Comparing Comprehension of Indirect Answers by

Children with Autism Spectrum Disorder to Children with Typical Development

A Dissertation

SUBMITTED TO THE FACULTY OF

UNIVERSITY OF MINNESTOA

BY

Chunyang “Timothy” Huang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Lizbeth H. Finestack

July 2020

Acknowledgements

I am extremely grateful for everyone who supported me in this journey. Here is the short list:

My incredible research mentor, Dr. Liza Finestack, for her continuous guidance and encouragement. She was always willing and ready to assist in any way she could throughout my doctoral program. She was also patient and not afraid of letting me learn from my mistakes so that I could become a better researcher and teacher. I hope she did not regret taking a chance on me five years ago.

My awesome committee members, Dr. Benjamin Munson, Dr. Brian Reese, and Dr. Yang Zhang, for their thoughtful feedback on my dissertation project. I was lucky to find such a committee that is pure genius. I felt a little smarter every time I met with them. I hope I will also radiate wisdom to my students in the future.

Student research assistants, Ms. Julia Olenchek, Ms. Isabel Visuri, and Ms. Oralyn Weah, for spending many, many hours on transcribing, scoring, and coding participant responses; Mr. Robin Kelly, Ms. Megan Larsen, and Ms. Sydney Nygaard, for their bravery of recruiting participants at the Minnesota State Fair. I am eager to see the contributions they will make to our profession.

Friends and colleagues in SLHS, Shriya Basu, Minky Choi, Coral Dirks, Heekyung Han, Hannah Julien, Chieh Kao, Tess Koerner, Kirstin Kuchler, Xiaofan Lei, Kristi Oeding, Trever Perry, Joslyn Yu, and Luodi Yu for making Shevlin Hall warm and joyful. Special thanks to Katie Bangert for providing many life-saving statistics consultations. I would have not survived the first-year statistics courses without her.

Finally, my family, Diane, Kelcee, and Cloee, for inspiring my research and pushing me to where I am. I can’t wait to spend more time with them!

Abstract

Purpose

Children with autism spectrum disorder (ASD) often show difficulties understanding non-literal language, such as metaphors and irony, because of their impaired theory of mind ability. The current study investigated comprehension of indirect answers as an understudied form of non-literal language by this population. There were three primary aims. The first aim compared performance on comprehension of indirect answers by children with ASD to their peers with typical development (TD). The second aim examined theory of mind and other potential contributors to comprehension of indirect answers. The third aim explored erroneous interpretations of speaker intentions by the two groups.

Method

Nineteen 5- to 8-year-old children with ASD and forty-eight 5- to 10-year-old children with TD participated in the study. Participants with ASD completed various standardized, norm-referenced assessments of their non-verbal IQ, receptive and expressive language ability, severity of autism symptomology, theory of mind, world knowledge, and pragmatic skills. Participants with TD completed an abbreviated protocol with one assessment of non-verbal IQ and one of expressive language ability. All participants completed an experimental task designed to measure comprehension and explanation of indirect answers. Participant responses were scored and coded using a newly established coding scheme for comparisons.

Results

Children with ASD performed at a level similar to their peers with TD matched on age and expressive language ability. Hierarchical linear regression analysis indicated that receptive and expressive language ability and world knowledge were significant contributors to comprehension of indirect answers, whereas non-verbal IQ, severity of autism symptomology, theory of mind, and pragmatic skills might not be as important in accounting for task performance. Examination of inadequate explanations of indirect answers revealed that children with ASD had significantly more responses that were characterized as Irrelevant to Context than their peers with TD.

Conclusion

The novel finding that the two groups performed at a similar level suggests that comprehension of indirect answers is not a consistent weakness in language for children with ASD. Instead, it may be a strength for a subgroup with a particular cognitive and language profile. Clinically, it may be useful for speech-language pathologists to target receptive and expressive language skills as well as world knowledge to facilitate comprehension of indirect answers. Explanations that are irrelevant to context should be investigated further as a potential error type unique to this population's metapragmatic weaknesses in interpreting speaker intentions.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

When my 2-year-old daughter was making cupcakes with Play-Doh, I asked her to give me a hug. Being interrupted, she looked at me and said, "But my hands are dirty." Her response is an example of an indirect answer because it is not a "yes" or "no." Nonetheless, a clear rejection to the request can be inferred from her response. Indirect answers are considered a form of non-literal language, such as figures of speech, because of the discrepancy between a speaker's intended message and the explicit linguistic expression. Indirect answers are natural and common in everyday communication, accounting for 13%-38% of responses to yes-no questions (de Marneffe et al., 2009; Hockey et al. 1997; Stenstrom, 1984).

The intended meanings of indirect answers are known as *conversational implicatures* in the field of linguistic pragmatics. In his "Logic and Conversation," Grice (1975) described the phenomenon of meaning one thing while saying something else and explained how speakers manage to understand each other. Grice postulated a general principle that speakers are "cooperative" with the intention to achieve effective communication. Assuming my daughter intended to answer my request of hugging in the exchange above, her response that her hands are dirty must be relevant to accepting or rejecting the request. Because it is commonly known that people do not want to be touched by dirty hands, her utterance intentionally communicated (or "implicated" per Grice's terminology) a negative answer to the request.

Children with typical development (TD) begin to comprehend indirect answers consistently around the age of 6 years, and this development continues to improve steadily throughout the early school years (Bernicot, Laval, & Chaminaud, 2007; Bucciarelli, Colle, & Bara, 2003; de Villiers, de Villiers, Coles-White, & Carpenter, 2009; Loukusa, Leinonen, & Ryder, 2007; Verbuk & Shultz, 2010). Some evidence suggests that when supported with joint attention, children as young as 3 years of age may draw appropriate inferences from indirect comments, such as interpreting “I find the broccoli disgusting” as the speaker’s refusal to buy it (see Schulze, Grassmann, & Tomasello, 2013 and Tribushinina, 2012). However, it is important to note that the experimental stimuli used in these studies are solely preference-based utterances (e.g., “I find X nice” and “I find X boring”) that make the speaker’s communicative intent of wanting the item salient. Nevertheless, these results provide preliminary evidence that children with TD start to understand speaker intentions from indirect answers at an early age.

One defining feature of children with autism spectrum disorder (ASD) is their weaknesses in social communication and social interaction (American Psychiatric Association, 2015). Although there is considerable heterogenity in language abilities among this population, research has consistently reported that children with ASD show difficulties using language for social purposes (Dennis, Lazenby, & Lockyer, 2001; Ellawadi & Weismer, 2015; Happé, 1993, 1995; Tager-Flusberg, Paul, & Lord, 2005). That is, they demonstrate weaknesses in the comprehension and use of the literal aspects of language to show proper etiquette, maintain normal back-and-forth conversation, request communicative repairs when needed, and adjust vocabulary to suit various social contexts (Baron-Cohen, 1988; Boucher, 2003; Eales, 1993; Julien, 2018; Martin & McDonald, 2004; Ozonoff & Miller, 1996; Tager-Flusberg, 1996). Children with ASD also have difficulties understanding the non-literal aspects of language, such as figures of speech, because they tend to focus on the semantic meaning of an utterance rather than the intended message that goes beyond the explicit linguistic expression (Dennis et al., 2001; Emerich, Creaghead, Grether, Murray, & Grasha, 2003; Happé, 1993, 1995; Martin & McDonald, 2004; Mitchell, 1997).

Weakness in the comprehension of non-literal language among children with ASD is often attributed to their underlying deficits in theory of mind (ToM; Andrés-Roqueta & Katsos, 2020; Baron-Cohen, 1988; Capps, Losh, & Thurber, 2000; Tager-Flusberg, 1995; Tager-Flusberg & Sullivan, 1995). ToM is the ability to understand that someone else's mental states, such as beliefs, preferences, and intentions, may contrast with their own (Bauminger-Zviely, 2013; Premack & Woodruff, 1978). Thus, children with ASD are presumably less capable of detecting speaker intentions, which adversely affects their ability to interpret non-literal language. However, emerging evidence suggests that core language ability (e.g., grammar and vocabulary) may be a more reliable predictor. Particularly, children with ASD whose syntactic language ability is within normal range have been found to perform similar to their peers with TD on comprehension of non-literal language (Norbury, 2004; Pouscoulous, 2014; Whyte, Nelson, & Scherf 2014).

Among various forms of non-literal language, comprehension of metaphors and irony have often been cited as a challenge for children with ASD (e.g., Colich et al., 2012; Deliens, Papastamou, Ruytenbeek, Geelhand, & Kissine, 2018; Happé, 1993, 1995; Kalandadze, Norbury, Nærland, & Næss, 2018; Norbury, 2005; Rundblad & Annaz, 2010). In contrast, research focused on comprehension of indirect answers, another form of non-literal language, by children with ASD is scarce. Two existing studies by de Villiers et al. (2009) and Chin (2017) suggest that children with ASD perform significantly lower than their peers with TD on measures indexing the comprehension of indirect answers. Both studies also found that ToM and syntactic language ability were highly correlated with task performance. However, little is known about the independent contributions of the two variables to the comprehension of indirect answers due to the shared language component in ToM and syntactic language tasks (Norbury, 2004).

Following de Villiers et al. (2009) and Chin (2017), the current study had three primary aims. The first aim was to examine comprehension of indirect answers by children with ASD and compare their performance to children with TD. The purpose of this aim was to provide empirical evidence to the claim that children with ASD perform significantly lower than their peers with TD on a measure indexing the comprehension of indirect answers. The second aim was to evaluate cognitive and language variables, including non-verbal IQ, receptive and expressive language ability, severity of autism symptomology, ToM, world knowledge, and pragmatic language skills, that may be important contributors to the comprehension of indirect answers by children with ASD. Increased understanding of cognitive and language abilities necessary to foster comprehension of indirect answers is fundamental to the development of intervention approaches tailored to this communication weakness. Given the novelty of the current study, the third aim was to examine erroneous interpretations of speaker intentions by children with ASD and compare their error patterns to children with TD. Results from the error analyses would inform the reasoning and metapragmatic skills of this specific population.

Chapter 2 provides a review of literature relevant to the acquisition and development of non-literal language by children with TD and those with ASD. Chapter 3 describes the current study method, including the participants, assessments, experimental stimuli, overall procedure, and statistical analyses. Chapter 4 presents results from the analyses. Chapter 5 discusses implications of the study findings and study limitations. Chapter 6 presents an overall conclusion for the study.

Chapter 2

Literature Review

In this chapter, a description of Grice's (1975) theory of conversational implicatures is provided to ground indirect answers as a type of conversational implicature (i.e., relevance implicature) in his theoretical framework. Next, studies are reviewed that have examined acquisition and development of various forms of non-literal language, including (a) figurative language, (b) scalar implicature, and (c) relevance implicature, in both children with TD and those with ASD. This is followed by a review of potential contributors, including ToM, syntactic language ability, and world knowledge, to comprehension of indirect answers. The chapter ends with three research questions and predictions based on the studies reviewed.

Theory of Conversational Implicatures

Grice (1975) introduced the notion of conversational implicature in characterizing the phenomenon of meaning one thing while saying something else. Under the assumption that speakers are cooperative with the intention to achieve effective communication, Grice described the difference between "what is said," which is closely related to the conventional meaning of the words uttered, and "what is meant," which is the non-literal message that can only be inferred from the utterance. Grice postulated a general principle that speakers are "cooperative" with the intention to achieve effective communication. He further specified the following maxims to detail what it means to be cooperative:

Maxim of Quality, where speakers try to be truthful to avoid lying or making unsupported claims;

Maxim of Quantity, where speakers try to be informative to make their contributions neither more nor less than the conversation requires;

Maxim of Relation, where speakers try to be relevant to make their contribution pertinent to the discussion; and

Maxim of Manner, where speakers try to be clear, brief, and orderly to avoid ambiguity and obscurity.

These maxims provide the premises to describe how conversational implicatures arise and how the intended messages can be restored. Table 1 overviews different types of conversational implicatures based on Grice's maxims. The subsequent sections will further discuss implicatures based on the Maxims of Quality, Quantity, and Relation. Implicatures based on the Maxim of Manner is beyond the scope of the current study; however, more discussion can be found in Levinson (2000).

Table 1

Implicature Types based on Grice's (1975) Maxims

Maxim	**Type**	**Example (non-literal message)**
Quality	Figurative language	The river is a snake (moving in a windy way).
Quantity	Scalar implicature	He ate some pie (not all of it).
Relation	Relevance implicature	The dog looks happy (because he ate the ham).
Manner	M-implicature	He made the car stop (in an unusual way).

In addition to the cooperative principle, Grice broadly distinguished implicatures into two categories, generalized conversational implicatures and particularized conversational implicatures, by whether the implicated messages can be drawn across contexts. That is, generalized conversational implicatures are inferences that occur by default in any type of context. It is information that can be inferred in a prototypical way, as long as there is no specific information that denies or contradicts it (Levinson, 2000).

In the following utterances, the non-literal messages [in square brackets] are generalized conversational implicatures as they arise in most conversations regardless of context.

(1) You can have milk or juice [but not both].

(2) He ate some pie [but not all of it].

(3) The coffee is warm [but not hot].

The non-literal messages in these utterances are known as scalar implicatures because the speaker is assumed to observe the Maxim of Quantity to make his communication as informative as required (see Table 1). Thus, the use of an informationally weaker term implicates the negation of a stronger alternative on the same scale. For example, the weaker conjunction “or” in (1) is taken to mean the negation of both denoted by the stronger term “and.” Similarly, the weaker degree word “some” in (2) is taken to convey the negation of the stronger alternative “all.” In the case of (3), the speaker suggests the negation of the stronger adjective “hot” on the temperature scale. These inferences are triggered by the use of scalar words (see Horn, 1972 and Levinson, 1983 for further discussion) without referencing to any specific context.

In contrast, particularized conversational implicatures are closely linked to specific or particular contexts. That is, successful derivation of these inferences is linked to mutual beliefs or shared knowledge about very specific contextual information. Consider the following exchange, borrowed from de Villiers et al. (2009), between two speakers:

(4A): What happened to the ham?

(4B): The dog looks happy.

In the exchange above, 4A would likely derive the implicature "the dog ate the ham" from 4B's indirect answer. This is due to 4A's belief that 4B is cooperative with the intention to achieve effective communication by observing Grice's Maxim of Relation (i.e., responding to 4A's question). Thus, 4B's response would not be meaningful unless the happy-looking dog is relevant to what happened to the ham. Because it is commonly known that dogs are happy when they eat, especially when they eat human food, 4B's indirect answer is implicating that the dog ate the ham. The proposition "the dog looks happy" would typically not convey anything about the result of a dog eating the ham, so the derived implicature in this case is unique, or particularized, to the context as well as the utterance itself. This type of implicature is known as relevance implicature because the speaker is assumed to observe the Maxim of Relation to provide information pertinent to the question (see Table 1).

In the following exchange, Levinson (2000) made a clear distinction between generalized and particularized conversational implicatures that can be derived from the same utterance:

(5A): Did the children's summer camp go well?

(5B): Some of them got stomach flu.

The generalized conversational implicature that can be obtained from 5B is that "not all of the children got stomach flu." This implicature is lexically determined and would arise across contexts. However, the particularized conversational implicature from 5B's response is that "the summer camp did not go well." Such an implicature is contextually determined based on 5A's question and would not arise in another conversation. Because of this distinction, Levinson argued that generalized and particularized conversational implicatures require differential cognitive processing. More specifically, generalized conversational implicatures are the result of fast and frugal cognitive mechanisms because the implicated meanings are embedded in certain linguistic forms (e.g., scalar terms “some” = not all; “or” = not both). Particularized conversational implicatures, however, may involve more cognitive demand in terms of evaluating contextual information, shared beliefs, and world knowledge concurrently, which results in slower and more effortful calculation of proper interpretation. The next section reviews studies focused on the acquisition and development of figurative language, which is a type of non-literal language based on the Maxim of Quality, in children with TD and children with ASD.

Comprehension of Figurative Language

Figurative language is the most recognized form of non-literal language in literature as well as daily conversation (Colston & Kuiper, 2002; Kerbel & Grunwell, 1997). Common examples include metaphor (e.g., using “The weather is a roller coaster” to describe rapid changing temperatures), irony (e.g., “The weather is lovely” to describe a stormy day), overstatement (also known as *hyperbole*, e.g., “The weather is the worst in history” to describe drizzling rain), understatement (also known as *meiosis* or *litotes*, e.g., “The weather is not good” to describe a heavy downpour), and idiom (e.g., “It’s raining cats and dogs” to describe heavy rain). Grice (1975) viewed figurative language as a type of particularized conversational implicature because of its characteristic of being context dependent. When taken literally, figurative language violates the Maxim of Quality that charges speakers to say only what they believe to be true. To restore the intended meaning, the listener must assume that the speaker indeed complies with the cooperative principle and must recognize the discrepancy between the literal and figurative meanings of the expression in order to arrive at the desired interpretation (Gibbs & Colston, 2012; Gluksberg, 2001).

