Gaze Allocation to the Speaker's Face during Ironic and Sincere Statements: Relations to Features of the Broad Autism Phenotype in Typically Developing Children and Adults

by

Gayle Serlin

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, the City University of New York

2012

This manuscript has been read and accepted for the
Graduate Faculty in Psychology in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Michael Siller, Ph.D.

Date

Chair of Examining Committee

Maureen O'Connor, Ph.D., J.D.

Date

Executive Officer

Michael Siller, Ph.D.
Gary Winkel, Ph.D.
A. Ting Wang, Ph.D.
Thomas Preuss, Ph.D.
Latha Soorya Ph.D.
Paul Cascella, Ph.D.
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

Gaze Allocation to the Speaker's Face during Ironic and Sincere Statements: Relations to Features of the Broad Autism Phenotype in Typically Developing Children and Adults

by

Gayle Serlin

Advisor: Professor Michael Siller

The human ability to navigate our highly social world is reflected in the richness of our interactions with others. Using specialized skills for language and theory of mind, i.e., interpreting the minds of the self and others, we can communicate with others verbally and nonverbally to convey our intentions. During healthy development, these social skills emerge naturally but they can be deficient or absent in individuals diagnosed with autism. The goal of the current project is to examine several aspects of social cognition in 50 typical children ages 3.2-9.3 years old ($M = 6.3$, $SD = 1.8$) and 42 typical adults ages 18-30 ($M = 22.4$, $SD = 4.7$) and to identify autism endophenotypes revealed by the performance of our samples. We administered the Nonliteral Language eye-tracking paradigm to children and adults to ascertain patterns of gaze allocation to expressive faces of speakers delivering ironic, sincere or neutral remarks. Each participant completed an inventory targeting social deficits to determine the presence of broad autism phenotype (BAP), or traits associated with autism that occur in the general population. Children also received a Theory of Mind Battery and a narrative competence task to provide a measure of internal state language. We discovered that when controlling for global development in the correlations between theory of mind performance and measures of narrative competence, some aspects of linguistic complexity remained significantly correlated to theory of mind but internal state language did not. Results concerning all measures revealed that both adults and children allocated more gaze to expressive faces delivering ironic remarks, yet gaze allocation was not predicted by expression of the BAP in either sample. However in adults, greater expression of the BAP predicted less gaze allocation to all expressive faces regardless of the type of remark a speaker made. In children, we found inverse effects for the relation between Theory of Mind Battery and the use of internal state language on gaze allocation, respectively: Decreased performance predicted greater gaze allocation to all expressive faces regardless of the type of remark that was delivered. These contrasting findings in adults and children suggest that the optimal strategies socially impaired individuals utilize may vary across development. The impact of this possibility should be considered in the study of autism endophenotypes related to social behavior.

Acknowledgments

I am delighted to acknowledge Minorities in Biomedical Research Support-Research Initiative for Scientific Enhancement (NIH), the program that awarded me my doctoral fellowship. Thanks to Victoria Luine and Janerie Rodriguez for championing a great cause!

I would like to begin by thanking Michael Siller, my dissertation supervisor, for inviting me into his lab and introducing me to developmental research. Michael's flair for design and analysis are absolutely unmatched and I have become a more precise scientist as a result.

Gary Winkel has been gracious and exceptional in the time and consideration he's given me throughout my dissertation process. The statistical knowledge he has imparted on me is a tremendous asset that I will take with me in my future endeavors. I admire Gary's sharp perspective and vision.

Ting Wang was kind of enough to share her wonderful irony paradigm with me and I certainly got a lot of mileage out of it! She truly helped me flesh things out along the way, especially during my first steps into this project, and I am thankful for her ear and advice.

Paul Cascella and Latha Soorya made themselves readily available to me and were warm and generous with their advice regarding my study and my future plans in the field.

Special thanks to Thomas Preuss, the final addition of my committee. Thomas willingly jumped in without missing a beat and his thoughtful feedback goes a long way.

A few faculty members outside my committee offered their expertise to help me parse through the data. Virginia Valian treated me as though I was one of her students and carefully guided me in my analysis of children's narratives. Without her tutelage, I am sure I would still be formulating a coding scheme! Martin Chodorow graciously shared statistical and theoretical suggestions that allowed me to return to my data with fresh eyes. I would also like to acknowledge Mark Hauber for his remarkable care and concern for my dissertation process.

My diligent research partner Meghan Swanson helped me run a very tight ship from organizing our research team to running efficient assessments of each child. I also found a gem in Kristen Berry, formerly of the Siller lab. Kristen's dedication to sharing all she knows from her vast experience working with children with autism is unparalleled. I am grateful for her patience with my questions and selfless investment in my pursuits and me.

Our lab had no shortage of hardworking students who were thoroughly excited about building this study! Sunita Nair, Valentyna Erstenyuk and Ioan Cozma gracefully worked with Meghan and me as we gained our footing with our first participants. I am especially moved by the contributions of Melinda Cornwell and Matt Edelstein who were vital in the conception and realization of the ever-nebulous Theory of Mind Battery. Leehee Agur was brilliant in paving the way for the tedious coding of children's narratives. Viktoriya Novikova and Manya Jyotishi, two extraordinarily bright honors undergraduates, continued this coding and I have never had the pleasure of working with more committed and intelligent young students. Lastly I would like to thank Frieda Masry for her hand in coding children's Theory of Mind Batteries.

There are a handful of special people in my life who have ensured that I have remained happy and healthy throughout my doctoral journey. My Mom and Dad have always lovingly been there to take my calls, share their wisdom and crack a few jokes. Gomez has been like a brother and unwavering in his support of me. My personal mentor and biggest fan, Ron DeAngelo has taken me under his wing and allowed me to grow at my own pace. I hope to achieve half of his wisdom in my lifetime. Endless thanks to Matthew Kennedy whose love and support makes me a finer and stronger person each day.

Finally, I would like to dedicate my dissertation to my Grandma Shirley. Her light shines in my heart and inspires me to smile through it all and give this world the best ‘me’ I possibly can. I am forever grateful for our time Grammy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

I. General Introduction

General Overview

In this study, we examined developmental milestones as they occur in a sample of 40 typically developing children ages 3.2 to 9.3 years of age (M = 6.3, SD = 1.8). We evaluated children's theory of mind and also considered the influence of two additional developmental markers, expressive and receptive language. Previous studies have focused only on behavioral assessments targeting these abilities but we also incorporated an eye-tracking paradigm that asked participants to interpret nonliteral language and determine a speaker's communicative intent. In order to establish a reference for mature social development, we also assessed a sample of 42 typical adults ages 18-30 years old (M = 22.4 years, SD = 4.7 years) on our eye-tracking paradigm. Using selected inventories, we collected information from all participants concerning their presentation of traits associated with autism.

Many of the current research discussions addressing these developmental milestones are central to the investigation of autism. As such, there is a gap in the research and corresponding body of literature solely dedicated to the examination of how we expect developmental skills to evolve in neurotypical children. One of our goals was to fill in that gap; a preliminary step vital to a thorough understanding of the atypicalities observed in children with autism.

Recent research has been fueled by the search for autism endophenotypes, or identifiable and measurable components of a disorder that extend into the general population and originate from common genetic anomalies (Gottesman & Gould, 2003). The additional goals of the project were to establish clearer links among the broad autism phenotype, global development, theory of mind and visual attunement to cues in the social environment. In light of current research trends, we were interested in discovering if gaze patterns were reflective of expression of the broad autism phenotype as indicated by of our measures, thus suggesting an autism endophenotype.

Typical development of social cognition

Eye gaze is readily observable behavior that has been studied extensively during early development to gauge infants' attention and interest in the world around them. Eye trackers have the technology to collect eye gaze data concerning length of gaze allocation, reaction time and path of gaze to regions of interest across images. Analysis of these data can reveal gaze patterns that tell us how different populations preferentially scan the surrounding social environment.

Previous studies that have investigated typical development using eye-tracking indicate an early biological attunement toward relevant social cues in others such as, facial expression or gaze direction (Boraston & Blakemore, 2007). This is a biological advantage in our world that is extremely dependent on social interaction (Hoehl, Parise, Palumbo, Reid, Handl & Striano, 2009). This selective ability begins immediately as numerous studies have shown that newborns show social orienting, an unlearned a preference for social stimuli (Dawson, Meltzoff, Osterling, Rinaldi & Brown, 1998). Meltzoff and Moore (1983) elicited facial imitation in neonates, an ability contingent on social orienting, despite the presumption that early learning was necessary to condition this behavior. These researchers favor the hypothesis that neonates are driven by a representational system leading them to imitate observed behaviors which creates a framework to unify observed behavior of others with their own motor behavior. Farroni et al. (2002) first demonstrated that in contrast to images of faces with averted gaze, 5-day old neonates preferred faces displaying direct eye gaze that appeared to reciprocate their gaze. As early as 3 months of age, infants can detect someone else's eye gaze in their immediate environment. This triggers unconscious attention shifts as infants' gaze showed a natural tendency to match the direction of another's gaze and serve as precursors to the emergence of joint attention. This lends support to Baron-Cohen's assertion that there is an intrinsic Eye Direction Detector (EDD) that attunes the infant to relevant eye gaze in their immediate field of vision (Hood, Willen & Driver, 1998; Farroni, Csibra, Simion & Johnson, 2002).

Sensitivity for spontaneous gaze-following behavior, in which the infant follows another's line of sight towards an external target object also begins emerging at approximately 3 months. This adaptive behavior is known as triadic communication and reflects an interest in what others are observing and attending to. Triadic communication is more reliably seen in children 6 months and older and is a skill that helps them monitor their world during the first year of life (Striano & Stahl, 2005; Striano, Chen, Cleveland & Bradshaw, 2006).

In typical development, the biological attunement to socially relevant cues may be refined over time. In an eye-tracking study that examined infant's attention to faces during the first year of life, Frank et al. (2009) found that in comparison to 3 or 6-month olds, 9-month olds spent the most time dwelling on the faces of characters versus background objects in an animated movie. However, 3- month olds spent ~50% of viewing time on faces and 6-month olds spent ~60% of time on faces which points toward a trend that grows over time. Adults spent an average of 80% of time on faces leading Frank et al. to hypothesize that the association between faces and their value as a rich source of social information grows over the course of time. Motivation to cue into faces may also become more robust as attention to faces is steadily reinforced throughout their first year (Frank, Vul & Johnson, 2009).

Autism spectrum disorder: A brief history

Parents who have children that are eventually diagnosed with autism are often sensitive to the lack of these early social behaviors that normally emerge in infancy and toddlerhood. They often describe an intuition about their baby's atypical initiative which prompts them to increase soliciting behaviors toward their child (Saint-George, Mahdhaoui, Chetouani, Cassel, Laznik, Apicella, Muratori P., Maestro, Muratori F. & Cohen, 2011). Autism is defined as a pervasive developmental disorder that lies on a continuum comprised of Asperger's syndrome, a mild form of autism and pervasive developmental disorder not otherwise specified (PDD-NOS). Together these conditions are referred to as autism spectrum disorder (ASD). While the DSM-IV-TR (4^{th} ed., 2000) identifies core features that characterize the disorder, the variability of autism's symptoms results in a heterogeneous phenotype (Abrahams & Geschwind, 2008). For an autism diagnosis to occur, the DSM-IV-TR criteria indicate the following (DSM-IV-TR, 4^{th} ed., 2000):

1.) **Difficulty in social interaction** manifested by at least two of the following: impairment in the use of multiple nonverbal behaviors (eye contact/gaze, facial expression, body gestures); lack of development of peer relationships; lack of bids to others in order to share interest/enjoyment and a lack of social and emotional reciprocity.

2.) **Difficulty communicating and using language** as demonstrated by at least one of the following: a delay or lack of language; impairment in the ability to initiate and sustain a conversation with others; stereotyped language and a lack of pretend/spontaneous play.

3.) **Performance of repetitive and stereotyped behaviors, interests or activities** that are displayed by at least one of the following: a preoccupation with one or more stereotyped patterns of interest that is extreme in intensity; an inflexible adherence to nonfunctional routines; repetitive and stereotyped motor movements; persistent preoccupation with parts of objects

Autism symptoms in the general population: The basis for studying endophenotypes

Broad autism phenotype (BAP) is a term that describes the presentation of milder symptoms of autism within the general population. Individuals that display these traits do not meet diagnostic criteria for autism, yet their behavioral tendencies are characteristic of the core deficits of autism spectrum disorder (Hurley, Losh, Parlier, Reznick & Piven, 2007). Research examining individuals with BAP and healthy individuals from the general population is key to the classification of endophenotypes, measurable components of a disorder. Pursuing the endophentype approach is attractive since it deconstructs ASD by concentrating on the genetic liability of discrete symptoms. Thus far, evidence points toward a myriad of genes that converge to cause autism versus a single genetic disruption (Happe, Ronald & Plomin 2006).

In 1977, Folstein and Rutter first suggested the idea that perhaps genetics were responsible for subclinical autism traits seen in nonautistic individuals who were related to someone with a confirmed diagnosis. Later research that focused on monozygotic twins and family studies revealed a greater incidence of core autism symptomatology among these individuals (Bailey, Le Couteur, Gottesman, Bolton, Simonoff, Yuzda and Rutter, 1995; Bolton, Macdonald, Pickles, Rios, Goode, Crowson, Bailey & Rutter, 1994). Using the Social Responsiveness Scale, Constantino (2006) showed that nonautistic siblings of individuals with autism demonstrated impairment in social language, communication and interaction while cognitive and verbal abilities remained intact. This was particularly true for multiplex families; i.e., families that had more than one person diagnosed with autism (Constantino, Lajonchere, Lutz, Gray, Abbachi, McKenna, Singh & Todd, 2006).

At the neural level, fMRI studies have shown differential patterns of activation in the brains of nonautistic siblings of individuals with autism. Dalton et al. (2007) compared teenagers with autism, their siblings and a typical control group on a face-processing task. Participants performed a facial recognition task and were presented pictures of their own family members as well as pictures of other participants' relatives. They were asked to determine if the face was familiar or not by pressing a button. Brain scans indicated that when compared to the control group, both teenagers with autism and their siblings spent significantly less time on the eye region of all faces. No significant differences existed between the autism and sibling group in their gaze allocation to eye regions of faces. Additionally, the autism and sibling samples displayed hypoactivation in the right fusiform gyrus, a primary site of activation during face processing, when viewing all faces. A positive correlation was discovered between eye fixations and activation of the fusiform gyrus implying that scanning strategies and not an abnormal fusiform area were responsible for atypical patterns of activation (Dalton, Nacewicz, Alexander, Davidson, 2007).

Research investigating nonautistic parents of children with autism has highlighted the difficulties they may face when it comes to mentalizing. Baron-Cohen et al. (1997) used the ‘Reading the Mind in the Eyes Test’ to measure participants’ prediction of the mental state of others based solely on an image of the eye region. When compared to mothers and fathers of typical children, parents of children with Asperger’s Syndrome performed significantly worse at this mindreading task (Baron-Cohen & Hammer, 1997). A similar study by Gokcen and colleagues (2009) also found that parents of children with autism experienced difficulty reasoning about the emotional states of others. This was indicated by their performance on the Unexpected Outcomes Test; a 12-item vignette measure that tests an individual’s reasoning about the emotional states of others (Gokcen, Bera, Erermis, Kesikci & Aydin, 2009).

Recently, Hurley et al. (2007) introduced the Broad Autism Phenotype Questionnaire (BAPQ) as an instrument to reliably detect these subclinical features in nonautistic individuals. The BAPQ is a self-report containing 36 statement items that provides subscores across three categories: aloofness, rigidity and pragmatic language. Results from validity testing of the BAPQ indicated that the BAPQ was equally specific for signaling the absence of BAP in parents who were clinically assessed and deemed negative for BAP. Hurley and colleagues suggest that in the absence of a clinical assessment, the BAPQ is an excellent way to determine an individual's status regarding the broader phenotype.

Theory of mind and traditional behavioral assessments

A seminal paper by Premack and Woodruff (1978) first described theory of mind as an individual's ability to 'impute mental states to him/herself and others'. They tested this skill in an adult female chimpanzee by presenting her with a videotape of an actor who was caught in a sequence of predicaments such as, not being able to reach a food source and being locked in a cage. After viewing this sequence, the chimp was presented with various solutions to each of these dilemmas in the form of pictures depicting possible actions toward resolution. The chimp successfully chose the correct solution for each predicament. This led Premack and Woodruff to conclude that their subject correctly identified the situations as problems for the actor who presumably needed solutions to achieve his purposes (Premack & Woodruff, 1978).

This research was soon followed by investigations of theory of mind in children. Do they understand that others have separate perspectives and can they attribute belief, desires, intentions and emotions to others (Astington, 1998)? To test this, Wimmer et al. (1983) devised the classic false belief task that targets children's knowledge about the perspectives and beliefs of others. In this scenario, Mother returns home from shopping for groceries and her son Maxi helps put away the groceries. Maxi asks where he should put the chocolate and Mother tells him to put it in the blue cupboard. After Maxi does this, he goes outside to play. While Maxi is gone, Mother takes the chocolate out of the blue cabinet, uses some for a recipe and then returns it to the green cabinet. Mother leaves to run an errand and Maxi returns for a chocolate snack. Participants are asked, ‘Where will Maxi look for the chocolate?’. Theory of mind is traditionally tested using this task as: (1) performance illustrates the degree to which a child can maintain basic representations of someone else’s mind, (2) it measures an individual’s understanding that others can have perspectives that do not match reality and (3) it measures an individual’s capacity to simultaneously represent competing perspectives (Wimmer & Perner, 1983; Frith & Frith, 2006; Saxe & Baron-Cohen, 2006). A meta-analysis of 178 studies has shown that the preponderance of empirical evidence agrees that during the preschool years, significant developmental change occurs and by the age of 6, most typical children across various cultures pass first-order false belief tests (Wellman, Cross & Watson, 2001).

Accurately characterizing an individual’s knowledge about another’s knowledge is termed second-order false belief. Perner and Wimmer (1985) demonstrated that by the age of 6-7, most children succeed at second-order false belief tasks. Wimmer and Perner (1983) also tested children on more complex theory of mind skills including a deceptive task to see if children could identify the deceptive action in a story and then justify why they did or did not believe a character was acting spitefully. By the age of 4.5, children were not consistently able to recognize and justify deception in contrast to 5.5-year olds who were able to do so. Additionally, in measures that ask for moral judgments concerning the behavior of others, 4-year olds were able to identify the flawed morality behind intentional lying. However, only 9 and 10-year olds were able to distinguish nuances in more challenging tasks that ask children to make moral judgments about breaking commitment due to either intentional or unavoidable circumstances (Mant & Perner, 1988). An investigation of children’s understanding of figurative language yielded a similar pattern. When presented with different types of figurative language (i.e., sarcasm, metaphor, understatement, hyperbole and irony), 8-year olds but not 6-year olds were able identify the discrepancies in the statements or the speaker's communicative intention. In their attempts to describe the purpose of these statements, 8-year olds performed at chance. Eleven year olds, however, succeeded at identifying discrepancies and explaining them (Demorest, Silberstein, Gardner & Winner, 1983). Thus, growing abilities for theory of mind may yield increased performance on ToM tasks which may positively correlate with visual attunement to social cues that help children mentalize about others.

Measures of theory of mind have informed developmental psychology by revealing the degree to which a child is attuned to the social world around him/her. As highly social creatures, we are constantly regulating our interactions by performing online interpretations about the perspectives of others. This allows for opportunities to reason about the meaning behind their behavior (Baron-Cohen, Wheelwright & Joliffe., 1997). For example, Baron-Cohen (2000) reported that cueing into someone else's gaze direction can let us know something as simple as the fact that someone is in thought if they are looking up and away. Following someone's gaze can also provide more complex information such as, which of several objects a person wants (Baron-Cohen, 2000). Another aspect of theory of mind includes our ability for detecting and decoding affect in others. Rochat and Striano (1999) suggested that this attunement begins with infants' attention to the primary caretaker's face that acts as a 'primordial theater' staging emotional and intentional cues used to monitor others and predict their behavior.

Narrative as a tool to measure theory of mind

In order to adopt characters' perspectives and relate them to a listener in an engaging way, theory of mind is believed to be integral to narrative discourse (Tager-Flusberg, 2000). Cross-culturally, narrative ability has been shown to appear in early development. Storytelling serves the unifying purpose of communicating experiences. In doing so, the narrator is challenged with the simultaneous demands of providing order to otherwise disjointed events and conveying a sense of continuity among different points in time. Within this temporal framework, the storyteller must offer a causal schema for the events being reported. Through the use of numerous linguistic devices, the narrator can achieve this and deliver a coherent story that maintains listener interest. Traditionally, personal narratives have served the purpose of revealing the self by inviting the listener into the narrator's subjective world (Ochs & Capps, 1996; Labov, 2006). Storybook narratives can do this by offering insight into the narrator's mind by way of their attributions of mental and emotional states to the story's characters (Losh & Capps, 2003). To tell a good story, the storyteller uses a hierarchical organizational structure to weave together these internal aspects of the characters across a sequence of events. To measure how sensitive a storyteller is to elements that contribute to a complete narrative, linguistic and pragmatic devices have been identified and quantified in comprehensive coding schemes. Measuring these devices can illustrate the narrator's interpretation of events and ability to build the suspense around the story's point (Goffman, 1974).

The research investigating neurotypical children on storybook narrative measures mostly exists within the context of comparative samples, for example, autism or speech impaired children (Tager-Flusberg & Sullivan, 1995; Capps et al., 2000; Losh et al., 2003). Results of these studies vary and the age range of children exceeds the developmental period that the current study will investigate however, these studies mark notable contributions to the field.

In an experiment that used very similar behavioral measures as the current study, Capps et al. (2000) recruited 13 children with autism (Mean age = 12.6 years) and developmental delay (DD) (Mean age = 9.8 years), respectively, and compared them to 13 typical children (Mean age = 6 years) on theory of mind and narrative measures. Results showed that both the ASD and DD groups were less likely than typical children to use complex syntax; a device that facilitates the description of linear events and ties together a series of episodes leading to a well-integrated story. In children with autism, the relationship between ToM and syntactic diversity was mediated by language skill. Children with autism also showed positive correlation between ToM and their use of mental state language which lends further support to the role ToM plays in identifying the cognitive state of another individual. Surprisingly, a negative correlation between ToM and the use of emotional state language was found for children with autism. Capps and colleagues explained that this might be due to a compensatory strategy of 'emotion-labeling' that children with fewer theory of mind abilities used to sustain their narratives (Capps et al., 2000).

In a related study comparing 28 high functioning autism/Asperger's adolescents (Mean age = 11.3 years) and 22 typical adolescents (Mean age = 10.6 years), Losh et al. (2003) examined the relationship between theory of mind and internal state language, i.e., cognitive and emotional language. Language and theory of mind tasks were administered and emotional competence was determined using the Berkeley Empathy Measure (labeling a child protagonist's emotion on a video clip) and an emotion definition task. In children with ASD, Losh et al. found no correlation between the use of internal state language and a theory of mind measure. Alternatively, they found a positive correlation with this group's ability to define 12 emotions and: (1) theory of mind, (2) internal state language and (3) complex syntax. These correlations were not found in the typical sample. Children with autism also lacked causal explanations for the events they described in their storybook narratives. Additionally, verbal ability was also not significantly associated with measures of narrative, theory or mind or emotional understanding. Losh and colleagues suggested that in a high functioning, verbally fluent autism sample, emotional competence may be a better predictor of narrative proficiency than theory of mind or verbal ability. They also theorized that younger, less developmentally mature samples might demonstrate different associations among the behavioral variables measured (Losh et al., 2003).

