to the effect of gesture and executive function on common ground, we also looked at the contribution of more general skills, including IQ and language level, in both our TD and ASD samples. Full scale IQ and receptive vocabulary did not correlate with common ground use in either group, suggesting that differences in general cognitive and linguistic skills were not sufficient to explain why some participants showed a common ground effect while others did not, possibly because the variability in our sample was too small to show a relationship. These findings are largely consistent with the results of a recent study on audience design in ASD. In this study, children with ASD who exhibited more sophisticated levels of audience design were found to have higher overall language abilities than children with ASD who exhibited no audience design or lower levels only (Nadig et al., 2009). However, no differences were observed in terms of nonverbal IQ, age, ASD symptom severity, or adaptive functioning, suggesting that these factors had little influence on audience design in ASD. One important contrast between this dataset and our own is that we did not observe a relationship between language skills and common ground, whereas Nadig and colleagues found that language was the only individual factor distinguishing children with ASD who exhibited audience design from those who did not. This difference is most likely due to the difference in our choice of measures. While we used the PPVT-3, a test of receptive vocabulary, they used the CELF-4, a comprehensive battery of structural language use. Thus, the contrast between our findings and theirs suggests that some aspect of language functioning other than receptive vocabulary must underlie audience design abilities in ASD.

4.2d Common ground and age. Finally, we looked at the effect of age on the common ground effect. Although common ground appears to be in place by mid-childhood in TD samples (Nadig & Sedivy, 2002; Nilsen & Graham, 2009), its use in adolescence has not specifically been explored in the literature. Further, our task relied on a rather subtle, implicit measure of common ground that may still have been undergoing development at this age. In our TD sample, age did *not* correlate with the common ground effect, suggesting that even relatively subtle uses of common ground during discourse are stable within the 12- to 16-year age range. For the ASD sample, however, age was strongly and significantly correlated with the common ground effect, such that older participants showed a greater common ground effect than younger participants. When we looked at this relationship in more detail, we found that participants in the ASD sample above the age of 15 were likely to show a common ground effect, while participants below the age of 15 were not. These findings suggest that, even as late as early adolescence, pragmatic language skills, including the implicit common ground effect that we assess here, are not fully in place in individuals with ASD, and that, in fact, they are likely on an upward trajectory. That is, far from being a skill that is completely out of reach for high-functioning individuals with ASD, common ground appears to be a skill that simply develops later for them, and may fall into place by late adolescence. As described above, adolescence is a period of great social and neuropsychological change. It stands to reason that while our participants with TD had already developed enough of the underlying skills necessary to tackle the complex pragmatics of common ground, that our ASD sample was in the process of doing so, but had not yet arrived. This could be due to general delays in social and neuropsychological functioning, or perhaps those with ASD require even greater cognitive resources than those with TD to approach common ground, because of their significant social impairments.

In fact, this is not the first study of pragmatic language to find age-related differences in task performance in an ASD sample when no differences were apparent in control groups. Arnold and colleagues (2009) looked at pronoun usage in 9- to 17-year-olds with TD and ASD. Although no differences in pronoun use were observed in the TD sample by age, the authors found that the youngest group of participants with ASD (aged 9- to 12-years-old) were more likely to use overly explicit terms (e.g., a character name, '*Tweety*', or description, '*the bird*') rather than pronouns. A parallel can be drawn to the fact that our younger participants with ASD may have failed to show a common ground effect because they gave overly explicit descriptions of the cartoon stimuli, particularly in the shared condition. Taken together, these findings suggest that some pragmatic language skills that may have been thought to be absent in ASD samples may simply be later emerging, either through the influence of effective interventions, or through the delayed acquisition of cognitive, linguistic, and social skills. These findings provide hope that a wider range of pragmatic language skills can be mastered and spontaneously employed, especially with the help of appropriate interventions.

4.3 Gesture in ASD

The literature on the frequency of gesture use in ASD has been inconsistent to date. For example, many studies have found reductions in protodeclarative pointing, but not protoimperative pointing (Baron-Cohen, 1989; M. Bono, T. Daley, & M. Sigman, 2004; Camaioni et al., 1997; Mundy et al., 1986). Gestures delays have been consistently observed in studies using developmental questionnaires, such as the CDI (Charman, Drew, Baird, & Baird, 2003; Luyster et al., 2007), and gesture reductions have been observed in comparison to control groups (Bartak et al., 1975). However, several studies have also failed to find differences in gesture frequency between samples with ASD and control groups (Attwood et al., 1988; Capps et al., 1998; de Marchena & Eigsti, 2010; García Pérez et al., 2007). These mixed results suggest that there is likely something either about the about the specific samples included in the above studies, or about the specific tasks used, that influences the amount of gestures produced by participants with ASD.

The results of the current study suggest, consistent with Bartak (1975), but inconsistent with previous work from our own lab (de Marchena & Eigsti, 2010), that adolescents with ASD use fewer gestures than adolescents with TD on a semi-structured narrative task, even after controlling for story length. In our previous study, participants also told brief narratives, but the narratives were produced in response to six stimulus cards that depicted a story using black and white line drawings, rather than in response to 60-second cartoons. The stimuli were less complex and more structured than the cartoon stimuli used in the current study. In the previous study we found that participants (also adolescents) with ASD used as many gestures as participants with TD; however, their gestures were less integrated with their speech. The contrast between these two sets of findings suggests that when tasks are more structured, participants with ASD may be more inclined to use gesture. In fact, this is exactly what Bartak (1975) found in his early study of gesture in ASD: in his sample of children with ASD, gestures in a spontaneous free play task were reduced relative to participants' own gesture use during a structured pantomime test.

4.3a Gesture in a non-communicative context. An interesting finding from our data also suggests that gesture use in ASD may be related to specific task demands. We found that, although our ASD sample used fewer gestures than the TD sample during the narrative task, they actually used more gestures than the TD sample during an executive function task. This task, the *inhibition* task from the NEPSY-2 (Korkman et al., 2007), includes three different subtasks, which measure speeded naming, inhibition itself, and set-shifting. The experimenter models pointing to the stimuli when teaching the task, but never explicitly instructs participants themselves to point. We found that participants with

ASD on average pointed to the stimuli twice as often as participants with TD. This finding was reliable and robust, and was not influenced by the effect of order. To our knowledge, this is the first study to examine gesture use in ASD in a completely non-communicative context. The contrast between gesture use in the narrative task versus gesture use in the executive function task suggests that individuals with ASD may be more inclined to gesture in very structured contexts. The contrast between the two narrative studies conducted in our lab, and the findings of Bartak's (1975) study, suggests that the external structure imposed on a task likely plays a role in how frequently gesture is used in this population. The classic studies demonstrating increased protoimperative pointing relative to protodeclarative pointing in children with ASD (e.g., Baron-Cohen, 1989) suggest that the social context or goal of a gesture is also a major influence on its use. Future studies exploring gesture use across a range of more and less structured tasks, as well as more and less social tasks, will be needed to explore these two hypotheses.