Children with TD. Many studies reviewed here focus on metaphor and irony because of their frequent occurrence in communication. In general, comprehension of figurative language begins at the age of 5 to 6 years and improves steadily throughout childhood and adolescence in children with TD (Dews et al., 1996; Hancock, Dunham, & Purdy, 2000; Nippold, 1985; Pexman & Glenwright, 2007; Rundblad & Annaz, 2010; Semrud-Clikemen & Glass, 2010; Vosniadou & Ortony, 1983). To examine comprehension of metaphors by younger children, Siltanen (1986) included 3- to 5-year-olds as a subgroup in a large-scale study. The researcher read stories containing metaphoric sentences and asked participants to explain what it means to say, for example, "Butterflies are rainbows." Responses were scored using a coding scheme that ranked answers from "No comprehension" to "Metaphor comprehension based on perceptual and conceptual grounds." The researcher compared mean scores by the 3-, 4-, and 5-year-olds and found that only the 5-year-old children were able to comprehend easy metaphors that were perceptually grounded (e.g., "Raindrops are the sky's tears" and "The Scioto River is a snake"). However, all children in this age group failed to interpret more abstract metaphors that were conceptually grounded (e.g., "Suspicion is quicksand" and "Silence is cancer"). Thus, the researcher concluded that children with TD begin to comprehend basic metaphors that are perceptually grounded at the age of 5 years.

Winner, Rosenstiel, & Gardner (1976) examined comprehension of metaphors by older children with TD between the ages of 6 and 14 years. The researchers asked participants to explain metaphoric sentences, such as “After many years of working at the jail, the prison guard had become a hard rock that could not be moved.” Results indicated that comprehension increased gradually with age, and a higher level of metaphoric understanding emerged in early adolescence. Responses by the youngest participants showed little to no signs of metaphoric understanding, for example, explaining that the prison had hard rock walls or the guard used to stay on a rock. Eight-year-old children demonstrated initial understanding of metaphors by commenting that the guard had muscles as hard as rocks, linking physical similarities between the guard and rock. By 10 years of age, children began to provide genuine metaphoric responses, interpreting the guard as hard rock because he did not care about anybody. Such progressive comprehension continues into early adolescence. In another experimental sentence where participants had to explain “The taste was a sharp knife,” a 10-year-old interpreted it to mean “It was spicy,” while a 14-year-old explained with more enriched descriptions that “The taste was a shocking flavor, hitting all of my senses at once” (p. 296).

Glenwright and Pexman (2010) investigated children’s ability to distinguish whether the same utterance was sarcasm (remarks directed at targets; e.g., “This is a great bike” as a sarcastic criticism to someone’s possession) or verbal irony (remarks not directed at specific targets; e.g., “This is a great bike” as an ironic criticism to the condition of a bike). The researchers presented participants with puppet shows where a puppet provided a personal or non-personal criticism. After viewing each show, participants answered whether the puppet made a positive or negative evaluation of the target and then rated the puppet’s attitude in making the remark on a 6-point scale with “Nice” and “Mean” on the two ends. The researchers found that 5- to 6-year-olds began to understand that the sarcastic and ironic speakers meant the opposite of what they said. However, their ability to distinguish sarcastic and ironic speakers' intentions was not fully developed until the age of 9 to 10 years, as evident by the older children's ratings of the sarcastic criticisms being more "mean" than the ironic criticisms. Thus, these studies provide an important reference for comprehension of figurative language by children with TD. That is, 5-year-olds begin to understand basic metaphors that are perceptually grounded, and this ability continues to grow as they grasp the meanings of more abstract metaphors and other forms of figurative language such as sarcasm and irony.

Children with ASD. MacKay and Shaw (2004) examined comprehension of six forms of figurative language by 8- to 11-year-old children with ASD and their age-matched peers with TD. These forms included hyperbole (e.g., "I've been there millions of times."), indirect requests (e.g., "That cake looks delicious."), irony (e.g., "Great singing, Jim."), metonymy that used a linked term to represent an object or concept (e.g., "Green corner! Work more quietly, please."), rhetorical questions that did not require an answer (e.g., "How could you?"), and understatement (e.g., "It's a bit sore."). After listening to short stories, participants answered questions about the target sentences containing figurative language. Responses were scored for correct interpretation of the literal meaning and speaker intention. Results revealed that the TD group outperformed the ASD group on every form of figurative language, with hyperbole and metonymy showing statistically significant differences between groups.

Rundblad and Annaz (2010) compared comprehension of metaphors by 5- to 11-year-old children with ASD to children with TD. Using picture stories containing a metaphor at the end, the researchers asked participants to complete each story by answering what the metaphor means (e.g., “a flood outside the museum” means lots of people come to the museum). The researchers found significantly lower performance on comprehension of metaphors by children with ASD compared to their age-matched controls with TD. Additionally, while performance increased with age for the TD group, performance was close to floor level across ages for the ASD group. Even when matched on mental age [using the Pattern Construction subtest of the British Ability Scales-Second Edition (Elliot et al., 1997)], the ASD group showed significantly lower level of performance than the TD group. The findings further support previous studies by Dennis et al. (2001) and MacKay and Shaw (2004) that suggest group differences based on comprehension of metaphors.

Despite the consistent findings for metaphors, examinations of other forms of figurative language have indicated greater variability in performance by individuals with ASD. For example, studies examining comprehension of irony in children with ASD have yielded inconsistent findings. While some studies document difficulties in comprehension of irony by children and adolescents with ASD (Adachi et al., 2004; Happé, 1993; Kaland et al., 2002; MacKay & Shaw, 2004; Wang, Lee, Sigman, & Dapretto, 2006) and even in adults with ASD (Deliens et al., 2018; Martin & McDonald, 2004; Saban-Bezalel & Mashal, 2015), there is evidence showing negligible differences between children with ASD and their peers with TD (Pexman et al., 2011; Saban-Bezalel, Dolfin, Laor, & Mashal, 2019). This inconsistency highlights the possibility that even though the deficit of comprehension of non-literal language is a hallmark of individuals with ASD, some skills may be preserved (Deliens et al., 2018). Another form of non-literal language for which study results have been inconsistent is scalar implicature, arising from the Maxim of Quantity that charges speakers to make their communication as informative as required. The next section reviews the acquisition and development of scalar implicatures in children with TD and children with ASD.

Comprehension of Scalar Implicatures

Children with TD. Earlier experimental studies by Noveck (2001) and Papafragou and Musolino (2003) found that children younger than 7 years old have significantly more difficulty comprehending scalar implicatures than adults. Even when provided with training prior to the experimental task, 5- to 6-year-old children in the Papagragou and Musolino study failed to achieve adult levels of performance, with an increase from 12% to 52% accuracy of interpreting "some" to mean "not all."

Guasti et al. (2005) replicated Noveck's (2001) study to compare comprehension of scalar implicatures by 7-year-olds to adults. The researchers auditorily presented a series of statements containing the scalar terms "all" or "some" and asked the participants to judge whether each statement was true or false. Target sentences such as "Some giraffes have long necks" and "Some airplanes have wings" should be judged false if the pragmatic inferences "Not all giraffes have long necks" and "Not all airplanes have wings" were derived. The results showed a significant difference between groups in judging the target sentences with 87% accuracy for adults and 50% for 7-year-olds.

Pouscoulous, Noveck, Politzer, and Bastide (2007) found consistent results when comparing comprehension of scalar implicatures between slightly older children of 9 and 10 years of age and adults. The researchers presented a series of pictures of four boxes with animals in or out of the boxes. Then, the researchers produced an utterance, such as "Some turtles are in the boxes," and asked the participants to determine whether the statement was a true or false description of the picture. The researchers found that 9- and 10-year-old children tend to obtain the logical reading of “some, and possibly all, turtles are in the boxes” and thus judge the picture of all turtles in the boxes to be true. In contrast, adults are able to derive the pragmatic meaning of “some, but not all, turtles are in the boxes” and thus judge the same picture as false. Compared to previous studies, the results further suggest that children with TD may not be able to make scalar implicatures in the same way adults do even at the age of 10 years.

Children with ASD. There are few studies that compare comprehension of scalar implicatures by children with ASD to children with TD. Pastor-Cerezuela, Tordera Yllescas, González-Sala, Montagut-Asunción, & Fernández-Andrés (2018) utilized stimuli composed of a statement paired with three possible interpretations (e.g., “Some guests came to Maria’s party” paired with the following choices: “All the people Maria invited came,” “Not all the guests Maria expected came,” and “Exactly three guests came”) for participants to choose the best-corresponded interpretation. The researchers compared performance by children 6 to 13 years of age across three groups: children with ASD, children with TD matched on age, and children with TD matched on language ability. The results indicated that the age-matched TD group performed better than the language-matched TD group (but without a significant difference), and the two groups performed significantly better than the ASD group.

Although the Pastor-Cerezuela et al. (2018) study indicates that children and young adolescents with ASD demonstrate weaknesses in comprehending scalar implicatures, studies that have included adolescents with ASD have found the opposite results. Chevallier, Wilson, Happé, & Noveck (2010), for example, found that both adolescents with ASD and those with TD between the ages of 10 and 16 years derived the exclusive interpretation of the scalar disjunctive "or" (i.e., not both) at similar rates and comparable speeds. More specifically, participants in the study looked at a picture showing two objects, listened to an auditory stimulus describing the picture, and judged whether the description was true or false (e.g., a picture of a sun and a train with the auditory prompt "There is a sun or a train."). The results indicated no difference between the rates of deriving scalar implicatures (i.e., judging the above example false) by the two groups. Therefore, the researchers concluded that adolescents with ASD are as capable of deriving scalar inferences as their peers with TD.

Similar results were found in an adult population with ASD. Pijnacker, Hagoort, Buitelaar, Teunisse, and Geurts (2009) examined comprehension of the scalar quantifier "some" and the scalar disjunctive "or" across three conditions by 19- to 40-year-old adults with high-functioning autism, Asperger syndrome, or TD. The participants read a statement, such as "some sparrows are birds", on a computer screen and judged whether the sentence was true or false. The results indicated that participants with ASD (high-functioning autism and Asperger syndrome combined) and those with TD had similar rates of pragmatic interpretations for the targeted sentences. Because there were no significant differences between group performance, the researchers concluded that participants with ASD were as good as those with TD at deriving scalar inferences.

In summary, the Pastor-Cerezuela et al. (2018) study indicates that children and young adolescents with ASD (ages 6 to 13 years) performed significantly below their TD peers on comprehension of scalar implicatures as the most common type of generalized conversational implicature. However, the studies by Chevallier et al. (2010) with adolescents (ages 10 to 16 years) and Pijnacker et al. (2009) with adults (ages 19 to 40 years) suggest that there are no noteworthy differences in the comprehension of scalar implicatures between individuals with and without ASD. Thus, it is possible that the acquisition of scalar implicatures is slower in children with ASD than in children with TD during childhood, but these differences are diminished as they enter adolescence and adulthood.

It is important to note that none of the reviewed studies reported the severity of autism symptomology of their participants. Pastor-Cerezuela et al. (2018) recruited participants with a clinical diagnosis of ASD and used the Gilliam Autism Rating Scale-Second Edition (GARS-2; Gilliam, 2006) to determine inclusionary eligibility. The GARS-2 is a parent survey designed to screen for ASD in individuals between the ages of 3 and 22 years. A cut-off score of 85 or higher indicates a high likelihood of ASD (*M* = 100, *SD* = 15). The researchers reported that participants in the ASD group received a GARS-2 score ranging from 85-135 without details of the distribution. Chevallier et al. (2010) confirmed individuals' diagnosis of ASD by existing school files and medical diagnoses made by a clinical psychologist or psychiatrist. Participants in the Pijnacker (2009) study had an established clinical evaluation of ASD, and some of them completed the Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003) for the study. However, the researcher did not specify any obtained scores related to the measure. The next section turns to the acquisition and development of relevance implicatures, which is another type of non-literal language based on the Maxim of Relation, in children with TD and children with ASD.

Comprehension of Relevance Implicatures

Children with TD. Research on the acquisition and typical development of relevance implicatures reports that children begin to derive appropriate relevance inferences consistently around the age of 6 years (Bernicot et al., 2007; Bucciarelli et al., 2003; de Villiers et al., 2009; Loukusa et al., 2007; Verbuk & Shultz, 2010). There are four recent studies that depict this development from 2 to 10 years of age. First, Bucciarelli et al. (2003) used video-taped stories to test 2- to 7-year-old children on their comprehension of relevance implicatures (named "complex indirects" in the study). After viewing a story, the children had to choose a possible ending from four pictures. For example, in one scenario, two siblings stop in front of a doll shop. The brother asks, "Would you get me that game?" and the sister answers, "We don't have any money." In this example, selecting the picture of the siblings walking away from the store empty handed would be scored correct. Results indicated that accuracy increased with age from 38% for 2.6- to 3-year-olds, 42% for 3.6- to 4-year-olds, 43% for 4.6- to 5.6-year-olds, to 68% for 6- to 7-year-olds. That is, only children older than 6 years showed relatively reliable performance on deriving relevance inferences.

Second, Loukusa et al. (2007) tested children with TD between the ages of 3 and 9 years on their ability to derive relevance implicatures. The researchers verbally presented a scenario such as: "A man is mowing, and a woman says to him, 'There are flowers growing in the middle of the lawn so remember to be careful'" followed by a question prompt, "Why does the woman say this?" Responses were judged as correct/appropriate (e.g., "So that the flowers wouldn't be cut") or incorrect/inappropriate (e.g., "She doesn't want to do it") based on whether the implicated meaning was derived.

The researchers found that the mean score of correct/appropriate answers increased with age. Moreover, there was a significant difference in mean scores between 3- and 4-year-olds as well as between 5- and 6-year-olds. Examination of percent correct by age groups revealed that 3-year-olds drew relevance inferences from 21% of the questions, and this percentage increased to 77% by the age of 6 years. Eight- and 9-year-old children performed near ceiling.

Third, Bernicot et al. (2007) used a computer-based story completion task to test comprehension of relevance implicatures, among other non-literal language forms, by children between the ages of 6 and 10 years. In one story, Donald and Daisy are in the yard. Donald asks Daisy, “Should I mow the lawn?” and Daisy replies, “The nephews are taking a nap.” The children had to pick a picture from two possible endings: one indicates the inference is understood (i.e., Donald waters the flowers) and the other indicates the inference is not understood (i.e., Donald mows the lawn). The researchers found that 75% of the 6-year-old children were able to correctly select the implicated ending in three or four of four tested items. Percentages for older children were near ceiling, with 95% for the 8-year-olds and 100% for the 10-year-olds.

Finally, de Villiers et al. (2009) investigated comprehension of relevance implicatures by children aged 3 to 10 years. The researchers presented pictures with short question-answer pairs (e.g., Adult: “What happened to the ham?” Child: “The dog looks happy.”) to children and asked them to explain what the speaker meant (e.g., “What did the boy mean?”, “Why did he say that?”). Their answers were coded as adequate or inadequate based on whether the implicated message was derived (e.g., “Because the dog ate the ham”; “Because the dog looked happy,” respectively). Results showed that performance increased with age, with 4-year-olds providing adequate answers about 25% of the time and 9-year-olds providing adequate answers 90% of the time. Six-year-olds were able to provide adequate answers about half of the time.

These four studies report a range of success rates for comprehending relevance implicatures by children with TD aged 2 to 10 years and indicate that this pragmatic language skill grows steadily with age. Particularly, children at the age of 6 years appear to be capable of drawing relevance inferences more consistently, ranging from 50% (de Villiers et al., 2009) to 77% (Loukusa et al., 2007). However, Schulze et al. (2013) argued that the finding that only children older than 5 or 6 years show adequate comprehension of relevance implicatures may be attributed to the methods used in these studies. Specifically, children in these studies had to judge the appropriateness of a target utterance to reason its meaning based on a given context, and previous research has shown that such advanced metalinguistic skills are not developed until school years (Ackerman, 1982; Anderson-Wood & Smith, 1997; Bernicot et al., 2007).

To determine whether relevance implicatures can be acquired at an earlier age, Tribushinina (2012) further investigated 2- to 5-year-old children's ability to draw appropriate pragmatic inferences based on the speaker's likes and dislikes in a shopping-themed experiment (e.g., interpreting "I find X delicious" as the speaker's intention to buy X). The researcher found that children as young as 3 years of age are able to comprehend such relevance implicatures at 80%-100% accuracy when they are supported by joint attention, defined as a shared focus with the experimenter on the same object. They also found that at age 5, children are able to make relevance inferences similar to adults without joint attention.