Theory of mind: Relationships to chronological age and developmental age (language)

When considering children's performance on these behavioral tasks, chronological age must be considered as a potential explanatory variable. Intuitively, it might seem that performance on a social cognitive task may be approximated by chronological age since a child at any age will be presumed to have reached a certain set of milestones ascribed to their current age. Children's achievement of milestones begins with arousal and attention; natural mechanisms that appear in infancy to guide learning through social interaction. These mechanisms orient children toward referential information in the environment, e.g., facial expression, pointing of a speaker. Attunement to these cues coincides with the acquisition of social abilities such that the development of these abilities has been reliably mapped onto specific age ranges (Rochat & Striano, 1999; Kuhl, 2007; Striano & Reid, 2006,). As discussed earlier, it appears to predict the performance of typically developing children with respect to theory of mind milestones.

It is also plausible that a third variable acts as a moderator between cognitive performance and chronological age. Language ability, a critical aspect of developmental age, may underlie children's individual abilities and has been considered a marker for performance in behavioral measures. Within the relevant literature, there is no shortage of dialogue concerning the role language plays in children's abilities to represent the world and how essential it is to mentalizing. In a study that assessed 3 to 5-year olds on standard false belief tests and multiple language measures, Jenkins and Astington (1996) found high correlations between ToM task performance and language competence. Kleinknecht et al. (2004) also lend support to this theory as they found positive correlations between evaluative linguistic devices and narratives as told by preschoolers ages 3 to 5. It makes intuitive sense to look for a link between language and ToM since they are both acquired in the same period in early development (Gopnik, 1990).

J.W. Astington has raised important issues concerning the nature of the relationship that exists between language and theory of mind. She proposed that narrative is a fundamental mediator between language and ToM but has noted that it remains unclear whether theory of mind is scaffolded by narrative or if theory of mind is a prerequisite for the formation of narratives. Additionally, Astington asserted that metacognition is central to theory of mind in that it is the purest representation of truly knowing as it incorporates knowledge about what is known by the individual as well as others. For preschoolers, theory of mind is implicit as it is encompassed within their social understanding (Astington, 1990). In a later study, Astington et al. (1997) discovered that as children's competence with metacognitive verbs (think, know, believe) shows improvement, so does the tendency to coherently link action to consciousness in their narratives. This competence also increased in proportion to children's performance on false belief tasks. In light of these findings, Astington et al. suggested that for preschoolers, metacognitive language is the vehicle that makes theory of mind explicit (Astington & Pelletier, 1997). Astington and colleagues (1999) performed a longitudinal study of 59 neurotypical 3-year olds over a span of seven months to investigate direction of the relationship between theory of mind and language. A second research aim was to determine the contributions of syntactic and semantic abilities to this relationship. Children were assessed three times during their third year on a battery of basic theory of mind tasks and their expressive and receptive language skills to determine their proficiency for syntax and semantics. It was found that language ability predicted later theory of mind performance but theory of mind performance did not predict later language test performance. On account of this, Astington et al. argued for a strong version of the hypothesis that theory of mind initially depends on language. In their view, syntax in particular allows the child to symbolically track the details of a false belief task and to mentally represent a character's perspective of their situation (Astington et al., 1999).

These theories are given further support but Hughes et al. (1997) who observed 50 preschoolers (Mean age = 47 months) while they were playing with a friend for a period of 20 minutes. Children received several basic theory of mind tasks, verbal ability was assessed and children's mean length utterance during play was quantified. After coding videos of the dyads at play, a significant positive correlation emerged between rates of pretend play and frequency of mental state language. A significant positive correlation was also discovered between level of performance on ToM tasks and the (1) frequency of mental state talk and (2) frequency of pretend play. Lastly, Hughes et al. reported that individual differences in verbal ability were only positively correlated with frequency of mental state talk. This leads to speculation that verbal ability directly subserves mental state talk which is dependent on adeptness with language. Thus, verbal ability as indicated by lexical knowledge may be an underlying mechanism that facilitates theory of mind and pretend play by supporting mental state talk. It is important to note that Hughes et al. quantified mental state talk by coding children's references to thoughts, beliefs and metacognitive language (e.g., I *think*, You *know*) and did not consider emotional language (Hughes & Dunn, 1997). The importance of metacognitive language is echoed in earlier work by Olson (1988) who has argued that the development of theory of mind is contingent upon the child's fluency with such terms as *think, know* and *remember*. Astington et al. (1999) hypothesized that the use of these linguistic terms provides a framework for children to map their subjective feelings onto the experiences of others to internalize the experiences of others.

The Nonliteral Language eye-tracking paradigm and appreciation of irony

The Nonliteral Language eye-tracking paradigm draws on theory of mind, language and attunement to social cues. It was conceived by Wang and colleagues to examine the neural correlates underlying individuals' perception and interpretation of others using potentially ironic cartoons (Wang, Lee, Sigman & Dapretto, 2007). Eighteen typical boys (Mean age = 11.8 years) and 18 boys with autism spectrum disorder (Mean age = 12.5 years) participated in this study. Participants viewed audiovisual social stories that ended with a speaker delivering either an ironic or sincere statement. Speaking characters wore facial expressions that matched the tone of their comment. Participants received these vignettes according to three conditions of instructions: no instructions (neutral condition); pay attention to face; and pay attention to the voice. Using fMRI, Wang and colleagues measured brain activity as participants interpreted the speakers' communicative intent. In the neutral condition during ironic scenarios, boys with autism showed less selective activation in the medial prefrontal cortex (mPFC) when compared to the typical sample. That is, they did not recruit this additional brain region while interpreting ironic scenarios versus unambiguous, sincere scenarios. This suggested that they were not spontaneously integrating multiple salient cues including facial expression and tone of voice (Wang, Lee, Sigman & Dapretto, 2006). Subsequently, Wang et al. (2007) discovered that when children with ASD received instructions to attend to the speakers' facial expression or tone of voice, activity was elicited in the mPFC and group differences disappeared. This is the first study to demonstrate that explicit instructions to attend to salient social stimuli can normalize hypoactivated centers of the brain in an autism sample. Using an adapted version of this cartoon paradigm, we aimed to highlight the salience of emotionally expressive faces by demonstrating that they are targets of social information when determining a speaker's intention.

Gaze allocation patterns to such nonverbal cues may change over time which may reflect the development of cognitive abilities. For example, it has been shown that the capacity to understand nonliteral language such as sarcasm and irony emerges around the age of 5 or 6 for typical children (Ackerman, 1981; Dews, Winner, Kaplan, Rosenblatt, Hunt, Lim, McGovern, Qualtier & Smarsh, 1996; Glenwright & Pexman, 2010). While researchers have hypothesized that at the age of 5, children are able to determine that a remark is nonliteral they are not able to simultaneously grasp an ironic speaker's *actual* belief and the listener's belief that the speaker does not mean exactly what they say (Dews et al., 1996). This reflects an inability to appreciate the discrepancy between the speaker's choice of words and cues such as prosody or tone that would lead to a full understanding of the speaker's communicative intent (Dews et al., 1996; Morton & Treghub, 2001; Filippova & Astington, 2008). Results of converging studies have revealed that even by the age of 9-10, children are able to identify subtle nuances that distinguish sarcasm (a mean utterance directed at a target) and irony (a non-directed mean utterance) yet, they are not consistently able to provide complete justifications of their choices (Glenwright et al., 2010). These findings speak to the notion that full understanding of nonliteral language is a developmental process even though the basic building blocks are in place by an early age.

In contrast, developmentally mature adults possess the ability to juxtapose language and paralinguistic cues which allows them to express their true attitudes about people or situations using indirect modes of communication (Creusere, 1999). Their understanding of the incongruence between their word choice and the prosody or tone of their delivery affords adults the opportunity to use irony for purposes of 'mean' humor, muted criticism and signaling annoyance or superiority (Dews et al., 1996; Creusere, 1999). To investigate the approach to the problem of verbal irony in typical children 5 to 8 years old versus adults, Climie et al. (2008)

presented participants with a puppet show featuring a character that delivered either ironic remarks or literal remarks. Participants were asked to judge the speaker's intent while response latency and direction of eye gaze during the task were measured. Results revealed that adults outperformed children in determining a speaker's meaning and when children accurately determined ironic remarks, they took longer to respond than adults. Additionally, eye gaze results signified that during ironic trials, children appropriately referenced the 'ironic' versus the 'literal' response object. This result demonstrates that when faced with irony, children were not attempting literal first processing, rather, their system of coordination toward understanding nonliteral language was slower due to inexperience (Climie & Pexman, 2008).

Contrary to these findings, Demorest et al. (1984) demonstrated an opposite effect in a study that addressed the verbal irony dilemma. A sample of 6-year old children received different versions of a story about a boy's reaction to his friend's new haircut. The boy's reaction was either sincere, deceptive, sarcastic or neutral. Across conditions, children's responses to the boy's reaction revealed that 6-year olds were inclined to take the boy's words literally and did not seem to recognize the deliberate falsehood and its possible purpose. Performance on the present eye-tracking task may clarify the discrepancies between these results.

The current study is comprised of three experiments that examined social cognition in typical children and developmentally mature typical adults. Although each experiment considers a discrete set of research questions, they draw on the same concepts: theory of mind, language and visual attunement to the social world.

Aims and Hypotheses

ADULTS:

AIM 1: To validate the use of the Nonliteral Language paradigm on an eye-tracker by determining that viable gaze data can be collected from a neurotypical adult sample ages 18-30.

> **HYPOTHESIS:** It is predicted that across experimental conditions, a majority of adults will be on-task, i.e., and attending to the stimuli (AOI plus background screen) longer than 50% of the 3.5 second window during a character's potentially ironic statement.

AIM 2: To determine adults' gaze allocation patterns to faces across conditions on the Nonliteral Language paradigm. To determine the relations between adults' gaze allocation patterns on the Nonliteral Language paradigm and their performance on the Broad Autism Phenotype Questionnaire (BAPQ).

> **HYPOTHESIS:** The association between gaze allocation and behavioral measures is stronger in the ironic condition than the neutral and sincere condition. Adults' BAPQ scores will predict gaze allocation patterns such that a negative correlation exists between subscale score(s) and/or total average score and gaze allocation to ironic faces.

CHILDREN:

AIM 3: To determine if there is a correlation between performance on the ToM Battery and the use of internal state language during a spontaneous storybook narrative in a sample of neurotypical children ages 3.2 to 9.3 years.

> **HYPOTHESIS:** A positive correlation exists between performance on the ToM Battery and linguistic complexity, the use internal state language and range of evaluative devices used during a storybook narrative. These relationships will remain significant after controlling for chronological age and developmental age.

AIM 4: To determine children's gaze allocation patterns upon faces across conditions in the Nonliteral Language paradigm. To determine the relation between gaze allocation patterns on the Nonliteral Language paradigm and performance on the Social Reciprocity Scale (SRS) in a sample of neurotypical children ages 3.2 to 9.3 years.

HYPOTHESIS: The association between gaze allocation and behavioral measures is stronger in the ironic condition than the neutral and sincere condition. Children's SRS scores will predict gaze allocation patterns such that a negative correlation exists between subscale score(s) and/or total average score and gaze allocation to ironic faces.

AIM 5: To determine the relations between children's performance on the Theory of Mind Battery (ToM) and their gaze allocation patterns during the Nonliteral Language Paradigm. To determine the relations between children's use of internal state language during a spontaneous storybook narrative and their gaze allocation patterns.

HYPOTHESIS: Performance on behavioral measures that evoke ToM predict gaze allocation to faces in ironic scenarios such that a positive correlation exists between ToM Battery and use of internal state language, respectively, with gaze allocation to ironic faces.

II. Visual Processing of Social Information and Expression of the Broad Autism Phenotype in Typical Adults

Abstract

Recent research investigating the etiology of autism has focused on the identification of autism endophenotypes to narrow the search for the genetic liability of discrete symptoms. Impairments characteristic of autism have been observed in the general population. In the current study, we assessed neurotypical adults on the Nonliteral Language eye-tracking paradigm and the Broad Autism Phenotype Questionnaire to determine the relation between gaze allocation patterns to expressive faces, a socially salient cue, and presentation of characteristics associated with autism. When asked to determine a speaker's communicative intent, we found that adults spent more time gazing upon faces of speakers that delivered ironic remarks versus faces of sincere or neutral speakers. Secondly, we found that individuals who showed higher expression of traits associated with autism spent less time gazing at all faces regardless of the type of remark a speaker delivered. Possible reasons for this global finding are discussed. Greater expression of autism-like traits in typical adults may suggest an endophenotype characterized by a strategy of less gaze allocation to a range of salient expressive faces in social situations.

Introduction

Eye gaze patterns upon relevant social information can give clues about an individual's attunement to the external world. Atypical gaze allocation to social cues has been reported in Autism spectrum disorder (ASD) and is an example of a deficit that can segregate into the general population. This is known as an endophenotype, or a 'measurable component of a disorder unseen by the unaided eye' that exists midstream between the illness and its genotype. While ASD endophenotypes can occur in any undiagnosed individual, the endophenotype found in affected family members is found in unaffected relatives at a higher rate than in the general population (Gottesman et al., 2003). Often, these family members are symptomatic of the broad autism phenotype (BAP) which is a subclinical condition related to ASD characterized by the presentation of subtler symptoms of autism (Hurley et al., 2007). Eye-tracking is a tool that allows us to quantify this otherwise unobservable gaze phenomenon and using this technology, the current study explores abnormal gaze allocation upon social material as an endophenotype in a community sample.

Social cognitive deficits in ASD are often observable hallmarks of the disorder and abnormal patterns of eye gaze in response to social stimuli have been well documented in individuals with ASD (Boraston et al., 2007). In a notable eye-tracking study, Klin et al. (2002) used clips from *Who's Afraid of Virginia Woolf*, an emotional and dramatic movie, to compare gaze patterns between adolescent and young adult neurotypical males and males with autism. Significant differences were discovered for fixation time as the autism group spent twice as much time gazing upon actors' mouth and body regions, two times more on the object region and two times less on the eye region relative to the typical sample. Further analysis revealed that percentages of fixation time on the mouth and object regions, but not the eyes, were the best predictors of higher levels of social competence. This intriguing finding might point to the disproportional significance that the mouth has for individuals and a compensatory strategy for understanding the social world (Klin, Jones, Schultz, Volkmar & Cohen, 2002).

Abilities for pragmatic language, the appropriate social use of language, are compromised in individuals with autism and in conversation they struggle to operate in ways that acknowledge the needs of the listener. For example, they may violate expectations of reciprocal dialogue and be unable to offer relevant information (Tager-Flusberg, 2000). Irony is a universal communicative device used to assert beliefs that contrast with the literal meaning of the speaker's chosen words. This type of nuanced language can easily be lost on individuals with autism (Creusere, 1999). This is likely due to a lack of intuition that communication is predicated on intended versus literal meaning (Tager-Flusberg, 2000). These problems exist even for high-functioning (HF) people with autism despite their mastery of advanced language skills (Young, Diehl, Morris, Hyman & Bennetto, 2005). A study of 37 HF children with autism showed that performance on verbal and nonverbal measures explained a majority of the variance associated with children's scores on the Test of Pragmatic Language. However, 30% of the variance remained unexplained by cognitive nonverbal ability or structural language skills. While formal language abilities may strongly relate to how a speaker functions socially, it does not unilaterally determine competence for pragmatic language (Volden, Coolican, Garon, White & Bryson, 2009).

Cross-sectional research has supported a higher incidence of BAP in relatives of individuals with ASD. Using the Social Responsiveness Scale, Constantino (2005) showed that neurotypical siblings of individuals with autism demonstrated impairment in social language, communication and interaction while cognitive and verbal abilities remained intact. This was particularly true for multiplex families; i.e., families that had more than one person diagnosed with autism. This was followed by the development of the Broad Autism Phenotype Questionnaire (BAPQ) designed to detect autism-like traits in the neurotypical population. During extensive clinical assessments, parents of children with autism were labeled as either BAP positive or negative. Across all three subscales and total score, BAP positive parents were found to have significantly higher scores than BAP negative and control parents (Hurley et al., 2007). The nature of the link between BAP and Autism spectrum disorder has been rigorously explored and continues to be elucidated through family studies.

At the neural level, fMRI studies have shown differential patterns of activation in the brains of nonautistic siblings of individuals with autism when engaged in a visual task. Dalton et al. (2007) compared teenagers with autism and their unaffected siblings on a face recognition task. Brain scans indicated that when compared to the control group, teenagers with autism and their siblings spent significantly less time on the eye region of all faces when determining if the face was familiar or not. No gaze pattern differences existed between the autism and sibling groups. The autism and sibling samples also displayed hypoactivation in the right fusiform gyrus, the face processing area, when viewing all faces. Belmonte et al. (2010) extended these findings in their study assessing boys with ASD and their unaffected brothers. Gratings of various colors, orientations and locations were used in their visual attention task that asked participants to identify target stimuli. The autism sample was the least accurate in this behavioral component of the task followed by the sib sample while the controls were the most accurate. fMRI results revealed that controls showed prefrontal cortical activation and a pervasive network of neural activation, while children with autism showed hypoactivated prefrontal cortices and impoverished neural networks. In contrast, siblings showed typical neural networks but abnormally low prefrontal cortical activation (Belmonte, Gomot & Baron-Cohen, 2010). These results point to the existence of familial endophenotypes in ASD (Wallace, Sebastian, Pellicano, Parr & Bailey, 2010).

The broad search for ASD biomarkers has been confounded by the phenotypic heterogeneity it presents and the associated genetic variability. By focusing on smaller more identifiable components of the disorder, it may be easier to trace and map links between ASD's plethora of candidate genes and its core symptoms (Kaiser, Hudac, Shultz, Lee, Cheung, Berken, Deen, Pitskel, Sugrue, Voos, Saulnier, Ventola, Wolf, Klin, Vander Wyk & Pelphrey, 2010). As basic constituents of psychological disorders, endophenotypes are also believed to more strongly correspond to genetic underpinnings, an implicit advantage in adopting this approach. Moreover, results from Ronald and colleagues found that the triad of core deficits in autism are likely fractionable in the general population. By definition, these symptoms in individuals diagnosed with autism cannot be teased apart. However, their population based studies that analyzed data from more than 3,000 MZ and DZ child twins revealed relatively low correlations between autism characteristics from the three core areas. Using model fitting analyses cross-twin and cross-trait correlations, they determined that more than half the genes responsible for the variation in social skills, for example, are independent of genes that contribute to variation in autism's other domains. Children tended to show impairments in one area of the autism triad as 59% of children displayed social deficits and their difficulties were confined to this one domain (Happe et al., 2006; Ronald, Happe, Price, Baron-Cohen & Plomin, 2006; Ronald, Happe, Bolton, Butcher, Price, Wheelwright, Baron-Cohen & Plomin, 2006). These discoveries offer convincing evidence that each core deficit can be disentangled from the autism symptom triad and can reasonably be examined independently.

fMRI research by Wang et al. (2007) investigated neural responses in HF adolescents with autism by presenting participants with a battery of social vignettes and asking them determine whether a speaker's communicative intent was sincere or ironic. In contrast to typical participants, HF adolescents showed decreased activity in the superior temporal gyrus and medial prefrontal cortex when considering ironic scenarios. Neural activation differences between groups disappeared when HF adolescents were explicitly instructed to pay attention to the speakers' voice or face. Our paradigm adapted Wang et al.'s stimuli and aimed to highlight differences in the visual strategies used by 42 neurotypical adults with varying levels of social aptitude as indicated by the BAPQ. The relationship between participants' BAPQ results and their gaze patterns as they determined potentially ironic communicative intent were evaluated for performance that might suggest the presence of an endophenotype. We hypothesized that adults would allocate the greatest amount of gaze to faces of ironic speakers. We also expected to find an interaction between gaze allocation and BAPQ score such that adults with higher BAPQ scores would spend less time gazing upon ironic faces to determine communicative intent.

Method

Participants

Our sample included 42 neurotypical adults comprised of 31 males and 11 females. Participants were between 18-30 years old (*M* = 22.4, *SD* = 4.7). A majority of participants were recruited using Sona Systems database at Hunter College that targets Psychology 100 students who are required to participate in research as a course requirement. Other participants including friends and colleagues were recruited word-of-mouth.

All recruitment procedures were approved by the Hunter College Institutional Review Board.

All participants were free of cognitive, hearing and visual impairment. Adults who required corrective prescription lenses were included as Tobii hardware is designed to track eye gaze through glasses.

Materials

Informed Consent & Background Questionnaire

Adults reviewed and completed a consent form that described materials, procedures and explanations of participant rights and resources. Hunter College students participating for course requirements acknowledged they would receive research credits. Other participants acknowledged they would receive $10. Consent was obtained from all participants according to Hunter College Institutional Review Board protocol.

The Background Questionnaire developed by our lab contains eight items and took inventory of participants' basic personal and demographic information such as race, current living location and education level.

Broad Autism Phenotype Questionnaire

All adults completed the Broad Autism Phenotype Questionnaire (BAPQ; Hurley et al., 2007) which is a 36-item inventory used to assess the presence of traits characteristic of autism expressed by non-autistic individuals. Questionnaires were labeled 'Personality Styles and Preferences Questionnaire' to avoid influencing responses. Participants responded to each item using a Likert scale ranging from one to six (1 = very rarely and 6 = very often). Twelve items were dedicated to each subscale: aloofness, rigidity and pragmatic language, and the average for all items yielded the total average score. Sensitivity and specificity percentages are 70% or above for all subscales (including total average) and above 80% in two out of three subscales. Scoring above cutoff on two out of three subscales qualified an individual as expressing BAP; eight people met this criterion. In the rigid, aloof and pragmatic languages subscales, nine, eight and

19 participants scored above cutoff, respectively. Six individuals scored above threshold on total average score. Pearson's bivariate correlations did not show any significant trends between age or gender and subscale scores or total average score of the BAPQ.

Eye-Tracking Stimuli

Audiovisual black and white cartoon scenarios were adapted from Wang et al. (2007) and modified for presentation on an eye-tracker. These static stimuli were divided into three conditions: 12 scenarios in the neutral condition, 10 scenarios in the ironic condition and 10 in the sincere condition. All scenarios were narrated by a female speaker. Neutral scenarios ended with a speaker making a neutral comment that lacked emotional valence (see Figures 2.1 and 2.2). Ironic scenarios concluded with a speaker delivering a remark in a sarcastic tone and sincere scenarios concluded with a speaker delivering a remark in a genuine tone (see Figures 3 and 4). In each condition, speakers had an emotionally expressive face as they made their remarks. At the end of each vignette, participants were asked to determine if the speaker meant what they said. Wang and colleagues determined that characters' remarks were perceived as sincere or ironic by analyzing ratings from 12 blind coders. These adult volunteers were only exposed to the audio statements and were asked to rate remarks on a scale of one to seven (1 = ironic and 7 = sincere). Mean ratings were 1.4 for ironic remarks (SD = 0.7) and 6.6 (SD = 0.7) for sincere statements. Wang et al. matched sincere and ironic remarks on the following: syntactic structure, semantic complexity and length (Wang et al., 2007).

To determine the effect of various emotional expressions of characters' faces on participants' gaze allocation, we obtained face ratings from 41 Psychology 100 students at Hunter College who were blind to the paradigm. Students rated the degree of emotional expressivity on the faces of all characters (32) that participants would view when asked to judge communicative intent. Students viewed the faces on slides (Microsoft Office PowerPoint 2007) presented by the instructor. Faces were shown for 10s. and students scored them from one to nine (1 = extremely negative expression; 5 = neutral and 9 = extremely positive expression). Mean emotional expressivity score was 5.22 ($SD = .25$). Independent raters' scores were highly reliable: $g = .989$. Examples of our ratings can be seen in Figure 2.1 below.

2.05 **5.61** **7.95**

Figure 2.1. Examples of mean face ratings according to emotional expressivity.

Faces were designated as 'Areas of Interest' (AOIs) and we analyzed the amount of time participants allocated their gaze to faces. However, the size of characters' faces differed within the stimulus set and as a result, AOI size varied accordingly. AOIs included a character's entire face plus two centimeters added to the perimeter of the face from any given point. Additional space was added in accordance with Tobii technicians' recommendations to account for possible differences in actual gaze fixation and gaze fixation as recorded by Tobii Studio. Across all faces, mean AOI diameter = 4.53cm (*MIN* = 2.92cm, *MAX* = 6.13cm).