Another intriguing possibility is that individuals with ASD gesture more when their gestures are not means to a social or communicative end, but rather serve themselves in some way. That is, those with ASD may spontaneously use more gestures when the gestures have a self-serving function, such as helping direct the internal focus of attention, scaffolding problem solving, or retrieving words from the lexicon. There is abundant evidence that gesture fills these cognitive roles in typical populations (e.g., Goldin-Meadow, 2005; Morsella & Krauss, 2005); however, this possibility has not been investigated in ASD samples. Insight into strategies that may scaffold cognitive processes in ASD is important for the development of interventions targeting cognitive development. Although participants with ASD gestured more on the inhibition task than participants with TD, our data did not conclusively demonstrate that gesture served as a cognitive aid in our ASD group. We found that TD participants who gestured more performed better on the speeded naming component of the task, suggesting that gesture facilitated task performance to a certain extent. However, gesture frequency was not correlated with task performance in the ASD sample, suggesting that, although they gestured frequently, participants with ASD did not benefit from their gesture use.

It remains an open question why our ASD participants gestured so much on the inhibition task when, as a group, they did not appear to benefit from this behavior. One possibility is that some adolescents, but not others, benefited from pointing to the stimuli, obscuring any group-level relationships between gesture and task performance. Another possibility is that the adolescents in our sample, who have received years of services targeting a wide range of social, cognitive, and academic skills, were *taught* to use pointing as a strategy to facilitate visual tracking, for example, during reading. Thus, they generalized a strategy they learned elsewhere that may in fact not have been advantageous during the inhibition task. The current study was not specifically designed to look at the potential benefits of gesture on cognition in ASD. Future studies should directly investigate this interesting area by using tasks in which gesture is known to facilitate cognition in TD samples. This work is currently being done in our lab.

4.4 Limitations and Future Directions

This study had several limitations that should be addressed in future studies of common ground in ASD. Most noteworthy was the order effect that we observed. The effect of study order made some of our findings, particularly those having to do with the gesture constraint, difficult to interpret. Because we did not anticipate that order would influence the difference between the gesture constraint conditions to such an extent, this was not fully controlled in our counterbalancing. In fact, as noted in the *Order Effect* section, few if any studies using gesture constraints have explored the effect of order. Our finding of a strong order effect suggests that all future studies using a gesture constraint manipulation should carefully counterbalance the order of the constraint with other conditions *and* report any order effects (including null effects) in their analyses and conclusions.

In the current study, we chose to use a narrative format based on cartoons rather than a conversation or personal narrative. Structured narratives based on experimenter-provided stimuli appear to be easier than personal narratives for individuals with ASD (Losh & Capps, 2003). Our choice of procedure is advantageous in that narratives are based on identical stimuli, allowing us to closely compare narratives across participants; however, it also could potentially have obscured interesting group differences that may have been apparent with a more challenging task. A future study could design a common ground task based on personal narratives, for example, by asking participants to discuss a favorite vacation, as in Losh and Capps (2003). In the shared condition, the experimenter would say, for example, "oh, I've been to Disneyworld lots of times!" and in the private condition, "I've never been there," resulting in a difference in common ground inferences between the two conditions. This type of design would also allow us to see if the use of common ground in ASD samples differs when shared knowledge must be inferred rather than made explicit by the experimental setup (as in the current study). Such a task may also require less processing, as it is dependent on gestalt adjustments in speech rather than requiring speakers to track specific knowledge held by the listener, so it may in fact emerge earlier in development, both in TD and ASD samples. On the other hand, such a task may in fact require *more* processing, as it requires both commonality assessment and message formation components of audience design (Horton & Gerrig, 2005a), whereas our task only required message formation.

The discussion of common ground presented here has focused primarily on the role of the speaker and the role of the listener as separate processes, and in our experimental setup, as in many others, this truly was the case, as the participant's interlocutor was always a research assistant. This method is useful for disentangling what the speaker and listener each bring to the communicative table. However, this method necessarily oversimplifies the communicative process as well. The early referential communication studies beautifully demonstrated that communication is in fact a back-and-forth dynamic process, and that what the "speaker" contributes can never be fully separated from what the "listener" does, since, in reality, the two roles are constantly in flux. Future studies of common ground in ASD may be able to utilize referential communication tasks to examine the more dynamic processes of common ground. The listeners in the current study were research assistants who heard the same stories told many times. Although all listeners who assisted in data collection were attentive and engaged, the repetition of hearing the same stories, and seeing the same cartoon previews, many times could have affected their ability to attend closely to the previews, as well as the way in which they listened to the stories, in a manner potentially discernible to participants.

In terms of our gesture analyses, we only looked at gestures produced during the unconstrained conditions, in which participants were able to gesture freely. This approach may in fact have limited what our data are able to tell us about gesture in ASD. For example, typical adults make more facial expressions and in general move their faces more when their arm/hand and foot/leg movements are restricted, (Rime et al., 1984). In the current study, we only restricted participants' hands, and not any other body part, so our gesture restriction condition could potentially provide insight as to what participants do with their arms, legs, and faces when they do not have access to manual gestures. Future analyses will look at other movements made by participants under conditions of gesture restriction. Although it did not happen often, some participants with ASD occasionally lifted the Velcro desk off their laps when particularly animated, and head movements were also noticeably present, both with and without gesture restriction. These expressive movements were not analyzed in the current study.

Anecdotally, participants in both groups did not tend to explicitly express the fact that they shared common ground with the listener as often as had been expected. However, some participants (from both groups) did at times include phrases such as, "like you saw in the preview," or, "after the part you saw." Our finding that participants with ASD did not reduce their word count in response to shared common ground with the listener demonstrates that at an *implicit* level common ground did not affect the narratives produced by adolescents with ASD. However, these data do not tell us anything about how common ground may have been referenced by participants at an *explicit* level. It is certainly possible that participants with ASD may have explicitly acknowledged the common ground shared with their interlocutor, and even recognized how a nod to this common ground might facilitate listener comprehension. If this were in fact the case, it would suggest that pragmatic language processes, such as common ground, may have an impact at an explicit level, but not a deeper, potentially more automatic implicit level, consistent with observations that individuals with ASD often struggle more with implicit aspects of communication and social interaction as compared to explicit processes. This hypothesis can and will be tested with the existing dataset.