Consistent results were found in Schulze et al.'s (2013) study, in which the researchers investigated young children's comprehension of indirectly stated utterances (e.g., "I find X good" implies a request for X). In Experiment 1, one researcher told the child to give a toy animal to the other researcher. Then, that researcher replied with an indirect comment to show her acceptance or rejection of the offer. For example, the response "I find lions good" would indicate her desire to play with the animal. Experiment 2 replicated the same design with only one difference: the first researcher asked the other researcher whether the child should give her the toy (as opposed to asking the child to give the toy to the other researcher directly). The results of both experiments demonstrated that 3-year-old children could reliably draw relevance inferences from indirect comments regardless of the recipient.

It is important to note that both the Tribushinina (2012) and Schulze et al. (2013) studies utilized preference-based utterances (e.g., "I find the broccoli disgusting" and "I find lions good") that makes the speaker's communicative intent salient. As a result, such responses functioned similarly as speech acts in the testing contexts (i.e., shopping and offering toys) and the implicated meanings became conventionalized after a few trials. Thus, it is possible that the children treated these indirect requests as speech acts and simply complied with them without further reasoning. Additionally, the stimuli used in these studies assess a very narrow set of world knowledge (i.e., when the speaker comments on an object with a positive adjective, such as "good," he must want it, and vice versa). It remains unknown whether 3-year-olds would perform as well when the response is no longer about preferences (e.g., "The lion looks smiley" or "Lions have sharp teeth").

Children with ASD. To date, research focused on comprehension of relevance implicatures by children with ASD is scarce. One study conducted by de Villiers et al. (2009) compared the comprehension of indirect answers of ten 5- to 12-year-old children with ASD and 20 age-matched children with TD. To help characterize autism symptomology, the parents completed the Social Communication Questionnaire (Berument, Rutter, Lord, Pickles, & Bailey, 1999) for which scores greater than 15 indicate characteristics associated with pervasive developmental disorder and scores greater than 22 indicate characteristics associated with ASD. Of the 10 children with ASD in the study, two were < 15, three were between 15-22, and five were >22 (range = 13-31, *M* = 22). In the experiment, the researchers presented pictures with short question-answer pairs (e.g., Adult: "What happened to the ham?" Child: "The dog looks happy.") to the children and asked them to explain what the speaker meant (e.g., "What did the boy mean?" or "Why did he say that?"). Participant responses were judged as adequate or inadequate based on whether the intended message was derived. For example, the response "Because the dog ate the ham" would be marked adequate, and the response "Because the dog looks happy" would be marked inadequate. Results indicated a significant difference in performance between groups. The ASD group had a mean age of 8 years but performed at a level similar to the 5-year-old children in the TD group.

The other study was by Chin (2017), who investigated comprehension of indirect answers by children with TD and children with ASD aged 7 to 10 years (12 participants in each group). The researcher confirmed ASD diagnosis by administering the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000). All participants met the cut-off score of 7 (*M* = 15.45, *SD* = 4.66). The researcher adapted Bernicot et al.'s (2007) experimental paradigm by showing children a picture along with a verbal prompt (e.g., “Here’s Sandra and her mom relaxing in the living room.”). Then, participants listened to a short conversation in which Sandra asked for permission to engage in an activity and her mom responded with an indirect answer (e.g., Sandra: “I have a music competition tomorrow. Can I practice the drums?” Mom: “The neighbors have already gone to bed.”). After the presentation, participants were asked “What will Sandra do next?” and provided four possible endings. Selecting the picture of Sandra reading sheet music would be scored correct, whereas selecting the picture of Sandra playing drums and the other two distractors would be scored incorrect. Consistent with the results from de Villiers et al. (2009), Chin found that the ASD group performed significantly lower than their age- and non-verbal IQ-matched peers with TD. Thus, there appears to be a significant difference in comprehension of indirect answers between the two groups. The next section provides an overview of factors that support comprehension of indirect answers.

Potential Contributors

Weakness in the comprehension of non-literal language among children with ASD is often attributed to their underlying deficits in theory of mind (ToM; Baron-Cohen, 1988; Capps et al., 2000; Tager-Flusberg, 1995; Tager-Flusberg & Sullivan, 1995). ToM is the ability to understand that someone else’s mental states, such as beliefs, preferences, and intentions, may contrast with their own (Bauminger-Zviely, 2013; Premack & Woodruff, 1978). Thus, children with ASD are presumably less capable of detecting speaker intentions, which adversely affects their ability to interpret non-literal language. Happé (1993) investigated the connection between ToM reasoning and performance on metaphor and irony comprehension. The researcher found that only the individuals with ASD who passed the first-order ToM tasks (assessing an individual's ability to understand what another person thinks or feels, e.g., He thinks that X) were successful on metaphor tasks, and only those who passed the second-order ToM tasks (assessing an individual's ability to understand what one person thinks another person thinks or feels, e.g. He thinks that she thinks that X) performed well on irony tasks.

The studies by de Villiers et al. (2009) and Chin (2017) suggest a strong relationship between ToM and comprehension of indirect answers. To assess ToM, de Villiers et al. (2009) examined first- and second-order false belief by asking participants to explain the cognitive states and actions of the characters in picture-supported narratives. The materials were similar to those used by Sullivan, Zaitchik, and Tager-Flusberg (1994). Pearson correlation analyses revealed a significant positive relationship between age and comprehension of indirect answers and between ToM and comprehension of indirect answers. However, the researchers noted that it is not possible to attribute the ability to comprehend indirect answers exclusively to age or ToM because of the small sample size (n = 10) and because age and ToM were also significantly correlated. In Chin's (2017) study, the researcher used a similar narrative task to assess ToM ability and found that second-order false belief reasoning was a significant independent predictor of comprehension of indirect answers in the children with TD. Given the small sample size of the ASD group (n = 12), the researcher was underpowered to generalize the finding to this population.

Emerging evidence suggests that core language skills may be another reliable predictor of non-literal language comprehension. In Norbury's (2004) study, a subgroup of children with ASD whose vocabulary and syntax were within the average range performed significantly better than those with language impairments only and those with ASD and comorbid language impairments when providing figurative interpretations of idioms. Thus, the researcher concluded that syntactic language ability was a strong predictor, better than autism symptomology, of comprehension of idioms. Whyte et al. (2014) reported consistent results. In this study, children with ASD performed significantly lower on a measure of the comprehension of idioms than age-matched controls with higher syntax age equivalence scores, but their performance was comparable to syntax-matched controls. In another study by Norbury (2005), the researcher examined the importance of semantic knowledge and ToM in comprehension of metaphors by children with ASD. The findings indicated that semantic knowledge was a stronger predictor of task performance than ToM ability.

Finally, world knowledge is a factor that pragmatic theorists argue as a significant prerequisite for comprehension of indirect answers (Sperber & Wilson, 1986; Zufferey, Moeschler, & Reboul, 2019). World knowledge can be broadly defined as an individual's accumulative fund of general information and learned material acquired at home or school (Roid, 2003). It determines a child's ability to understand some indirect answers earlier than others as shown in the two indirect answers discussed previously: "I find lions good" (implicating that "I want the toy lion") and "The dog looks happy" (implicating that "The dog ate the ham"). There is a clear difference in complexity between the intended messages. In the first indirect answer, the connection between the speaker's comment on lions being "good" and his intention of "wanting" the toy lion is strong and thus, easier to reason. This is evident by Repacholi & Gopnik's (1997) study that by 3 years of age, children are able to reason about other people's desires. However, in the second indirect answer, it involves more knowledge of the world about dogs, such as their preferences for human food and the typical result of dogs eating food, to reason the indirect answer. Despite the apparent connection between world knowledge and comprehension of indirect answers, this variable has not been formally examined as a potential contributor in the literature.

This review of the literature indicates mix results of the comprehension abilities of figurative language and scalar implicatures among children with ASD. To date, two studies that have compared the comprehension of indirect answers (also known as relevance implicatures) by children with ASD and children with TD suggest a discrepancy in performance levels; however, more empirical evidence is needed to confirm this finding. Moreover, previous studies have found ToM and language ability to be significant contributors to comprehension of indirect answers, but the independent contribution of each factor remains unknown due to the shared language component in the tasks. Thus, there is a need to better understand the comprehension abilities of indirect answers among children with ASD as well as factors that have a significant influence on comprehension of indirect answers.

Current Study

The current study had three primary aims. The first aim was to examine comprehension of indirect answers by children with ASD and compare their performance to that of children with TD. The purpose of this aim was to provide empirical evidence to the limited literature that suggests comprehension of indirect answers by children with ASD is significantly lower than their peers with TD. One noticeable methodological difference between de Villiers et al. (2009) and Chin (2017) arises from how comprehension was measured in each study. While de Villiers et al. used open-ended why-questions to probe children's ability to explain indirect answers, Chin's study adapted a forced-choice format that assessed participants' judgement of indirect answers. The current study consolidated the two methodological approaches by examining comprehension with both forced-choice and open-ended questions. Observed differences would inform future studies regarding potential impact of the measures on task performance. Moreover, current literature has only investigated indirect answers that are contextually clear (e.g., Q: "Are you going to the party?" A: "I have to work late.") but not indirect answers that are ambiguous (e.g., Q: "Are you going to the party?" A: "Bob will be there."). The study included a novel category of indirect answers that are contextually ambiguous. Performance on these question-answer pairs by children with ASD and children TD will provide insight into their abilities to reason speaker intentions that are more complex and unclear and reveal the presence of group difference, if any.

The second aim of the current study was to explore variables that may be important contributors to comprehension of indirect answers by children with ASD. Among numerous potential contributors, ToM has been identified as having a strong link to comprehension of non-literal language, with syntactic language ability moderating this relationship (Norbury, 2004). That is, ToM tasks, especially when presented in narrative format, require well-developed language skills to understand sentences with complex syntactic structures (see Milligan, Astington, & Dack, 2007 for a complete meta-analysis). As a result, it is difficult to tease out the independent contributions of ToM and syntax ability in comprehension of non-literal language. The current study used a naturalistic ToM measure, the Theory of Mind Inventory- Second Edition (ToMI-2;

Hutchins, Prelock, & Bonazinga, 2010), which utilizes parent report and eliminates linguistic demands from the child to assess the unique contribution of ToM. World knowledge is another variable that pragmatic theorists argue to be a significant prerequisite of understanding indirect answers (Sperber & Wilson, 1986; Zufferey et al., 2019). However, this variable has not been formally examined in previous literature. The current study assessed this relationship using the non-verbal Knowledge subtest from the Stanford-Binet Intelligence Scales- Fifth Edition (SB-5; Roid, 2003). Results will serve as preliminary evidence to evaluate the role of world knowledge in comprehension of indirect answers.

The third aim of the current study was to analyze the error patterns of children's explanations of indirect answers. Given the novelty of investigating comprehension of indirect answers by children with ASD, the study attempted to provide insight to how this specific population reasons speaker intentions and to identify potential challenges they may encounter in this process. Increased understanding of the metapragmatic skills of children with ASD may benefit investigations on intervention approaches tailored to this communication weakness. Children with ASD who are able to understand and use indirect answers will have increased opportunities to engage in social communication with their peers, which will ultimately lead to improved social relationships and quality of life (Cummings, 2017). To address these aims, the current study included three research questions with predictions:

1. Is there a significant difference in comprehension of indirect answers between 5- to 8-year-old children with ASD and children with TD when matched on age and language ability?

Prediction: It was predicted that children with ASD would perform significantly lower than their peers with TD based on previous evidence by Chin (2017) and de Villiers et al. (2009). Language ability may moderate task performance as Norbury (2004) found that when vocabulary and syntactic language ability were controlled, children with ASD performed at a level similar to their peers with TD on comprehension of idioms. However, in Chin's study, there were no differences in vocabulary or syntactic language ability between the children with ASD and those with TD. De Villiers et al. did not compare language ability between groups but indicated that the children with ASD had age equivalence scores on language assessments close to their mean chronological age.

2. What are the relationships between comprehension of indirect answers and ToM, syntactic language ability, and world knowledge?

   Prediction: It was predicted that the three variables would be highly correlated with comprehension of indirect answers and would each account for a significant variance in task performance based on the studies reviewed.

3. When reasoning speaker intention from indirect answers, is there a significant difference in the error patterns between children with ASD and those with TD?

   Prediction: Inadequate explanations were coded as: I Don't Know/No Response, Repetition of Response, Irrelevant to Context, Made-up Interpretation that is appropriate to context but different from speaker intention, or Insufficient Explanation that fails to capture speaker intention. It was predicted that children with ASD would have significantly more responses characterized as I Don’t Knows and Repetitions than their peers with TD based on their presumably lower performance on comprehension of indirect answers.

Chapter 3

Method

The study implemented a within-subjects design to examine the relationships between participant characteristics (e.g., age, language ability, ToM, and world knowledge) and task performance (i.e., comprehension and explanation of indirect answers) in children with ASD. Between-subjects analyses were conducted to compare task performance by children with ASD and their age- and language-matched peers with TD.

Data collection for the TD group occurred in summer 2019 during Minnesota's annual State Fair. Participants were recruited through the University of Minnesota's research facility where interested fairgoers could sign up for a variety of research studies. Data collection for the ASD group occurred between December 2019 and March 2020 at the University of Minnesota. Participants were recruited through the University's Focus in Neurodevelopment (FiND) registry and flyers posted at private speech-language-hearing clinics and autism centers in the metropolitan Minneapolis-St. Paul area.

Participants in the ASD group completed a full protocol, and participants in the TD group completed an abbreviated version of the protocol. The study's protocol for the ASD group was completed in conjunction with Bangert's (in preparation) study that investigated how heart rate variability interacts with various cognitive and social activities. The study was approved by the University of Minnesota's Institutional Review Board for human subjects. Parents signed consent forms prior to participating in any study sessions.

Participants

The ASD group included a total of 19 children, 13 boys and 6 girls, between the ages of 5 years; 0 months and 8 years; 11 months (*M* = 7;6, *SD* = 12.47 months). Participants met the following inclusionary criteria: (a) have a documented medical diagnosis or educational qualification status of ASD, (b) be monolingual English speaker, (c) use at least 3-word utterances to communicate, and (d) have normal or corrected-to-normal vision and hearing per parent report. Upon enrollment, a licensed speech-language pathologist who was research reliable on the Autism Diagnostic Observation Schedule- Second Edition (ADOS-2; Lord et al., 2012) administered the test to each participant to confirm diagnosis and determine severity of autism symptomology, such as repetitive behaviors, rituals and routines, and presence of hyper- or hypo-sensitivity. All participants met the cut-off for ASD diagnosis (*M* = 13.68, *SD* = 6.55). Scores on each scale were interpreted relative to a clinical sample of individuals diagnosed with ASD to categorize a child's severity of autism symptomology as low, moderate, or high. Three children in the study scored in the low symptoms severity category, 10 children scored in the moderate symptoms category, and six children scored in the high symptoms category.

The TD group included 48 children, 23 boys and 25 girls, between the ages of 5 years; 0 months and 10 years; 11 months (*M* = 8;2, *SD* = 19.77 months). Of the 48 children, seven were 5-year-olds, eight were 6-year-olds, nine were 7-year-olds, seven were 8-year-olds, nine were 9-year-olds, and eight were 10-year-olds. Participants met the following inclusionary criteria: (a) be monolingual English speaker, (b) use at least 3-word utterances to communicate, (c) have normal or corrected-to-normal vision and hearing per parent report, and (d) receive a T-score lower than 60 on the Social

Responsiveness Scale- Second Edition (SRS-2; Constantino & Gruber, 2012). The SRS-2 identifies social impairments associated with ASD, and scores lower than 60 are considered within normal limits and not associated with clinical presentations of ASD. Additionally, participants could not have a history of language impairments or developmental delay per parent report. Five participants who reported receiving speech-language services, of which four were due to speech sound errors and one was due to stuttering, were included in the study. Participants completed the Matrices subtest of the Kaufman Brief Intelligence Test- Second Edition (KBIT-2; Kaufman & Kaufman, 2004) and the Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals- Fourth Edition (CELF-4; Semel, Wiig, & Secord, 2003) as indices of non-verbal cognitive ability and language ability, respectively. Table 2 provides a detailed summary of demographic and linguistic characteristics of the two groups. Distribution of age, nonverbal IQ, sex, and race did not differ significantly between groups (all *p*s > .11). Compared to the TD group, the ASD group had a significantly higher mean SRS-2 score [$U(65) = 8.5$, $p < .001$] and a significantly lower mean CELF-4 score [$U(65) = 285.5$, $p = .01$].