Procedures

All tasks were administered at the Communication and Play Lab at Hunter College, NY. Lab visits began by obtaining informed consent. Participants then completed the Background

Questionnaire and the BAPQ in a quiet assessment room. They alerted the examiner when they were finished and then the eye-tracking paradigm was administered.

Eye-tracking Paradigm

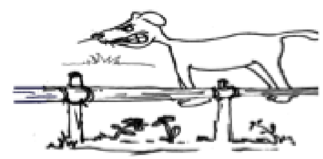
Examiners introduced the task with the following instructions:

> *Sometimes when people say things, they mean what they say and sometimes they're just being silly, mean or sarcastic. You're going to look at some pictures and listen to a short story and decide whether or not the person in the story means what they say. When the narrator asks you if the person meant what they said, simply say 'yes' or 'no' out loud and we'll record your answer.*

For the sincere and ironic conditions, two different versions of 10 black and white cartoon scenarios were presented. All contained three static images and were delivered according to the same structure: The first image introduced two characters and an activity they were engaged in. In the second image, an event occurred that resulted in an explicitly negative or positive outcome (e.g., a dog growls at two friends/a dog wags its tail at two friends). Characters' faces were not shown in the second image. The third image depicted only the second character's face and verbal response to the un/desirable outcome with a comment such as, 'He's a friendly one!'. In the nonliteral version of the story, the intonation of the comment was snarky and ironic. In the sincere version, the intonation was genuine and heartfelt. If the speaker made an ironic statement, he/she wore a look of disgust or disdain but if the speaker was being sincere, he/she smiled and appeared warm and complementary. After the final remark, the narrator asked the participant, 'Did X mean what they said?'. This last image featured on the second character's face and remained on the monitor for 8.5 seconds (see Figures 2.2 and 2.3).

Four practice scenarios, two ironic and two sincere, were administered to each participant to ensure that they could distinguish between ironic and sincere scenarios. Every adult accurately identified communicative intent during practice and subsequently received the entire paradigm.

Introduction: "Barbara and Julie are taking a walk."

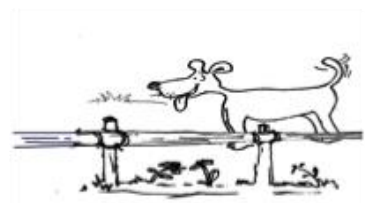


Event: "They pass a dog behind a fence that growls and bares it's teeth at them."

Ironic Statement: "Julie says, 'He's a friendly one!'"

Figures 2.2 and 2.3. Ironic and sincere versions of 'Dog' vignette.

Introduction: "Barbara and Julie are taking a walk."



Event: "They pass a dog behind a fence that barks and wags its tail at them."

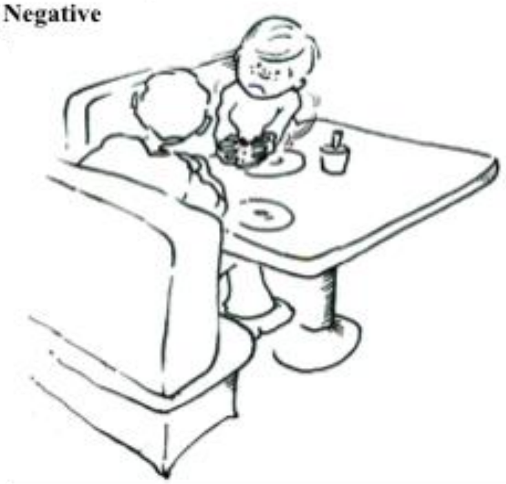
Sincere Statement: "Julie says, 'He's a friendly one!'"

Participants received both versions of each scenario. Versions were split between two experimental blocks so that both versions of a scenario never occurred in the same block. Within each block, scenarios were randomized. For every two to three vignettes shown, a rest screen featuring a popular cartoon was displayed for four to five seconds.

Introduction: "Jacob and his dad eating lunch."

A. Positive

B. Negative



Neutral Statement: "When Jacob can't eat anymore he says, 'Dad, I'm full.'"

Figure 2.4. Negative and positive versions of 'Lunch' vignette from the neutral condition.

Vignettes in the neutral condition contained only two images; the first was the introduction image depicting two characters engaged in an activity (e.g., eating lunch). The second portrayed a continuation of the activity as one character made a remark neutral in tone and prosody (e.g., ‘Dad, I’m full.’). There were two versions of six scenarios and in each, only one character’s face and verbal response were depicted. In one version, the character wore a face expressing some degree of happiness or joy and in the second, the character wore a face expressing some degree of anger or sadness (see Figure 2.4). Participants were asked ‘Did X mean what they said?’. The last image was remained on the monitor for a total of 13 seconds and participants responded in the manner described above. Neutral scenarios were divided between two blocks with each block containing one version of each scenario. For every two vignettes shown, a rest screen featuring a popular cartoon was shown for four to five seconds.

Blocks containing ironic-sincere scenarios (two blocks of ten) alternated with the presentation of neutral blocks (two blocks of six) on a rotating schedule across participants (e.g., *ABCD*, *BCDA*, *CDAB*, etc.,). For each condition, participants had approximately four seconds to respond to the question. If they did not respond within this time, E-Prime was programmed to continue to the next scenario. A lack of response was recorded as missing data.

Eye-tracker

Eye-tracking tasks were delivered using a Tobii T60 eye tracker. Participants were seated 60-70cm from the eye-tracking monitor and calibrated using five-point calibration; i.e., following a red ball that bounces around the screen and stops at five discrete points on the screen. The T60 has with a 17” screen and a data rate of 60 Hz. It is a noninvasive machine that emits infrared light from a horizontal strip of sensors located above and below the monitor. Patterns of infrared light reflect off of the participant’s corneas onto the eye tracker’s image which are interpreted by Tobii using mathematical and anatomical algorithms. This approach allows the T60 to make a three-dimensional representation about where each eyeball is located in space and subsequently, the path of their gaze. This hardware was supported by a Dell Inspiron 530 desktop, Microsoft Windows XP Professional 2002 operating system running Tobii Studio 2.1.13 to collect viewer data (www.tobii.com).

The experimental paradigm was delivered by a third server, a Dell Precision M4400 Workstation, Microsoft Windows XP Professional 2002 operating system running E-Prime 2.0 Professional (www.pstnet.com). Audio was delivered using Altec Lansing desktop speakers connected to the M4400 laptop.

Tobii Studio software recorded and translated participants' eye gaze to numerical data. Data included length of time a participant allocated gaze to AOIs during presentation of the entire paradigm which Tobii Studio recorded as video. Using time flags that indicated the beginning and end of portions of the video that were of interest, we designated selections in the video for Tobii to compute gaze allocation data. Tobii Studio provided raw data that were transferred to Microsoft Excel and uploaded into SAS 9.2 for analysis.

Results

Preliminary analysis: Evaluating the quality of the data

Preliminary analyses were conducted to evaluate the quality of the collected eye-tracking data. The On-task duration was computed for each trial, defined as the duration during which the participant gazed at any point on the eye-tracker screen. Gaze was measured within a 3.50s time window of interest. Forty participants were included in this analysis as technical difficulties with E-Prime prevented two participants from receiving the paradigm properly. Across all conditions, time spent gazing anywhere on the screen averaged 3.21s (*SD* = .26) with mean durations ranging from 2.27s to 3.50s for all participants.

To indentify predictors of On-task duration, we specified a series of mixed models using SAS Proc Mixed. The significance associated with individual predictors was evaluated by comparing fit statistics (-2 log likelihoods) across nested models using a chi-square test. To evaluate possible confounds, we tested the following parameters: (1) condition (i.e., neutral, sincere, ironic) and (2) BAPQ scores. Descriptive statistics for BAPQ performance are reported in Table 2.1. Condition marginally improved model fit, χ^2 (2) = 5.6, *p* = .06. On-task duration time was 3.26s (*SE* = .04), 3.19s (*SE* = .04) and 3.22s (*SE* = .04) in the neutral, sincere and ironic conditions, respectively. In a test of multiple comparisons, Tukey-Kramer contrasts showed that

Table 2.1

Descriptive statistics of adults' BAPQ performance

BAPQ Subscale	*M*	*SD*	Min	Max	Cutoff
Rigidity	3.09	0.69	2.00	5.09	3.50
Aloofness	2.68	0.76	1.33	4.73	3.25
Pragmatic language	2.67	0.54	1.75	4.17	2.75
Total average score	2.81	0.50	1.97	4.30	3.15

the only significant differences in gaze allocation to the screen existed between the neutral (*M* = 3.26s) and sincere conditions (*M* = 3.19s), *t* (78) = 2.39, *p* = .02. A nested model including total average BAPQ score showed that this parameter significantly improved model fit again, χ^2 (1) = 7.10, *p* = .01. There was no evidence of an interaction between condition and total average BAPQ score.

Behavioral Results

Judgments concerning communicative intent (‘*Did X mean what they said?’*) were evaluated for accuracy. Due to a data collection error, data were missing for one additional participant. In the ironic and sincere conditions, participants’ accuracy ranged between 90-100% (*M* = .98, *SD* = .04). According to the means, our prediction that adults would perform at ceiling on this measure was confirmed.

Preliminary Analysis: Addressing potentially confounding variables

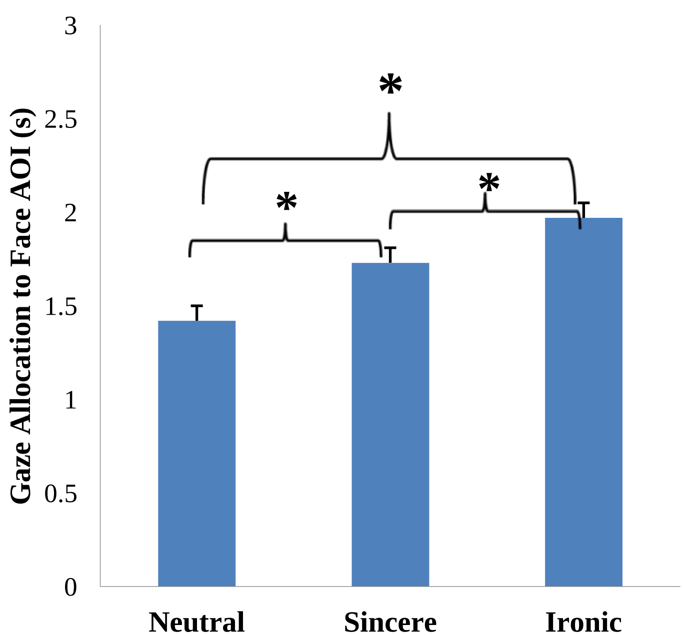
Our key analyses concerned gaze allocation to the Face AOI. Due to the naturalistic quality of our stimuli, three possible confounds were evaluated: (1) AOI size and (2) expressivity of faces, and (3) On-task duration. To evaluate these confounds, we fit a series of mixed models with Face AOI as the outcome. Results showed that gaze allocation to the Face AOI was independently predicted by AOI size, F (1,1203) = 39.04, $p < .00$, expressivity of faces, F (1, 1203) = 16.01, $p < .00$, and On-task duration, F (1, 1203) = 33.06, $p < .00$. Subsequent analyses controlled for these covariates. We found no evidence for significant interaction effects between AOI size and expressivity of faces. Neither age nor gender predicted gaze allocation to the Face AOI.

Will gaze allocation to Face AOI in the irony condition differ significantly from the sincere and neutral conditions?

Results thus far allowed us to develop a model that controlled for the independent contributions of three covariates discovered to influence gaze allocation to faces, our target stimuli. Table 2.2 outlines the next steps of our hierarchical mixed model analysis. We added condition to our multipredictor model to test our hypothesis that condition determines gaze allocation to characters’ expressive faces. This prediction was confirmed as adding condition to the model significantly improved model fit, χ^2 (2) = 92.0, *p* = .00. On average, participants attended to the Face AOI for an average duration of 1.42s (*SE* = .08) and 1.73s (*SE* = .09) in the neutral and sincere conditions, respectively. Faces in the ironic condition received the most attention as participants attended to the Face AOI for an average of 1.97s (*SE* = .09) (see Figure 2.5). Our hypothesis that gaze time allocation to the ironic Face AOI will differ significantly from the other two experimental conditions was supported. Tukey-Kramer post-hoc tests showed that all multiple comparisons differed significantly.

Is gaze allocation to the ironic Face AOI predicted by individual differences in the expression of traits associated with broad autism phenotype?

Our second hypothesis asserted that performance on the BAPQ predicts gaze allocation to faces such that individuals who score above cutoff (more symptomatic) spend less time gazing at ironic faces. Again, we began with a model that controlled for (1) AOI size, (2) expressivity of



Note. * *p* < .01

Figure 2.5. Mean gaze allocation to faces by condition.

faces and (3) On-task duration and tested the independent effects of each BAPQ subscale.

Pragmatic language subscale scores significantly improved model fit above and beyond condition, χ^2 (1) = 3.7, *p* = .05. Rigid subscale scores also significantly improved model fit, χ^2 (1) = 5.5, *p* = .02. Aloof subscale scores did not significantly predict gaze allocation to the Face AOI, *F* (1, 38) = 1.20, *p* = .28. Entering total average BAPQ score as a final parameter significantly improved goodness of model fit, χ^2 (1) = 5.0, *p* = .03 (see Table 1.2). Although BAPQ performance explained a portion of the variance associated with gaze allocation across the three conditions, no evidence existed suggesting an interaction between condition and (1) pragmatic language scores, (2) rigid scores or (3) total average BAPQ score. Our hypothesis that individuals who show traits associated with autism will selectively allocate their gaze less to the Face AOI in ironic scenarios was not supported.

Post-Hoc Analyses of our Experimental Stimuli

A portion of our experimental stimuli included ten scenarios that were presented as both ironic and sincere versions. The goal of this post-hoc analysis was to identify scenarios where gaze time allocation (a) differed significantly by condition and (b) was associated with individual differences in BAPQ scores. Results revealed that gaze allocation to the Face AOI in six out of ten scenarios was significantly predicted by condition. P levels ranged from $p < .00$ – $p = .02$. Gaze allocation was predicted by total average BAPQ score alone for the 'catch' and 'mall' scenarios, *F* (1, 38) = 7.83, $p < .00$ and *F* (1, 38) = 6.10, *p* = .02, respectively. One out of ten scenarios, 'swings', was significantly predicted by an interaction between condition and total average BAPQ score, *F* (1, 37) = 4.45, *p* = .04. Since these scenarios are most effective at

Table 2.2

Parameter estimates (standard errors in parentheses) and fit statistics from a hierarchical mixed model analysis to determine predictors of visual attention to Face AOI

Variable	Step 1	Step 2	Step 3
Intercept	1.663 (0.073)***	1.947 (0.084)***	2.882 (0.363)***
Screen	0.424 (0.074)***	0.462 (0.076)***	0.458 (0.075)***
AOI size	-0.177 (0.028)***	-0.232 (0.030)***	-0.232 (0.030)***
Expressivity	-0.044 (0.011)***	-0.005 (0.015)	-0.005 (0.015)
Condition: neutral		-0.559 (0.064)***	-0.559 (0.064)***
Condition: sincere		-0.245 (0.058)***	-0.246 (0.057)***
Condition: ironic		0	0
BAPQ total avg. score			-0.334 (0.118)**
Fit statistic			
- 2 log likelihood	3107.1	3015.1	3010.1

Note. Step 1 includes the three covariates: On-task duration, AOI size and expressivity of faces. Step 2 adds experimental condition and Step 3 adds total average score on the BAPQ. For Step 2 and Step 3, the parameters and standard errors associated with the ironic condition are reported as the intercept.

*** $p < .001$, ** $p < .01$

eliciting differential responses by condition and highlighting individual differences, they are ideal stimuli for condensed stimuli blocks. Going forward, it will be useful to shorten stimuli blocks for participant populations including children with autism, for example, who will likely have decreased attention spans.

Discussion

A considerable body of research has been dedicated to the understanding of face perception. This research has focused on comparative studies examining how delayed samples perceive faces differentially when compared to typically developing individuals. The impetus behind the recent surge in these studies is to unveil atypical eye gaze strategies shared by the

ASD population and general population to discover endophenotypes. The current experiment focused on face perception in typically developed adults to gain a greater sense of the variation of eye gaze patterns in the general population. This variation may be attributable to the presence of characteristics related to autism that appear in individuals without an ASD diagnosis. We predicted that when determining communicative intent in the Nonliteral Language eye-tracking paradigm, adults would allocate gaze significantly longer to characters' faces when they made an ironic statement versus a sincere or neutral statement. The data reflected this prediction. Our first hypothesis was fully supported.

Our second prediction asserted that as adults' presentation of traits associated with autism increase, their gaze allocation to ironic faces would significantly decrease compared to adults whose average score fell in the normal range on the Broad Autism Phenotype Questionnaire. . Instead, we discovered a significant negative relationship between total average BAPQ score and gaze allocation to faces across conditions such that adults with higher scores gazed less at all faces regardless of condition. Adults' scores on the pragmatic language subscale and the rigid subscale of the BAPQ also predicted less looking time to all faces.

The rationale for our first prediction was based on the idea that correct interpretation of a character's nonliteral communicative intent requires greater attention to the speaker's face, a salient social cue. Presumably, the salience of faces operates to guide humans towards becoming 'face experts' who can recognize faces and the messages they carry. In this way, we can mentalize reliably and predict others' behavior (Haxby, Hoffman & Gobbini, 2000; Wang et al., 2007). Our finding supports this idea of facial salience. Our finding also agrees with a host of studies that agree that typical individuals' focus on faces is a natural tendency that is central to interpreting and appreciating social situations (Sabbagh, 2004; Tottenham, Leon & Casey, 2006; Boraston et al., 2007).

Social cues, especially nonverbal indications, become even more critical when the potential for nonliteral language arises. Such inferences about other people's internal states are essential in these situations as they serve to guide our behavior accordingly (Haxby et al., 2000). When language is ambiguous, words cannot simply be taken at face value and the listener must be able to negotiate a representation of the speaker's words and their masked intention. Adults have been shown to take longer to process ironic statements than literal statements to account for the ambiguity of nonliteral language (Dews & Winner, 1999). This phenomenon has been explained by the indirect-negation model which states that a speaker's chosen words and the contrasting intended meaning of their nonliteral statement must be simultaneously processed so that the discrepancy between the two can be identified (Giora, 1995). Another theory of irony, the parallel-race model supposes that both the literal and nonliteral meanings of an ironic statement must be computed for the listener to recognize the statement's humor, a device inherent in irony. As such, adults allocated gaze to ironic faces the most which on average, was over 50% of the time window we measured. Adults reliably judged these speakers as having ironic intent; a process that may have required more visual attention to faces than determining that our sincere characters truly meant what they said. There is no incongruence of spoken word and hidden meaning to decode and referencing the face need only be brief. Accordingly, the decision that neutral speakers lack ambiguity and are being literal in the strongest sense, also commands less visual attention.

Unexpectedly, increased BAP expression did not lead adults to preferentially allocate less eye gaze to faces of ironic speakers. The rationale for our second hypothesis falls in line with extensive research indicating that relatives of individuals with autism and individuals with autism show atypical face processing strategies (Losh, Childress, Lam & Piven, 2008; Wallace et al., 2010). For example, prior research by Baron-Cohen et al. (1997) has shown that parents of children with Autism spectrum disorder perform significantly worse than control parents when asked to label emotional states of others on the Eyes Task (Baron-Cohen et al., 1997). An additional study by Gokcen et al. (2009) discovered that in comparison to control parents, parents of children with autism were impaired in reasoning about characters' emotional states and emotional reactions in a battery of vignettes (Gokcen, Bora, Erermis, Kesikci & Aydin, 2009). These findings suggests these individuals had difficulty using the face as a reference for interpreting affect and analyzing context clues to identify emotions and their social meaning. Our paradigm taps a similar skill set as these tasks and we believed that our results would be consistent with these findings. With regards to our choice of measures, this suggests that the BAPQ may be a better predictor of global social impairment but a weaker tool when used to distinguish nuanced responses such as gaze allocation when considering irony. Such responses may correlate more strongly with performance on an instrument that directly targets the social capacities being tested, in this case, emotional perception and interpretation (Gokcen et al. 2009). The choice of measurements used in the study of autism endophenotypes requires careful study of features of the broad autism phenotype to determine the instruments that assess the traits that closely approximate the social skill in question (Losh & Piven, 2007).

It is possible that the failure to find an interaction BAP and gaze allocation to ironic faces is attributable to the quality of characters' voices during ironic statements. Our audiovisual cartoons were designed to be accessible to children as well as individuals with autism. As such, ironic statements are delivered in an exaggerated somewhat campy manner. This may preclude extended gaze allocation to the face and after a handful of repeated trials, the pattern of ironic voices alone may serve as a sufficient clue for any typical adult to understand the speaker's communicative intent. Adults that are symptomatic of the BAPQ may even have adopted a 'hacking' strategy whereby they can avoid visual fixation to faces and rely more heavily on other cues to interpret social information. In the case of our paradigm, the tone and prosody of characters' voices were strong cues that could have aided such an alternative strategy (Frith, Happe & Siddons, 1994). This is one limitation of our study.

It is not a surprising finding that more symptomatic adults, as reflected by BAPQ total average score, were less visually attuned to expressive faces across our social scenarios. It is well documented that individuals who express the broad autism phenotype or are biologically related to someone with ASD (and untested for BAP) show similar phenotypical traits including social and communication impairments. With regard to our findings concerning subscales, it also makes intuitive sense that difficulties with pragmatic language would predict less decreased looking time to faces. Interestingly, rigidity also predicted less gaze allocation to faces. According to BAPQ items targeting rigid personality, these individuals are uncomfortable changing their routines and being flexible around how things should be done, despite acknowledging that there could be a better way to do things (Hurley et al., 2007). Perhaps this rigid approach also applies to the aberrant visual scanning strategies that these adults have adapted and grown accustomed to over time. The eye gaze patterns of high scoring adults suggests a social cognitive endophentype defined by less visual attention to faces which may be genetically traced to a common link among the general and clinical populations.

Lastly, we performed a final analysis to determine the vignettes that would be ideal for presentation to various samples of children for future studies. To do this, we identified the vignettes that were best distinguished by condition by examining differences in gaze allocation to vignettes as a function of condition. We also evaluated scenarios for gaze allocation differences as predicted by the interaction between condition and BAP score to discern the stimuli that would best identify differences among samples. Our results showed that gaze allocation in six out of our ten pairs of vignettes was significantly predicted by condition and best suited to evaluate gaze allocation going forward. In one out of ten pairs of scenarios, gaze allocation was predicted by the interaction between condition and BAP score. This is in line with our findings concerning our second hypothesis. It would be premature to state that the traits associated with autism would not interact with condition to influence gaze allocation in another sample. A pilot study examining BAP children, for example, would confirm or refute the results of our findings.

III. The Relations among Narrative Competence, Theory of Mind and Global Development in Typical Children

Abstract

Narrative tasks are a dynamic tool to gauge an individual's fluency with language. An effective narrator's role is twofold: As they relate the events of their story, they must monitor a listener's attention to ensure sustained engagement, a skill that draws on theory of mind. In a sample of neurotypical children, we investigated the relationship between theory of mind and narrative competence with respect to global development. Children aged 3.2 to 9.3 years old received standardized language tests, a Theory of Mind (ToM) Battery and delivered a spontaneous narrative task using a wordless picture book. We found that utterance total, unique verb total unique adjective total, internal state language and range of evaluative devices were significantly positively correlated with ToM Battery scores. The relationship between unique verb total and ToM Battery was attenuated when controlling for expressive language age and receptive language age, respectively. When controlling for chronological age and expressive and receptive language, relations between ToM Battery score and internal state language and range of evaluative devices, respectively, were weakened. Our findings illustrate some of the limits that a sample of young, developing children face when asked to construct a spontaneous narrative. We discuss the link between narrative competence abilities and theory of mind in the context of global development.