The age effect observed in our ASD sample suggests that some aspect of social or cognitive development has led to a more normative use of common ground in these adolescents. One possibility is that specific interventions may have resulted in better use of common ground over time. Older participants would have had both more years of intervention, and exposure to a wider range of interventions targeting pragmatic language. Given the wide range of treatments available to children and adolescents with ASD, it would be very difficult, if not impossible, to quantify the relationship between the intervention(s) a given participant received and their use of common ground. However, this would be very interesting and important information to obtain. To gain insight into this possibility, treatment studies could include implicit measures of pragmatic language, such as common ground use, as outcome variables.

4.5 Concluding statements

This study sought to explore the common ground effect in adolescents with ASD and adolescents with a typical developmental history. Specifically, we were interested in whether or not this effect exists in ASD, and in what factors may underlie its use, or lack thereof. Adolescents with TD showed a clear common ground effect; however, the effect did not appear to be driven by any of the specific individual factors we measured, such as executive function, or demographic factors including age, IQ, or receptive vocabulary. There was some suggestion that gesture use may have affected common ground in TD, given that the common ground effect was enhanced when gestures were restricted; however, due to the order effect we observed, and its interaction with our gesture constraint conditions, most of the gesture findings from this study are difficult to interpret. Adolescents with ASD, in contrast to adolescents with TD, showed no common ground effect at the group level. However, two specific findings suggest that these adolescents did in fact respond to differences in common ground. For one, age was strongly related to the common ground effect in this sample, with older participants demonstrating a common ground effect. Second, naïve raters observed that narratives produced by participants with ASD were harder to follow under conditions of shared ground, suggesting that they may have omitted information that was shared with the listener, thus making their stories harder for raters to follow. Overall, the results of our common ground analyses suggest that while TD adolescents clearly have the subtle pragmatic language skills necessary to demonstrate a common ground effect, this process may be undergoing a period of development in adolescents with ASD. Future studies of common ground in ASD should be conducted to explore what specific aspects of its pragmatics are and are not being used by those with ASD.

Finally, our results provided some interesting data on gesture use in ASD. Adolescents with ASD gestured at a lower rate than adolescents with TD on our narrative task. However, the same group of adolescents with ASD spontaneously gestured at a much *higher* rate than controls during an executive function task, suggesting that gesture use in ASD is very task-dependent. Gesture use in ASD may depend on the task's inherent structure, on the social content of the gesture itself, or on the potential that the gesture has to benefit the speaker him or herself. The clear distinction between gesture use on these two very different tasks points to the importance of gesture research in ASD (which to date has been relatively minimal). Gesture research in ASD should focus on when it is and is not employed in this population, as well as its potential to benefit communication and cognitive functioning.

References

Ackerman, B. P., Szymanski, J., & Silver, D. (1990). Children's use of the common ground in interpreting ambiguous referential utterances. *Developmental Psychology, 26*(2), 234.

Alibali, M., Kita, S., & Young, A. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processes, 15*(6), 593-613.

Andersen, E. (1990). *Speaking with Style*. London: Routledge.

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

Arnold, J., & Griffin, Z. (2007). The effect of additional characters on choice of referring expression: Everyone counts. *Journal of Memory and Language, 56*(4), 521-536.

Attwood, H. H., Frith, U., & Hermelin, B. (1988). The understanding and use of interpersonal gestures by autistic and Down's syndrome children. *Journal of Autism and Developmental Disorders, 18*, 241-257.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556.

Baltaxe, C. (1977). Pragmatic deficits in the language of autistic adolescents. *Journal of Pediatric Psychology*, *2*(4), 176-180.

Baltaxe, C., & D'Angiola, N. (1996). Referencing skills in children with autism and specific language impairment. *International Journal of Language & Communication Disorders, 31*(3), 245-258.

Barnard, L., Muldoon, K., Hasan, R., O'Brien, G., & Stewart, M. (2008). Profiling executive dysfunction in adults with autism and comorbid learning disability. *Autism*, *12*(2), 125-141.

Baron-Cohen, S. (1989). Perceptual role-taking and protodeclarative pointing in autism. *British Journal of Developmental Psychology, 7*, 113-127.

Baron-Cohen, S. (2000). Theory of mind and autism: A 15-year review. *Understanding other minds: Perspectives from developmental cognitive neuroscience,* 3-21.

Barr, D. J., & Keysar, B. (2002). Anchoring comprehension in linguistic precedents. *Journal of Memory and Language, 46*(2), 391-418.

Bartak, L., Rutter, M., & Cox, A. (1975). Comparative study of infantile autism and specific developmental receptive language disorders. I. The children. *British Journal of Psychiatry*, *126*, 127-145.

Beattie, G., & Coughlan, J. (1999). An experimental investigation of the role of iconic gestures in lexical access using the tip-of-the-tongue phenomenon. *British Journal of Psychology, 90*(1), 35-56.

Beattie, G., & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany speech. *Journal of Language and Social Psychology*, *18*(4), 438-462.

Bennetto, L., Pennington, B. F., & Rogers, S. (1996). Intact and impaired memory functions in autism. *Child Development*, *67*, 1816-1835.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, *39*, 15-22.

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641-1660.

Bishop, D., & Norbury, C. (2005). Executive functions in children with communication impairments, in relation to autistic symptomatology: 2: Response inhibition. *Autism*, *9*(1), 29-43.

Bono, M., Daley, T., & Sigman, M. (2004). Relations among joint attention, amount of intervention and language gain in autism. *Journal of autism and developmental disorders*, *34*(5), 495-505.

Bono, M. A., Daley, T., & Sigman, M. (2004). Relations among joint attention, amount of intervention and language gain in autism. *Journal of Autism and Developmental Disorders, 34*, 495-505.

Boucher, J. (1988). Word fluency in high-functioning autistic children. *Journal of Autism and Developmental Disorders*, *18*(4), 637-645.

Brennan, S., & Clark, H. (1996). Conceptual pacts and lexical choice in conversation. *Learning, Memory, 22*(6), 1482-1493.