Table 2

Characteristics of Participant Groups

Variable		ASD[a] *n* = 19	TD[b] *n* = 48	*P*[c] (*d*)
Age (year;month)				
	Mean	7;6	8;2	.2
	SD	12.47	19.77	(.48)
	Min-Max	5;4-8;11	5;1-10;9	
Sex				
	Male:Female	13:6	23:25	.12
Race				
	White	18	38	.11
	Hispanic	1	4	
	African American	0	1	
	Asian Pacific Islander	0	3	
	American Indian/Alaskan Native	0	2	
KBIT-2[d]				
	Mean	105	104	.75
	SD	20.33	15.56	(.05)
	Min-Max	69-137	68-134	
CELF-4[e]				
	Mean	8.36	11.25	.01**
	SD	4.34	3.13	(.76)
	Min-Max	1-15	5-16	
SRS-2[f]				
	Mean	73	47	< .001***
	SD	9.44	5.28	(3.39)
	Min-Max	53-90	39-59	

Note. [a]ASD = children with autism spectrum disorder; [b]TD = children with typical development; [c]group comparisons using chi-square analyses for sex (male vs. female) and race (white vs. other), or Wilcoxon Mann-Whitney *U*-tests for age, KBIT-2, CELF-4, and SRS-2; [d]KBIT-2 = Matrices subtest of the Kaufman Brief Intelligence Test- Second Edition (Kaufman & Kaufman, 2004), mean standard score = 100, *SD* = 15; [e]CELF-4 = Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals-Fourth Edition (Semel et al., 2003), mean scaled score = 10, *SD* = 3; [f]SRS-2 = Social Responsiveness Scale- Second Edition (Constantino & Gruber, 2012), T-scores < 60 are considered within normal limits and scores ≥ 60 are associated with clinical presentations of ASD. *indicates that *p*-value is significant at .05; **indicates that *p*-value is ≤ .01; ***indicates that *p*-value is ≤ .001.

To compare comprehension of indirect answers between children with ASD and those with TD when matched on age and language ability, the researcher matched each participant in the ASD group to a child with TD on age (within 8 months) and expressive language ability (within 2 points of the CELF-4 Recalling Sentences scaled score). Four participants with ASD were excluded because of their low language scores (ranging from 1-4) for which no children from the TD group could be matched. As a result, two subgroups of 15 participants each were created. Table 3 provides a detailed summary of the demographic and linguistic characteristics of the two subgroups. Except the SRS-2 score, distribution of all other variables did not differ significantly between groups (all *p*s > .18).

Table 3

Characteristics of Participant Subgroups

Variable		ASD[a] *n* = 15	TD[b] *n* = 15	*P*[c] (*d*)
Age (year;month)				
	Mean	7;7	7;7	.84
	SD	13.10	11.85	(.01)
	Min-Max	5;4-8;11	6;0-8;10	
Sex				
	Male:Female	9:6	7:8	.46
Race				
	White	15	12	.28[g]
	Hispanic	0	1	
	African American	0	1	
	American Indian/Alaskan Native	0	1	
KBIT-2[d]				
	Mean	110	99	.18
	SD	19.37	16.98	(.59)
	Min-Max	69-137	70-127	
CELF-4[e]				
	Mean	10.13	10.33	.33
	SD	2.82	3.15	(.06)
	Min-Max	6-15	5-16	
SRS-2[f]				
	Mean	71	48	< .01*
	SD	9.19	5.96	(.04)
	Min-Max	53-85	39-59	

Note. [a]ASD = children with autism spectrum disorder; [b]TD = children with typical development; [c]group comparisons using chi-square analyses for sex (male vs. female) and race (white vs. other), or Wilcoxon Mann-Whitney *U*-tests for age, KBIT-2, CELF-4, and SRS-2; [d]KBIT-2 = Matrices subtest of the Kaufman Brief Intelligence Test- Second Edition (Kaufman & Kaufman, 2004), mean standard score = 100, *SD* = 15; [e]CELF-4 = Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals-Fourth Edition (Semel et al., 2003), mean scaled score = 10, *SD* = 3; [f]SRS-2 = Social Responsiveness Scale- Second Edition (Constantino & Gruber, 2012), T-scores < 60 are considered within normal limits and scores ≥ 60 are associated with clinical presentations of ASD. [g]Modified *p*-value by adding .5 to each cell value. [*]indicates that *p*-value is significant at .05; [**]indicates that *p*-value is ≤ .01; [***]indicates that *p*-value is ≤ .001.

Procedure

Participants in the ASD group completed a research protocol in conjunction with the Bangert (in preparation) study requiring a total of 3.5-4 hours. Participants were scheduled to start in midmorning, complete a 2-hour block of assessments, take a lunch break, and then finish the other 2-hour block. During each study session, participants were encouraged to take breaks as needed. Some participants chose to complete the two sessions on two separate visits scheduled within one week. Participants in the TD group completed an abbreviated research protocol that was approximately 20-30 minutes. All sessions were audio recorded using a digital audio recorder.

Assessments

Parent-report assessments. All parents or caregivers of children in the ASD and TD groups completed a Family Background Questionnaire (FBQ; adapted from Bangert, Halverson, & Finestack, 2019) and the Social Responsiveness Scale- Second Edition (SRS-2; Constantino & Gruber, 2012). Parents or caregivers of children with ASD completed an additional measure: the Theory of Mind Inventory- Second Edition (ToMI-2; Hutchins et al., 2010).

The FBQ (adapted from Bangert et al., 2019) included a series of questions about family demographic variables including race, ethnicity, maternal education, employment, and household income. The variables were used to characterize participants' demographic background. The FBQ also asked about the child's medical history, diagnosis of neurodevelopmental disorders, and special services (e.g., speech, occupational, and physical therapy) that they were receiving. The information obtained was used to exclude participants from the TD group if language impairments or developmental delays were reported.

The SRS-2 (Constantino & Gruber, 2012) is a measure of social behaviors in children and adults across three different age groups: preschool-age, school-age, and adult. The current study used the school-age version appropriate for individuals between 4-18 years of age. Parents or caregivers rated their children's reciprocal social behaviors (e.g., "Plays appropriately with children his/her own age," "Has an unusually narrow range of interests") on a 4-point Likert scale (i.e., "not true," "sometimes true," "often true," and "almost always true"). Scores from the SRS-2 served as an exclusionary criteria for participants in the TD group. Scores lower than 60 are considered within normal limits and not associated with clinical presentations of ASD. Thus, participants in the TD group with an SRS-2 score of 60 and higher were excluded from the study.

The ToMI-2 (Hutchins et al., 2010) is a measure of theory of mind competencies, consisting of 60 items designed to measure a wide range of social cognitive understanding. For each item, parents or caregivers read a statement (e.g., "My child understands whether someone hurts another on purpose or by accident") and rated their child on a 5-point scale ranging from "definitely not", "probably not", "undecided", "probably", to "definitely." Standard scores (*M* = 100, *SD* = 15) represented the child's overall ToM ability in natural environments and were used to examine the relationship between ToM and comprehension of indirect answers. Only parents or caregivers of children with ASD completed this measure.

Assessments for children with ASD. In addition to the ADOS-2 (Lord et al., 2012), which was used to confirm diagnosis and measure the severity of autism symptomology, participants in the ASD group also completed parts of four standardized, norm-referenced assessments: the Kaufman Brief Intelligence Test- Second Edition (KBIT-2; Kaufman & Kaufman, 2004), the Clinical Evaluation of Language Fundamentals- Fourth Edition (CELF-4, Semel et al., 2003), the Stanford-Binet Intelligence Scales- Fifth Edition (SB-5; Roid, 2003), and the Comprehensive Assessment of Spoken Language- Second Edition (CASL-2; Carrow-Woolfolk, 2017). Additionally, participants with ASD completed a criterion-referenced false-belief task (modeled after Hogrefe, Wimmer, & Perner, 1986 and Tager-Flusberg & Sullivan, 2000) designed to screen development of first-order ToM.

The KBIT-2 (Kaufman & Kaufman, 2004) is a measure of verbal and non-verbal intelligence for individuals between the ages of 4 years; 0 months and 90 years; 11 months. Participants completed the Matrices subtest of the assessment during which they viewed a pair of pictures that were related in some way (e.g., a rabbit and a carrot) and a third picture (e.g., a dog) paired with a question mark. Then, they were asked to select a picture from five possibilities that would best match with the third picture (i.e., a bone) in a way similar to the first set of pictures. As participants progressed, the test items became more difficult transitioning from relationships between people and objects to abstract symbols and designs with more pictures within an array to analyze. Participants established basal by obtaining three consecutive correct answers and reached ceiling by four consecutive scores of 0. Standard scores (M = 100, SD = 15) were calculated and used to characterize participants' non-verbal IQ and examine its relationship with comprehension of indirect answers.

The CELF-4 (Semel et al., 2003) is a measure of expressive and receptive language abilities for individuals aged 5 years; 0 months to 21 years; 11 months. Participants in the ASD group completed four core subtests of the CELF-4: Concepts and Following Directions, Word Structure, Recalling Sentences, and Formulated Sentences. For the Concepts and Following Directions subtest, participants identified objects in response to oral commands with increasing length and complexity (e.g., "Point to the apple after you point to the fish"). This subtest assesses receptive language ability. Examinees do not need to establish basal and reach ceiling after seven consecutive scores of 0. For the Word Structure subtest, participants looked at pictures and completed orally presented sentences that assess knowledge of grammatical rules (e.g., "Here's a horse. Here are two ___"). This subtest assesses expressive language ability and has no basal or ceiling rules. For the Recalling Sentences subtest, participants repeated sentences with varying length and syntactic complexity (e.g., "My mom is the nurse who works in the community clinic"). This subtest assesses expressive language ability. Examinees do not need to establish basal and reach ceiling after five consecutive scores of 0. For the Formulated Sentences subtest, participants formulated a sentence using target words and pictures (e.g., "Use the word 'quickly' to make a sentence about this picture"). This subtest assesses expressive language ability. Examinees do not need to establish basal and reach ceiling after five consecutive scores of 0. Scaled scores for each subtest (M = 10, SD = 3) were calculated by converting raw scores to a standardized scale. The Core Language scores (M = 100, SD = 15) were derived from all four subtests, and the Expressive Language scores (M = 100, SD = 15) were derived from the Word Structure, Recalling Sentences, and Formulated Sentences subtests. All scores were used to examine the relationships between language abilities and comprehension of indirect answers.

The SB-5 (Roid, 2003) is a measure of intelligence and cognitive abilities for individuals aged 2 years; 0 months through 85 years; 11 months. Participants completed the Non-verbal Knowledge subtest, which assesses procedural knowledge (e.g., using gestures to show how to blow a whistle or sweep with a broom) and picture absurdities (e.g., man cutting the tree branch that he is sitting on or ice cubes on the bottom of a glass of water). To score correct for the first example question, a participant could point to the man sitting on the branch and slide their finger to the ground or verbally describe, “He is going to fall.” The Non-verbal Knowledge subtest has six testlets with increasing difficulty. Each level has six questions, and examinees must obtain at least two correct answers to establish basal and move to the next level. They reach ceiling by having two or fewer correct answers in the testlet. In the standard testing procedure, examinees complete a routing task to determine the starting level. Given the age range, participants in the current study all started at Level 3, which is typical for preschoolers. All participants established basal at this level. Scaled scores (M = 10, SD = 3) for the subtest represented world knowledge and were used to examine its relationship with comprehension of indirect answers.

The CASL-2 (Carrow-Woolfolk, 2017) is a measure of oral language processing skills of comprehension and expression for individuals aged 3 years; 0 months to 21 years; 11 months. Participants completed the Pragmatic Language subtest, which measures appropriate use of language in a variety of social contexts (e.g., “What four details about a party would you tell your friend when you invite him?” “What would you say to a friend who is wearing a heavy jacket on a hot summer day?"). An appropriate response to the second question would be "Aren't you hot?" or "Why are you wearing a jacket?" and an inappropriate response would be "You are weird." or "Take it off." Examinees establish basal by obtaining four consecutive correct answers and reach ceiling after four consecutive scores of 0. Standard scores (*M* = 100, *SD* = 15) were calculated and used to examine the relationship between pragmatic language skills and comprehension of indirect answers.

The false-belief task, modeled after Hogrefe et al. (1986) and Tager-Flusberg & Sullivan (2000), is a criterion-referenced assessment of representational ToM. During the task, the researcher presented a crayon box containing a lollipop and asked the children, "What do you think is inside?" and provided the expected answer (i.e., crayons). Then, the researcher showed the children that there was actually a lollipop inside the crayon box. After closing the box, the researcher asked the following questions: 1) "What's really in the box?" (control question), 2) "If I show the crayon box to your mom, will she know what is in here?" (ignorance question), and 3) "What will she think is in the box?" (false-belief question). The control question was not counted toward the score. The ignorance and false-belief questions were scored as 0 or 1, resulting in the maximum score of 2 for this task. The scores were used to examine the relationships between first-order ToM ability and comprehension of indirect answers. Table 4 summarizes these assessment scores by participants with ASD.

Table 4

Assessment Scores by Participants with ASD

Assessment		ASD[a] (n = 19)
ADOS-2[b]		
	Mean	13.68
	SD	6.55
	Min-Max	7-32
KBIT-2[c] Matrices		
	Mean	105
	SD	20.33
	Min-Max	69-137
CELF-4[d] Concepts/Following Directions		
	Mean	7.94
	SD	3.99
	Min-Max	1-13
CELF-4[d] Word Structure		
	Mean	7.84
	SD	3.2
	Min-Max	3-13
CELF-4[d] Recalling Sentences		
	Mean	8.36
	SD	4.34
	Min-Max	1-15
CELF-4[d] Formulated Sentences		
	Mean	8.63
	SD	5.24
	Min-Max	1-16
CELF-4[d] Expressive Language Score		
	Mean	89.26
	SD	22.88
	Min-Max	49-124
CELF-4[d] Core Language Score		
	Mean	88.68
	SD	22.36
	Min-Max	44-120
SB-5[e] Non-verbal Knowledge		
	Mean	8.78
	SD	2.67
	Min-Max	5-16
CASL-2[f] Pragmatic Language		
	Mean	95.68
	SD	21.43
	Min-Max	63-139

False-belief Task[g]		
	Score of 0	*n* = 4
	Score of 1	*n* = 4
	Score of 2	*n* = 11
ToMI-2[h]		
	Mean	72.26
	SD	10.95
	Min-Max	57.6-96.3
SRS-2[i]		
	Mean	73
	SD	9.44
	Min-Max	53-90

Note. [a]ASD = children with autism spectrum disorder; [b]ADOS-2 = Autism Diagnostic Observation Schedule- Second Edition (Lord et al., 2012); [c]KBIT-2 = Kaufman Brief Intelligence Test- Second Edition (Kaufman & Kaufman, 2004); [d]CELF-4 = Clinical Evaluation of Language Fundamentals- Fourth Edition (Semel et al., 2003); [e]SB-5 = Stanford-Binet Intelligence Scales- Fifth Edition (Roid, 2003); [f]CASL-2 = Comprehensive Assessment of Spoken Language- Second Edition (Carrow-Woolfolk, 2017); [g]criterion-referenced false-belief task modeled after Hogrefe et al. (1986) and Tager-Flusberg & Sullivan (2000) with a maximum score of 2; [h]ToMI-2 = Theory of Mind Inventory- Second Edition (Hutchins et al., 2010); [i]SRS-2 = Social Responsiveness Scale- Second Edition (Constantino & Gruber, 2012).

Assessments for children with TD. Participants in the TD group completed an abbreviated protocol containing only two of the standardized, norm-referenced assessments administered to the children with ASD: the Matrices subtest of the KBIT-2 (Kaufman & Kaufman, 2004) and the Recalling Sentences subtest of the CELF-4 (Semel et al., 2003). Scores were used to characterize participants' non-verbal IQ and expressive language ability (see Tables 2 and 3).