Introduction

Theory of mind (ToM) is the ability to infer and appreciate a range of internal states, cognitive and emotional, as they occur in the self and others. Reflecting on one's own state of mind as well as that of others is often referred to as 'mentalizing' and serves to guide our understanding of intentionality (Baron-Cohen, 2001). Children's use of internal state language, i.e. terms describing cognitive and emotional states, has been examined in relation to theory of mind abilities. The association between theory of mind and internal state language has not been fully explored in typically developing children and the present study aims to highlight the relationship between these abilities with respect to global development.

Acquisition of theory of mind is an additive, developmental process comprised of several measurable milestones (Wellman & Lagattuta, 2000). At 2 years of age, healthy children engage in pretend play that signals the child's capacity to create representations that are 'decoupled' from reality (Leslie, 1987). By the age of 3, children begin talking about basic mental states and can recognize the unique intentions and desires of others (Shatz, Wellman & Silber, 1983; Schultz, Wells & Sarda, 1980; Repacholi & Gopnik, 1997). Within the next two years by the ages of 4-5, children typically master first-order false belief tasks and can accurately perceive and report a situation from a character's perspective despite differences from the child's point of view (Wimmer & Perner, 1983). As early as 6 years of age, children properly interpret second-order belief and can report embedded mental states (Perner & Wimmer, 1985). Years later at approximately 11, children recognize the incongruence fundamental to a speaker's ironic statements and also provide viable explanations for communicative intent (Demorest, Silberstein, Gardner & Winner, 1983). Intact theory of mind is integral for fluency in our complex social world. The precise role of language in the development of theory of mind is unclear however, and researchers continue to explore the nature of this relationship.

The development of language and theory of mind occur together in the early years but that is not enough to explain the assumption that there is an important relationship between the two. There is convincing evidence that exposure to language provides the child with a way to learn about mental states such as what another might be thinking or feeling, actions that are unobservable (Gleitman, 1990). Verbal communication may also be necessary to negotiate an event that occurred in the past in order to mentally represent the contrast of the past and present state of affairs (Lagutta & Wellman, 2001, Harris, 2005). Another idea supposes that language is a symbolic tool to mentally represent and track useful information that gives clues about the reality of the state of things (Karmiloff-Smith, 1992). This leads to the controversial question of whether language development primarily scaffolds theory of mind or vice-versa. In an exhaustive review of the literature addressing this issue, Garfield et al. (2001) argued that humans' innate hardwiring for acquisition of global social cognition and language are individually causally necessary and jointly causally sufficient for the development of theory of mind (Garfield, Peterson & Perry, 2001). Others have proposed that language and ToM skills feedback upon each other in a bidirectional relationship in the context of early experience in a social environment (Saxe et al., 2006).

With respect to specific components of language that promote ToM, a seminal longitudinal study performed by Astington and Jenkins (1999) that examined 3-year olds showed that syntax but not semantics consistently predicted several early ToM skills. They did not find any evidence that ToM performance predicted language. Another longitudinal investigation by de Villiers and Pyers (2002) found that preschoolers' comprehension of embedded complements within sentences predicted false belief performance 4 months later. This suggests that the ability to represent potentially contrasting information provided in the complement versus the complete sentence was necessary to arrive at the true value of the full sentence. Conversely, work by Ruffman et al. discovered the opposite: Three year olds' performance on semantic, but not syntactic measures predicted later ToM performance at ages 3.5, 4 and 5.5 (Ruffman, Slade, Rowlandson, Rumsey & Garnham, 2003). Like Astington et al. (1999), de Villiers et al. (2002) discovered that language fostered theory of mind competence but false belief performance did not predict language ability. While it is quite clear that language and theory of mind are intertwined, the research still seeks to illuminate the specific mechanisms that drive the relationship.

Telling narratives is a sophisticated skill that depends on the coordination between cognitive and language skills. As children deliver a story, they are simultaneously describing and sequencing events, referencing cause and effect relationships and conforming to a universal storytelling structures that allow the listener to follow along (Smith, 1993). Generally by 3 years of age, toddlers have enough language and understanding of temporal relations to practice translating memories into stories (Markowiak, 2005). Kleinknecht and Beike (2004) conducted an investigation into the relations among memory and: (1) proficiency for telling an autobiographical story, (2) narrative competence using a storybook and (3) performance on a ToM Battery. In 22 children ages 3 to 5 years old, these researchers found that memory was not significantly related to ToM but a relation between higher ToM and a wider range of narrative devices used during the storybook task was identified. Kleinknecht et al. also found that children with high ToM performance showed an increased use of action propositions but not evaluations (narrative device category that includes mental state language). This finding is explained by the young age of the sample and their limited knowledge and inexperience around the use of evaluative information in delivering a narrative. Furthermore, children with stronger performance on ToM tasks were more engaged during storytelling and were more successful at conveying the gist of the story to the experimenter. These researchers theorized that more developed theory of mind skills impart greater social awareness and insight on an individual which generates an interest in storytelling with others. In turn, narrative discourse with others probably influences and promotes ToM understanding (Kleinknecht et al., 2004).

The use of metacognitive and emotional language within a story reflects an understanding of ToM and may also be related to narrative competence. Bamberg and Damrad-Frye (1991) studied narratives collected from 5-year olds, 9-year olds and college age adults who told stories using wordless picture books. When compared to adults, both groups of children used references to cognitive and emotional frames of mind significantly less. Five-year olds used references to these frames of mind as often as other evaluative devices (e.g., character speech, hedges, etc.,) but 9-year olds and adults used these references with significantly increased frequency than other devices. These findings echo those of Kleinknecht et al. and point to how improved theory of mind likely relates to richer narratives.

The goal of the current study is to further examine the relation between theory of mind and narrative competence in a sample of 50 neurotypical children ages 3.2 to 9.3 years of age ($M = 6.3$, $SD = 1.8$). To examine this, we collected spontaneous narratives from children and coded them for aspects of linguistic complexity and instances evaluative devices used to engage a listener. Children's' performance on this narrative task and a ToM Battery were evaluated with regard to their performance on standardized expressive and receptive language tests. We hypothesized that our measures of narrative competence and ToM Battery score will be positively correlated with each other and with global development. We predict that that the relationship between narrative competence and ToM will remain significant when controlling for global development.

Method

Participants

Our sample included 50 neurotypical children, 23 females and 27 males, between the ages of 3.2 to 9.3 years of age (*M* = 6.3, *SD* = 1.8). Sample characteristics are outlined in Table 3.1. Using the Differential Ability Scales II, all participants were screened for intellectual disability by confirming they had an IQ greater than 70. Participants were recruited using Craig's List, Hunter College's Listserv, Big Apple Parent Magazine, informational sessions at schools and events and through word-of-mouth. All recruitment procedures were approved by the Hunter College Institutional Review Board.

Using our Medical Questionnaire, children were screened for medical conditions linked to Autism spectrum disorder that might preclude typical performance on behavioral measures (Fragile X, Tuberous Sclerosis, Cerebral Palsy, and hydrocephalus).

Materials

Informed Consent & Entrance Evaluations

Children's parents reviewed and completed consent forms that described materials, study procedures, video recording procedures and explanations of participant rights and resources. Parents acknowledged they would receive $25. Consent was obtained from all participants according to Hunter College Institutional Review Board protocol. Assent was obtained from all children according to guidelines approved by Hunter College Institutional Review Board.

Table 3.1

Sample characteristics (N = 50)

Variable	
Gender	
Male	27
Female	23
Race/ethnicity	
White	6
African-American	17
Asian	8
Hispanic/Latino	15
Mixed Heritage	4
Languages spoken at home	
English only	33
English/Spanish	11
English/other	5
Predominantly another language	1
Maternal education	
Did not complete high school	4
High school graduate	3
Partial college	19
Standard college graduate	9
Graduate professional training	10
Paternal education	
Did not complete high school	0
High school graduate	6
Partial college	5
Standard college graduate	9
Graduate professional training	9
Household income	
Above $90k	6
$50k to $90k	9
$20k to $50k	14
Below $20k	16

Note. Due to incomplete or missing information from participants' parents, not all category values total 50.

To rule out the possibility of autism, parents completed several inventories on behalf of their children including the Social Responsiveness Scale (SRS; Constantino & Gruber, 2005). This questionnaire consists of 65 items answered according to a Likert scale from one to six (1 = not true and 4 = almost always true). The SRS is designed to gauge a child's expression of the broad autism phenotype by quantifying behaviors according to the three domains that show deficits in autism: social behavior, language/communication and repetitive and stereotyped behaviors and interests.

The Background Questionnaire was developed by our lab and contains eight items and that took inventory of participants' basic personal and demographic information such as race, current living location and parent's education and income level.

Lastly, parents completed a Medical Questionnaire that asked them to report their child's known neurological conditions, genetic disorders and developmental disorders. Parents were also asked to provide their family's relevant medical history (e.g., autism, depression, intellectual disability, tuberous sclerosis, etc.,).

Developmental Age Tests

Table 3.2 provides a summary of children's performance on all standardized measures. Children were tested on two standardized measures to gauge verbal abilities. The Early One Word Picture Vocabulary Test, Fourth Edition (EOWPVT-IV; Brownell, 2000) uses full-page color drawings depicting objects, actions and concepts. As the examiner points to each picture, he/she asks the child to name what the picture describes. The EOWPVT-IV provides standardized scores and age equivalents for expressive language.

The Peabody Picture Vocabulary Test, Fourth Edition (PPVT-IV; Dunn & Dunn, 2007) uses a series of color drawings depicting objects and actions. Each page presented contains four pictures and the examiner calls out the name of one of the pictures and asks the child to point to the or verbally indicate the number of the picture being named. The PPVT-IV provides standardized scores and age equivalents for receptive language.

Children also received the Differential Ability Scales (DAS-II; Elliot, 1990) to gauge nonverbal abilities. The DAS-II is an individually administered battery of cognitive and achievement tests that are divided into three levels: Lower Preschool, Upper Preschool, and School age. Similar to an IQ test, the DAS-II yields composite nonverbal scores and nonverbal age equivalents.

Table 3.2

Descriptive statistics for global development and Social Responsiveness Scale

Measure	*M*	*SD*
Chronological Age	74.9	20.1
Expressive Language Age Equivalent	75.5	26.2
Receptive Language Age Equivalent	78.0	26.8
Intelligence Quotient	82.5	26.8
SRS T-Score	51.9	8.91

Note. Intelligence Quotient was determined using nonverbal and spatial subtests from the DAS II

Behavioral Measures

Theory of Mind Battery

We adapted the Theory of Mind Battery from Steele, Joseph and Tager-Flusberg (2003) to measure each child's understanding of the perspectives, desires and thoughts of others that may differ from their own. Tasks within this assessment asked the child to anticipate, justify or demonstrate what idea or behavior a story's character would likely experience. Test questions were worth varying amounts of points and for the first two batteries, children were required to pass the 'gatekeeper' or core task to advance. Once children reached the Advanced Battery, they received all of its tasks. The three levels of the assessment and corresponding tasks are as follows and gatekeeper tasks are highlighted (see Appendix A).

- The Early Battery included the ***Desire Task***. This exercise targeted the concept of personal agency by asking the child to report what a character would reasonably choose to do next based on the desire of the character presented in a short story.
- The Basic Battery was comprised of the ***First-Order False Belief Task*** and the *Unexpected Contents-Object Change Task*. These tasks required the child to consider competing mental representations about their own knowledge versus a character's limited knowledge about a situation. Children must were asked to report the character's false belief.
- The Advanced Battery consisted of the *Second-Order False Belief Task,* the *Lies and Jokes Task* and the *Moral Responsibility Task*. These tasks depended on the child's understanding of sophisticated social conventions. Children were asked to identify characters' intentions and behaviors and explain why a character might have deceived, committed a moral transgression or told a white lie. Once a child reached this level of the battery, they received all three tasks.

To evaluate the ToM Battery score in our analyses, we constructed a composite variable comprised of scores from the all subtests from the ToM Battery. Our SAS code stipulated that scores would be entered and analyzed by the model only if the child passed the *Desire Task*, the gatekeeper task from the Early Battery. The score from this subtest was only used for this condition and excluded from the composite score.

Two independent raters achieved reliability according to a composite score that summed scores across the tasks that each child completed. The intraclass correlation coefficient was g = .98.

Narrative Competence

Based on the work of Capps, Losh and Thurber (2000) and Losh and Capps (2003), we used two wordless picture books, *Frog on his own* (Mayer, 1973) and *Frog goes to dinner* (Mayer, 1974) to record each child's spontaneous narrative. Our aim was to obtain measures of children's level of linguistic complexity, use of internal state language and use of evaluative devices employed to describe characters' experiences while actively engaging a listener. We adapted several aspects of Losh et al.'s coding scheme and added several original elements of our own.

Mayer's books range from 28-32 pages and each illustrates the story of a frog that gets separated from his boy companion. The frog has many misadventures as he gets himself into compromising situations, such as, getting into customers' food at a restaurant or incurring the wrath of a nasty cat that pursues him. The nature of these dramatic episodes provided the narrator with many chances to describe sequences of action and cognitive and emotional states of the characters.

Procedure

All tasks were administered in the assessment room at the Communication and Play Lab at Hunter College, New York, NY. Lab visits began as one examiner reviewed informed consent and the entrance evaluations with the parent(s) while the other examiner guided the child through assent procedures. Using simple terms, the examiner showed the child the study materials and equipment while describing what would transpire. Assessments began only when the child displayed clear agreement and comfort with the described procedures. Parents completed the SRS, and the Background and Medical Questionnaires as the examiner started administering standardized tests. Audiovisual recording began at this point in the session.

Children were seated across from one examiner at a small table that was central to several video cameras mounted on the walls. The PPVT-IV, EOWPVT-IV and the DAS-II were administered first. For all tests, participants began at the age appropriate level and continued until they performed at ceiling. Next, children were offered a short break and upon their return, they received the ToM Battery from a second examiner. All participants began with the Early Battery and the examiner continued as long as the child passed the gatekeeper items within each section of the battery.

Children took another short break and returned to the table to complete their visit with the narrative competence task. The second examiner gave the child a choice between the two wordless picture books and then eased the child into the task by narrating the first one or two episodes of the book. Other children were comfortable beginning the task immediately. After the examiner turned narration over to the child, the examiner only communicated prompts for elaboration (e.g., ‘And then what?’) or prompts for clarification (e.g., ‘What do you mean by this?’).

Coding protocol for narratives

I. Utterance coding

The first coder to transcribe each story was required to determine the end of each utterance and insert the appropriate punctuation: period, exclamation point or question mark for the purposes of providing something reasonably intelligible to the subsequent transcribing coders. However, these punctuation decisions were rudimentary and only served as placeholders.

After the coders agreed on transcript content for every child’s narrative, utterances were evaluated by two independent raters: a doctoral student and an undergraduate research assistant.

The doctoral student listened to every narrative twice and inserted punctuation where the end of an utterance was perceived.

Several aspects of children's speech were considered to determine utterance boundaries. Utterance segmentation can be mapped by prosody across languages since the two are frequently coterminal (Vaissiere, 2006). Accordingly, speech was considered for consistent prosodic patterns that might indicate an individual's style of narration. For some children, their intonation tended to fall at the end of utterances while we detected that for others, their intonation would rise as they completed utterances. Secondly, we considered the individual's breaks in speech for a discernible pattern of pauses that might hint at utterance boundaries. Breaks in speech were particularly useful for cases in which prosodic patterns seemed erratic and unreliable. We used this cue with caution since in some cases a pause occurred so the speaker could breathe, deliberately accent their preceding words or, plan their following words. Thus, such breaks between words are not always in a one to one correlation with grammatical units. Lastly, we looked to the content of what the child was saying to provide clues if prosody patterns and pauses still did not seem to clearly signal the end of an utterance. We relied on this aspect of children's narratives as a last resort since what might objectively be determined as the end of a coherent thought might not always correspond with the child's communicative intent as narrator (Stockman, 2010).

Utterance agreements were deemed to be instances in which the coders agreed on the end of an utterance by placing a period, exclamation point or question mark within three words of each other. Agreements were not contingent on the type of punctuation selected to signal the end of the sentence. Disagreements existed when one of the coders determined the end of an utterance and the second coder did not have such punctuation within a three-word range, either before or after the first coder's decision.

Coders achieved reliability on twenty percent of the pool of narratives that they coded, 58-100% (*M* = .87, *SD* = .12). In light of the poor rating for one of these reliability files (58%), coders resolved utterance discrepancies by reviewing this child's recording together. Once reliability was achieved at the level of utterances, the building blocks of the narrative, stories were coded for content.

II. Linguistic complexity coding

To determine children's linguistic sophistication and their ability to actively engage a listener we quantified the following aspects of each child's narrative: length, use of unique verbs and adjectives and use of evaluative devices.

Utterance total: We quantified the length of each narrative by calculating utterance total, i.e., the total frequency of utterances in the child's entire narrative. Invalid utterances contained unintelligible speech and were not included in the participant's final count. Utterances beginning with false starts were only included if the utterance ended with an intelligible independent clause.

Unique verb total: Every unique verb that the child used throughout their narrative was recorded. Verbs were classified according to subcategory by type: (1) emotional verbs, (2) cognitive verbs, (3) action verbs, (4) narrative verbs, (5) possessive verbs, (6) non-thematic verbs and (7) auxiliary verbs (see Table 3.3).

Repetitions of any given verb were not recorded, however, if the child used a root verb in various forms, their use of different forms was recorded and the child was awarded credit for use of a new verb (e.g., look, looked, looking). For each child, verbs across the seven categories

Table 3.3

Verb subcategories

Category	Definition	Examples
Action	Describes something that was performed or done	'The frog **jumped.**' 'The boy **looked** for the frog.'
Emotional	Refers to feeling/ affect	'The turtle **worried** about his missing friend.' 'Froggy **loved** his freedom.'
Cognitive	Describes a mental state	'The lady **realized** the frog was mischievous.' 'The cat **knew** what the frog was up to.'
Narrative	Indicates a character's communication of a specific message	'The boy's mom **yelled**, "Go to your room!"' 'He **screamed** for the cat to back off.'
Possessive	Refers to a character's possession of something	'The frog **took** the baby's bottle!' 'The waiter **caught** the frog on the table.'
Non-thematic	Verb that is preceded by 'it' which acts as a placeholder for a subject	'It **seemed** like everyone was shocked.' 'It **happened** that the frog escaped.'
Auxiliary	Any linking or helping verb	'The couple **doesn't** see the frog in the basket.' 'He **should** bring his sweater for the picnic.'

described above were summed and the composite frequency was the child's unique verb total. Three independent raters achieved reliability coding for the frequency of each unique verb across subcategories. The intraclass correlation coefficient was g = .97.

There is dissension in the linguistics field as to whether past participles (e.g., broken, ruined, worried, etc.,) function as verbs or adjectives. We classified such terms as verbs if some form of the auxiliary verb 'have' preceded the past participle, thus reflecting an active state (e.g., 'They had chased the frog all around the restaurant!').

Unique adjective total: Every novel adjective that the child used throughout their narrative was recorded. Adjectives were classified according to subcategory by type: (1) emotional adjectives, (2) cognitive adjectives, (3) physical adjectives, (4) quality adjectives, (5) quantity adjectives and (6) type/kind adjectives. As with verbs, repetitions of any given adjective were not recorded. Only if the child used a root verb in various forms was credit awarded for use of a novel adjective (e.g., happy, happier, happiest). Additionally, if some form of the auxiliary verbs 'be' or 'get' preceded a past participle thus reflecting the passive state, we categorized the terms as adjectives (e.g., 'The frog was/got stuck in the man's saxophone!'). For each child, adjectives across the six categories described above were summed and the composite frequency was the child's unique adjective total (see Table 3.4).

Table 3.4

Adjective subcategories

Category	Definition	Examples
Emotional	Describes feeling/ affect	'The turtle was **sad** about his missing friend.' 'At first, Froggy was **thrilled** to be free.'
Cognitive	Describes a mental state	'His pets were **curious** when he left.' 'The **wily** cat spotted the frog.'
Physical	Describes observable traits	'The **young** boy sailed his boat on the lake.' 'He saw the **tall** man enter the restaurant.'
Quality	Refers to the extent that a feature is/ is not an attribute	'The frog crashed the **fancy** restaurant!' 'The man's meal was **nasty**!'
Quantity	Refers to an amount of something	'**Two** people arrived to have a picnic.' '**Hundreds** of flies buzzed around the frog.'
Type/Kind	A general reference that distinguishes something/someone	'**Another** couple sat down for dinner.' 'The lady pushed her **baby** carriage.'

Three independent raters achieved reliability coding for the frequency of each unique adjective across subcategories. The intraclass correlation coefficient was *g* = .92.

Internal state language

To evaluate children's use of language that reflects character knowledge, thought or emotion we created a discrete internal state language variable. For each child, we summed the frequencies of emotional and cognitive verbs and adjectives to compute a total internal state language score. Three independent raters achieved reliability coding for the frequency of these internal state terms. The intraclass correlation coefficient was *g* = .92.

Evaluative device coding

In the next phase of our coding scheme, we scored the use of several devices designed to express a narrator's interpretation of events and to keep a listener engaged (Capps et al., 2000): (1) causal explanations of emotion, (2) causal explanations of cognition (3) causal explanations of physical states/action, (4) character speech, (5) onomatopoeia and sound effects and (6) subjective remarks related to the story (see Table 3.5).

In some instances, an utterance contained more than one device and was coded twice:

e.g.) The mom said, "Shhhh! Behave in the restaurant!" (sound effect and character speech)

We created a variable that accounted for the *range of evaluative devices* that a child could potentially use during their narrative. Scores on this variable ranged from zero to six. Three independent raters achieved reliability coding for the frequency of each evaluative device. Intraclass correlation coefficients ranged from *g* = (.86 – 1.00).

Table 3.5

Evaluative Devices

Category	Definition	Examples
Causal explanations of emotion	Explanations of affective states	'The turtle was sad *because his friend was gone.*' 'Froggy was thrilled *since he got free.*'
Causal explanations of cognition	Explanations of mental states	'His pets were curious *because he left dressed up.*' 'Mom grounded him *so then he was upset.*'
Causal explanations of physical states/action	Explanations describing the physical state or activity of a thing or person	'The boat broke *when the frog dove on it.*' 'The frog ran away *because the cat chased him.*'
Character speech	Creating character voice or dialogue	'He said, *"Lookout!! Frog!"*' 'He saw the **tall** man enter the restaurant.'
Onomatopoeia/ sound effects	Use of either device to draw attention	'The frog *crashed* the fancy restaurant!' '"*Shhhhh*! Be quiet!" said the boy's sister.'
Subjective remarks (related)	Explicit observations, evaluations or opinionated commentary related to the content of the story	'*Look! Look at that fat man* playing the tuba!' 'I'll name the boy Bill since that's my favorite name.'

Results

Preliminary analysis of descriptive statistics for narrative variables and ToM Battery

A summary of children's performance on our narrative competence measures is reported in Table 3.6. Average narrative length was approximately 33 utterances however, the large standard deviation points to a large variability around this mean. The data also show that children relied on verbs more heavily than adjectives to construct their narratives. Of the subcategories we quantified, action verbs constituted the majority of verbs children employed and the frequency of internal state verbs was much less frequent. In general, our sample's use of internal state language and a range of evaluative devices was conservative.

Table 3.6

Descriptive statistics for narrative competence variables (N = 45)

Variable	*M*	*SD*
Utterance total	33.40	15.09
Unique verb total	33.36	17.68
Unique adjective total	6.00	4.20
Internal state language	6.36	4.33
Range of evaluative devices	2.64	1.45
Proportion of verb types	*M*	*SD*
Action verbs	0.68	0.09
Internal state verbs	0.12	0.05

Table 3.7 provides a summary of descriptive statistics displaying children's performance on each of the tasks from the Theory of Mind Battery. All children received the battery. Fifteen children who failed the desire task were excluded from further analysis. Thirty children, just over half of our sample, progressed to the Advanced Battery. There was substantial variability in the ToM composite score which illustrates the incremental nature of theory of mind development in preschool and school aged children

Table 3.7

Descriptive statistics for Theory of Mind Battery

Variable	*N*	*M*	*SD*
Unexpected contents-object change	35	4.46	2.16
First-order false belief	35	4.91	1.75
Second-order false belief	30	4.63	1.81
Lies & jokes	30	2.87	1.24
Moral responsibility	30	2.57	1.61
ToM Battery composite	35	18.0	7.89

Will a positive relationship exist between performance on the Theory of Mind Battery and a narrative competence task?