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, *136*, 539-550.

Brown, P., & Dell, G. (1987). Adapting production to comprehension: The explicit mention of instruments. *Cognitive Psychology, 19*(4), 441-472.

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

Butterworth, B., & Hadar, U. (1989). Gesture, speech, and computational stages: A reply to McNeill. *Psychological Review, 96*(1), 168-174.

Camaioni, L., Perucchini, P., Muratori, F., & Milone, A. (1997). Brief report: A longitudinal examination of the communicative gestures deficit in young children with autism. *Journal of Autism and Developmental Disorders, 27*, 715-725.

Capps, L., Kehres, J., & Sigman, M. (1998). Conversational abilities among children with autism and children with developmental delays. *Autism, 2*, 325-344.

Capps, L., Losh, M., & Thurber, C. (2000). The Frog Ate the Bug and Made his Mouth Sad: Narrative Competence in Children with Autism. *Journal of Abnormal Child Psychology, 28*(2), 193-204.

Cassell, J., McNeill, D., & McCullough, K.-E. (1998). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics and Cognition, 6*, 1-24.

Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language, 30*(01), 213-236.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. *Cognition, 23*, 43-71.

Clark, E., & Bernicot, J. (2008). Repetition as ratification: How parents and children place information in common ground. *Journal of Child Language, 35*(2), 349-371.

Clark, H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1-39.

Clark, H. H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

Clark, H. H., & Carlson, T. B. (1982). Hearers and speech acts. *Language, 58*(2), 332-373.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. *Advances in Psychology, 9*, 287-299.

Constantino, J., Abbacchi, A., Lavesser, P., Reed, H., Givens, L., Chiang, L., Gray, T., Gross, M., Zhang, Y., & Todd, R. (2009). Developmental course of autistic social impairment in males. *Development and Psychopathology, 21*(1), 127-138.

Constantino, J., & Gruber, C. (2005). *The Social Responsiveness Scale*. Los Angeles: Western Psychological Services.

Constantino, J., Przybeck, T., Friesen, D., & Todd, R. (2000). Reciprocal social behavior in children with and without pervasive developmental disorders. *Journal of Developmental & Behavioral Pediatrics, 21*(1), 2-11.

Constantino, J. N., & Todd, R. D. (2003). Autistic traits in the general population: A twin study. *Archives of General Psychiatry, 60*(5), 524-530.

Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: Speakers' gestures affect listeners' action. *Cognition, 113*(1), 98-104.

de Marchena, A., & Eigsti, I. M. (2010). Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency. *Autism Research, 3*(6), 311-322.

Diehl, J. J., Bennetto, L., & Young, E. C. (2006). Story recall and narrative competence in high-functioning children with autism spectrum disorders. *Journal of Abnormal Child Psychology, 34*(1), 83-98.

Dunn, L., & Dunn, L. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Eales, M. (1993). Pragmatic impairments in adults with childhood diagnoses of autism or developmental receptive language disorder. *Journal of Autism and Developmental Disorders, 23*(4), 593-617.

Eigsti, I. M. (2011). Executive functions. In D. Fein (Ed.), *Neuropsychology of Autism* (pp. 185-204). New York: Oxford University Press.

Eigsti, I. M., de Marchena, A., & Dixon, J. (2010). Conversational gestures in autism spectrum disorders: A qualitative difference, *Boston University Conference on Language Development*. Boston, MA.

Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective taking as egocentric anchoring and adjustment. *Journal of Personality and Social Psychology, 87*(3), 327-339.

Frick-Horbury, D., & Guttentag, R. (1998). The effects of restricting hand gesture production on lexical retrieval and free recall. *The American Journal of Psychology*, *111*(1), 43-62.

Fussell, S., & Krauss, R. (1992). Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology, 62*(3), 378-391.

Galati, A., & Brennan, S. E. (2010). Attenuating information in spoken communication: For the speaker, or for the addressee? *Journal of Memory and Language, 62*(1), 35-51.

Garber, P., & Goldin-Meadow, S. (2006). Gesture offers insight into problem-solving in adults and children. *Cognitive Science, 26*, 817-831.

García-Pérez, R. M., Lee, A., & Hobson, R. P. (2007). On intersubjective engagement in autism: A controlled study of nonverbal aspects of conversation. *Journal of Autism and Developmental Disorders, 37*, 1310-1322.

Geller, E. (1998). An investigation of communication breakdowns and repairs in verbal autistic children. *British Journal of Developmental Disabilities, 44*, 71-85.

Glucksberg, S., & Krauss, R. M. (1967). What do people say after they have learned to talk? Studies of the development of referential communication. *Merrill-Palmer Quarterly: Journal of Developmental Psychology, 13*(4), 309-316.

Goldin-Meadow, S. (2005). *Hearing Gesture: How Our Hands Help Us Think.* Cambridge: Harvard University Press.

Goldin-Meadow, S., Nusbaum, H. C., Garber, P., & Church, R. B. (1993). Transitions in learning: Evidence for simultaneously activated strategies. *Journal of Experimental Psychology: Human Perception and Performance, 19*(1), 92-107.

Goldin-Meadow, S., Nusbaum, H. C., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516-522.

Goldstein, G., Johnson, C., & Minshew, N. (2001). Attentional processes in autism. *Journal of autism and developmental disorders, 31*(4), 433-440.

Gopnik, A., Capps, L., & Meltzoff, A. (2000). Early theories of mind: What the theory theory can tell us about autism. *Understanding other minds: Perspectives from developmental cognitive neuroscience*, 50-72.

Gordon, P., Grosz, B., & Gilliom, L. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive science, 17*(3), 311-347.

Graham, J., & Heywood, S. (1975). The effects of elimination of hand gestures and of verbal codability on speech performance. *European Journal of Social Psychology, 5*(2), 189-195.

Grice, H. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics: Vol. 3: Speech acts* (pp. 41-58). New York: Seminar Press.

Grosz, B., Weinstein, S., & Joshi, A. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics, 21*(2), 203-225.

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*(2), 274-307.

Hanna, J., Tanenhaus, M., & Trueswell, J. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*(1), 43-61.

Hill, E. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences, 8*(1), 26-32.

Hill, E. L., & Bird, C. M. (2006). Executive processes in Asperger syndrome: Patterns of performance in a multiple case series. *Neuropsychologia, 44*(14), 2822-2835.

Holler, J., & Beattie, G. (2003). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener? *Gesture*, *3*(2), 127-154.

Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior, 33*(2), 73-88.

Holler, J., & Wilkin, K. (2009). Communicating common ground: How mutually shared knowledge influences speech and gesture in a narrative task. *Language and Cognitive Processes, 24*(2), 267-289.

Horton, W., & Gerrig, R. (2005a). Conversational common ground and memory processes in language production. *Discourse Processes, 40*(1), 1-35.

Horton, W., & Gerrig, R. (2005b). The impact of memory demands on audience design during language production. *Cognition, 96*(2), 127-142.

Horton, W., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition*, *59*(1), 91-117.

Hudson-D'Zmura, S., & Tanenhaus, M. K. (1998). Assigning antecedents to ambiguous pronouns: The role of the center of attention as the default assignment. In M. Walker & A. Joshi & E. Prince (Eds.), *Centering Theory in Discourse* (pp. 199-226). Oxford: Oxford University Press.

Hughes, C., Russell, J., & Robbins, T. W. (1994). Evidence for executive dysfunction in autism. *Neuropsychologia, 32*, 477-492.

Hupet, M., Berrewaerts, J., & Feyereisen, P. (2007). Pragmatic skills in the early stages of Alzheimer's disease: An analysis by means of a referential communication task. *International Journal of Language and Communication Disorders, 42*(1), 1-17.

Isaacs, E., & Clark, H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General, 116*(1), 26-37.

Iverson, J., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature, 396*, 228.

Johnston, J., Miller, J., & Tallal, P. (2001). Use of cognitive state predicates by language-impaired children. *International Journal of Language & Communication Disorders*, *36*(3), 349-370.

Joseph, R., Steele, S., Meyer, E., & Tager-Flusberg, H. (2005). Self-ordered pointing in children with autism: failure to use verbal mediation in the service of working memory? *Neuropsychologia*, *43*(10), 1400-1411.

Kaland, N., Smith, L., & Mortensen, E. L. (2008). Brief report: Cognitive flexibility and focused attention in children and adolescents with asperger syndrome or high-functioning autism as measured on the computerized version of the wisconsin card sorting test. *Journal of autism and developmental disorders, 38*(6), 1161-1165.

Kelley, E., Paul, J., Fein, D., & Naigles, L. (2006). Residual language deficits in optimal outcome children with a history of autism. *Journal of Autism and Developmental Disorders*, *36*(6), 807-828.

Keysar, B., Barr, D., Balin, J., & Paek, T. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language, 39*, 1-20.

Keysar, B., Barr, D., & Horton, W. (1998). The egocentric basis of language use: Insights from a processing approach. *Current Directions in Psychological Science*, *7*(2), 46-50.

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

Kita, S. (2000). How representational gestures help speaking. In D. McNeill (Ed.), *Language and Gesture* (pp. 162-185). Cambridge: Cambridge University Press.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Defining and quantifying the social phenotype in autism. *American Journal of Psychiatry, 159*, 895-908.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II*. San Antonio, TX: Psychological Corporation.

Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, *4*(3), 343-346.

Kronmuller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language, 56*(3), 436-455.

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

Landa, R. J., & Goldberg, M. C. (2005). Language, social, and executive functions in high functioning autism: A continuum of performance. *Journal of Autism and Developmental Disorders, 35*(5), 557-573.

Liles, B., Duffy, R., Merritt, D., & Purcell, S. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech and Hearing Research, 38*(2), 415-425.

Liss, M., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., Waterhouse, L., & Rapin, I. (2001). Executive functioning in high-functioning children with autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*(02), 261-270.

Liss, M., Harel, B., Fein, D., Allen, D., Dunn, M., Feinstein, C., Morris, R., Waterhouse, L., & Rapin, I. (2001). Predictors and correlates of adaptive functioning in children with developmental disorders. *Journal of Autism and Developmental Disorders, 31*(2), 219-230.

Lloyd, P., Mann, S., & Peers, I. (1998). The growth of speaker and listener skills from five to eleven years. *First Language, 18*(52), 81-103.

Lockridge, C., & Brennan, S. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review, 9*(3), 550-557.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (2002). *Autism Diagnostic Observation Schedule (ADOS)* Los Angeles: Western Psychological Services.

Lord, C., Rutter, M., & LeCouteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24,* 659-685.

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

Loveland, K. A., & Landry, S. H. (1986). Joint attention and language in autism and developmental language delay. *Journal of Autism and Developmental Disorders, 16*, 335-349.

Loveland, K. A., McEvoy, R., & Tunali, B. (1990). Narrative story telling in autism and Down's syndrome. *British Journal of Developmental Psychology, 8*(1), 9-23.

Luyster, R., Lopez, K., & Lord, C. (2007). Characterizing communicative development in children referred for Autism Spectrum Disorders using the MacArthur-Bates Communicative Development Inventory (CDI). *Journal of Child Language, 34*(3), 623-654.

Martin, I., & McDonald, S. (2003). Weak coherence, no theory of mind, or executive dysfunction? Solving the puzzle of pragmatic language disorders. *Brain and Language, 85*(3), 451-466.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

Minshew, N., & Goldstein, G. (1998). Autism as a disorder of complex information processing. *Mental Retardation and Developmental Disabilities Research Reviews, 4*(2), 129-136.

Minshew, N., & Goldstein, G. (2001). The pattern of intact and impaired memory functions in autism. *Journal of Child Psychology and Psychiatry, 42*(8), 1095-1101.

Minshew, N., Goldstein, G., Muenz, L., & Payton, J. (1992). Neuropsychological functioning nonmentally retarded autistic individuals. *Journal of Clinical and Experimental Neuropsychology, 14*(5), 749-761.

Minshew, N., Goldstein, G., & Siegel, D. (1995). Speech and language in high-functioning autistic individuals. *Neuropsychology, 9*, 255-255.

Minshew, N., Goldstein, G., & Siegel, D. (1997). Neuropsychologic functioning in autism: Profile of a complex information processing disorder. *Journal of the International Neuropsychological Society*, *3*(04), 303-316.

Morsella, E., & Krauss, R. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology, 117*(3), 411-424.

Morsella, E., & Krauss, R. (2005). Muscular activity in the arm during lexical retrieval: Implications for gesture-speech theories. *Journal of psycholinguistic research, 34*(4), 415-427.

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*, 115-128.