Experimental Tasks

Participants in both the TD and ASD groups also completed an experimental task designed to measure comprehension of indirect answers and compare performance. The experimental stimuli consisted of 30 question-answer pairs designed to assess children's ability to comprehend and explain indirect answers. Similar to those found in the Bernicot et al. study (2007), each item included two images with audio stimuli of two people having a conversation. One person asked the question and the conversational partner responded with an indirect positive answer (e.g., (e.g., Q: “Are you feeling cold?” A: “I should have worn a sweater.”), an indirect negative answer (e.g., Q: “Are you feeling hungry?” A: “I just came from a pizza party.”), an ambiguous answer (e.g., Q: “Are you feeling hot?” A: “It feels like yesterday.”) or a direct answer (e.g., Q: “Are you feeling tired?” A: “I am feeling tired.”). After viewing the conversation on an iPad, the researcher pointed to the responder and asked the participant whether the speaker meant yes or no (Comprehension Task). Then, the researcher asked the participant “Why?” or “How did you know that?” (Explanation Task) to explain his or her answer. If the participant simply repeated the second person’s utterance (e.g., “He said he just came from a pizza party”), the researcher prompted the child by asking “Tell me more.” or “Why did you think s/he meant yes/no?” The researcher recorded child responses verbatim and scored them online. All sessions were audio recorded using a digital audio recorder for further coding and reliability purposes.

Development of Experimental Task. The 30 question-answer pairs were split across four conditions based on how the conversation partner responded to the question posed: Indirect Yes (10 items), Indirect No (10 items), Ambiguous Response (5 items), and Direct Response (5 items). Indirect Yes answers provided a positive response to the yes-no question without stating “yes.” Indirect No answers provided a negative response to the yes-no question without stating “no.” Ambiguous Response were designed to provide an unclear answer that could be interpreted either way to the yes-no question. Direct Response provided a clear “yes” or “no” to the question. The Ambiguous and Direct Response conditions had fewer items because the former served as an exploratory condition and the latter served as a control condition for comparison. Appendix A provides a complete list of the items.

Prior to testing, the researcher created a total of 48 question-answer pairs (e.g., Q: "Are you feeling cold?" A: "I should have worn a sweater."), evenly split across the four conditions. The researcher asked 20 adults who were native English speakers between 19 and 45 years of age to read these pairs in written form and judge whether the answer meant yes or no. For Indirect Yes and Indirect No conditions, the researcher selected the items with the highest agreement. The consensus indicated a mean agreement of 97% for the Indirect Yes items and 99% for the Indirect No items, with agreements ranging from 90%-100% for all items. The researcher selected the items with an agreement close to 50% for the Ambiguous Response items. The mean agreement for this condition was between 40%-45%, and the mean agreement for the Direct Response items was 100%.

The researcher controlled the syntactic complexity and semantic difficulty of the stimuli. The mean length of utterance (MLU; Brown, 1973) in morphemes for all question and answer prompts ranged from 5 to 7 morphemes with a mean of 5.8 and *SD* = .79. Vocabulary used in all prompts were acquired by age four, according to the age-of-acquisition norms created by Kuperman, Stadthagen-Gonzalez, & Brysbaert (2012). There were no suggestive words (e.g., good, bad, favorite, hate, like) that might reveal preferences and thus bias judgement. The researcher avoided contractions (e.g., she's, isn't, can't) to maximize the clarity of utterances and avoided other forms of non-literal language (e.g., idiom and sarcasm) to ensure the validity of the experimental stimuli.

The auditory stimuli were recorded by two native English speakers, one male and one female, with a mid-western dialect. They were naïve to the study aims and were instructed to read through a list of statements followed by a list of questions (i.e., the experimental stimuli). Two individual recording sessions took place in a quiet therapy room using a microphone and digital recorder. The speakers were instructed to keep the same volume and speech rate and use a neutral tone when reading the sentences. The researcher edited the sound files so that the male and female alternated to ask and answer questions. The number of items with the male versus female asking the question in each category was counterbalanced. The researcher created two randomized sequences of the 30 question-answer pairs. For each sequence, no more than two of the Indirect Yes or Indirect No items appeared consecutively.

Scoring and Coding

Prior to scoring and coding, the researcher and a research assistant listened to the audio recordings and transcribed all child explanations in an Excel spreadsheet. For the Comprehension Task, the researcher scored Indirect Yes, Indirect No, and Direct Response items as correct (1) or incorrect (0). The Ambiguous Response items were not scored because the answers could be interpreted either way (e.g., Q: "Did you have fun playing baseball?" A: "I tossed the ball."). Thus, the maximum score for the Comprehension Task was 25 (i.e., 10 Indirect Yes, 10 Indirect No, and 5 Direct Response items). For the Explanation Task, two trained research assistants, undergraduate majors in Speech-Language-Hearing Sciences, who were not involved in the transcription and naïve to the purpose of the study and participant condition (i.e., TD or ASD) independently judged whether the explanation for an indirect answer was adequate (1) or inadequate (0).

All research assistants completed the online Basic Course for Social/Behavior or Humanist Research training through the Collaborative Institutional Training Initiative Program. The researcher trained the two coding assistants through direct instruction on definitions of an adequate explanations and characteristics of each error code for inadequate explanations, utilizing participant responses as examples. The research assistants completed approximately 1 hour of training with the researcher. Next, the research assistants independently scored and coded 150 child responses (i.e., responses by three children with TD and two children with ASD). The researcher coded the same responses and calculated interrater reliability between the researcher and the assistants for each test item. Reliability on the Comprehension Task across all test items ranged from 80%-100% between the researcher and one assistant and 60%-100% between the researcher and the other assistant. Reliability on the Explanation Task across all test items ranged from 80%-100% between the researcher and the two assistants. The researcher and the research assistants met a second time to discuss instances of disagreement item by item and used the coding manual (Appendix B) and the error code worksheet (Figure 1) as a reference.

The researcher defined an adequate explanation as a response that links the speaker's utterance to his/her intention or provides an alternative reason that is appropriate to the context. For example, adequate explanations to the conversation, Q: "Are you hungry?" A: "I just came from a pizza party", included "He was not hungry because he already ate at the party." Inadequate explanations to the same question included "He said he went to a pizza party" or "Parties are fun." Scoring of the Explanation Task was independent from the Comprehension Task. That is, a participant could score 1 for the Comprehension Task but 0 for the Explanation Task and vice versa. For the "pizza party" question, a participant could score 0 by answering "Yes (he is hungry)" for the Comprehension Task but 1 for the Explanation Task by reasoning that "He didn't eat anything because he didn't like pizza, so probably hungry."

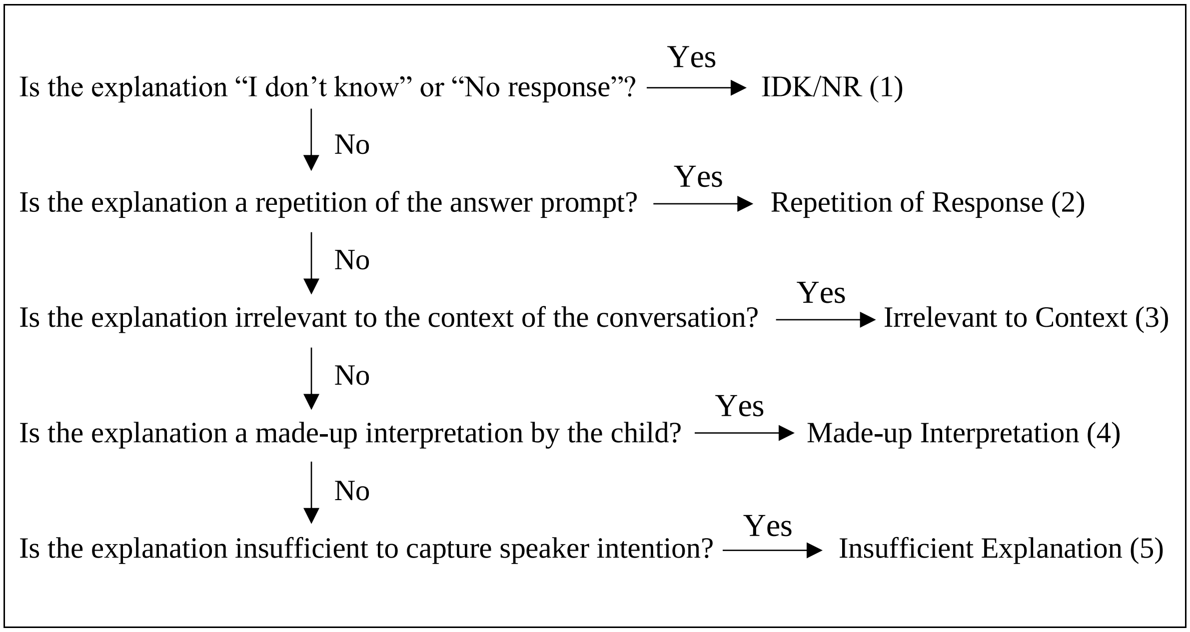
Scoring for the Ambiguous and Direct Response items was different from the other two conditions. Because the Ambiguous answer prompts were designed to be unclear, explanations were judged based on whether an appropriate or adequate speaker intention was provided. For example, taking the conversation, Q: "Did you have fun playing baseball?" A: "I tossed the ball," adequate explanations included "(Yes) because he might have tossed the ball really high and got a home run" and "(No) all he did was tossing the ball." Inadequate explanations included "(Yes) he tossed the baseball" and "(No) he didn't have fun." For Direct Response items, repetition of the answer prompt was considered adequate. For example, in the conversation: Q: "Did you go to the garden." A: "I did not go to the garden," "He said he did not go to the garden" or "He said so" would be judged as adequate (see Appendix B for more examples).

The research assistants double scored all responses. Interrater reliability was computed by dividing the number of instances of agreement by the total number of opportunities and multiplying by 100. Reliability across Indirect Yes, Indirect No, and Ambiguous Response items ranged from 82%-94% with an overall average of 87%. Instances of disagreement arose from the coders' subjective judgment of whether a response "adequately" explained speaker intentions. Take the item Q: "Have you finished your homework?" A: "I just got home from school." for example, the explanation "He doesn't want to do it." was coded as adequate by one coder but inadequate by the other coder. The former argued that the response showed that the child interpreted the indirect answer as an excuse, and the latter argued that the response failed to provide a convincing explanation (e.g., “He did not have time to do it.”). All disagreements were subsequently resolved by the researcher as a third coder. Interrater reliability was 100% for Direct Response items as all explanations were judged adequate.

After consensus on adequate versus inadequate responses was attained, the research assistants further assigned an error code to the inadequate responses as I Don’t Know/No Response (1), Repetition of Response (2), Irrelevant to Context (3), Made-up Interpretation that is appropriate to context but different from the speaker’s intention (4), or Insufficient Explanation that fails to capture speaker intention (5). The researcher created the error codes based on an initial evaluation of 150 child explanations and the coding scheme by Nippold and Martin (1989) for idiom interpretation. The researcher developed a worksheet that outlined a binary approach for the research assistants to categorize participant explanations (Figure 1). The research assistants double coded all inadequate explanations. Interrater reliability across Indirect Yes, Indirect No, and Ambiguous Response items ranged from 88%-98% with an overall average of 95%. Interrater reliability for Direct Response items was not calculated as all explanations for this category were judged adequate and thus, did not receive an error code. Appendix B provides descriptions and example responses assigned to each error code.

Figure 1

Error Code Worksheet



Statistical Analyses

To address Research Question 1, the research assistants scored participants' answers as correct (1) or incorrect (0) for the Comprehension Task. The researcher calculated percent correct for the Indirect Yes, Indirect No, and Direct Response conditions for every participant. Percent correct for the Ambiguous Response condition was not calculated because the answers could be interpreted either way (e.g., Q: "Are you going to the party?" A: "Bob will be at the party."). Given that all participants achieved 100% accuracy for the Direct Response items, the researcher created an Overall percent correct variable by averaging the accuracy of the two experimental conditions (i.e., Indirect Yes and Indirect No). For the Explanation Task, the research assistants scored participants' responses as adequate (1) or inadequate (0). The researcher calculated percent "adequate" for all four conditions and created an Overall percent "adequate" variable by averaging the accuracy of the three experimental conditions (i.e., Indirect Yes, Indirect No, and Ambiguous Response).

Next, the researcher conducted the Wilcoxon Mann-Whitney *U*-tests on the two matched subgroups (n = 15 each) to examine mean differences in the Comprehension Task and the Explanation Task across all experimental conditions. The Wilcoxon Mann-Whitney *U*-tests provided a more conservative non-parametric approach to test whether two independent groups had been sampled from the same population (Siegel & Castellan, 1988). The researcher also evaluated effect sizes using Cohen's *d*, with .2, .5, and .8 representing small, medium, and large effect sizes, respectively (Howell, 2016).

Because the study implemented the 50-50 "yes or no" forced-choice format for the Comprehension Task, the researcher calculated *d′* to index participants' sensitivity and response bias to the Indirect Yes and Indirect No items. The statistic *d′* is a measure of the distance between hit rate (i.e., proportion of Indirect Yes items to which participants responded "yes") and false alarm rate (i.e., proportion of Indirect No items to which participants responded "yes"). When the hit rate or false alarm rate was 0, .01 was added to calculate a modified *d′*; when the hit rate or false alarm rate was 1, .01 was subtracted to calculate a modified *d′*. Thus, the *d′* values ranged from 0 to 4.65 (greatest sensitivity), and 69% correct for both Indirect Yes and Indirect No items corresponded to a *d′* of 1 (MacMillan & Creelman, 1991).

Finally, the researcher created two scatter plots to visualize age as a function in task performance. The first figure shows children's performance on the Comprehension Task. The researcher plotted each participant's overall percent correct individually based on age on the x-axis and accuracy on the y-axis. The researcher also created a trend line for each group. The second figure shows children's performance on the Explanation Task. The researcher followed the same procedure to show performance by the two groups.

To address Research Question 2, the researcher conducted Pearson's correlational analyses to examine the relationships between the following participant characteristics: age, non-verbal IQ, receptive and expressive language ability, world knowledge, ToM ability, severity of autism symptomology, and pragmatic skills. Then, the researcher conducted another set of correlational analyses to examine performance on both the Comprehension Task and Explanation Task and the participant characteristics. Based on the derived correlation coefficients, the researcher built several hierarchical linear models using the lm() function in R (R Core Team, 2016) to identify explanatory variables that accounted for a significant amount of variance. Then, the researcher used the anova() function to compare change in effect size between different models and determine if fitting an additional explanatory variable significantly increased the variance in the new model.

To address Research Question 3, the research assistants coded inadequate explanations with one of the five error codes: I Don't Know/No Response, Repetition of Response, Irrelevant to Context, Made-up Interpretation, or Insufficient Explanation. The researcher tallied the number of instances of each error code and calculated the percentage of each error code for the two groups. The researcher conducted the Wilcoxon Mann-Whitney *U*-tests to examine differences in the distribution of error codes between the two groups.

Chapter 4

Results

Research Question 1

The first research question focused on comparing comprehension of indirect answers between the children with ASD and their peers with TD matched on age (± 8 months) and language ability (± 2 points of the Recalling Sentences subtest scaled score on the CELF-4). Table 5 summarizes performance by the two subgroups. No significant differences were found in either task across all experimental conditions (all *p*s > .13). However, medium effect sizes emerged from Comprehension of Indirect No answers [$d = .51$, $U(28) = 87$, $p = .29$] and Explanation of Ambiguous answers [$d = .51$, $U(28) = 80$, $p = .18$].

Table 5

Percent Correct/Adequate of ASD and TD Subgroups

Task		ASD[a] *n* = 15	TD[b] *n* = 15	*p*[c] (*d*)
Comprehension				
Indirect Yes				
	Mean	80	84	.49
	SD	13.62	13.52	(.29)
	Min-Max	60-100	60-100	
Indirect No				
	Mean	86	92	.29
	SD	14.04	8.61	(.51)
	Min-Max	60-100	70-100	
Direct/Control				
	Mean	100	100	-
	SD	0	0	
Overall[d]				
	Mean	83	88	.2
	SD	10.14	8.19	(0.54)
	Min-Max	65-95	75-100	
Explanation				
Indirect Yes				
	Mean	53.33	48.74	.58
	SD	23.8	18.78	(.21)
	Min-Max	20-90	20-80	
Indirect No				
	Mean	54.66	54.16	.98
	SD	31.13	24.11	(.01)
	Min-Max	0-100	0-90	
Ambiguous				
	Mean	37.33	51.33	.18
	SD	31.95	21.58	(0.51)
	Min-Max	0-100	0-80	
Direct/Control				
	Mean	100	100	-
	SD	0	0	
Overall[e]				
	Mean	48.44	51.41	.61
	SD	26.95	17.4	(.13)
	Min-Max	6-93	6-73	

Note. [a]ASD = children with autism spectrum disorder; [b]TD = children with typical development; [c]group comparisons using Wilcoxon Mann-Whitney *U*-tests to evaluate significance at .05 and Cohen’s ***d*** to measure effect sizes; [d]Overall percent correct calculated by averaging Indirect Yes and Indirect No; [e]Overall percent "adequate" calculated by averaging Indirect Yes, Indirect No, and Ambiguous Response.