We examined the relationship between several aspects of the narrative competence task and composite score on the ToM Battery. Our prediction was that children's score on the ToM Battery would be positively correlated with children's linguistic complexity, use of internal state language and range of evaluative devices. Four children were unable to provide a narrative due to challenging behaviors and one persisted in telling a personal narrative, however, these children completed the standardized language measures and the ToM Battery so we did not exclude them from any of the subsequent analyses.

An initial series of Proc GLM regressions ruled out the potential confounds of gender or bilingualism on any aspect of narrative outcomes or ToM Battery score.

Preliminary Analyses: Exploring the influence of chronological age and global development on narrative competence

We began by exploring the link between linguistic complexity and global development. Results are reported in Table 3.8. We measured children's linguistic complexity by quantifying: (1) utterance total, (2) unique verb total and (3) unique adjective total. In SAS 9.2, we evaluated this with a series of Pearson Partial Correlations controlling for narrative length (utterance total) among: (1) chronological age, (2) expressive language age equivalent and (3) receptive language age equivalent and these linguistic complexity variables. Results showed that unique verbs and unique adjectives were significantly related to all aspects of global development. However, none of these aspects of age correlated with utterance total (Pearson correlations testing utterance total were conducted without a control variable).

Table 3.8

Partial correlations between global development and (1) narrative competence variables and (2) ToM Battery when controlling for narrative length

Variable	Chronological Age	Expressive Language Age Equivalent	Receptive Language Age Equivalent
Utterance total	0.12	0.17	0.06
Unique verb total	0.52**	0.62**	0.60**
Unique adjective total	0.51**	0.59**	0.58**
Internal state language	0.45**	0.49**	0.44**
Range of Evaluative devices	0.42**	0.45**	0.46**
ToM Battery scores	0.61**	0.62**	0.69**

Note. Pearson correlations testing Utterance total did not include a control variable
** $p < .01$

Next, we considered the relations between internal state language and global development and discovered that when controlling for narrative length, it was also significantly positively correlated with chronological age and expressive and receptive language age equivalents.

Another series of Pearson Partial Correlations showed that chronological age and expressive and receptive language age equivalents were also significantly positively correlated with range of evaluative devices used.

Correlations between narrative performance and ToM Battery score when controlling for global development

We aimed to determine if children's narrative performance was related to ToM Battery performance when controlling for global development. Statistical results reported in this section are outlined in Table 3.9. Using a standard Pearson Correlation, we found that utterance total was

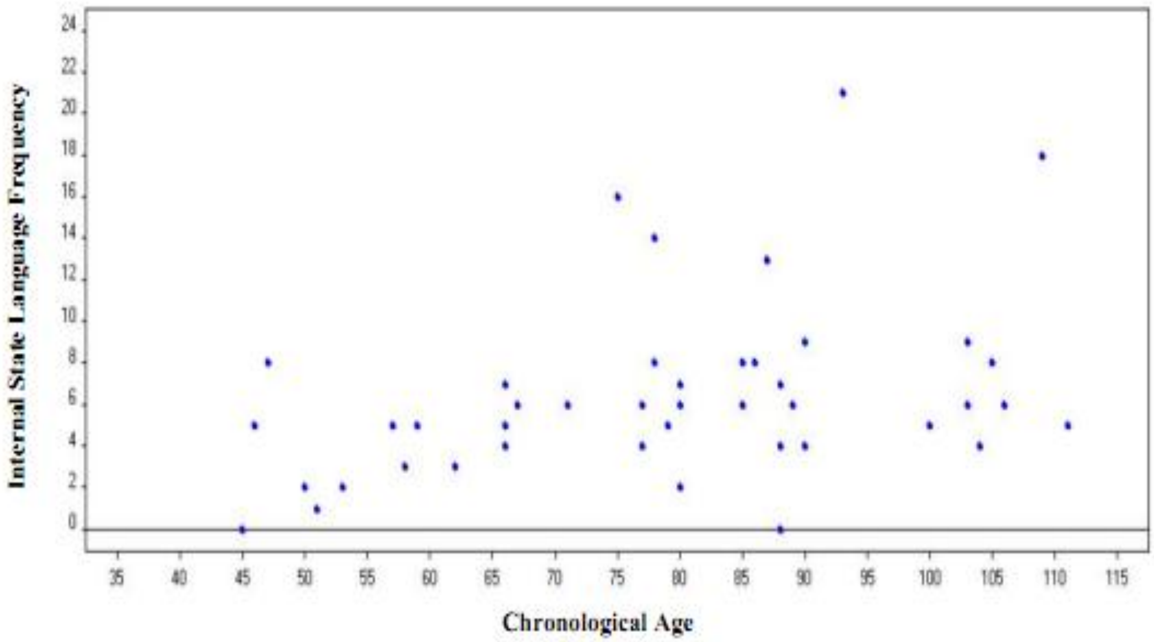


Figure 3.1. Distribution of internal state language composite scores by chronological age

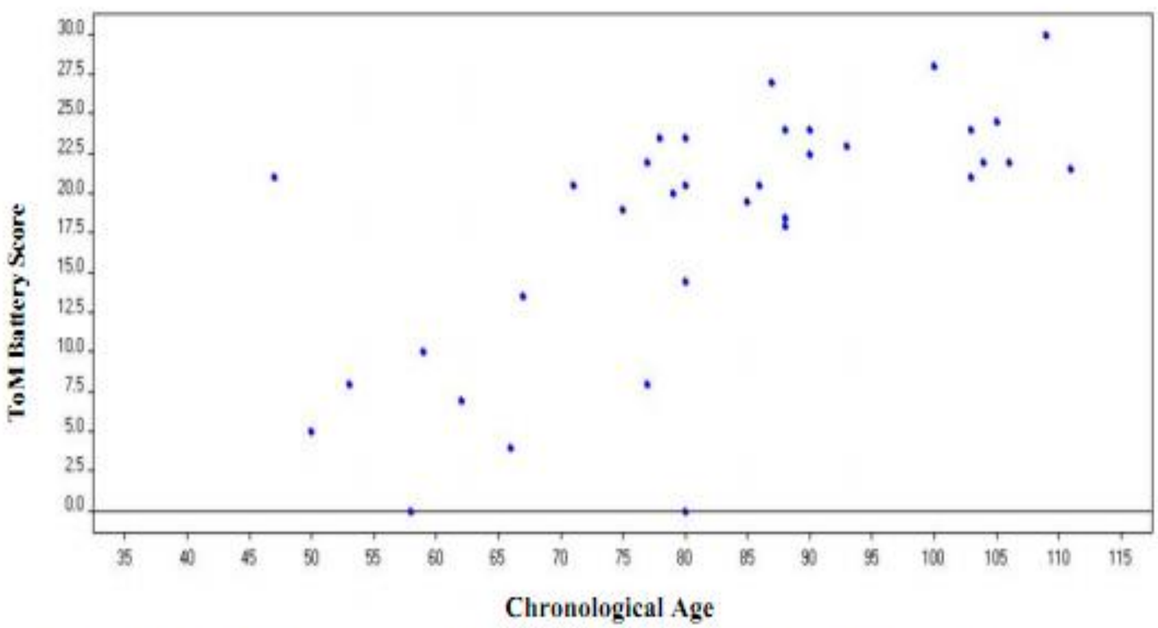


Figure 3.2. Distribution of Theory of Mind Battery composite scores by chronological age

significantly positively correlated with ToM Battery scores. Significance remained when controlling for chronological age and receptive language age equivalent. When controlling for expressive language age equivalent, this relationship became more modest *r* (33) = .32, *p* = .075.

Next, we fit a series a Pearson Partial Correlations that controlled for narrative length and chronological age, expressive language age equivalent and receptive language age equivalent, respectively, to determine the relations between our narrative variables and ToM Battery scores.

Unique verb total was significantly positively correlated with ToM Battery performance and remained so when controlling for chronological age. Significance decreased slightly when controlling for expressive language age equivalent, *r* (33) = .35, *p* = .057 and again when controlling for receptive language age equivalent, *r* (33) = .33, *p* = .073.

Unique adjective total maintained a significantly positive correlation with ToM Battery scores when we controlled for each aspect of global development.

Internal state language showed a marginal positive correlation with ToM Battery score, *r* (33) = .34, *p* = .059. However, upon testing this relationship when controlling for each aspect of global development, this relationship was attenuated.

Lastly, range of evaluative devices was significantly positively correlated with ToM Battery score. This significant relationship was not maintained when we partialed out the effects of global development.

Table 3.9

Partial correlations between narrative competence variables and Theory of Mind Battery when controlling for narrative length and global development

	ToM	ToM	ToM	ToM
Control Variable	Utterance total	Utterance total	Utterance total	Utterance total
Age Covariate		Chronological Age	Expressive Language Age	Receptive Language Age
Utterance total	0.36*	0.36*	0.32	0.44*
Unique verb total	0.55**	0.46**	0.35	0.33
Unique adjective total	0.55**	0.44*	0.36*	0.36*
Internal state language	0.34	0.28	0.19	0.25
Range of Evaluative devices	0.37*	0.26	0.19	0.18

Note. A standard Pearson correlation was used to test the correlation between Utterance total and ToM Battery. For Utterance total's correlations with ToM Battery, only respective age variables were partialed out.
** $p < .01$, * $p < .05$

Discussion

The current study examined the relationships among theory of mind, chronological age and developmental age in a sample of 50 neurotypical children 3.2 – 9.3 years old. Our goal in concentrating on this group alone was to outline how this relationship emerges in the context of healthy development. Narrative abilities and skills for theory of mind are both related to chronological and developmental age. We first identified the extent of these relationships in order to appropriately unpack the relationship between theory of mind and narrative competence. Our findings concerning the relations between our narrative variables and global development yielded anticipated results. As hypothesized, we discovered that global development was significantly related to all measures of narrative competence except utterance total. We then discovered that ToM Battery score was also significantly correlated with global development. With this in mind, we considered the relations between our measures of narrative competence and ToM Battery score. Unique adjective total remained highly correlated with ToM Battery score when controlling for all aspects of global development. Utterance total also had a strong relationship with ToM Battery score however, we discovered that expressive language skills also accounted for children's capacity to tell longer narratives. Both expressive language and receptive language emerged as influential skills beyond theory of mind when considering children's ability to use a dynamic set of verbs. Internal sate language showed a marginal relationship with ToM Battery score that also weakened as we considered the impact of chronological age and expressive and receptive language age. Finally we found that children's use of a wide range of evaluative devices was significantly related to ToM Battery score but all aspects of global development explained some of the variance associated with the relation between the range children used and ToM Battery score.

As documented in previous research, in the absence of a mastery of advanced grammar, children in our sample's age range chiefly rely on their vocabularies' diversity to tell cohesive stories (Bamberg et al., 1991; Bishop & Donlan, 2005; Heilmann, Miller, Nockerts & Dunaway, 2010). Our findings agree with these prior studies that indicate younger, less experienced children depend on their lexicon to provide a framework for their narratives. It has been shown that younger children's narratives tend to be constructed as a sequence of utterances that are essentially a description of the story's events in the order they occurred (Berman, 1988). Our data also show that the majority of the unique verbs we recorded from children were action verbs. A majority of unique verbs used by our sample described action thus, children conformed to the strategy of episodic description outlined by Berman (1988). These results are also in agreement with Kleinknecht et al. (2004) who found that as preschoolers' theory of mind grew, their use of action verbs but not mental state terms in a narrative task significantly increased. Overall, our sample appeared to have sufficient theory of mind skills to appreciate that they should engage their listener using the method that was most available to them. Given their language abilities, children's narratives relied on strings of utterances characterized by dynamic event descriptions that depended on vocabulary above other grammatical skills.

Utterance total was our only measure of linguistic complexity that was not driven by any age related changes. Remarkably, this variable was determined by gains in theory of mind. There may be link such that children who were confident in participating in this type of task were comfortable due to increased theory of mind. In her discussion concerning the interface of language and theory of mind, de Villiers asserted that children's spontaneous speech generally reflects what they are confident they can say (de Villiers, 2007). Increased theory of mind provides a child with better access to interpretations of their own thoughts and beliefs as well as those of others. In turn, this gives them confidence to actively engage their world which may have moderated the strong positive relation between narrative length and theory of mind.

The results of our sample's use of internal state language and evaluative devices, qualitatively higher-level narrative components than our linguistic complexity measures, suggest some limitations in performance. In our consideration of children's use of internal state language, we found that it was predicted by global development. This reflected our expectations however, we then discovered only a modest relationship when we evaluated theory of mind as a predictor of internal state language. This was contrary to our hypothesis. Further analysis of this marginal relationship between internal state language and theory of mind revealed that chronological age and expressive and receptive language age, respectively, contributed to children's ability to utilize cognitive and emotional state terms. Overall, there was not a high frequency of these terms, which is likely explained by young theory of mind and nascent language skills that are unable to further bolster their theory of mind (Astington et al., 1999). The lack of this type of language points to a novice style of storytelling that indicates a restricted appreciation for the social problem-solving quality of narratives (Capps et al., 2000). Through the role of narrator, the child has the opportunity to experience him/herself as a character faced with unique circumstances. Failure to take on the characters' attitudes may lead to an inability to recognize and/or express their internal states. Use of internal state terminology is crucial to forging links between self and others and a shortage of it is likely due to immature theory of mind (Ochs, Taylor, Rudolph & Smith, 1992; Fivush, 2011). In sum, our sample's basic abilities for theory of mind were unable to foster consistent use of cognitive and emotional states.

Our analysis of range of evaluative devices supported our assumption that they would be significantly related to global development. However, range of evaluative devices did not remain significantly correlated with theory of mind when controlling for all aspects of global development. It appears that the overall level of global development within our sample only provided for rudimentary theory of mind skills. These low-level abilities were insufficient in providing children with an appreciation for and access to sophisticated storytelling devices as reflected by their average range of evaluative devices. Without consistently drawing on evaluative devices throughout a narrative, a story may seem disjointed, lacking suspense or drama and without a driving causal framework (Capps et al., 2000).

As demonstrated by the narrative performance of our young sample, elements of linguistic complexity were more observable phenomena than use of internal state language and range of evaluative devices. Literacy and syntax, abilities that older children draw on to produce more complex narratives, are still emerging in preschool and school age children. Astington et al. (1999) and de Villiers et al. (2002) have argued that such tools are central to young children's developing theory of mind. These skills allow children to form a macrostructure for their stories exemplified by a discussion of the problem, attempts at solving the problem and consequences/outcomes (Strong, 1988; Bamberg et al., 1991; Bishop et al., 2005; Heilmann et al., 2010). Without a mature repertoire of linguistic tools and incomplete theory of mind, younger children cannot adhere to this overarching structure by detailing cause and effect relationships.

There are extremely few studies that have investigated global development, theory of mind and spontaneous narrative tasks in typical children, however. It is possible the individual relationships between internal state language and range of evaluative devices with ToM Battery score were so tenuous due to our choice of ToM assessment. Perhaps the addition of a behavioral measure that draws on theory of mind but targets recognition of specific psychological states would have improved our chances of elucidating predictive relationships. Several 'mentalizing' measures including the Eyes Test, Psychological-Intentional Picture Stories, the Berkeley Empathy Measure or an emotion definition task may have served to illuminate behavioral links (Baron-Cohen, Leslie & Frith, 1986; Baron-Cohen et al., 1997; Losh et al., 2003). Prior work has shown that these tests are correlated with proficiency for recognizing mental states and emotions. Specifically, for children with autism, performance on the Psychological-Intentional Stories task and the emotion definition task, respectively, has demonstrated a positive relationship with their use of internal state language (Baron-Cohen et al., 1986; Losh et al., 2003).

Another possibility is that in our young sample, children’s use of internal state terms, although rare, may be associated with what they are learning from their family. Research has also shown that children’s use of internal state language is often predicted by how often a child’s mother mentions cognition and emotional states during narratives to her child (Slaughter, Peterson & Mackintosh, 2007; Guajardo & Watson, 2002). Family discourse containing discussions concerning emotions has also been shown to determine children’s development of theory of mind (LaBounty, Wellman & Olson, 2008; Taumoepeau & Ruffman, 2008). Children’s exposure to this type of language may also be a predictor of their use of internal state language and future studies might examine the influence of family experiences in the development of theory of mind.

IV. Visual Processing of Social Information and Expression of the Broad Autism Phenotype in Typical Children

Abstract

Autism presents with a heterogeneous phenotype that has complicated attempts to identify the network of genes responsible for the disorder. Recent research investigating endophenotypes in the general population may help clarify autism's genetic etiology. In the current study, we assessed neurotypical children on the Nonliteral Language eye-tracking paradigm and the Social Responsiveness Scale to determine the relation between gaze allocation patterns to expressive faces, a socially salient cue, and presentation of characteristics associated with autism. Children also received a Theory of Mind (ToM) Battery and a narrative competence task. When asked to determine a speaker's communicative intent, we found that children spent more time gazing upon faces of speakers that delivered ironic remarks versus faces of sincere or neutral speakers. However, we found that higher expression of traits associated with autism did not predict a strategy of less gaze allocation to speakers' faces. Next we discovered that performance on the ToM Battery had an inverse relationship with children's gaze allocation to all faces, regardless of the type of remark a speaker delivered. We found that children's increased use of internal state language during a spontaneous narrative task also had an inverse relationship with gaze allocation to and predicted less looking to all faces across all conditions. Implications of our results are discussed with regard for optimal strategies that can reliably identify autism endophenotypes when studying typical children.

Introduction

Eye-tracking research has been instrumental in studying the anomalies in Autism spectrum disorder (ASD) by highlighting differences in the way that these individuals visually scan social scenes when compared to typical controls. However, the scope of social and communication impairments varies greatly among individuals with autism, the most severe form of ASD (DSM-IV-TR, 4th ed., 2000). Autism's heterogeneity challenges researchers in their efforts to identify clear paths that outline the disorder from etiology to phenotypical presentation. Moreover, these impairments bear on a host of cognitive skills such as, attention, motivation, expressive language, receptive language and theory of mind that have intricate interdependent relationships. Thus, considering autism as a whole has proven overwhelming and impractical in the search for answers (Happe et al., 2006; Eigsti, Marchena, Schuh & Kelley, 2011). In response, researchers have turned their focus toward an exploration of possible endophenotypes. Classifying individual symptoms into subtypes yields homogenous populations and greatly reduces noise in the investigation of genetic liability for symptoms (Gottesman et al., 2003; Tager-Flusberg, 2006). In the current study, we measured children's expression of the broad autism phenotype (BAP). We evaluated children's gaze allocation patterns on a potentially ironic eye-tracking paradigm and considered these patterns with respect to children's abilities for theory of mind and use of internal state language. By profiling emerging social cognition and BAP traits, we aimed to identify possible autism endophenotypes.

Potential for healthy development begins with a nascent system that is primed to discern relevant environmental cues. Within months of birth, infants show sensitivity to vital social cues such as eye contact and vocal signals and they also naturally detect unusually delayed responses during dyadic interaction with their mothers (Striano & Bertin, 2004; Striano, Henning & Stahl, 2006). Infants show a preference for faces with direct gaze which guides visual attention toward others' faces thus providing a source of information about the internal states of others (Farroni, Menon & Johnson, 2006). This attunement paves the way for development of theory of mind, i.e., the ability to infer and appreciate a range of internal states, cognitive and affective, as they occur in the self and others. Reflecting on one's own state of mind as well as that of others is often referred to as 'mentalizing' and serves to guide our understanding of intentionality (Baron-Cohen, 2001). Theory of mind is a cornerstone of social cognition that fosters proficiency for reciprocal relationships (Hernandez, Metger, Magne, Brilhault, Roux, Barthelemy & Martineau, 2009). One way this skill manifests in our relationships is through pragmatic language, the use of language in communication. We constantly monitor our behavior and that of the person/people we are speaking to in order to regulate social interaction. This includes adjusting the tone and level of our voice appropriately, turn-taking in conversation, showing attentiveness and interest and use of communicative gestures (Ben-Yizhak, Yirmiya, Seidman, Alon, Lord & Sigman, 2011). Irony is a sophisticated pragmatic language device that incorporates nonliteral language, nuanced prosody and subtle humor. Together these features combine to mute the criticism that is generally inherent to ironic remarks (Dews & Winner, 1995). In typical development between the ages of 5 and 8, sensitivity to irony also develops but a full understanding of its use does not begin until the age of 9 (Ackerman, 1981; Glenwright & Pexman, 2010).

In recent years, a sizable body of research and literature has addressed how the acquisition of these social abilities to goes awry in Autism spectrum disorder (ASD). Two of the three core deficits of autism include difficulty engaging in social interactions and difficulty with communication (DSM-IV-TR, 4th ed., 2000). fMRI research by Wang and colleagues used an innovative paradigm to examine the neural correlates underlying perception of nonliteral language (Wang et al., 2007). Eighteen boys with ASD and typical matched controls viewed and heard potentially ironic cartoon vignettes. The investigators were interested in brain activity during participants' interpretation of the speaker's communicative intent. Wang et al. discovered that when children with autism received instructions to attend to the speakers' face or voice during ironic vignettes, activity was elicited in the mPFC. In contrast to controls, this pattern of activation was absent during the 'no instruction' condition. This is the first study to demonstrate that explicit instructions to attend to salient social stimuli can normalize hypoactivated centers of the brain in an autism sample.

Atypical eye gaze patterns have also been well documented in autism yet, there are findings to the contrary that prevent researchers from drawing broad conclusions. For example, it has been found that autism populations (representing a range of ages) generally focus more on the mouth when decoding emotional expressions (Harms, Martin & Wallace, 2010). This tendency has been identified early on as 2-year old children with ASD have demonstrated atypical gaze patterns across faces during a facial recognition task. During the same task, four-year olds with ASD allocated gaze even less to core facial features and for both samples, face encoding was compromised (Chawarska & Shic, 2009). Others have failed to find evidence of abnormal gaze patterns including the predictive value of such patterns as an early biomarker for an autism diagnosis (Young, Merin, Rogers & Ozonoff, 2009; Harms et al., 2010). These researchers have suggested that perhaps individuals with autism employ a top-down strategy causing them to focus longer on features of the face considered less integral to face processing (Neumann, Spezio, Piven & Adolphs, 2006). It is also possible that results fluctuate because the extent of deviant gaze patterns varies systematically with severity of individual symptoms, with age or with both (Riby & Hancock, 2008; Kirchner, Hatri, Heekeren & Dziobek, 2011).

The examination of the neurotypical population to help establish endophenotypes has become increasingly more common. Undiagnosed individuals have been shown to display autism-like traits suggesting a universal spectrum of symptomalogy at the behavioral level (Constantino, Przybeck, Friesen & Todd, 2000; Happe et al., 2006). Studies involving nonautistic relatives of individuals with autism who express broad autism phenotype reveal a high heritability of behavioral impairments and indicate common genetic underpinnings across populations (Hurley et al., 2007). This effect is frequently studied by comparing social abilities between children with autism at their at-risk younger siblings. Pragmatic language performance has been shown to be impoverished in children with autism and their at-risk siblings and serves as an example of a deficit that can appear in the undiagnosed (Ben-Yizhak et al., 2011). However, within both the general and the autism populations, there is only modest overlap of autism’s three core deficits. Evidence consistently points to discrete genetic influences for each impaired domain (Ronald et al., 2006). Pursuing this deconstructed approach to the study of ASD can also result in treatment plans that more precisely target the needs of the individual (Gottesman et al., 2003; Tager-Flusberg, 2006).