Mundy, P., Sigman, M., Ungerer, J., & Sherman, T. (1986). Defining the social deficits of autism: The contribution of non-verbal communication measures. *Journal of Child Psychology and Psychiatry, 27*, 657-669.

Nadig, A., & Sedivy, J. (2002). Evidence of perspective-taking constraints in children's on-line reference resolution. *Psychological Science*, *13*(4), 329-336.

Nadig, A., Vivanti, G., & Ozonoff, S. (2009). Adaptation of object descriptions to a partner under increasing communicative demands: a comparison of children with and without autism. *Autism Research, 2*(6), 334-347.

Nelson, E. E., Leibenluft, E., McClure, E., & Pine, D. S. (2005). The social re-orientation of adolescence: A neuroscience perspective on the process and its relation to psychopathology. *Psychological Medicine, 35*(2), 163-174.

Nilsen, E., & Graham, S. (2009). The relations between children's communicative perspective-taking and executive functioning. *Cognitive Psychology, 58*(2), 220-249.

Norbury, C., & Bishop, D. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders, 38*(3), 287-313.

O'Neill, D. K., & Holmes, A. C. (2002). Young preschoolers' ability to reference story characters: The contribution of gestures and character speech. *First Language, 22*, 73-103.

O'Neill, D. K., & Topolovec, J. C. (2001). Two-year-old children's sensitivity to the referential (in)efficacy of their own pointing gestures. *Journal of Child Language, 28*(1), 1-28.

Ozonoff, S., Cook, I., Coon, H., Dawson, G., Joseph, R., Klin, A., McMahon, W., Minshew, N., Munson, J., & Pennington, B. (2004). Performance on Cambridge Neuropsychological Test Automated Battery subtests sensitive to frontal lobe function in people with autistic disorder: evidence from the Collaborative Programs of Excellence in Autism network. *Journal of Autism and Developmental Disorders, 34*(2), 139-150.

Ozonoff, S., & Jensen, J. (1999). Specific executive function profiles in three neurodevelopmental disorders. *Journal of Autism and Developmental Disorders, 29*, 171-177.

Ozonoff, S., Pennington, B., & Rogers, S. (1991a). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32*(7), 1081-1105.

Ozonoff, S., Pennington, B. F., & Rogers, S. (1991b). Executive function deficits in high-functioning autistic individuals: Relationship to theory of mind. *Journal of Child Psychology and Psychiatry, 32*, 1081-1105.

Ozonoff, S., & Strayer, D. L. (1997). Inhibitory function in nonretarded children with autism. *Journal of Autism and Developmental Disorders, 27*(1), 59-77.

Ozonoff, S., & Strayer, D. L. (2001). Further evidence of intact working memory in autism. *Journal of Autism and Developmental Disorders, 31*(3), 257-263.

Ozonoff, S., Strayer, D. L., McMahon, W. M., & Filloux, F. (1994). Executive function abilities in autism and Tourette syndrome: An information processing approach. *Journal of Child Psychology and Psychiatry, 35*(6), 1015-1032.

Paul, R. (2007). Communication and its development in autism spectrum disorders. In F. Volkmar (Ed.), *Autism and Pervasive Developmental Disorders* (pp. 129-155). Cambridge: Cambridge University Press.

Perkins, M. (1998). Is pragmatics epiphenomenal? Evidence from communication disorders. *Journal of Pragmatics, 29*(3), 291-311.

Perry, M., Church, R. B., & Goldin-Meadow, S. (1988). Transitional knowledge in the acquisition of concepts. *Cognitive Development, 3*(4), 359-400.

Pine, K. J., Bird, H., & Kirk, E. (2007). The effects of prohibiting gestures on children's lexical retrieval ability. *Developmental Science, 10*(6), 747-754.

Prior, M., & Ozonoff, S. (2007). Psychological factors in autism. In F. Volkmar (Ed.), *Autism and Pervasive Developmental Disorders* (2nd ed., pp. 69-128). New York, NY: Cambridge University Press.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7,* 226-231.

Rime, B., Schiaratura, L., Hupet, M., & Ghysselinckx, A. (1984). Effects of relative immobilization on the speaker's nonverbal behavior and on the dialogue imagery level. *Motivation and Emotion, 8*(4), 311-325.

Robbins, T., James, M., Owen, A., Sahakian, B., McInnes, L., & Rabbitt, P. (1994). Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. *Dementia and Geriatric Cognitive Disorders, 5*(5), 266-281.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 31*, 131-144.

Robinson, S., Goddard, L., Dritschel, B., Wisley, M., & Howlin, P. (2009). Executive functions in children with autism spectrum disorders. *Brain and Cognition, 71,* 362-368.

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition*. Itsaca, IL: Riverside Publishing.

Roxβnagel, C. (2000). Cognitive load and perspective taking: applying the automatic controlled distinction to verbal communication. *European Journal of Social Psychology, 30*(3), 429-445.

Rumsey, J., & Hamburger, S. (1988). Neuropsychological findings in high-functioning men with infantile autism, residual state. *Journal of Clinical and Experimental Neuropsychology, 10*(2), 201-221.

Rumsey, J. M., & Hamburger, S. D. (1990). Neuropsychological divergence of high-level autism and severe dyslexia. *Journal of Autism and Developmental Disorders, 20*(2), 155-168.

Russell, J. (1997a). *Autism as an Executive Disorder*. Oxford: Oxford University Press.

Russell, J. (1997b). How executive disorders can bring about an inadequate 'theory of mind'. In J. Russell (Ed.), *Autism as an Executive Disorder* (pp. 256-304). Oxford: Oxford University Press.

Russell, J., Jarrold, C., & Henry, L. (1996). Working memory in children with autism and with moderate learning difficulties. *Journal of Child Psychology and Psychiatry, 37*(6), 673-686.

Rutter, M., Bailey, A., & Lord, C. (2003). *The Social Communication Questionnaire*. Los Angeles: Western Psychological Services.

Sachs, J., & Devin, J. (1976). Young children's use of age-appropriate speech styles in social interaction and role-playing. *Journal of Child Language, 3*(1), 81-98.

Schober, M., & Clark, H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology, 21*(2), 211-232.

Schuh, J. M. (2010). *Pragmatic Language Abilities: Working Memory Influences on Mutual Information.* (Doctoral dissertation) University of Connecticut, Storrs, CT.

Schuh, J., Mirman, D., & Eigsti, I.M. (under review). Finding common ground: The impact of individual differences in working memory on referential communication for children and adults.