The statistic *d′* indexes participants' sensitivity and response bias to the Indirect Yes and Indirect No items and ranges from 0 to 4.65 (greatest sensitivity). Approximately 69% correct for both Indirect Yes and Indirect No items corresponds to a *d′* of 1, and approximately 84% correct for both Indirect Yes and Indirect No items corresponds to a *d′* of 2 (MacMillan & Creelman, 1991). Table 6 summarizes *d′* values for the ASD and TD subgroups.

Table 6

Statistic d′ for ASD and TD Subgroups

d′	**ASD[a]** ***n* = 15**	**TD[b]** ***n* = 15**
Mean	2.31	2.84
SD	0.98	1.04
Min-Max	0.77-3.61	1.36-4.65

Note. [a]ASD = children with autism spectrum disorder; [b]TD = children with typical development.

Figure 2 shows the scatter plot of the overall performance on the Comprehension Task by children with TD and children with ASD.

Figure 2

Performance on Comprehension Task by Children with TD and Children with ASD

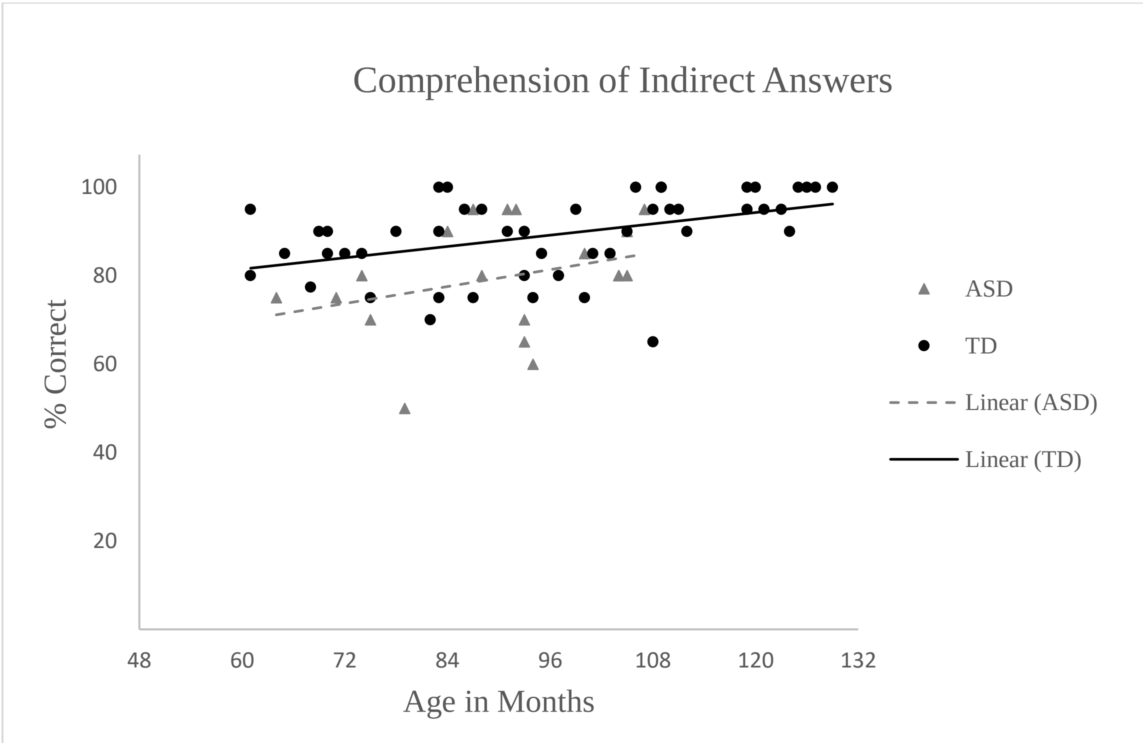
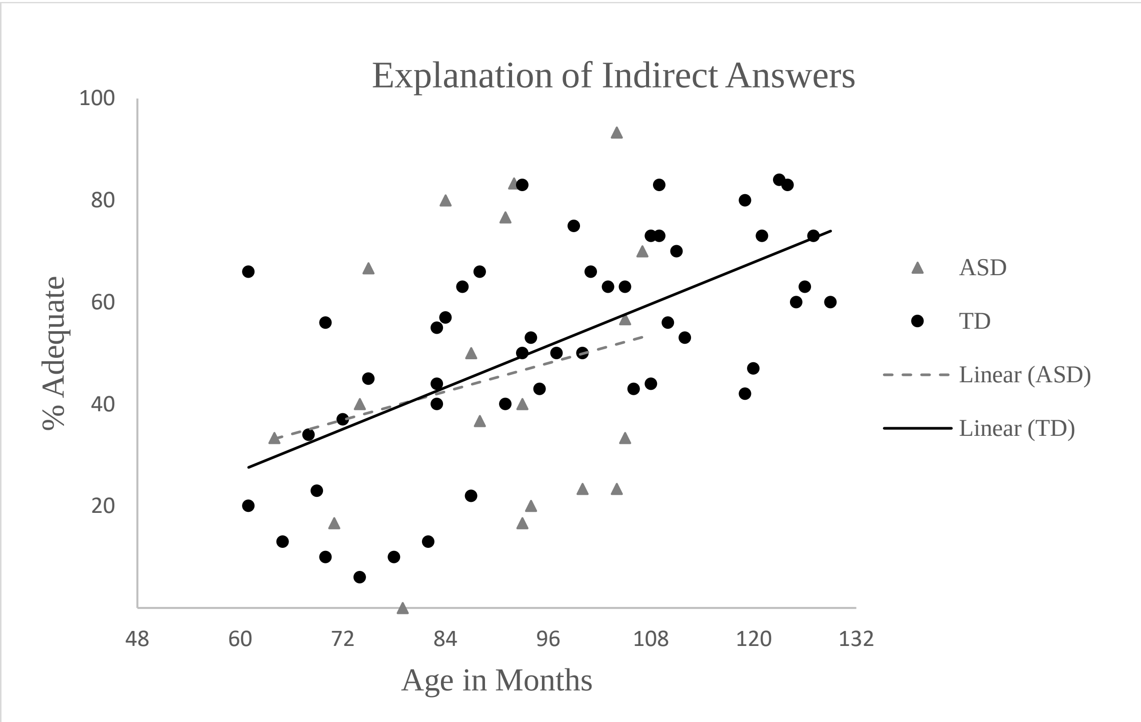


Figure 3 shows the scatter plot of the overall performance on the Explanation Task by children with TD and children with ASD.

Figure 3

Performance on Explanation Task by Children with TD and Children with ASD



Research Question 2

The second research question examined the relationships between comprehension of indirect answers and various participant characteristics as supporting factors to task performance among children with ASD. Table 7 shows correlation coefficients between the following participant characteristics: age, language ability (measured by the four Core Language subtests of the CELF-4), severity of autism symptomology (measured by the ADOS-2), non-verbal IQ (measured by the Matrices subtest of the KBIT-2), world knowledge (measured by the non-verbal Knowledge subtest of the SB-5), theory of mind ability (measured by the ToMI-2 and the false-belief task), and pragmatic skills (measured by the Pragmatic Language subtest of the CASL-2). Table 8 shows correlation coefficients between the two tasks and the above variables. The correlation coefficient ranged between -1 and +1, where ± .4, ± .6, and ± .8 represent weak, moderate, and strong relationship, respectively (Howell, 2016).

Table 7

Correlations between Participant Characteristics for Children with ASD

	Age	**Language Ability[a]**	**Autism Symptomology[b]**	**Non-verbal IQ[c]**	**World Knowledge[d]**	**Naturalistic ToM[e]**	**First-order ToM[f]**	**Pragmatic Skills[g]**
Age	1	-	-	-	-	-	-	-
Language Ability[a]	.17	1	-	-	-	-	-	-
Autism Symptomology	.07	**-.72**	1	-	-	-	-	-
Non-verbal IQ	.43	**.6**	-.51	1	-	-	-	-
World Knowledge	.09	**.63**	-.54	.56	1	-	-	-
Naturalistic ToM	-.3	.4	-.19	.04	.22	1	-	-
First-order ToM	.13	.25	-.08	-.16	.06	.02	1	-
Pragmatic Skills	.19	**.8**	-.52	.47	**.75**	.51	.23	1

[a]Language ability was measured by the four Core Language subtests of the CELF-4; [b]Autism symptomology was measured by the ADOS-2), [c]Non-verbal IQ was measured by the Matrices subtest of the KBIT-2; [d]World knowledge was measured by the non-verbal Knowledge subtest of the SB-5; [e]Naturalistic ToM was measured by the ToMI-2; [f]First-order ToM was assessed by a criterion-referenced false-belief task modeled after Hogrefe, Wimmer, & Perner (1986) and Tager-Flusberg & Sullivan (2000); [g]Pragmatic skills was measured by the Pragmatic Language subtest of the CASL-2. Bold text indicates moderate or strong relationships.

Table 8

Correlations between Task Performance and Potential Contributors for Children with ASD

Variable	Comprehension	Explanation
Age	.32	.34
Language Ability		
Concepts and Following Directions	**.67**	.57
Word Structure	.50	.52
Recalling Sentences	**.76**	**.62**
Formulated Sentences	.46	**.74**
Expressive Language Score[a]	**.64**	**.72**
Core Language Score[b]	**.72**	**.73**
Autism Symptomology	-.40	-.58
Non-verbal IQ	.40	.44
World Knowledge	.20	.52
ToM		
Naturalistic ToM	.09	.17
First-order ToM	.15	.19
Pragmatic Skills	.45	**.66**

Note. [a]Expressive language score was derived from the Word Structure, Recalling Sentences, and Formulated Sentences subtests; [b]Core language score was derived from the Concepts and Following Directions, Word Structure, Recalling Sentences, and Formulated Sentences subtests. Bold text indicates moderate relationships.

Comprehension Task. The researcher conducted a hierarchical linear regression analysis to evaluate the relationship between comprehension of indirect answers and age, language ability, severity of autism symptomology, non-verbal IQ, world knowledge, ToM, and pragmatic skills. For the first block analysis, the predictor variables age and language ability (measured by the Core Language score of the CELF-4) were analyzed. Although the relationship between age and task performance was weak (r = .32), the variable was included in the model to account for the small sample size with a wide age range. The results of the hierarchical linear regression analysis revealed a statistically significant model: F(2,16) = 10.19, $p < .01$ with an R^2 of .56, suggesting that age and language ability account for 56% of the variance in performance on the Comprehension

Task. Table 9 summarizes the results from the first block analysis as well as models that included additional variables as explanatory variables to the first model.

Table 9

Hierarchical Linear Regression Models to Predict Comprehension of Indirect Answers

Model	R^2	ΔR^{2a}	p^b
1: Age and Language	.56	-	-
2: adding ASD symptomology	.57	.01	.5
3: adding non-verbal IQ	.58	.02	.4
4: adding world knowledge	.66	.1	.04*
5: adding naturalistic ToM	.58	.02	.4
6: adding first-order ToM	.56	0	.76
7: adding pragmatic skills	.6	.04	.18

Note. [a]ΔR^2 indicates change in R^2 when compared to Model 1; [b]model comparisons using *t*-tests to evaluate significance. *indicates that *p*-value is significant at .05.

Of the seven hierarchical linear regression models, the fourth model using age, language ability, and world knowledge to predict performance on the Comprehension Task accounted for the most variance. Participants' predicted performance is equal to [31.85 + .19 (age) + .54 (language ability) - 2.02 (world knowledge)] percent correct when age is measured in months, language ability is measured by the standard score of the CELF-4 Core Language score, and world knowledge is measured by the scaled score of the SB-5 Non-verbal Knowledge subtest. Language ability ($b = .54$, $t(15) = 4.92$, $p < .01$) and world knowledge ($b = -2.02$, $t(15) = -2.21$, $p = .04$) were found to be significant independent predictors, whereas age was not a significant predictor ($b = .19$, $t(15) = 1.27$, $p = .22$).

Explanation Task. The researcher conducted a hierarchical linear regression analysis to evaluate the prediction of explanation of indirect answers from age, language ability, the severity of autism symptomology, non-verbal IQ, world knowledge, ToM, and pragmatic skills. For the first block analysis, the explanatory variables age and language ability (measured by the Core Language score of the CELF-4) were analyzed. The results of the hierarchical linear regression analysis revealed a statistically significant model: F(2,16) = 11.62, $p < .01$ with an R^2 of .59, suggesting that age and language ability account for 59% of the variance in performance on the Explanation Task. Table 10 summarizes the results from the first block analysis as well as models that included additional variables to the first model as predictors.

Table 10

Hierarchical Linear Regression Models to Predict Explanation of Indirect Answers

Model	R^2	ΔR^{2a}	p^b
1: Age and Language	.59	-	-
2: adding ASD symptomology	.61	.02	.39
3: adding non-verbal IQ	.6	.01	.52
4: adding world knowledge	.59	0	.67
5: adding naturalistic ToM	.59	0	.78
6: adding first-order ToM	.59	0	.92
7: adding pragmatic skills	.6	.01	.56

Note. [a]ΔR^2 indicates change in R^2 when compared to Model 1; [b]model comparisons using *t*-tests to evaluate significance at .05.

Of the seven hierarchical linear regression models, the first model using age and language ability to predict performance on the explanation was the best fit because adding additional variables did not significantly increase the variance. Participants' predicted performance is equal to [-60.82 + .43 (age) + .78 (language ability)] percent "adequate" when age is measured in months and language ability is measured by the standard score of the CELF-4 Core Language score. Language ability ($b = .78$, $t(16) = 4.31$, $p < .01$) was found to be a significant independent predictor, whereas age was not a significant predictor ($b = .43$, $t(16) = 1.35$, $p = .19$).

Research Question 3

The third research question further examined participants' inadequate responses by categorizing them into five errors types: I Don't Know/No Response, Repetition of Response, Irrelevant to Context, Made-up Interpretation, and Insufficient Explanation. Table 11 shows the distribution of the error types in percentage by the two groups.

Table 11

Percentage of Error Types by Participant Groups

Error Type	**ASD[a]** ***n* = 19**	**TD[b]** ***n* = 48**	***P*[c]** **(*d*)**
I Don't Know/No Response			
Mean	9.06	6.17	.58
SD	23.66	14.87	(.14)
Min-Max	0-36.36	0-72.72	
Repetition of Response			
Mean	33.64	41.42	.29
SD	26.12	26.13	(.29)
Min-Max	0-88.88	0-95.83	
Irrelevant to Context			
Mean	17.96	8.81	.02*
SD	15.39	11.66	(.67)
Min-Max	0-36.36	0-44.44	
Made-up Interpretation			
Mean	4.91	5.69	.81
SD	6.33	7.43	(.11)
Min-Max	0-15.78	0-27.27	
Insufficient Explanation			
Mean	34.4	37.9	.42
SD	24.65	22.67	(.14)
Min-Max	0-83.33	0-85.71	

Note. [a]ASD = children with autism spectrum disorder; [b]TD = children with typical development [c]group comparisons using Wilcoxon Mann-Whitney *U*-tests to evaluate differences and Cohen's *d* to measure effect sizes; *indicates that *p*-value is significant at .05.

For the ASD group, the most frequent error type was Insufficient Explanation (34%), followed by Repetition of Response (34%), Irrelevant to Context (18%), I Don't Know/No Response (9%), and Made-up Interpretation (5%). For the TD group, Repetition of Response was the most frequent error type (41%), followed by Insufficient Explanation (38%), Irrelevant to Context (9%), I Don't Know/No Response (6%), and Made-up Interpretation (6%). The orders are similar with the exception of the first two error types. Examination of differences in distribution between groups reveals that the children with ASD had significantly more explanations that were Irrelevant to Context (M = 17.96, SD = 15.39) than their peers with TD [M = 8.81, SD = 11.66, $U(65) = 293$, $p = .02$]. Distribution of the other error types did not differ significantly between groups (all *p*s > .27 and had small effect sizes).

Chapter 5

Discussion

The purpose of this study was to compare comprehension of indirect answers between children with ASD and their peers with TD matched on age and language ability. The results indicated that the two groups did not perform differently on measures of comprehension of indirect answers. This finding is contradictory to the prediction based on de Villiers et al. (2009) and Chin (2017) who reported that children with ASD performed significantly lower than their matched controls with TD. While differences in participant characteristics across samples may be the reason, Table 12 shows that this is not likely the case. Even though the children with ASD in the current study had the youngest mean age among the three studies as well as slightly lower language ability and autism severity than the participants in the Chin (2017) study, they performed at a similar level as their peers with TD. The participant characteristics could not be compared with the de Villiers et al. (2009) study because of different measures for language ability and ASD diagnosis.