The next step according to this conceptual framework is to determine the distinct behaviors, or lack thereof, that constitute potential risk factors. Identifying subclinical factors that may interfere with the evolution of these relationships in the context of healthy children is essential to toward an understanding of impaired social development in disorders such as autism (Eigsti et al., 2011). The goal of the current study that assessed 50 neurotypical children between the ages of 3.2 to 9.3 years of age (*M* = 6.3, *SD* = 1.8) was to investigate the relationship between visual perception of social material with traits associated with autism and theory of mind, respectively, to elucidate if and how they relate. To do this, we administered the ‘Nonliteral

Language' eye-tracking paradigm that targets understanding of ironic statements and considered children's gaze allocation patterns in the context of their SRS scores, a battery designed to measure autism-like characteristics. Children also completed a Theory of Mind Battery (ToM) and delivered a spontaneous narrative to determine if eye gaze patterns were associated with ToM Battery scores and/or use of internal state language. We hypothesized that children would show significantly greater gaze allocation to faces in the ironic condition versus the sincere or neutral condition. We also hypothesized that higher SRS scores would predict less gaze allocation to ironic faces and conversely, higher ToM Battery scores and increased use of internal state language would reflect greater social attunement and predict greater gaze allocation to ironic faces.

Method

Participants

Our sample included 50 neurotypical children, 23 females and 27 males, between the ages of 3.2 to 9.3 years of age (M = 6.3, SD = 1.8). All participants were free of intellectual disability. Participants were recruited using Craig's List, Hunter College's Listserv, Big Apple Parent Magazine, informational sessions at schools and events and through word-of-mouth. All recruitment procedures were approved by the Hunter College Institutional Review Board.

Using our Medical Questionnaire, children were screened for medical conditions linked to Autism spectrum disorder that might preclude typical performance on behavioral measures (Fragile X, Tuberous Sclerosis, Cerebral Palsy, and hydrocephalus). Children who required corrective prescription lenses were included as Tobii hardware is designed to track eye gaze through glasses.

Materials

Informed Consent & Entrance Evaluations

Children's parents reviewed and completed consent forms that described materials, study procedures, video recording procedures and explanations of participant rights and resources. Parents acknowledged they would receive $25. Consent was obtained from all participants according to Hunter College Institutional Review Board protocol. Assent was obtained from all children according to guidelines approved by Hunter College Institutional Review Board.

Parents completed several inventories on behalf of their children including the Social Responsiveness Scale (SRS; Constantino et al., 2005). This questionnaire consists of 65 items answered according to a Likert scale from one to six (1 = not true and 4 = almost always true). The SRS is designed to gauge a child's expression of the broad autism phenotype by quantifying behaviors according to the three domains that show deficits in autism: social behavior, language/communication and repetitive and stereotyped behaviors and interests.

The Background Questionnaire was developed by our lab and contains eight items and that took inventory of participants' basic personal and demographic information such as race, current living location and parent's education and income level.

Lastly, parents completed a Medical Questionnaire that asked them to report their child's known neurological conditions, genetic disorders and developmental disorders. Parents were also asked to provide their family's relevant medical history (e.g., autism, depression, intellectual disability, tuberous sclerosis, etc.,).

Eye-Tracking Stimuli

Audiovisual black and white cartoon scenarios were adapted from Wang et al. (2007) and modified for presentation on an eye-tracker. These static stimuli were divided into three conditions: 12 scenarios in the neutral condition, 10 scenarios in the ironic condition and 10 in the sincere condition. All scenarios were narrated by a female speaker. Neutral scenarios ended with a speaker making a neutral comment that lacked emotional valence. Ironic scenarios concluded with a speaker delivering a remark in a sarcastic tone and sincere scenarios concluded with a speaker delivering a remark in a genuine tone. In each condition, speakers had an emotionally expressive face as they made their remarks. At the end of each vignette, participants were asked to determine if the speaker meant what they said. Wang and colleagues determined that characters' remarks were perceived as sincere or ironic by analyzing ratings from 12 blind coders. These adult volunteers were only exposed to the audio statements and were asked to rate remarks on a scale of one to seven (1 = ironic and 7 = sincere). Mean ratings were 1.4 for ironic remarks (*SD* = 0.7) and 6.6 (*SD* = 0.7) for sincere statements. Wang et al. matched sincere and ironic remarks on the following: syntactic structure, semantic complexity and length (Wang et al., 2007).

To determine the effect of various emotional expressions of characters' faces on participants' gaze allocation, we obtained face ratings from 41 Psychology 100 students at Hunter College who were blind to the paradigm. Students rated the degree of emotional expressivity on the faces of all characters (32) that participants would view when asked to judge communicative intent. Students viewed the faces on slides (Microsoft Office PowerPoint 2007) presented by the instructor. Faces were shown for 10s and students scored them from one to nine (1 = extremely negative expression; 5 = neutral and 9 = extremely positive expression). Mean emotional expressivity score was 5.22 (*SD* = .25). Independent raters' scores were highly reliable: *g* = .989.

Faces were designated as 'Areas of Interest' (AOIs) and we analyzed the amount of time participants allocated their gaze to faces. However, the size of characters' faces differed within the stimulus set and as a result, AOI size varied accordingly. AOIs included a character's entire face plus two centimeters added to the perimeter of the face from any given point. Additional space was added in accordance with Tobii technicians' recommendations to account for possible differences in actual gaze fixation and gaze fixation as recorded by Tobii Studio. Across all faces, mean AOI diameter = 4.53cm (*MIN* = 2.92cm, *MAX* = 6.13cm).

Developmental Age Tests

Descriptive statistics for children's performance on measures of global development and the SRS are reported in Table 3.2 (p. 58). Children were tested on two standardized measures to gauge verbal abilities. The Early One Word Picture Vocabulary Test, Fourth Edition (EOWPVT-IV; Brownell, 2000) uses full-page color drawings depicting objects, actions and concepts. As the examiner points to each picture, he/she asks the child to name what the picture describes. The EOWPVT-IV provides standardized scores and age equivalents for expressive language.

The Peabody Picture Vocabulary Test, Fourth Edition (PPVT-IV; Dunn & Dunn, 2007) uses a series of color drawings depicting objects and actions. Each page presented contains four pictures and the examiner calls out the name of one of the pictures and asks the child to point to the or verbally indicate the number of the picture being named. The PPVT-IV provides standardized scores and age equivalents for receptive language.

Children also received the Differential Ability Scales (DAS-II; Elliot, 1990) to gauge nonverbal abilities. The DAS-II is an individually administered battery of cognitive and achievement tests that are divided into three levels: Lower Preschool, Upper Preschool, and School age. Similar to an IQ test, the DAS-II yields composite nonverbal scores and nonverbal age equivalents.

Behavioral Measures

Narrative Competence

Based on the work of Capps. Losh and Thurber (2000) and Losh and Capps (2003), we used two wordless picture books, *Frog on his own* (Mayer, 1973) and *Frog goes to dinner* (Mayer, 1974) to record each child's spontaneous narrative. Mayer's books range from 28-32 pages and each illustrates the story of a frog that gets separated from his boy companion. The frog has many misadventures as he gets himself into compromising situations, such as, getting into customers' food at a restaurant or incurring the wrath of a nasty cat that pursues him. The nature of these dramatic episodes provided the narrator with many chances to describe sequences of action and cognitive and emotional states of the characters.

We quantified use of internal state language by summing the frequencies of children's use of the following to create a composite score: cognitive verbs (e.g., think, know, believe) and adjectives (e.g. wily, befuddled) and emotional verbs (e.g. cry, laugh, love) and adjectives (e.g. happy, angry). Three independent raters achieved reliability coding for the frequency of these internal state terms. The intraclass correlation coefficient was $g = .92$.

Theory of Mind

We adapted the Theory of Mind Battery from Steele et al. (2003) to measure each child's understanding of the perspectives, desires and thoughts of others that may differ from their own. Tasks within this assessment asked the child to anticipate, justify or demonstrate what idea or behavior a story's character would likely experience. Test questions were worth varying amounts of points and for the first two batteries, children were required to pass the 'gatekeeper' or core task to advance. Once children reached the Advanced Battery, they received all of its tasks. The three levels of the assessment and corresponding tasks are as follows and gatekeeper tasks are highlighted (see Appendix A).

- The Early Battery included the ***Desire Task***. This exercise targeted the concept of personal agency by asking the child to report what a character would reasonably choose to do next based on the desire of the character presented in a short story.
- The Basic Battery was comprised of the ***First-Order False Belief Task*** and the *Unexpected Contents-Object Change Task*. These tasks required the child to consider competing mental representations about their own knowledge versus a character's limited knowledge about a situation. Children must were asked to report the character's false belief.
- The Advanced Battery consisted of the *Second-Order False Belief Task,* the *Lies and Jokes Task* and the *Moral Responsibility Task.* These tasks depended on the child's understanding of sophisticated social conventions. Children were asked to identify characters' intentions and behaviors and explain why a character might have deceived, committed a moral transgression or told a white lie. Once a child reached this level of the battery, they received all three tasks.

To evaluate the ToM Battery score in our analyses, we constructed a composite variable comprised of scores from the all subtests from the ToM Battery. Our SAS code stipulated that scores would be entered and analyzed by the model only if the child passed the *Desire Task*, the gatekeeper task from the Early Battery. The score from this subtest was only used for this condition and excluded from the composite score.

Two independent raters achieved reliability according to a composite score that summed scores across the tasks that each child completed. The intraclass correlation coefficient was $g = .98$.

Procedure

All tasks were administered in the assessment room at the Communication and Play Lab at Hunter College, New York, NY. Lab visits began as one examiner reviewed informed consent and the entrance evaluations with the parent(s) while the other examiner guided the child through assent procedures. Using simple terms, the examiner showed the child the study materials and equipment while describing what would transpire. Assessments began only when the child displayed clear agreement and comfort with the described procedures. Parents completed the SRS, and the Background and Medical Questionnaires as children began to receive the nonliteral language eye-tracking task in the adjoining room.

Eye-tracking Paradigm

Examiners introduced the task with the following instructions:

> *Sometimes when people say things, they mean what they say and sometimes they're just being silly, mean or sarcastic. You're going to look at some pictures and listen to a short story and decide whether or not the person in the story means what they say. When the narrator asks you if the person meant what they said, simply say 'yes' or 'no' out loud and we'll record your answer.*

For the sincere and ironic conditions, two different versions of 10 black and white cartoon scenarios were presented. All contained three static images and were delivered according to the same structure: The first image introduced two characters and an activity they were engaged in. In the second image, an event occurred that resulted in an explicitly negative or positive outcome (e.g., a dog growls at two friends/a dog wags its tail at two friends). Characters' faces were not shown in the second image. The third image depicted only the second character's face and verbal response to the un/desirable outcome with a comment such as, 'He's a friendly one!'. In the nonliteral version of the story, the intonation of the comment was snarky and ironic. In the sincere version, the intonation was genuine and heartfelt. If the speaker made an ironic statement, he/she wore a look of disgust or disdain but if the speaker was being sincere, he/she smiled and appeared warm and complementary. After the final remark, the narrator asked the participant, ‘Did X mean what they said?’. This last image featured on the second character’s face and remained on the monitor for 8.5 seconds (see Figures 2.2 and 2.3, p. 36).

Four practice scenarios, two ironic and two sincere, were administered to each child to ensure that they could distinguish between ironic and sincere scenarios. Children fell into four categories according to their level of understanding upon receiving practice trials:

(1) Children who demonstrated full comprehension of the task by accurately identifying communicative intent according to story condition and were administered the entire task.

(2) Children who demonstrated confusion and did not correctly report communicative intent beyond chance until receiving the practice block two or three times and were then administered the entire task.

(3) Children who received the practice block at least three times, did not perform better than chance during any presentation yet, were given a chance to complete one full experimental block to determine whether they should continue the task.

(4) Children who could not or refused to sit and view the task; children who viewed the task but refused to respond to the behavioral question; children who demonstrated no understanding by performing below chance after several attempts at the practice block.

Experimental versions were split between two experimental blocks so that both versions of a scenario never occurred in the same block. Within each block, scenarios were randomized. For every two to three vignettes shown, a rest screen featuring a popular cartoon was displayed for four to five seconds.

Vignettes in the neutral condition contained only two images; the first was the introduction image depicting two characters engaged in an activity (e.g., eating lunch). The second portrayed a continuation of the activity as one character made a remark neutral in tone and prosody (e.g., 'Dad, I'm full.'). There were two versions of six scenarios and in each, only one character's face and verbal response were depicted. In one version, the character wore a face expressing some degree of happiness or joy and in the second, the character wore a face expressing some degree of anger or sadness. Participants were asked 'Did X mean what they said?'. The last image was remained on the monitor for a total of 13 seconds and participants responded in the manner described above. Neutral scenarios were divided between two blocks with each block containing one version of each scenario. For every two vignettes shown, a rest screen featuring a popular cartoon was shown for four to five seconds (see Figure 2.4, p. 37).

Blocks containing ironic-sincere scenarios (two blocks of ten) alternated with the presentation of neutral blocks (two blocks of six) on a rotating schedule across children (e.g., *ABCD, BCDA, CDAB,* etc.,). For each condition, participants had approximately four seconds to respond to the question. If they did not respond within this time, E-Prime was programmed to continue to the next scenario. A lack of response was recorded as missing data.

Developmental Age Tests and Behavioral Assessments

We began audiovisual recording at this point in the session when we rejoined children's parent(s) in the adjacent assessment room to began our series of standardized assessments. Children were seated across from one examiner at a small table that was central to several video cameras mounted on the walls. The PPVT-IV, EOWPVT-IV and the DAS-II were administered first. For all tests, participants began at the age appropriate level and continued until they performed at ceiling. Next, children were offered a short break and upon their return, they received the ToM Battery from a second examiner. All participants began with the Early Battery and the examiner continued as long as the child passed the gatekeeper items within each section of the battery.

Children took another short break and returned to the table to complete their visit with the narrative competence task. The second examiner gave the child a choice between the two wordless picture books and then eased the child into the task by narrating the first one or two episodes of the book. Other children were comfortable beginning the task immediately. After the examiner turned narration over to the child, the examiner only communicated prompts for elaboration (e.g., 'And then what?') or prompts for clarification (e.g., 'What do you mean by this?').

Eye-tracker

Eye-tracking tasks were delivered using a Tobii T60 eye tracker. Children were seated 60-70cm from the eye-tracking monitor and calibrated using five-point calibration; i.e., following a red ball that bounces around the screen and stops at five discrete points on the screen. The T60 has with a 17" screen and a data rate of 60 Hz. It is a noninvasive machine that emits infrared light from a horizontal strip of sensors located above and below the monitor. Patterns of infrared light reflect off of the participant's corneas onto the eye tracker's image which are interpreted by Tobii using mathematical and anatomical algorithms. This approach allows the T60 to make a three-dimensional representation about where each eyeball is located in space and subsequently, the path of their gaze. This hardware was supported by a Dell Inspiron 530 desktop, Microsoft Windows XP Professional 2002 operating system running Tobii Studio 2.1.13 to collect viewer data (www.tobii.com).

The experimental paradigm was delivered by a third server, a Dell Precision M4400 Workstation, Microsoft Windows XP Professional 2002 operating system running E-Prime 2.0 Professional (www.pstnet.com). Audio was delivered using Altec Lansing desktop speakers connected to the M4400 laptop.

Tobii Studio software recorded and translated participants' eye gaze to numerical data. Data included length of time a participant allocated gaze to AOIs during presentation of the entire paradigm which Tobii Studio recorded as video. Using time flags that indicated the beginning and end of portions of the video that were of interest, we designated selections in the video for Tobii to compute gaze allocation data. Tobii Studio provided raw data that were transferred to Microsoft Excel and uploaded into SAS 9.2 for analysis.

Results

Preliminary analysis: Evaluating the overall quality of the data

Eleven of the 50 participants failed to complete the eye-tracking measure due to insufficient understanding of irony (N = 8), challenging behaviors (N = 1), time constraints (N = 1), and recording software malfunction (N = 1). To evaluate differences in global development between the 39 children who completed the eye-tracking measure and the 11 children who did not, we conducted a series of t-tests. Results showed that children who were unable to complete the measure had lower age equivalent scores for expressive language, t (48) = -3.11, $p < .01$, receptive language, t (48) = -2.67, $p = .01$, nonverbal skills, t (48) = -2.05, $p = .05$, and spatial skills, t (48) = -2.35, $p = .02$. No significant differences were found for children's chronological age, developmental standard scores, or SRS T-scores. The means of these age equivalents ranged from 55.5 months for expressive language age equivalent to 68.8 months for nonverbal age equivalent. This suggests that this paradigm is optimal for children at approximately five years of age and older.

First, we assessed children's overall performance on the behavioral question asked at the end of every trial. Judgments concerning communicative intent ('Did X mean what they said?') were evaluated for accuracy. Due to a data collection error, data were missing for one additional participant. Across the ironic and sincere conditions, participants' accuracy ranged between 70-100% (*M* = .95, *SD* = .08).

Preliminary analyses were also conducted to evaluate the quality of the collected eye-tracking data for the 39 children who completed the measure. On-task duration was computed for each trial, defined as the duration during which the participant gazed at the eye-tracker screen. Gaze was measured within a 3.50s time window of interest. Across all conditions, time spent gazing anywhere on the screen averaged 2.87s (*SD* = .47) with mean durations ranging from .99s to 3.49s for all participants. No significant differences in on-task percentage were found between the three experimental conditions (i.e., neutral, sincere, ironic), F (2, 76) = .63, p = .54.

To identify global developmental predictors of On-task duration, we specified a series of mixed models using SAS Proc Mixed. Results showed revealed no significant associations between On-task duration and developmental age equivalent scores, developmental standard scores or SRS T-scores, neither as a main effect nor as a child characteristic by condition interaction effect.

Preliminary Analysis: Evaluating developmental characteristics and addressing potentially confounding variables

Our key analyses concerned the duration of gaze allocation to the Face AOI. In addition to children's on-task duration, we considered two possible confounds associated with our experimental stimuli that may explain variation in gaze allocation to the Face AOI: (1) AOI size, and (2) expressivity of faces. To evaluate these confounds, we fit a series of Proc Mixed models with Face AOI as the outcome. Results showed that gaze allocation to the Face AOI was independently predicted by AOI size, F (1, 990) = 9.33, $p < .00$, expressivity of faces, F (1, 990) = 16.58, $p < .00$ and On-task duration, F (1, 990) = 84.83, $p < .00$. Results also revealed a significant interaction between AOI size and expressivity of faces, F (1, 990) = 10.20, $p < .01$. Faces in small AOIs received considerably less gaze allocation as expressivity increased, i.e., faces displaying more positive affect. Faces in large AOIs received marginally less gaze allocation as expressivity increased. Subsequent analyses controlled for: (1) AOI size, (2) expressivity of faces, (3) On-task duration and (4) AOI size and expressivity interaction.

Next, we evaluated the extent to which children's gaze allocation to the Face AOI was predicted by global developmental measures (e.g., age, age equivalent scores, standard scores), gender, and bilingualism. Results revealed that expressive language standard scores significantly predicted gaze allocation to the Face AOI, F (1, 37) = 10.51, $p < .01$ ($\beta = 0.01$). No other significant relations emerged, as main effect or as child-characteristic by condition interaction effect.

Will gaze allocation to Face AOI in the irony condition differ significantly from the sincere and neutral conditions?

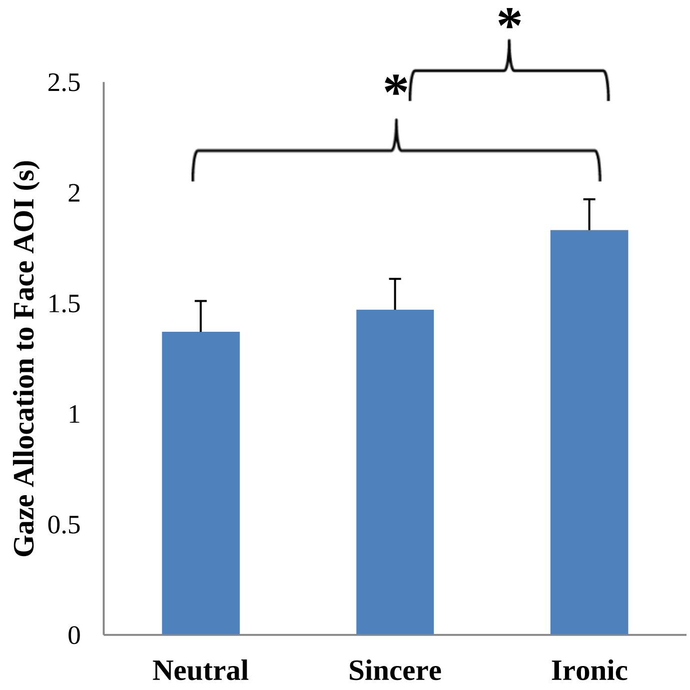
Results thus far allowed us to develop a model that controlled for the independent contributions of four covariates discovered to influence gaze allocation to faces, our target stimuli: (1) AOI size, (2) expressivity of faces, (3) On-task duration and (4) AOI size and expressivity interaction. We added condition to our multipredictor model to test our hypothesis that condition determines gaze allocation to characters' expressive faces. This prediction was confirmed as adding condition to the model significantly improved model fit, χ^2 (2) = 52.5, $p < .01$. On average, participants attended to the Face AOI for an average duration of 1.37s (*SE* = .06) and 1.47s (*SE* = .06) in the neutral and sincere conditions, respectively. Faces in the ironic condition received the most attention as participants attended to the Face AOI for an average of 1.83 (*SE* = .06). Our hypothesis was supported: Tukey-Kramer post-hoc tests of multiple comparisons showed that gaze time allocation to ironic faces differed significantly when compared to neutral and sincere faces, t (74) = -9.22, $p < .01$ and t (74) = -5.34, $p < .01$, respectively (see Figure 4.1).

Is gaze allocation to the ironic Face AOI predicted by individual differences in the expression of traits associated with broad autism phenotype?

Our next hypothesis asserted that performance on the SRS would predict gaze allocation to faces such that individuals who score above cutoff (more symptomatic) would gaze less upon ironic faces. Again, we began with a model that controlled for (1) AOI size, (2) expressivity of faces, (3) On-task duration, (4) interaction between AOI size and expressivity and (5) condition. We tested SRS T-score and it did not prove to be a significant predictor of the Face AOI, F (1, 37) = .42, $p = .52$ and there was no evidence of an interaction between condition and SRS T-score, F (2, 986) = .19, $p = .82$. Our hypothesis that individuals who show traits associated with autism will selectively allocate their gaze less to the Face AOI in ironic scenarios was not supported. Results from our hierarchical analyses thus far are summarized in Table 4.1.

Is gaze allocation to Face AOI across conditions predicted by performance on the Theory of Mind Battery?

We also posited that performance on the Theory of Mind (ToM) Battery would predict attunement to faces such that children who score higher on this battery, i.e., show stronger ToM abilities, demonstrate increased gaze allocation to faces in the ironic condition when determining a speaker’s communicative intent. Given that skills for theory of mind are highly correlated with receptive language, we tested a modified ToM Battery parameter. We began by conducting a simple regression equation testing receptive language age equivalent as a predictor of ToM



Note. * *p* < .01

Figure 4.1. Mean gaze allocation to faces by condition

Battery score to obtain the residuals associated with our model. By disentangling the variance portion of ToM score explained by receptive language age equivalent, we hoped to create a cleaner predictor. When we added our new predictor to a model comprised of our four covariates plus condition, ToM Battery proved significant, F (1, 29) = 5.59, p = .03, and the parameter estimated associated with our new ToM score was negative, β = -.018. A chi-square calculation showed that adding this ToM Battery parameter significantly improved model fit, χ^2 (1) = 509.3, p = .00. We did not discover an interaction between this modified ToM score and condition, F (2, 767) = .99, p = .37.