Scott, F., & Baron-Cohen, S. (1996). Imagining real and unreal things: Evidence of a dissociation in autism. *Journal of Cognitive Neuroscience, 8*(4), 371-382.

Shatz, M., & Gelman, R. (1973). The development of communication skills: Modifications in the speech of young children as a function of listener. *Monographs of the Society for Research in Child Development, 38*(5), 1-38.

Shu, B. C. (2001). Executive function deficits in non-retarded autistic children. *Autism, 5*(2), 165-174.

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences, 9*(2), 69-74.

Surian, L. (1996). Are children with autism deaf to Gricean Maxims? *Cognitive Neuropsychiatry, 1*(1), 55-72.

Szatmari, P., & Tuff, L. (1990). Asperger's syndrome and autism: Neurocognitive aspects. *Journal of the American Academy of Child & Adolescent Psychiatry, 29*(1), 130-136.

Tager-Flusberg, H. (1995). 'Once upon a ribbit': stories narrated by autistic children. *British Journal of Developmental Psychology, 13*, 45-59.

Tager-Flusberg, H., Paul, R., & Lord, C. E. (2005). Language and communication in autism. In F. Volkmar & R. Paul & A. Klin & D. J. Cohen (Eds.), *Handbook of Autism and Pervasive Developmental Disorders* (Fourth ed., Vol. 1, pp. 335-364). New York: Wiley.

Tager-Flusberg, H., & Sullivan, K. (1995). Attributing mental states to story characters: A comparison of narratives produced by autistic and mentally retarded individuals. *Applied Psycholinguistics, 16*(3), 241-256.

Tantam, D., Holmes, D., & Cordess, C. (1993). Nonverbal expression in autism of Asperger type. *Journal of Autism and Developmental Disorders, 23*, 111-133.

Turner, M. (1997). Towards an executive dysfunction account of repetitive behaviour in autism. In J. Russell (Ed.), *Autism as an Executive Disorder* (pp. 57-100). Oxford: Oxford University Press.

Turner, M. (1999). Generating novel ideas: Fluency performance in high-functioning and learning disabled individuals with autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*(2), 189-201.

Volden, J. (2004). Conversational repair in speakers with autism spectrum disorder. *International Journal of Language & Communication Disorders, 39*(2), 171-189.

Volden, J., Magill-Evans, J., Goulden, K., & Clarke, M. (2007). Varying language register according to listener needs in speakers with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 37*(6), 1139-1154.

Volden, J., & Sorenson, A. (2009). Bossy and nice requests: Varying language register in speakers with autism spectrum disorder (ASD). *Journal of Communication Disorders, 42*(1), 58-73.

Warren, A. R., & Tate, C. S. (1992). Egocentrism in childrenís telephone conversations. In R. M. Diaz & L. Berk, E. (Eds.), *Private speech: From social interaction to self-regulation* (pp. 245-264). Hillsdale, NJ, England: Lawrence Elrbaum Associates, Inc.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children: Fourth Edition (WISC-IV).* San Antonio, TX: Psychological Corporation.

Wesp, R., Hesse, J., Keutmann, D., & Wheaton, K. (2001). Gestures maintain spatial imagery. *The American Journal of Psychology, 114*(4), 591-600.

Wetherby, A. M., & Prutting, C. A. (1984). Profiles of communicative and cognitive-social abilities in autistic children. *Journal of Speech and Hearing Research, 27*, 364-377.

Ziatas, K., Durkin, K., & Pratt, C. (2003). Differences in assertive speech acts produced by children with autism, Asperger syndrome, specific language impairment, and normal development. *Development and Psychopathology, 15*(1), 73-94.

Appendix A

Example quiz (given after the practice trial) with correct answers circled and filled in

<u>Practice Quiz</u>

1) What happens when the roadrunner first runs really fast?
    a. He takes off like an airplane
    b. The road wobbles
    c. He becomes invisible
    d. He erases the lines on the road

2) What does the coyote use to watch the roadrunner?
    a. Infrared glasses
    b. A telescope
    c. Binoculars
    d. A webcam

3) What is the coyote holding while he's chasing the roadrunner?

    <u>A fork and a knife</u>

4) Describe the expression on the coyote's face after the roadrunner runs off:

    <u>His jaw drops to the ground</u>

5) Which event <u>was</u> in the preview?
    a. The introduction of the roadrunner's scientific name ("Accelleratii Incredibus")
    b. The coyote watching the roadrunner from a cliff
    c. The coyote putting on his bib
    d. The coyote dashing off to the hills

Appendix B

Questions given to undergraduate story raters

1. Did the storyteller watch this cartoon alone or with the listener?

❑ ALONE ❑ WITH THE LISTENER

2. How well were you able to visualize the story as you read it?

1	2	3	4	5	6	7
Poorly – hardly pictured anything						Very well - pictured every detail

3. How easy was it to follow and understand the story?

1	2	3	4	5	6	7
Very difficult; confusing, the plot didn't make sense						Totally coherent, a very clear plot

Appendix C
Study results presented split by order

Due to the observed order effect, we present here the analyses described in the main body of the document, split by Order 1 and Order 2, for completeness. Please note that dividing our sample by order, a between-subjects factor, significantly limits the power of our study to detect group differences, thus increasing the chance of a Type II error.

Baseline word count and gesture frequency. For word count by group, there was a trend for participants with ASD to tell shorter stories, t (15) = 2.11, p = .051, Cohen's d = 1.02, in Order 1; however, for Order 2, there was no difference in word count between groups, t (17) = 0.47, p = .65, Cohen's d = 0.22. In the collapsed analysis, participants with TD gestured at a higher rate than participants with ASD. After splitting by order, this difference only reached significance in Order 2, t (17) = 2.16, p = .045, Cohen's d = 1.05, and not in Order 1, t (15) = 1.47, p = .16, Cohen's d = 0.76; however, the direction of the effect was the same in both orders: participants with TD showed a higher rate of gestures than participants with ASD.

Enjoyment ratings and quiz performance. Order did not affect participants' subjective ratings of how much they enjoyed the cartoons: there was no group difference in enjoyment of the cartoons, either for Order 1, t (15) = 0.75, p = .47, Cohen's d = .39, or Order 2, t (17) = -1.04, p = .31, Cohen's d = -0.50. With respect to performance on the cartoon quizzes, there was a significant main effect of group in Order 1, F(1,15) = 5.08, p = .04, partial η^2 = 0.25, but this effect was not significant in Order 2, F(1,17) = 0.06, p = .82, partial η^2 = 0.003. Finally, for questions asking specifically about the cartoon previews, there was no significant difference between groups for either Order 1, *t* (15) = 1.54, *p* = .14, Cohen's *d* = 0.80, or Order 2, *t* (17) = 1.06, *p* = .30, Cohen's *d* = 0.51.