Table 12

Participant Age, Language Ability, and Severity of ASD in Current Study, Chin (2017), and de Villiers et al. (2009)

Characteristics	Current Study	Chin (2017)	de Villiers et al. (2009)
Sample Size			
ASD[a]:TD[b]	15:15	12:12	10:20
Mean Age (Range)	7;6 (5-8)	9;1 (7-10)	8;3 (5-12)
Mean Language Score[c]			
ASD:TD	10.13:10.33	10.92:11	Did not report[d]
Severity of ASD[e]			
Mean	13.68	15.45	Can not compare[f]
SD	6.55	4.66	

Note. [a]ASD = children with autism spectrum disorder; [b]TD = children with typical development; [c]language ability measured by the Recalling Sentences scaled scores of the Clinical Evaluation of Language Fundamentals- Fourth Edition (Semel et al., 2003) in the current study and by the same subtest from the Fifth Edition (Semel et al., 2013) in Chin (2017); [d]de Villiers et al. did not report specific scores, but indicated that the children with ASD received mean age equivalence scores of 8;3 on the Expressive Vocabulary Test (Williams, 1998) and 7;9 on the Peabody Picture Vocabulary Test - Third Edition (Dunn & Dunn, 1997), both of which were close to their chronological age. [e]Severity of ASD measured by the ADOS-2 (Lord et al., 2012) in the current study and the ADOS (Lord et al., 2000) in Chin (2017); [f]de Villiers et al. used the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) to measure behaviors that are characteristic of ASD. Scores > 15 are associated with pervasive developmental disorder, and scores > 22 are associated with ASD. Of the ten children in the study, two had a score < 15, three between 15-22, and five >22.

One possible explanation for the inconsistent results may stem from methodological differences. The experimental task used in the Chin (2017) study is similar to the Comprehension Task in the current study, both of which implemented a forced-choice format to assess participants' judgement of indirect answers. Instead of the 50-50 "yes or no" used in the current study, Chin provided two additional foils in the story completion task to decrease the chance of guessing the right answer. Nevertheless, the mean *d*′ of 2.31 suggests that the ASD group in the current study scored higher than 84% for both Indirect Yes and Indirect No items (MacMillan & Creelman, 1991), which is significantly above chance. The experimental task used in the de Villiers et al. (2009) study is similar to the Explanation Task in the current study, both of which used open-ended why-questions that asked participants to explain indirect answers. However, the current study focused solely on indirect answers to yes-no questions that might have made reasoning speaker intentions easier than the indirect answers to wh-questions presented in the de Villiers et al. (2009) study because developmentally children understand and produce yes-no questions earlier than wh-questions (Tyack & Ingram, 1977). Nevertheless, the current study included a novel category of ambiguous answers that are more difficult to reason speaker intentions, and the children with ASD still performed at a level close to their peers with TD. Despite the methodological differences, results from the current study provide counterevidence to Chin (2017) and de Villiers et al. (2009), suggesting that comprehension of indirect answers varies within children with ASD, and this ability may be a strength for a subgroup with a particular cognitive and language profile.

Data from the analyses addressing the first research question indicate a noticeable performance gap between the two tasks. The children with ASD achieved near 80% accuracy for the Comprehension Task but only 42% "adequacy" for the Explanation Task. Similarly, the children with TD achieved above 80% accuracy for the Comprehension Task but only 31%-68% "adequacy" for the Explanation Task. Only the 9- and 10-year-olds were able to provide adequate explanations to indirect answers at above 60%. This performance gap is expected as previous research found that school-aged children have better developed metapragmatic skills to explain non-literal language (Ackerman, 1982; Anderson-Wood & Smith, 1997; Bernicot, 1991; Bernicot et al., 2007). For example, Laval (2003) examined both comprehension and metapragmatic knowledge (i.e., the ability to justify their chosen answers) of idioms in children with TD. She found that while 6-year-olds are able to understand idioms, the corresponding metapragmatic knowledge is not matured until after age 9. In the current study, the metapragmatic knowledge of indirect answers among children with TD follows a similar developmental trajectory. More specifically, 5- and 6-year-olds are only able to provide adequate explanations to indirect answers for approximately 30% of opportunities, whereas 7- and 8-year-olds were adequate for 50% of opportunities, and 9- and 10-year-olds were above 60%.

The second research question focused on factors supporting comprehension of indirect answers by children with ASD. The researcher conducted regression analyses to pinpoint the unique contributions of various potential contributors, including age, language ability, severity of autism symptomology, non-verbal IQ, world knowledge, naturalistic ToM ability, first-order ToM ability, and pragmatic skills. It was predicted that language ability, world knowledge, and ToM would each account for a significant variance in task performance. For the Comprehension Task, language ability and world knowledge, as measured by the CELF-4 Core Language subtests and the SB-5 Non-verbal Knowledge subtest, predicted a significant amount of variance in task performance. For the Explanation Task, language ability alone predicted a significant amount of variance in task performance. Therefore, language ability appears to be the key determinant of comprehension of indirect answers. This is consistent with Norbury's (2004, 2005) studies, which concluded that for children with ASD, language ability is a significant contributor, even more important than the ASD diagnosis, to comprehension of idioms and metaphors. Thus, language ability may be a more reliable factor to predict comprehension of non-literal language than the ASD diagnosis and may explain the cross-study variability in comprehension of non-literal language, particularly irony and scalar implicatures, among this population (see Chevallier et al., 2010; Kaland et al., 2002; MacKay & Shaw, 2004; Pastor-Cerezuela et al., 2018; Pexman et al., 2011; Pijnacker et al., 2009; Saban-Bezalel et al., 2019; Wang et al., 2006).

World knowledge is a novel factor that has been discussed in theory but not examined in experimental studies. The current study used the Non-verbal Knowledge subtest on the SB-5 as the target measure of world knowledge and found that it predicted a significant amount of additional variance in the Comprehension Task after language ability and age had been accounted for. This preliminary evidence shows that world knowledge is a significant factor that must be used as premises to facilitate comprehension of indirect answers, which is in line with Sperber & Wilson's relevance theory (1986) and Zufferey et al. (2019). The reason why world knowledge accounted for a significant amount of additional variance in the Comprehension Task but not the Explanation Task may be explained by the increased variance that language ability had accounted for in the Explanation Task. Nevertheless, world knowledge had a moderate correlation with the Explanation Task ($r = .52$). Thus, further investigation on its independent contributions under different testing formats is warranted.

ToM ability, as measured by a first-order false belief task and a naturalistic parent survey, did not contribute significantly to comprehension of indirect answers once age and language ability had been accounted for. This finding is contradictory to the long-standing claim that ToM is the foundation of understanding non-literal language (Happé, 1993; Martin & McDonald, 2004; Mitchell, 1997; Tager-Flusberg, 1995), including indirect answers (Chin, 2017; de Villiers et al., 2009). The difference may arise from how ToM was assessed in the current study. Specifically, the two measures eliminated heavy linguistic demand compared to the traditional narrative-based ToM tasks used in previous research. Norbury (2004) argued that the importance of ToM in comprehension of idioms may be influenced by the shared language component in both tasks. This argument is further supported by the current study because after language ability is accounted for, measures that assess ToM skills in naturalistic environments (i.e., the ToMI-2) or require minimal language demand (i.e., the first-order false belief task) reveal insignificant contributions to comprehension of indirect answers.

The third research question examined and compared inadequate explanations produced by children with ASD to those with TD. It was predicted that the ASD group would have significantly more responses characterized as I Don't Know/No Response or Repetition of Response than the TD group. However, the frequency of these two error types as well as the Made-up Interpretation and Insufficient Explanation did not differ significantly between groups. The only significant difference was the Irrelevant to Context, which accounted for nearly 18% of erroneous responses for children with ASD versus 8% for children with TD. The results indicate that when reasoning speaker intentions from indirect answers, children with ASD are more likely to provide an explanation that is irrelevant to context. This type of response comments on the general subject that is under discussion but fails to integrate the contextual information for proper interpretation (see Appendix B for examples). The data may serve as preliminary evidence for the error pattern unique to children with ASD. In clinical practice, speech-language pathologists may focus on this type of error by teaching children with ASD how to identify and evaluate contextual information to derive appropriate speaker intentions.

Limitations and Future Directions

Although the current study provides novel evidence to the comprehension abilities of children with ASD, one limitation to these findings stems from the small sample size and uneven number of participants in each age group. Specifically, the ASD group included a total of 19 participants with two 5-year-olds, three 6-year-olds, eight 7-year-olds, and six 8-year-olds. Therefore, results from the current study are more appropriate to be generalized to 7- to 8-year-old children with ASD. Future studies should more closely examine and compare comprehension of indirect answers between preschool-aged children with ASD and TD and include a larger sample to increase external validity.

Another limitation relates to the ToM measures used in this study (i.e., the ToMI-2 and the first-order false belief task). They are a novel method to measure the development of perspective-taking abilities different from the traditional narrative-based method. While a narrative-based ToM task might have yielded opposite results with respect to the link between ToM and comprehension of indirect answers, the current study did not include this type of measure to allow for direct comparisons. Future studies that investigate the influence of ToM ability on non-literal language should consider using both naturalistic and narrative-based measures to evaluate the effect of the language component in ToM tasks on comprehension of non-literal language.

A third limitation arises from the exploratory nature of the error analyses. While the researcher created the coding scheme based on Nippold and Martin's (1989) classification system for idiom interpretation, the five error types are not an exhaustive list. Additionally, the researcher attempted to organize the error codes hierarchically from least to most satisfactory, albeit being inadequate explanations [i.e., I Don't Know/No Response (1), Repetition of Response (2), Irrelevant to Context (3), Made-up Interpretation (4), and Insufficient Explanation (5)]. This framework is not evidence-based and requires further investigation to validate its methodological soundness.

There is much work to be done to determine the comprehension ability of indirect answers among children with ASD. Similar to many other forms of non-literal language, there is evidence and counterevidence to this population's ability to achieve adequate performance as their peers with TD. Studies in this area should carefully document their participants' demographic characteristics as well as related cognitive and linguistic abilities so that a subgroup of children with ASD may be identified as equally good as or significantly poorer than those with TD. In the current study, the language ability of the children with ASD was within normal range as evident by their mean CELF-4 Core Language score of 88.68. However, there is another subgroup of children with ASD and comorbid language impairments (Condouris, Meyer, & Tager-Flusberg, 2003; Huang & Finestack, 2020; Rice, Warren, & Betz, 2005; Roberts, Rice, & Tager-Flusberg, 2004) who may not perform as well on the same tasks. Future studies may also consider comparing performance across different neurodevelopmental conditions (e.g., developmental language disorder, Down syndrome) with varying cognitive and language profiles to determine if comprehension of indirect answers is more susceptible to a specific diagnosis, language profile, or cognitive functioning.

The current study did not examine suprasegmental or paralinguistic factors that may facilitate comprehension of indirect answers. For example, intonation contours are known for their pragmatic function of suggesting or imposing an alternative message that is different from the literal utterance (de Marneffe & Tonhauser 2019; Dennison & Schafer, 2017; Kurumada, Brown, Bibyk, Pontillo, & Tanenhaus, 2014; Pierrehumbert & Hirschberg, 1990). Facial expressions can provide visual cues to determine whether there is a discrepancy between the literal message and speaker intention (Attardo, Eisterhold, Hay, & Poggi, 2003; Caucci & Kreuz, 2012; Deliens, Antoniou, Clin, Ostashchenko, & Kissine, 2018). Indirect answers are naturally accompanied by these multimodal cues in everyday communication. Thus, future research may investigate their influences on comprehension of indirect answers of children with ASD.

Finally, variability in the comprehension of different forms of non-literal language among children with ASD also remains much unknown. For example, Chin (2017) found that comprehension of scalar implicatures is easier than indirect answers and metaphors, and Deliens et al. (2018) found that comprehension of indirect requests (e.g., "It's cold in here" indirectly requests the hearer to close the window) is easier than irony. More studies that utilize within-group comparisons are needed to gain a full picture of this population's comprehension abilities of various forms of non-literal language. Results will reveal the challenges children with ASD may experience and guide assessment and treatment to address their communication weaknesses.

Chapter 6

Conclusion

The main contribution of the current study is that it provides a detailed comparison of the comprehension and explanation abilities of indirect answers between children with ASD and their age- and language-matched peers with TD. The results show that children with ASD perform at a level similar to those with TD on both tasks. This is novel evidence contradictory to previous research by Chin (2017) and de Villiers et al. (2009) that indicated a significant difference between the two groups. Thus, comprehension of indirect answers may be heterogenous among children with ASD, and this ability may be a strength for a subgroup with with a particular cognitive and language profile.

Looking at various factors supporting comprehension of indirect answers, language ability and world knowledge are crucial contributors, whereas non-verbal IQ, severity of autism symptomology, ToM ability, and pragmatic skills may not be as important in accounting for task performance. The clinical implication is that receptive and expressive language skills are the foundation of comprehension of indirect answers. Additionally, speech-language pathologists may consider teaching world knowledge and explaining how it relates to reasoning speaker intentions in conjunction with existing language goals. For example, if the child misinterprets "I have an exam tomorrow" as an indirect yes to going to a party, the clinician may discuss why people prioritize personal responsibilities over leisure activities in most circumstances.

The error analysis indicates that when interpreting speaker intentions, children with ASD are more likely to provide an explanation that is irrelevant to context than their peers with TD (see Appendix B for examples). Although these data are preliminary, metapragmatic skills in children with ASD may worth further investigation. If the findings can be replicated in future studies, this error pattern may be identified as unique to children with ASD. A more complete understanding of the nature of their struggle with indirect answers will allow clinicians and researchers to develop intervention approaches tailored to this communication skill.

References

Ackerman, B. P. (1982). On comprehending idioms: Do children get the picture? *Journal of Experimental Child Psychology*, *33*, 439–454.

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

American Psychiatric Association. (2015). *Neurodevelopmental Disorders: DSM-5® Selections*. American Psychiatric Pub.

Anderson-Wood, S., & Smith, B. R. (1997). Working with Pragmatics, Bicester: Winslow.

Andrés-Roqueta, C., & Katsos, N. (2020). A distinction between linguistic and social pragmatics helps the precise characterization of pragmatic challenges in children with autism spectrum disorders and developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 1-15.

Attardo, S., Eisterhold, J., Hay, J., & Poggi, I. (2003). Multimodal markers of irony and sarcasm. *Humor*, *16*(2), 243-260.

Bangert, K. J., Halverson, D. M., & Finestack, L. H. (2019). Evaluation of an explicit instructional approach to teach grammatical forms to children with low-symptom severity autism spectrum disorder. *American Journal of Speech-Language Pathology*, *28*(2), 650-663.

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

Bauminger-Zviely, N. (2013). *Social and academic abilities in children with high-functioning autism spectrum disorders*. Guilford Press.

Bernicot, J. (1991). French children's conception of requesting: The development of metapragmatic knowledge. *International Journal of Behavioral Development*, *14*(3), 285-304.

Bernicot, J., Laval, V., & Chaminaud, S. (2007). Non-literal language forms in children: In what order are they acquired in pragmatics and metapragmatics?. *Journal of Pragmatics*, *39*(12), 2115-2132.

Berument S. B., Rutter M., Lord C., Pickles A., Bailey A., (1999). Autism screening questionnaire: Diagnostic validity British Journal of Psychiatry, 175: 444–451.

Boucher, J. (2003, November). Language development in autism. In *International Congress Series* (Vol. 1254, pp. 247-253). Elsevier.

Brown, R. (1973). A first language: The early stages. Oxford, England: Harvard U. Press.

Bucciarelli, M., Colle, L., & Bara, B. G. (2003). How children comprehend speech acts and communicative gestures. *Journal of Pragmatics*, *35*(2), 207-241.

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.

Carrow-Woolfolk, E. (2017). Comprehensive Assessment of Spoken Language, Second Edition (CASL-2) [Manual]. Torrance, CA: Western Psychological Services.

Caucci, G. M., & Kreuz, R. J. (2012). Social and paralinguistic cues to sarcasm. *Humor*, *25*(1), 1-22.

Chevallier, C., Wilson, D., Happé, F., & Noveck, I. (2010). Scalar inferences in autism spectrum disorders. *Journal of autism and developmental disorders*, *40*(9), 1104-1117.

Chin, I. (2017). Variability in Pragmatic Abilities in Children with Autism Spectrum Disorder.

Colich, N. L., Wang, A. T., Rudie, J. D., Hernandez, L. M., Bookheimer, S. Y., & Dapretto, M. (2012). Atypical neural processing of ironic and sincere remarks in children and adolescents with autism spectrum disorders. *Metaphor and symbol*, *27*(1), 70-92.