Table 4.1

Parameter estimates (standard errors in parentheses) and fit statistics from a hierarchical mixed model analysis to determine predictors of visual attention to Face AOI

Variable	Step 1	Step 2	Step 3
Intercept	1.356 (0.045)***	1.642 (0.060)***	1.644 (0.059)***
Screen	0.420 (0.045)***	0.429 (0.048)***	0.423 (0.048)***
AOI size	-0.083 (0.027)**	-0.120 (0.028)***	-0.120 (0.028)***
Expressivity	-0.044 (0.010)***	0.015(0.016)	0.015 (0.016)
AOI size*expressivity	0.042 (0.013)**	0.030 (0.013)**	0.030 (0.013)*
Condition: neutral		-0.461 (0.050)***	-0.462 (0.050)***
Condition: sincere		-0.362 (0.068)***	-0.362 (0.068)***
Condition: ironic		0	0
SRS T-score			0.004 (0.006)
Fit statistic			
- 2 log likelihood	2327.3	2274.7	2274.2

Note. Step 1 includes the four covariates: On-task duration, AOI size, expressivity and AOI*expressivity. Step 2 adds experimental condition and Step 3 adds SRS T-score. For Step 2 and Step 3, the parameters and standard errors associated with the ironic condition are reported as the intercept.

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

Is gaze allocation to Face AOI across conditions predicted by use of internal state language?

Our final hypothesis stated that children's increased use of internal state language would predict greater gaze allocation to faces in the ironic condition when determining ironic intent. To remain consistent with the parameter we used to analyze the effect of ToM Battery on gaze allocation, we created a residual internal state variable that factored out the influence of receptive language. When we added our new predictor to a model comprised of our four covariates plus condition, we found that internal state language had a main effect on gaze allocation to all faces, F (1, 34) = 5.61, p = .02, β = - .044. Children who showed more use of internal state language gazed less at all expressive faces across conditions. A chi-square calculation showed that adding this internal state language parameter significantly improved model fit, χ^2 (1) = 209.2, p = .00. We did not discover an interaction between use of internal state language and condition, F (2, 900) = .90, p = .41.

Discussion

One of the primary aims of this study was to fill in the gap of eye-tracking research dedicated to describing normative eye gaze patterns in response to social material. Children were screened for traits associated with autism and assessed for their theory of mind including their mental state language abilities. They also received the Nonliteral Language paradigm that asked participants' to determine a character's communicative intent at the end of every trial. We sought to create a profile of the visual strategies used by typical children to interpret potentially ironic situations according to social impairments expressed by the general population. We believe testing using this behavioral approach can help to uncover endophenotypes of autism that can lead researchers toward the genetic origins of social deficits.

The data confirmed our first hypothesis that when determining communicative intent, children allocated gaze significantly longer to characters' faces when they made an ironic statement versus a sincere or neutral statement. Our second prediction asserted that as children's presentation of traits associated with autism increase, their gaze allocation to ironic faces would significantly decrease compared to children whose scores fell in the normal range on the Social Responsiveness Scale. Our theory was not supported and there were no individual differences in gaze allocation as a function of SRS score. Our third hypothesis predicted that ToM Battery score would predict children's gaze allocation to faces such that children with increased theory of mind would be more visually attuned to ironic faces. This prediction was not supported yet, we found a significant relationship in the opposite direction of what we anticipated: Children with increased theory of mind skills gazed less upon all faces across conditions. Lastly, we discovered that children who demonstrated increased use of internal state language also gazed less upon all faces across conditions. The latter two findings point to interesting implications for the study of autism endophenotypes in the population of typical children.

The tendency for the eye to be drawn to faces when presented with a social material fits the early profile of healthy development (Hoehl et al., 2009). Our finding concerning increased gaze allocation to ironic faces reflects previous research reporting that children in our sample's age range can accurately distinguish nonliteral irony and report that a character did not mean what they said, even if they are still a few years away from providing coherent justifications for their decisions (Ackerman, 1981; Winner & Leekam, 1991; Glenwright & Pexman, 2010). Notably, a new longitudinal paradigm that investigated children's appreciation of irony through observation sessions in the home indicated that as early as four years of age, children can understand and sometimes even use irony (Recchia, Howe, Ross & Alexander, 2010). This accounts for the ability of a portion of the youngest children in our sample to follow our paradigm and accurately identify communicative intent. Children's capacity to do so may lie in increased gaze allocation to the faces of ironic speakers versus literal speakers (e.g., sincere and neutral speakers). Young children's impulse when faced with irony is the tendency to take a speaker's words literally or interpret them as lies, particularly with sarcasm, by failing to decouple semantics from intention (Climie et al., 2008; Winner, Brownell, Happe, Blum & Pincus, 1998; Recchia et al., 2010). Our sample's strategy of greater visual attention to the ironic face, an available salient cue, appeared to compensate for this potential pitfall.

Our second hypothesis was motivated by research showing that individuals with autism and individuals who express the broad autism phenotype show atypical face processing strategies. Generally, these approaches are characterized by allocating increased gaze to areas thought to offer little to no social salience such as, a speaker's mouth, hair or the backdrop of a social scene (Osterling, Dawson & Munson, 2002; Klin et al., 2002; Chawarska & Shic, 2009; Wallace et al., 2010). A possible explanation for the lack of a relation between the two may be due to the nature of our ironic stimuli. Very often, the intention of an ironic remark is to convey an insult or a bit of criticism. It has been proposed that the positive tone of the words mutes this negative underlying intention (Dews & Winner, 1995). Our characters' ironic statements are delivered in dramatic and campy prosody that may explicitly subvert this muting function. Prosody then becomes an equally strong or stronger cue than the speaker's expressive face. Consequently, a 'hacking' strategy may be available to children who might otherwise struggle to comprehend ironic remarks characterized by a deadpan tone and less exaggerated prosody. Children with autism-like traits may have learned to anchor their performance in socially related situations on alternative cues that allow them to avoid direct referencing to faces (Frith, Happe & Siddons, 1994; Dews et al., 1995; Dews, Winner, Kaplan, Rosenblatt, Hunt, Lim, McGovern, Qualtier & Smarsh, 1996; Milosky & Ford, 1997; Nakassis & Snekeker, 2002).

It is also possible that the items from the Social Responsiveness Scale do not collectively capture the social traits that optimally predict visual scanning strategies to expressive faces. SRS scores were based on a parent's report of their child's personal preferences and social abilities. One possible issue with using this inventory is that in contrast to earlier reports, parent SRS reports have been found to have remarkably low correlation with teacher SRS reports by time, symptom severity and subscale when evaluating the same child. This is suggestive of several possibilities. Perhaps response bias on the part of either the teacher or the parent is responsible for this discrepancy. It may also be the case that a child's social functioning varies based on the environment they are, i.e., the child may appear different at home versus school leading to conflicting reports from the parent and teacher. Previous reports of high parent-teacher agreement on the SRS have also not been replicated in socio-economically and culturally diverse settings. If the reliability of the SRS is in question, that would have critical implications for our intentions to utilize SRS scores as a pure measure of children's attunement to social stimuli (Reisinger, Ming, Harker and Mandell, 2011).

In order to administer the Nonliteral Language Paradigm in a naturalistic way, we did not systematically standardize several aspects of our cartoon images. Thus in our statistical analyses of gaze allocation to expressive faces, we were required to control for several covariates. For example, we allowed characters' expressions and the size of their faces to vary across conditions. We also had to account for individual differences in visual attention to the eye-tracker screen during viewing of the paradigm by controlling for 'On-task' duration. As such, our baseline model contained these three covariates plus the interaction between AOI size and expressivity of faces, a fourth covariate. We must consider the possibility that some degree of noise was introduced into our model in controlling for these potential confounds. This may have interfered with or masked our ability to detect the influence of children's SRS scores on gaze allocation to faces.

Our position that increased theory of mind skills would lead to higher attunement to ironic faces during the ironic condition was rooted in longstanding beliefs concerning the function of theory of mind. Happe (1993) asserted that theory of mind is related to abilities essential to understanding others' speech, goals, social action and social reasoning. These abilities are believed to be encouraged by social interaction which helps us learn to decode salient signals in the environment, provided we maintain attention on them (Kleinknecht & Beike, 2004). As such, we believed children would allocate more gaze to ironic faces, thus reflecting a strategy that likely contributed to their advanced skills for understanding social meaning in the first place (Hoehl et al., 2009). Instead, it may be that their increased skills make them more efficient face experts that can very quickly identify the meaning of facial expressions and a speaker's corresponding intentions. This may afford them greater freedom to peruse entire social scenes and still reliably report ambiguous communicative intent without intently fixating on speakers' faces (Ribiero & Fearon, 2010).

Based on the much of the same reasoning we presumed that children who used more internal state language as measured during a storybook narrative would also be more visually attuned to ironic faces to determine communicative intent. Children's ability to use internal state language has been shown to be positively correlated with a capacity for theory of mind (Capps et al., 2000; Moore, Pure & Furrow, 1990; Olineck & Poulin-Dubois, 2007; Baron-Cohen et al., 1986). It has also been argued that a child's knowledge for cognitive states (e.g., what others *think, know,* or *remember),* though less observable than emotional states, is essential to their development of theory of mind (Olson, 1988). All internal state terms, cognitive and emotional, share the quality of being able to be represented internally via language so as to be appreciated as the experience of another who is separate from us (Astington et al. 1999). In fact, use of internal state language can even be considered a high-level measure of children's theory of mind. The relation between our internal state language measure and gaze allocation to all faces showed was consistent with the effect we found between theory of mind and gaze allocation to faces. We propose that advanced knowledge and use of internal state language serves to make an individual more socially sophisticated. In turn, this leads to greater intuitive attunement to the internal states of others that dispenses with the need for intensive fixation on faces of others to interpret their communicative intentions. These individuals have the opportunity to allocate their visual attention more freely upon social scenes while still receiving important information conveyed by those involved in an interaction. In the relations between gaze allocation to faces on the Nonliteral Language paradigm and our theory of mind and internal state language measures, respectively, decreased performance on these behavioral measures predicted increased gaze allocation to expressive faces to determine communicative intent. This may be suggestive of an autism endophenotype characterized by difficulty interpreting social situations.

In the search for autism endophenotypes using a typical population, it is clear that the choice of inventory to determine the degree of impairment and corresponding behavioral assessments must be carefully matched. To reiterate, the SRS may be too comprehensive or reflect response bias that interferes with its ability to predict highly specific responses in social situations. On the other hand, the analyses of our theory of mind and internal state language measures demonstrated links to visual attention to socially salient cues however, these were limited. We did not uncover behavioral performance that selectively predicted gaze according to social condition. It is unclear if these limits were due to the qualities inherent in these behavioral measures or a consequence of the noise in our eye-tracking paradigm. Future research with a standardized version of the Nonliteral Language paradigm and a series of assessments targeting emotional understanding, theory of mind and internal state language can shed light on this question.

V. General Discussion

Healthy social development is contingent on the reciprocal relationships among many cognitive abilities from attention to linguistic competence. Humans are predisposed to acquiring these abilities that have enabled us to construct a world rife with a multitude of cultures, each governed by their own unique rules for behavior and communication. For example, a fascinating cross-species study showed that preliterate children aged 2.5 years, vastly outperformed chimpanzees and orangutans of various ages on tasks targeting theory of mind, communication and social learning thus, sharply distinguishing us from even our closest primate relatives. It is possible that a highly evolved understanding for hidden causal forces underlies the sophistication of our exceptional social skills and permits us to interact with the world in a way that is uniquely human (Herrmann, Call, Hernandez-Lloreda, Hare & Tomasello, 2007). The current studies show that humans' inherent abilities for acquiring and developing advanced theory of mind and language permit us to communicate verbally and nonverbally in complex ways.

Our results yielded interesting and important findings with respect to our hypotheses between social deficits associated with autism and gaze allocation to faces. At first glance, we believed that the SRS and the BAPQ for children and adults, respectively, would capture social impairments that predict gaze allocation to faces of speakers who had different communicative intention. Specifically, we posited that according to these inventories, greater expression of the broad autism phenotype would predict less gaze allocation to faces of speakers delivering ironic remarks. Our finding that neither the SRS nor the BAPQ were ideal tools to distinguish varying amounts of visual attention between conditions suggests that they may not capture the characteristics that may drive gaze allocation during these nuances social interactions. This finding has important implications for researchers pursuing endophenotypes: It suggests that the behavioral tool that is most sensitive to detecting the deficit in question may not be a global social assessment. Instead, a behavioral inventory targeting a narrower range of highly specific traits may be likely more to distinguish task performance on eye-tracking measures targeting social responses across individuals with varying skill sets.

Upon collecting gaze data on the Nonliteral Language paradigm from child and adult samples, we had the opportunity to explore an additional hypothesis concerning a comparison of the samples' eye gaze patterns. We believed that gaze allocation to the Face AOI would vary between samples such that children would spend more time fixating upon faces across all conditions to accurately interpret communicative intent. A Proc Mixed model that controlled for (1) AOI size, (2) expressivity of faces and (3) On-task duration across both samples revealed that adults attended to all faces more than children, F (1, 77) = 10.60, $p < .01$. On average, adults spent 1.77s (*SE* = .08) across all Face AOIs versus children's average of 1.47s (*SE* = .05). Condition did not interact with sample 'assignment' and neither group preferentially allocated more gaze to faces in one condition over another.

Our prediction that children would allocate more eye gaze to all expressive faces was based on the assumption that as novices in the social world, children would require concentrated visual attention upon a speaker's face to determine communicative intent. In contrast, we supposed that experienced adults would be more naturally attuned to socially salient cues and would require less intense focus upon faces than children. Results were contrary to our expectations and our adult sample spent more time attending to faces than children across all conditions. This is a challenging finding to address given that in comparison to children, adults have developmentally mature theory of mind. We discovered that children with higher ToM Battery scores actually gazed less at faces so we would expect that adults with full-fledged theory of mind would gaze at faces for even less than children. The relations may not be this straightforward, however, and these gaze allocation outcomes may be based on some other aspect of adults' and children's cognitive profile that we did not address.

To reiterate our findings with respect to autism-like traits and gaze allocation to expressive faces, we discovered that adults who displayed social impairment as indicated by a high total average score on the BAPQ, allocated less gaze to all expressive faces across conditions. Conversely, children who showed social impairment as indicated by decreased theory of mind and use of internal state language, allocated more gaze to all expressive faces across conditions. We propose the possibility that gaze allocation patterns for typical adults and children with social deficits take a different shape according to developmental shifts that occur over time. These impairments may predict that gaze allocation outcomes upon static social stimuli become less fixated with age. One possible explanation for this may be that children compensate for social difficulties using a strategy that emphasizes increased visual attention to faces to extract necessary information from their interactions with others. However, this may change as children age and gain experience in the world. They may learn to focus on alternative salient cues (e.g., tone of voice or body language) that can provide them with sufficient social information through different means. Consequently, socially impaired adults may have discovered how to avoid cueing into a speaker's face to discern their internal state and interpret their communicative intention (Harms, Martin & Wallace, 2010).

Several studies have arrived at findings that support our proposition. It has long been suggested that individuals affected with ASD may not experience the same intrinsic pleasurable reward from interaction with others as healthy typical individuals do. The social motivation hypothesis posits that in children with autism, motivation to attend to faces is low as a result of the absence of implicit reward. The young system may then go on to develop in an atypical fashion characterized by increasing disinterest in faces, a lack of expertise for interpreting faces and less efficient neural correlates underlying face processing (Dawson, Webb, McPartland, 2005). Two pioneering studies assessed responses to social reward using fMRI. Dichter et al. examined adults using a social reward paradigm and discovered that when compared to typical adults, adults with autism showed hyperactivation in bilateral amygdala (Dichter, Richey, Rittenberg, Sabatino & Bodfish, 2012). Upon investigating adolescents on another variation of a reward paradigm, Scott-Van Zeeland et al. revealed group differences in autism and typical samples in frontostriatal activation during a social rewarded learning task (Scott-Van Zeeland, Dapretto, Ghahremani, Poldrack & Bookheimer, 2010). These results highlight the neural correlates that may account for the shift toward decreased gaze allocation to faces observed in adults with impoverished social skills.

It is important to note, however, that the suggestions concerning our findings be considered with utmost caution. While it is plausible that our results are suggestive of an autism endophenotype characterized by atypical gaze patterns to expressive faces, our study is a preliminary effort. The potential for this endophenotype to be further characterized by a developmental shift also warrants much further exploration. In our assessments, we did not perform exhaustive clinical examinations into the genetic histories of our participants. This would be an important element of future studies focusing on atypical gaze patterns and BAP. Also, we did not account for the effect of other influential variables such as social motivation or full-scale IQ, for example. Recent research by Rice and colleagues demonstrated that within an autism sample of children, cognitive profiles further distinguished the sample into subgroups. Rice et al. discovered that gaze patterns upon faces varied in their autism sample as a function of their IQ scores (Rice, Moriuchi, Jones & Klin, 2012). Consideration of factors such as these that can bear on performance within a single clinical group is essential to properly characterize performance in future studies.

Autism spectrum disorder research hailing from discrete disciplines is converging to suggest that the presence of atypical gaze patterns is a viable endophenotype. Based on the results we arrived at that are detailed above, the performance of our child an adult samples on the Nonliteral Language paradigm is suggestive of population-specific endophenotypes. In order for the examination of endophenotypes to meaningfully guide genetic analysis, Gottesman and Gould (2003) suggest criteria: The endophenotype is (1) associated with illness in the population, (2) it is heritable, (3) it manifests whether or not the illness is active and (4) within families, the endophenotype and illness co-segregate. Again, future studies investigating the performance of typical children and/or children with autism and their parents or siblings could confirm the existence of this putative endophenotype.

Appendix

Theory of Mind Battery

I. Early Battery

This module begins with the *Desire Task* (Wellman & Wooley, 1990) which is the gatekeeper task for the Early Battery. Using simple yet colorful drawings, the experimenter narrates two separate stories that describe a character's wishes to find an item for a specific purpose (e.g. 'Annie wants to find her crayons to take them to a friend's house'). The character fails to find this object but during their search, they discover another object. First, the child will be asked to recall what the character wanted to find. The test questions taps the child's knowledge of the agent's original desire and based on this, asks the child to predict the character's behavior. The participant is then asked to explain their prediction (Total points possible = 4 points):

- *Test question 1*: Will they continue searching or will they do something else (e.g., go to a friend's house)? (1 point)
- *Test question 2*: Why? (1 point)

II. Basic Battery

The gatekeeper task for this section is the *First-Order False Belief* task. This is based on the classic first-order false belief task introduced by Wimmer and Perner (1983) designed to gauge the recognition that another may have an incorrect belief that is incongruent with reality. The examiner leads the child through two separate scenarios using simple 2-D cutout characters manipulated upon drawings of common locations (e.g., living room, kitchen). Each vignette describes a situation that begins with two characters sharing knowledge about the location of a specific item. Next, one character exits the scene and the remaining character changes the location of the item to a new location. Each story ends with the return of the character that left. The participant receives two control questions asking: (1) ‘Where did the characters put the object before X left?’ and (2) ‘Where is the object now?’. If the child answers these incorrectly, the examiner reminds them of the correct locations. Two test questions target the child’s knowledge about what the protagonist believes about the object’s location. A third question asks the child to justify their prediction regarding where X will look (Total points possible = 6 points):

- *Test question 1*: Does X know the current location of the item? (1 point)
- *Test question 2*: Where will he/she look for it? (1 point)
- *Test question 3*: Why will X look there? (1 point)

The *Unexpected Contents Task* is a playful exercise in which a familiar labeled box/container is shown to the child and the examiner asks them about its contents (Perner, Leekam & Wimmer, 1987). Three boxed items are presented and the script is repeated for each container. Upon holding up the box to the child, the examiner asks, ‘What do you say is in here?’ The child often gives an answer according to the observable label or shape of the container (e.g., egg carton) and the examiner feigns surprise while she revealing the contents of the box to be something much different than expected (e.g., a matchbox car in a crayon box). The examiner returns the item to the box, closes it and asks a control question targeting memory: ‘What’s really in this box?’ The examiner reminds the child of the correct object if the child answers incorrectly. This is followed by two test questions that require the child to negotiate competing perspectives about the container’s contents: Their perspective versus that of someone who is not in the room (e.g., their best friend) (Total points possible = 6 points):

- *Test question 1*: If I show this box all closed up to X (a friend who is not present), will X know what is inside? (1 point)

- *Test question 2*: When X first looks at the box, what will X say is inside? (1 point)

III. Advanced Battery

This section contains three tasks and begins with *Second-Order False Belief Task*. Second-order false belief tests expand on the concept of false belief by probing a step further and evaluating whether a participant can conceive of someone's perspective of *another person's* perspective (Sullivan, Zaitchik & Tager-Flusberg, 1994). A flipbook with inked cartoon style drawings depicts two separate scenarios. Both stories involve a child inadvertently finding out about a surprise gift from their parent. The parent is ignorant about this and maintains the false belief that the child does not know about this forthcoming gift. The final scene of each story portrays the parent in conversation with a third party (e.g., grandma) and the participant is asked to predict what they what they think the parent believes about their child's knowledge. The first control question targets the participant's understanding of the reality of the situation: 'What is Dad really getting X?' Next, the participant is asked about the child's initial false belief (before the surprise gift was discovered):

- *Test question 1*: 'What does X think Dad got her for her birthday?' (1 point).

A second control question asks the participant to report whether or not parent knows the child discovered the gift: 'Does Dad know that X saw her gift in the garage?'. If the child answers any of the control questions inaccurately, the examiner reminds them of the correct answers. Finally, the participant is asked to describe and justify the parent's second-order false belief when the third party enters the story and asks the parent about the situation (Total points possible = 6 points):

- *Test question 2*: When grandma (third party) asks Dad, 'Does X know what you got her for her birthday?', what does dad say? (1 point)
- *Justification*: Why does Dad say that? (1 point)

The *Lies and Jokes Task* consists of four scenarios; two scenarios end in jokes and two end in lies. In all scenarios, a child character is asked by their parent to perform an undesirable chore. Each story ends with a false statement as the child declares that they've completed the chore (e.g., 'I did a really good job putting the dishes in the dishwasher!') (Sullivan, Winner & Hopfield, 1995). In the lie condition, this statement is uttered without the child knowing that their parent knows the truth. Conversely, in the joke condition, the child is being ironic as he/she delivers the statement with full knowledge that their parent is aware of the truth. Control questions assess if the participant recalls whether or not the child performed the chore as well as their knowledge concerning who knows what about the situation (e.g., 'Does Rob know that Dad sees the clothes on the floor?' and 'Does Rob know that Dad knows that he did not pick up his clothes?'). Test questions measures participants' understanding of the circumstantial contexts that distinguish the character's statement as a lie or joke and their ability to justify their response (Total points possible = 6 points):

- *Test question 1*: Was Rob lying or kidding? (1 point)
- *Justification*: How can you tell? (1 point)

The final test is the *Moral Responsibility Task* which contains four scenarios; two are characterized as 'commitment' stories and the other two are characterized as 'act of god' stories (Mant et al., 1988). Each story describes a situation in which two friends make plans to go out (e.g., movies, library) but only one friend actually keeps the commitment and shows up at the designated meeting place. In the commitment story, the absentee friend simply changes their mind and chooses not to go meet their friend. However, in the 'act of god' stories, the outcome is the same but the absentee friend fails to show due to an uncontrollable circumstance (e.g., the car won't start). Two control questions ascertain the child's knowledge about why one friend bailed

(‘Why didn’t X go?’) and if the friend that bailed knows that the other friend is still going to the planned meeting spot (‘Does X know that Y was going to the movies?’). A third control question taps the child’s memory for the reason that X did not go to the meeting spot (e.g. ‘Why didn’t X go to the movies?’) The child is reminded of the correct answers if they respond incorrectly. Next, two test questions are asked that target the child’s perception of the ethics of the character’s behaviors based on the circumstances of the story. A final test question asks the participant to justify their response (Total points possible = 6 points):

- *Test question 1*: When X didn’t go to the movies, was it good not to go, bad not to go or in between? (.5 points/1point)
- *Test question 2*: Was it really (good/bad), sort of (good/bad) or in between? (.5 points)
- *Justification*: Why?