Common ground. After splitting the sample by order, none of the effects observed in the collapsed analysis reached significance in either group. There was no main effect of group in either Order 1, $F(1,15) = 2.92$, $p = .11$, partial $\eta^2 = 0.16$, or Order 2, $F(1,17) = 0.72$, $p = .41$, partial $\eta^2 = .04$, and there was no main effect of ground in either Order 1, $F(1,15) = 0.003$, $p = .96$, partial $\eta^2 < .001$, or Order 2, $F(1,17) = 2.47$, $p = .13$, partial $\eta^2 = 13$. The interaction between group and ground also failed to reach significance in both Order 1, $F(1,15) = 3.39$, $p = .09$, partial $\eta^2 = 0.18$, and Order 2, $F(1,17) = 2.47$, $p = .14$, partial $\eta^2 = 0.13$, although there was a trend for the same interaction observed in the collapsed analysis in Order 1. It should be noted that, while not significant, the pattern of results for this interaction was similar for Order 1 and Order 2. Post-hoc *t*-tests revealed that although TD participants in Order 2 used significantly more words in the private as compared to the shared condition, *t* (8) = 2.48, *p* = .04, Cohen's *d* = 1.75, with a large effect size, as in the collapsed analysis, this difference did not reach significance in Order 1, *t* (8) = 1.38, *p* = .21, Cohen's *d* = 0.98, although the direction of the effect was in the predicted direction. There was no difference in word count between the private and shared conditions in the ASD group, either for Order 1, *t* (7) = -1.23, *p* = .26, Cohen's *d* = -0.93, or Order 2, *t* (9) = 0.003, *p* = .997, Cohen's *d* = 0.002.

Executive function and common ground. Split by order, there were no significant correlations between the executive function composite and the common ground effect variable. For Order 1, the collapsed correlation was not significant, *r*(17) = .11, *p* = .69, nor was the correlation for the ASD, *r*(8) = -.25, *p* = .55 or TD, *r*(9) = .41, *p* = .27, group.

For Order 2, the collapsed correlation was not significant, $r(19) = .21$, $p = .30$, nor was the correlation for the ASD, $r(10) = .29$, $p = .41$ or TD, $r(9) = .84$, $p = .08$, group.

Social skills and common ground. The correlation between SRS scores and the common ground effect variable did not reach significance in either order; however, the direction of the correlation was negative for both Order 1, $r(16) = -.30$, $p = .26$, and Order 2, $r(18) = -.37$, $p = .10$. For Order 1 there was no significant correlation in either the ASD group, $r(7) = .32$, $p = .49$ or the TD group, $r(9) = .58$, $p = .10$. For Order 2, there was a significant negative correlation in the ASD group, $r(9) = -.71$, $p = .03$ and no correlation in the TD group, $r(9) = .44$, $p = .24$.

Qualitative ratings. After the data was split by order (of the original participants, who provided the narratives, not of the participants who rated the narratives), the same pattern of effects emerged for the analyses of raters' gestalt impression of private or shared ground, although none of the analyses reached significance. For Order 1, there was a trend for a main effect of ground, $F(1,48) = 3.98$, $p = .052$, partial $\eta^2 = 0.08$, with stories produced during the private condition more likely be judged as based on stimuli watched with the listener. There was also a trend for a ground by group interaction, $F(1,48) = 3.82$, $p = .06$, partial $\eta^2 = 0.07$, with stories produced by the TD sample showing a greater difference in ratings between the private and shared conditions. There was no main effect of group, $F(1,48) = 2.40$, $p = .13$, partial $\eta^2 = 0.05$. For Order 2, the main effect of ground, $F(1,30) = 1.83$, $p = .19$, partial $\eta^2 = 06$, the main effect of group, $F(1,30) = 0.49$, $p = .49$, partial $\eta^2 = 0.02$, and the group by ground interaction, $F(1,30) = 1.40$, $p = .25$, partial $\eta^2 = 0.05$, were all found to be not significant.

Because of the way narratives were distributed to raters (based on constraint and ground conditions, and not based on original participant order), order was not distributed evenly among the other two conditions, or across participants. For this reason, Order 1 in particular was fairly underpowered in the analyses focusing on how well raters were able to follow the narratives. For Order 1, there were no significant main effects or interactions, including the main effect of group, $F(1,9) = 0.62$, $p = .45$, partial $\eta^2 = 0.07$, which was significant in the collapsed analysis. For Order 2, there was trend for participants in the TD sample to tell stories that were rated as easier to follow, $F(1,23) = 3.90$, $p = .06$, partial $\eta^2 = 0.15$. There was also a significant main effect of ground, $F(1,23) = 6.93$, $p = .02$, partial $\eta^2 = 0.23$, reflecting the fact that stories produced under conditions of private ground were rated as easier to follow than stories produced under conditions of shared ground, consistent with study hypotheses. For Order 2, a significant group by ground interaction was also observed, $F(1,23) = 10.02$, $p = .004$, partial $\eta^2 = 0.30$, with narratives produced by the ASD sample showing a bigger difference in ratings between private and shared ground than stories produced by the TD sample. No other main effects or interactions were significant, suggesting that gesture constraint had little effect on how easy the narratives were to follow.

For the ratings of story imageability, after splitting the narratives by original participant order, there were no significant main effects or interactions for either Order 1 or Order 2. The direction of the effect for both orders was, however, consistent with participants with ASD telling stories that were more difficult to visualize.

Non-social gestures. After splitting the sample by order, the results of the collapsed analysis showed the exact same pattern of findings for each order. For Order 1, there was no main effect of group, $F(1,15) = 0.19$, $p = .67$, partial $\eta^2 = 0.01$, but there was a significant task by group interaction, $F(1,15) = 10.49$, $p = .01$, partial $\eta^2 = 41$; for Order 2, there was also no main effect of group, $F(1,17) = 0.25$, $p = .62$, partial $\eta^2 = 0.02$, but there was a significant task by group interaction, $F(1,17) = 6.60$, $p = 0.02$, partial $\eta^2 =0.28$. The strong parallel between Order 1 and Order 2 reflects the robustness of this effect.