Colston, H. L., & Kuiper, M. S. (2002). Figurative Language Development Research and Popular Children's Literature: Why We Should Know," Where the Wild Things Are". *Metaphor and Symbol*, *17*(1), 27-43.

Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology*.

Constantino, J. N., & Gruber, C. P. (2012). *Social Responsiveness Scale Second Edition (SRS-2)*. Western Psychological Services (WPS).

Cummings, L. (2017). *Research in clinical pragmatics*, Vol. 11. Springer.

Deliens, G., Antoniou, K., Clin, E., Ostashchenko, E., & Kissine, M. (2018). Context, facial expression and prosody in irony processing. *Journal of memory and language*, *99*, 35-48.

Deliens, G., Papastamou, F., Ruytenbeek, N., Geelhand, P., & Kissine, M. (2018). Selective pragmatic impairment in autism spectrum disorder: Indirect requests versus irony. *Journal of autism and developmental disorders*, *48*(9), 2938-2952.

De Marneffe, M. C., & Tonhauser, J. (2019). Inferring meaning from indirect answers to polar questions: The contribution of the rise-fall-rise contour. In *Questions in discourse* (pp. 132-163). Brill.

De Villiers, P. A., de Villiers, J., Coles-White, D., & Carpenter, L. (2009). Acquisition of relevance implicatures in typically-developing children and children with autism. In *Proceedings of the 33th annual Boston university conference on language development* (pp. 121-132). Somerville, MA: Cascadilla Press.

Dennis, M., Lazenby, A.L. & Lockyer, L. (2001) Inferential Language in High Function Children with Autism, Journal of Autism and Developmental Disorders 31: 47–54.

Dennison, H., & Schafer, A. J. (2017). Processing intonationally implicated contrast versus negation in American English. *Language and Speech*, *60*(2), 174-199.

Dews, S., Winner, E., Kaplan, J., Rosenblatt, E., Hunt, M., Lim, K., ... & Smarsh, B. (1996). Children's understanding of the meaning and functions of verbal irony. *Child development*, *67*(6), 3071-3085.

Eales, M. J. (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.

Ellawadi, A. B., & Weismer, S. E. (2015). Using spoken language benchmarks to characterize the expressive language skills of young children with autism spectrum disorders. *American journal of speech-language pathology*, *24*(4), 696-707.

Elliott, C.D., Smith, P. & McCulloch, K. (1997). British Ability Scales II. Windsor: NFERNelson.

Emerich, D. M., Creaghead, N. A., Grether, S. M., Murray, D., & Grasha, C. (2003). The comprehension of humorous materials by adolescents with high-functioning autism and Asperger's syndrome. *Journal of autism and developmental disorders*, *33*(3), 253-257.

Gibbs R. W. & Colston H. L. (2012). *Interpreting Figurative Meaning*. Cambridge: Cambridge University Press.

Gilliam, J. E. (2006). *Gilliam Autism Rating Scale*, *2nd Edn.* Madrid: Symtéc.

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

Gluksberg S (2001) *Understanding Figurative Language: From Metaphor to Idioms,* Oxford Psychology Series. New York: Oxford University Press.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41-58). Brill.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes,* 20, 667.

Hancock, J. T., Dunham, P. J., & Purdy, K. (2000). Children's comprehension of critical and complimentary forms of verbal irony. *Journal of Cognition and Development*, *1*(2), 227-248.

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

Happé, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child development*, *66*(3), 843-855.

Hockey, B. A., Rossen-Knill, D., Spejewski, B., Stone, M., & Isard, S. (1997). Can you predict responses to yes/no questions? yes, no, and stuff. In *Fifth european conference on speech communication and technology*.

Hogrefe, G. J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. Child Development, 57, 567-582.

Horn, L. R. (1972). On the Semantic Properties of Logical Operators in English, PhD thesis, University of California, LA.

Howell, D. C. (2016). *Fundamental statistics for the behavioral sciences*. Nelson Education.

Huang, T., & Finestack, L. (2020). Comparing Morphosyntactic Profiles of Children With Developmental Language Disorder or Language Disorder Associated With Autism Spectrum Disorder. *American Journal of Speech-Language Pathology*, *29*(2), 714-731.

Hutchins, T. L., Prelock, P. A., & Bonazinga, L. B. (2010). Technical manual for the theory of mind inventory (tomi) and theory of mind task battery. *University of Vermont, Burlington*.

Julien, H. (2018). Communicative Repair Requests Produced by Typically-Developing Preschoolers and Preschoolers with Autism Spectrum Disorder.

Kaland, N., Moller-Nielsen, A., Callesen, K., Mortensen, E. L., Gottlieb, D., & Smith, L. (2002). A new 'advanced' test of theory of mind: Evidence from children and adolescents with Asperger syndrome. *Journal of Child Psychology and Psychiatry, 43*, 517–528.

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

Kaufman, A. S., & Kaufman, N. L. (2004). Kaufman Brief Intelligence Test, Second Edition. Bloomington, MN: Pearson, Inc.

Kerbel, D. & Grunwell, P. (1997). Idioms in the classroom: An investigation of language unit and mainstream teachers' use of idioms. *Child Language Teaching and Therapy*, *13*(2), 113-123.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior research methods*, *44*(4), 978-990.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D. F., & Tanenhaus, M. K. (2014). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition*, *133*(2), 335-342.

Laval, V. (2003). Idiom comprehension and metapragmatic knowledge in French children. *Journal of Pragmatics, 35*,723–739.

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

Levinson, S. C. (2000). Presumptive Meanings, MIT Press, Cambridge, MA.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S. (2000). Autism diagnostic observation schedule: Manual. Los Angeles, CA: Western Psychological Services.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., Bishop, S. (2012). Autism diagnostic observation schedule, second edition. Torrance, CA: Western Psychological Services.

Loukusa, S., Leinonen, E., & Ryder, N. (2007). Development of pragmatic language comprehension in Finnish-speaking children. *First Language*, *27*(3), 279-296.

MacKay, G., & Shaw, A. (2004). A comparative study of figurative language in children with autistic spectrum disorders. *Child Language Teaching and Therapy*, *20*, 13–32.

Macmillan, N. A., & Creelman, C. D. (1991). Detection theory: a user's guide. Cambridge UP, Cambridge, UK.

Martin, I., & McDonald, S. (2004). An exploration of causes of non-literal language problems in individuals with Asperger syndrome. *Journal of autism and developmental disorders*, *34*(3), 311-328.

Mitchell, P. (1997). *Introduction to theory of mind: Children, autism and apes*. Edward Arnold Publishers.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false-belief understanding. *Child development*, *78*(2), 622-646.

Nippold, M. A. (1985). Comprehension of figurative language in youth. *Topics in Language Disorders*.

Nippold, M. A., & Martin, S. T. (1989). Idiom interpretation in isolation versus context: A developmental study with adolescents. *Journal of Speech, Language, and Hearing Research*, *32*(1), 59-66.

Norbury, C. F. (2004). Factors supporting idiom comprehension in children with communication disorders. *Journal of Speech, Language, and Hearing Research*, *47*(5), 1179-1193.

Norbury, C. F. (2005). The relationship between theory of mind and metaphor: Evidence from children with language impairment and autistic spectrum disorder. *British Journal of Developmental Psychology*, *23*(3), 383-399.

Noveck, I. (2001). When children are more logical than adults: Experimental investigations of scalar implicature. *Cognition*, 78, 165–188.

Ozonoff, S., & Miller, J. N. (1996). An exploration of right-hemisphere contributions to the pragmatic impairments of autism. *Brain and language*, *52*(3), 411-434.

Pastor-Cerezuela, G., Tordera Yllescas, J. C., González-Sala, F., Montagut-Asunción, M., & Fernández-Andrés, M. I. (2018). Comprehension of generalized conversational implicatures by children with and without autism spectrum disorder. *Frontiers in psychology*, *9*, 272.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics-pragmatics interface. *Cognition*, 86, 253–282.

Pexman, P. M., & Glenwright, M. (2007). How do typically developing children grasp the meaning of verbal irony?. *Journal of Neurolinguistics*, *20*(2), 178-196.

Pexman, P. M., Rostad, K. R., McMorris, C. A., Climie, E. A., Stowkowy, J., & Glenwright, M. R. (2011). Processing of ironic language in children with high-functioning autism spectrum disorder. *Journal of autism and developmental disorders*, *41*(8), 1097-1112.

Pierrehumbert, J., & Hirschberg, J. B. (1990). The meaning of intonational contours in the interpretation of discourse.

Pijnacker, J., Hagoort, P., Buitelaar, J., Teunisse, J. P., & Geurts, B. (2009). Pragmatic inferences in high-functioning adults with autism and Asperger syndrome. *Journal of autism and developmental disorders*, *39*(4), 607.

Pouscoulous, N. (2014). The elevator's buttocks. *Metaphorical abilities in children. In D. Matthews (Ed.), Pragmatic development in first language acquisition*, 239-259.

Pouscoulous, N., Noveck, I. A., Politzer, G., & Bastide, A. (2007). A developmental investigation of processing costs in implicature production. *Language acquisition*, *14*(4), 347-375.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and brain sciences*, *1*(4), 515-526.

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from https://www.R-project.org/

Repacholi, B. M., & Gopnik, A. (1997). Early reasoning about desires: evidence from 14- and 18-month-olds. *Developmental psychology*, *33*(1), 12.

Rice, M. L., Warren, S. F., & Betz, S. K. (2005). Language symptoms of developmental language disorders: An overview of autism, Down syndrome, fragile X, specific language impairment, and Williams syndrome. *Applied Psycholinguistics*, *26*(1), 7-27.

Roberts, J. A., Rice, M. L., & Tager-Flusberg, H. (2004). Tense marking in children with autism. *Applied Psycholinguistics*, *25*(3), 429-448.

Roid, G. H. (2003a). Stanford-Binet Intelligence Scales, Fifth Edition. Itasca, IL: Riverside.

Rundblad, G., & Annaz, D. (2010). Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology*, *28*(3), 547-563.

Rutter, M., Le Couteur, A., & Lord, C. (2003). Autism diagnostic interview-revised. *Los Angeles, CA: Western Psychological Services*.

Saban-Bezalel, R., Dolfin, D., Laor, N., & Mashal, N. (2019). Irony comprehension and mentalizing ability in children with and without Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *58*, 30-38.

Saban-Bezalel, R., & Mashal, N. (2015). The effects of intervention on the comprehension of irony and on hemispheric processing of irony in adults with ASD. *Neuropsychologia*, *77*, 233-241.

Schulze, C., Grassmann, S., & Tomasello, M. (2013). 3-year-old children make relevance inferences in indirect verbal communication. *Child development*, *84*(6), 2079-2093.

Semel, E., Wiig, E. H., & Secord, W. A. (2003). Clinical evaluation of language fundamentals, fourth edition (CELF-4). Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Semrud-Clikeman, M., & Glass, K. (2010). The relation of humor and child development: Social, adaptive, and emotional aspects. *Journal of Child Neurology*, *25*(10), 1248-1260.

Siegel, S., & Castellan, N. J., JR. (1988). Nonparametric statistics for the behavioral sciences. New York, NY: McGraw-Hill, Inc.

Stenstrom, Anna-Brita. 1984. ¨ Questions and Responses in English Conversation. Lund: Lund University Press.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental psychology*, *30*(3), 395.

Tager-Flusberg, H. (1995). 'Once upon a ribbit': Stories narrated by autistic children. *British journal of developmental psychology*, *13*(1), 45-59.

Tager-Flusberg, H. (1996). Brief report: Current theory and research on language and communication in autism. *Journal of Autism and Developmental disorders*, *26*(2), 169-172.

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and communication in autism. *Handbook of autism and pervasive developmental disorders*, *1*, 335-364.

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.

Tager-Flusberg, H., & Sullivan K. (2000). A componential view of theory of mind: Evidence from Williams syndrome. Cognition, 76, 59 - 89.

Tribushinina, E. (2012). Comprehension of relevance implicatures by pre-schoolers: The case of adjectives. *Journal of Pragmatics*, *44*(14), 2035-2044.

Tyack, D., & Ingram, D. (1977). Children's production and comprehension of questions. *Journal of Child Language*, *4*(2), 211-224.

Verbuk, A., & Shultz, T. (2010). Acquisition of relevance implicatures: A case against a rationality-based account of conversational implicatures. *Journal of Pragmatics*, *42*(8), 2297-2313.

Vosniadou, S., & Ortony, A. (1983). The emergence of the literal-metaphorical-anomalous distinction in young children. *Child Development*, 154-161.

Siltanen, S. A. (1986). “Butterflies are rainbows?”: A developmental investigation of metaphor Comprehension. *Communication Education*, *35*(1), 1-12.

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

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2006). Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain*, *129*, 932–943.

Winner, E., Rosenstiel, A. K., & Gardner, H. (1976). The development of metaphoric understanding. *Developmental psychology*, *12*(4), 289.

Whyte, E. M., Nelson, K. E., & Scherf, K. S. (2014). Idiom, syntax, and advanced theory of mind abilities in children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research*.

Zufferey, S., Moeschler, J., & Reboul, A. (2019). *Implicatures*. Cambridge University Press.

Appendix A

Indirect Yes	Indirect No	Ambiguous Response	Direct Response
Are you going to the circus? I have my binoculars ready.	Are you going to the concert? I have an exam tomorrow.	Are you going to the party? Bob will be at the party.	Are you going to the movie? I am going to the movie.
Can you buy some ice cream? I have a coupon for ice cream.	Can you buy some popsicles? I am heading the other way.	-	-
Would you like some apple pie? I am always hungry for desserts.	Would you like some popcorn? I have a sore tooth.	-	-
Did you go to the museum? I saw real dinosaurs.	Did you go to the park? It was stormy all morning.	-	-
Are you feeling cold? I should have worn a sweater.	Are you feeling hungry? I just came from a pizza party.	Are you feeling hot? It feels like yesterday.	Are you feeling tired? I am feeling tired.
Have you done your chores? Mom will be happy with me.	Have you finished your homework? I just got home from school.	-	-
Did you water the flowers? My shoes are all wet.	Did you mow the lawn? Jonny was taking a nap.	Did you take out the trash? The garbage truck just came.	Did you go to the garden? I did not go to the garden.
Did you have fun diving? My dream is to do it everyday.	Did you have fun skiing? I sprained my ankle.	Did you have fun playing baseball? I tossed the ball.	Did you have fun swimming? I did not have fun swimming.
Are you going to the lake? I have packed my fishing rod.	Are you going to the beach? Getting sunburned hurts a lot.	Are you going to the mountains? Last time I saw a grizzly bear.	Are you going to the zoo? I am going to the zoo.
Do you have pets? Daisy is my best friend.	Do you have a brother or sister? I have two dogs.	-	-

Appendix B

Codes and Definitions	Indirect Yes Are you going to the circus? I have my binoculars ready.	Indirect No Are you feeling hungry? I just came from a pizza party.	Ambiguous Response Did you have fun playing baseball? I tossed the ball.	Direct Response Are you going to the movie? I am going to the movie.[b]
Adequate Explanation Child provided an explanation that links the speaker's utterance to his/her intention.	She will see better at the circus. *She is going bird watching, not circus.	He already "ated" at the party. *He didn't eat anything because he didn't like pizza, so probably hungry.	I think he got a home run. *All he did was tossing the ball.	He said I'm going to the movie. Yes means going. He said so.
I don't Know/No Response (1) Child stated "I don't know" or did not respond.	I don't know. He (silence).	I don't know. Because I know.[a]	I don't know how to explain.	
Repetition of Response (2) Child repeated the answer prompt or slightly changed it without adding new information.	She has her binoculars. She said I have my binoculars ready.	He just came from a pizza party. He said I just came from a party.	He said he tossed the ball. He had fun tossing the ball.	
Irrelevant to Context (3) Child commented on the general subject that is under discussion but failed to reason the speaker's intention with relevant contextual information.	She saw a lion. He wanted to play at the circus.	Because sometimes we have pizza parties. Pizza parties are the best.	I like watching baseball games.	
Made-up Interpretation (4) Child provided an explanation relevant to the context but was not based on the speaker's utterance or intention.	She likes animal tricks. She's afraid of clowns.	He doesn't eat much.	Because he likes to play baseball.	

Insufficient Explanation (5) Child response lacked a convincing explanation to the speaker's intention (i.e., missing the mark).	She packed.	Because pizza is food.	Maybe he tossed it hard.

Note. [a]Child did not respond after further prompting (e.g., "Tell me more."); [b]All participants scored "adequate" for Direct Response items. * indicates adequate explanations based on alternative assumptions.