There is a caveat in the scoring in the ‘act of god’ scenario as there are two ways to gain full credit. The child can answer ‘in between’ for one full point and then move onto the justification question. Or, the child can answer ‘bad’ and ‘sort of bad’ for .5 points on test questions one and two, respectively.

References

Abrahams, B. S., & Geschwind, D.H. (2008). Advances in autism genetics: On the threshold of new neurobiology. *Nature Reviews, 9*, 341-355.

Ackerman, B. (1981). Young children's understanding of a speaker's intentional use of a false utterance. *Developmental Psychology, 31*, 472-480.

Association, A. P. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., Text Rev. ed.). Washington D.C.: Author.

Astington, J. W. (1990). In B. P. Britton, A. (Ed.), *Narrative thought and narrative language*. Hillsdale, NJ: Erlbaum.

Astington, J. W. (1998). Theory of Mind, Humpty Dumpty, and the Icebox. *Human Development, 41*, 30-39

Astington, J.W., & Pelletier, J. (1997). *Metacognitive language in children's narrative comprehension and production.* Paper presented at the Annual Conference of the Canadian Society for the Study of Education, St. John's, NF.

Astington, J. W., & Jenkins, J.M. (1999). A longitudinal study of the relation between language and theory of mind development. *Developmental Psychology, 35*(5), 1311-1320.

Astington, J. W., & Baird, J.A. (2005). Introduction: Why language matters for theory of mind. In J. W. Astington, Baird, J.A. (Ed.), *Why language matters for theory of mind* (pp. 3-25). New York Oxford University Press, Inc.

Bailey, A., Le Couteur A., Gottesman, I., Bolton, P., Siminoff, E., Yuzda, E., & Rutter M. (1995). Autism as a strongly genetic disorder- Evidence from a British twin study. *Psychological Medicine, 25*(1), 63-77.

Bamberg, M., & Damrad-Frye, R. (1991). On the ability to provide evaluative comments: Further explorations of children's narrative competencies. *Journal of Child Language, 18*, 689-710.

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1985). Does the autistic child have a "theory of mind". *Cognition, 21*, 37-46.

Baron-Cohen, S., Leslie, A.M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology, 4*, 113-125.

Baron-Cohen, S., Wheelwright, S., & Jolliffe, T. (1997). Is there a 'language of the eyes'? Evidence from normal adults, and adults with autism or Asperger syndrome. *Visual Cognition, 4*(3), 311-331.

Baron-Cohen, S., & Hammer, J. (1997). Parents of children with Asperger Syndrome: What is the cognitive phenotype? *Journal of Cognitive Neuroscience, 9*(4), 548-554.

Baron-Cohen, S. (2000). In J. B. Hartley, A. (Ed.), *The Applied Psychologist*. U.K.: Open University Press.

Baron-Cohen, S. (2001). Theory of mind and autism: A review. *International Review of Research in Mental Retardation, 23*, 169-184.

Belmonte, M. K., Gomot, M., & Baron-Cohen, S. (2010). Visual attention in autism families: 'Unaffected' sibs share atypical frontal activation. *The Journal of Child Psychology and Psychiatry, 51*(3), 259-276.

Ben-Yizhak N., Y., N., Seidman, I., Alon, R., Lord, C., & Sigman, M. (2011). Pragmatic language and school related linguistic abilities in siblings of children with autism. *Journal of Autism and Developmental Disorders, 41*, 750-760.

Berman, R. (1988). On the ability to relate events in narrative. *Discourse Processes, 11*, 469-497.

Bernier, R., Dawson, G., Webb, S., & Murias, M. (2007). EEG mu rhythm and imitation impairments in individuals with autism spectrum disorder *Brain and Cognition 64*, 228-237.

Bishop, D., & Donlan, C. (2005). The role of syntax in encoding and recall of pictorial narratives: Evidence form specific language impairment. *British Journal of Developmental Psychology, 23*, 25-46.

Bolton, P., Macdonald, H., Pickles, A., Rios, P., Goode, S., Crowson, M., Bailey, A., & Rutter, M. (1994). A case-control family history study of autism. *Journal of Child Psychology and Psychiatry, 35*(5), 877-900.

Boraston, Z., & Blakemore, S-J. (2007). The application of eye-tracking technology in the study of autism. *The Journal of Physiology, 581*, 893-898.

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.

Carr, L., Iacoboni, M., Dubeau, M-C., Mazziota, J.C., & Lenzi, G.L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences, 100*(9), 5497-5502.

Chawarska, K., & Shic, F. (2009). Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year old children with Autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *39*, 1663-1672.

Climie, E. A., & Pexman, P.M. (2008). Eye gaze provides a window on children's understanding of verbal irony *Journal of cognition and development*, *9*(3), 257-285.

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

Constantino, J.N., & Gruber, C.P. (2005) *Social responsiveness scale (SRS) manual*. Los Angeles, CA: Western Psychological Services.

Constantino, J. N., & Todd, R.D. (2005). Intergenerational transmission of subthreshold autistic traits in the general population. *Biological Psychiatry, 57*, 655-660

Constantino, J. N., Lajonchere, L., Lutz, M., Gray, A., Abbachi, A., McKenna, K., Singh, D., & Todd, R.D. (2006). Autistic social impairment in the siblings of children with pervasive developmental disorder. *American Journal of Psychiatry, 163*, 294-296.

Dalton, K. M., Nacewicz, B.M., Alexander, A.L., & Davidson, R.J. (2007). Gaze-fixation, brain activation, and amygdala volume in unaffected siblings of individuals with autism. *Biological Psychiatry*, *61*(512-520).

Dapretto, M., Davies, M.S., Pfeifer, J.H., Scott, A.A., Sigman, M., Bookheimer, S.Y., & Iacoboni, M. (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorder. *Nature Reviews Neuroscience*, *9,* 28-30.

Dawson, G., Meltzoff, A.N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders, 28*(6), 479-485.

Dawson, G., Ashman, S.B., & Carver, L.J. (2000). The role of early experience in shaping behavioral and brain development and its implications for social policy. *Development and Psychopathology, 12*, 695-712.

Dawson, G., Webb, S.J., & McPartland, J. (2005). Understanding the nature of face processing impairment in autism: Insights from behavioral and electrophysiological studies. *Developmental Neuropsychology, 27*(3), 403-424.

de Villiers, J. (2007). The interface of language and theory of mind. *Lingua, 117*, 1858-1878.

de Villiers, J. G., & Pyers, J.E. (2002). Complements to cognition: A longitudinal study of the relationship between complex syntax and false-belief understanding. *Cognitive Development*, *17*, 1037-1060.

Demorest, A., Silberstein, L., Gardner, H., & Winner, E. (1983). Telling it as it isn't: Children's understanding of figurative language. *Journal of Developmental Psychology, 1*, 121-134.

Demorest, A., Meyer, C., & Phelps, E. (1984). Words speak louder than actions: Understanding deliberately false remarks. *Child Development, 55*, 1527-1534.

Devries, R. (1970). The development of role-taking as reflected by behavior of bright, average, and retarded children in a social guessing game. *Child Development, 41*, 759-770.

Dews, S., & Winner, E. (1995). Muting the meaning: A social function of irony. *Metaphor and Symbolic Activity, 10*(1), 3-19.

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

Dews, S., & Winner, E. (1999). Obligatory processing of literal and nonliteral meanings in verbal irony. *Journal of Pragmatics, 31*, 1579-1599.

Dichter, G. S., Richey, J.A., Rittenberg, A.M., Sabatino, A., & Bodfish, J.W. (2012). Reward circuitry function in autism during face anticipation and outcomes. *Journal of Autism and Developmental Disorders, 42*, 147-160.

Eigsti, I. M., de Marchena, A.B., Schuh, J.M., & Kelley, E. (2011). Language acquisition in Autism spectrum disorder: A developmental review. *Research in Autism spectrum disorder, 5*, 681-691.

Elliot, C. D. (1990). The nature and structure of children's abilities: Evidence from the Differential Ability Scales. *Journal of Psychoeducational Assessment, 8*, 376-390.

Farroni, T., Csibra, G., Simion, F., & Johnson, M.H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences, 99*(14), 9602-9605.

Farroni, T., Menon, E., & Johnson, M.H. (2006). Factors influencing newborns' preference for faces with eye contact. *Journal of Experimental Child Psychology, 95*(4), 298-308.

Filippova, E., & Astington, J.W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development, 79*(1), 126-138.

Fivush, R. (2011). The development of autobiographical memory. *Annual Review of Psychology, 62*, 559-582.

Frank, M. C., Vul, E., & Johnson, S.P. (2009). Development of infants' attention to faces during the first year. *Cognition, 110*, 160-170.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing *Neuron, 50*, 531-534

Frith, U., Happe, F., & Siddons, F. (1994). Autism and theory of mind in everyday life. *Social Development, 3*(2), 108-124.

Garfield, J. L., Peterson, C.C., & Perry, T. (2001). Social cognition, language acquisition and the development of theory of mind. *Mind & Language, 16*(5), 494-541.

Giora, R. (1995). On irony and negation. *Discourse Processes, 19*(239-264).

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*(1), 3-55.

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

Goffmann, E. (1974). *Frame analysis: An essay on the organization of experience*. New York: Harper & Row.

Gokcen, S., Bora, E., Eremis, S. Kesikci, H., & Aydin, C. (2009). Theory of mind and verbal working memory deficits in parents of autistic children. *Psychiatry Research, 166*(1), 46-53.

Gopnik, A. (1990). Developing the idea of intentionality: Children's theories of mind. *Canadian Journal of Philosophy, 20*, 89-114.

Gottesman, I. I., & Gould, T.D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry, 160*(4), 636-645.

Guajardo, N. R., & Watson, A.C. (2002). Narrative discourse and theory of mind development. *The Journal of Genetic Psychology, 163*(3), 305-325.

Happe, F., Ronald, A., & Plomin, R. (2006). Time to give up on a single explanation for autism. *Nature Neuroscience, 9*(10), 1218-1220.

Happe, F. G. E. (1995). The role of verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*(843-855).

Harms, M. B., Martin, A., & Wallace, G.L. (2010). Facial emotion recognition in Autism spectrum disorder: A review of behavioral and neuroimaging studies. *Neuropsychology Review, 20*, 290-322.

Harris, P. L. (2005). Convention, pretense and Theory of Mind. In J. W. B. Astington, J. (Ed.), *Why Language Matters for Theory of Mind* (pp. 70-83). New York: Oxford University Press.

Haxby, J. V., Hoffman, E.A., & Gobbini, M.I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223-233.

Heilmann, J., Miller, J.F., Nockerts, A., & Dunaway, C. (2010). Properties of the narrative scoring scheme using narrative retells in young school-age children. *American Journal of Speech-Language Pathology, 19*, 154-166.

Heiser, M., Iacoboni, M., Maeda, F., Marcus, J., & Mazziotta, J.C. (2003). The essential role of Broca's area in imitation *European Journal of Neuroscience, 17*, 1123-1128.

Hernandez, N., Metzger, A., Magne, R., Bonnet-Brilhault, F., Roux, S., Barthelemy, C., & Martineau, J. (2009). Exploration of core features of a human face by healthy and autistic adults analyzed by visual scanning. *Neuropsychologia, 47*, 1004-1012.

Herrmann, E., Call, J., Hernandez-Lloreda, M.V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*, 1360-1365.

Hoehl, S., Parise, E., Palumbo, L., Reid, V.M., Handl, A., & Striano, T. (2009). Looking at eye gaze processing and its neural correlates in infancy- Implications for social development and autism spectrum disorder. *Child Development, 80*(4), 968-985.

Hood, B. M., Willen, J.D., & Driver, J. (1998). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science, 9*(131), 131-134.

Hughes, D., & Dunn, C. (1997). "Pretend you didn't know": Young children's talk about mental states in pretend play. *Cognitive Development, 12*, 477-499.

Hurley, R. S. E., Losh, M., Parlier, M., Reznick, J.S., & Piven, J. (2007). The broad autism phenotype questionnaire. *Journal of Autism and Developmental Disorders, 37*, 1679-1690.

Jenkins, J. M., & Astington, J.W. (1996). Cognitive factors and family structure associated with theory of mind development in young children. *Developmental Psychology, 32*, 70-78.

Kaiser, M. D., Hudac, C.M., Shultz, S., Lee, S.M., Cheung, C., Berken, A.M., Deen, B., Pitskel, N.B., Sugrue, D.R., Voos, A.C., Saulnier, C.A., Ventola, P., Wolf, J.M., Klin, A., Vander Wyk, B.C., & Pelphrey, K.A. (2010). Neural signatures of autism. *Proceedings of the National Academy of Sciences, 107*(49), 21223-21228.

Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MA: MIT Press.

Kavanaugh, R. D., Eizenman, D.R., & Harris, P.L. (1997). Young children's understanding of pretense expressions of independent agency. *Developmental Psychology, 33*(5), 764-770.

Kirchner, J. C., Hatri, A., Heekeren, H.R., & Dziobek, I. (2011). Autistic symptomatology, face processing abilities, and eye fixation patterns. *Journal of Autism and Developmental Disorders, 41*, 158-167.

Kleinknecht, E., & Beike, D.R. (2004). How knowing and doing inform and autobiography: Relations among preschoolers' theory of mind, narrative and event memory skills. *Applied Cognitive Psychology, 18,* 745-764.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59*, 809-816.

Kuhl, P. (2007). Is speech learning 'gated' by the social brain? *Developmental Science, 10*(1), 110-120.

LaBounty, J., Wellman, H.M., Olson, S., Lagattuta, K., & Liu, D. (2008). Mothers' *and* Fathers' use of internal state talk with their young children. *Social Development, 17*(4), 757-775.

Labov, W. (2006). Narrative pre-construction. *Narrative Inquiry, 16*(1), 37-45.

Lagattuta, K. H., & Wellman, H.M. (2001). Thinking about the past: Young children's knowledge about links between past events, thinking, and emotion. *Child Development, 72*(82-102).

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind". *Psychological Review, 94*(4), 412-426.

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

Losh, M., & Piven, J. (2007). Social-cognition and the broad autism phenotype: Identifying genetically meaningful phenotypes. *Journal of child Psychology and Psychiatry, 48*(1), 105-112.

Losh, M., Sullivan, P.F., Trembath, D., & Piven, J. (2008). Current developments in the genetics of autism: From phenome to genome. *Journal of Neuropathology and Experimental Neurology, 67*(9), 829-837.

Mant, C. M., & Perner, J. (1988). The child's understanding of commitment. *Developmental Psychology, 24*(3), 343-351.

Markowiak, A. N. (2005). *Narrative comprehension in kindergarten: An analysis of talk about narratives by children differing in early literacy development.* University of Sydney, Sydney.

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

Mayer, M. (1973). *Frog on his own*. New York City: Dial Books for Young Readers.

Meltzoff, A. N., & Moore, K. (1983). Newborn infants imitate adult facial gestures. *Child Development*, *54*(3), 702-709.

Milosky, L. M., & Ford, J.A. (1997). The role of prosody in children's inferences of ironic intent. *Discourse Processes, 23,* 47-61.

Moore, C., Pure, K., & Furrow, D. (1990). Children's understanding of the modal expressions of speaker certainty and uncertainty and its relation to the development of theory of mind. *Child Development, 61*, 722-730.

Morton, J. B., & Treghub, S.E. (2001). Children's understanding of emotion in speech. *Child Development, 72*(3), 834-843.

Nakassis, C., & Snedeker, J. (2001). *Beyond sarcasm: Intonation and context as relational cues in children's recognition of irony*. Paper presented at the Annual Boston University Conference on Language Development.

Neumann, D., Spezio, M.L., Piven, J., & Adolphs, R. (2006). Looking you in the mouth: Abnormal gaze in autism resulting from impaired top-down modulation of visual attention. *Social Cognitive and Affective Neuroscience, 1*(3), 194-202.

Ochs, E., Taylor, C., Rudolph, D., & Smith, R. (1992). Storytelling as a theory-building activity. *Discourse Processes, 15,* 17-72.

Ochs, E., & Capps, L. (1996). Narrating the Self. *Annual Review of Anthropology, 25*, 19-43.

Olineck, K. M., & Poulin-Dubois, D. (2007). Imitation of intentional actions and internal state language in infancy predict preschool theory of mind skills. *European Journal of Developmental Psychology, 4*(1), 14-30.

Olson, D. R. (1988). On the origins and beliefs and other intentional states in children. In J. W. Astington, Harris, P.L. & Olson, D.R. (Ed.), *Developing theories of mind* (pp. 414-426). New York: Cambridge University Press.

Osterling, J., & Dawson, G. (1997). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders, 24*(3), 247-257.

Osterling, J., Dawson, G., & Munson, J. (2002). Early recognition of one year old infants with autism spectrum disorder versus mental retardation: A study of first birthday party home videotapes. *Development and Psychopathology, 14*, 239-252.

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

Paul, R., & Smith, R.L. (1993). Narrative skills in 4-year olds with normal, impaired, and late-developing language. *Journal of Speech and Hearing Research, 36*(3), 592-598.

Pelphrey, K. A., & Carter, E.J. (2008). Brain mechanisms for social perception: Lessons from autism and typical development. *Annals of the New York Academy of Sciences, 1145*, 283-299.

Perner, J., & Wimmer, H. (1985). John *thinks* that Mary *thinks* that...": Attribution of second-order beliefs by 5- to 10-year old children. *Journal of Experimental Child Psychology, 39*(3), 437-471.

Perner, J., Leekam, S.R., & Wimmer, H. (1987). Three-year olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*, 125-137.

Perner, J., & Ruffman, T. (1995). Episodic memory and autonoetic consciousness: Developmental evidence and a theory of childhood amnesia. *Journal of Experimental Child Psychology, 59*, 516-548.

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

Recchia, H. E., Howe, N., & Alexander, S.A. (2005). *You're determined to fall on your nose: Children's understanding and production of verbal irony in family conversations.* . Paper presented at the Society for Research in Child Development, Atlanta, GA.

Reisinger, E. M., Ming, X., Harker, C.M., & Mandell, D.S. (2010). *Parent-teacher agreement on the Social Responsiveness Scale for children in a large urban school district*. Paper presented at the International Meeting for Autism Research.

Repacholi, B., & Gopnik, A. (1997). Early reasoning about desires: Evidence form 14- and 18-month olds. *Developmental Psychology, 33*, 12-21.

Ribiero, L. A., & Fearon, P. (2010). Theory of mind and attentional bias to facial emotional expressions: A preliminary study. *Scandinavian Journal of Psychology, 51*, 285-289.

Riby, D. M., & Hancock, P.J.B. (2008). Viewing it differently: Social scene perception in Williams syndrome and Autism. *Neuropsychologia, 46*(11), 2855-2860.

Rice, K., Moriuchi, J.M., Jones, W., & Klin, A. (2012). Parsing heterogeneity in autism spectrum disorders: Visual scanning of dynamic social scenes in school-aged children. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(3) 238-248.

Rochat, P., & Striano, T. (1999). Early social cognition: Understanding others in the first months of life. In P. Rochat (Ed.), (pp. 3-34). Hillsdale, NJ: Erlbaum

Ronald, A., Happe, F., Bolton, P., Butcher, L.M., Price, T., Wheelwright, S., Baron-Cohen, S., & Plomin, R. (2006). Genetic heterogeneity between the three components of the autism spectrum: A twin study. *Journal of the American Academy of Child & Adolescent Psychiatry, 45*(6), 691-699.

Ronald, A., Happe, F., Price, T.S., Baron-Cohen, S., & Plomin, R. (2006). Phenotypic and genetic overlap between autistic traits at the extremes of the general population *Journal of the American Academy of Child & Adolescent Psychiatry, 45*(10), 1206-1214.

Rosenberg, R. E., Law, J.K., Yenokyan, G., McGready, J., Kaufmann, W.E., & Law, P.A. (2009). Characteristics and concordance of autism spectrum disorder among 277 twin pairs. *Archives of Pediatrics and Adolescent Medicine, 163,* 907-914.

Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotion understanding *Cognitive Development, 18,* 139-158.

Sabbagh, M. A. (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain and Cognition, 55,* 209-219.

Saint-Georges, C., Mahdhaoui, A., Chetouani, M., Cassel, R.S., Laznik, M.C., Apicella, F., Muratori, P., Maestro, S., Muratori, F., & Cohen, D. (2011). Do parents recognize autistic deviant behavior long before diagnosis? Taking into account interaction using computational methods. *PLoS One, 6*(7), 1-13. Retrieved from http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0022393

Saxe, R., & Baron-Cohen, S. (2006). Editorial: The neuroscience of theory of mind. *Social Neuroscience, 1*(3-4), i-ix.

Schultz, R. T. (2005). Developmental deficits in social perception in autism: the role of the amygdala and fusiform face area. *International Journal of Developmental Neuroscience, 23*, 125-141.

Schultz, T. R., Wells, D., & Sarda, M. (1980). The development of the ability to distinguish intended actions from mistakes, reflexes, and passive movements. *British Journal of Social and Clinical Psychology, 19*(301-310).

Scott-Van Zeeland, A. A., Dapretto, M., Ghahremani, D.G., Poldrack, R.A., & Bookheimer, S. (2010). Reward processing in autism. *Autism Research, 3,* 53-67.

Shatz M., W., H.M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition, 14*, 301-321.

Slaughter, V., Peterson, C.C., & Mackintosh, E. (2007). Mind what mother says: Narrative input and theory of mind in typical children and those on the autism spectrum *Child Development, 78*(3), 839-858.

Steele, S., Joseph, R.M., & Tager-Flusberg, H. (2003). Brief Report: Developmental change in theory of mind abilities in children with autism. *Journal of Autism and Developmental Disorders, 33*(4), 461-467.

Stockman, I. J. (2010). Listener reliability in assigning utterance boundaries in children's spontaneous speech. *Applied Psycholinguistics, 31*, 363-395.

Striano, T., & Bertin, E. (2004). Contribution of facial and vocal cues in the still-face response of 4-month old infants. *Infant Behavior and Development, 27*(4), 499-508.

Striano, T., & Stahl, D. (2005). Sensitivity to triadic attention in early infancy. *Developmental Science, 8*(4), 333-343.

Striano, T., Chen X., Cleveland, A., & Bradshaw, S. (2006). Joint attention social cues influence infant learning *European Journal of Developmental Psychology, 3*(3), 289-299.

Striano, T., Henning, A., & Stahl, D. (2006). Sensitivity to interpersonal timing at 3 and 6 months of age. *Interaction Studies 7*(2), 251-271.

Striano, T., & Reid, V.M. (2006). Social cognition in the first year. *Trends in Cognitive Sciences, 10*(10), 471-476.

Strong, C. (1998). *The Strong Narrative Assessment Procedure*. Eau Claire, WI: Thinking Publications.

Sullivan, K., Zaitchik D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30*(3), 395-402.

Sullivan, K., Winner, E., & Hopfield, N. (1995). How children tell a lie from a joke: the role of second-order mental state attributions. *British Journal of Developmental Psychology, 13*(2).

Tager-Flusberg, H., & Sullivan, K. (1995). Attributing mental states to story characters: A comparison of narrative produced by autistic and mentally retarded individuals. *Child Development, 63*, 161-256.

Tager-Flusberg, H. (2000). Language and understanding minds in autism: Connections in autism In S. Baron-Cohen, Tager-Flusberg, H. & Cohen, D.J. (Ed.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd edition ed.). Oxford, U.K.: Oxford University Press.

Tager-Flusberg, H. (2006). Defining language phenotypes in autism. *Clinical Neuroscience Research, 6*, 219-224.

Taumoepeau, M., & Ruffman, T. (2008). Stepping stones to others' minds: Maternal talk relates to child mental state language and emotion understanding at 15, 24, and 33 months. *Child Development, 79*(2), 284-302.

Tottenham, N., Leon, A.C., & Casey, B.J. (2006). The face behind the mask: A developmental study. *Developmental Science, 9*(3), 288-294.

Vaissiere, J. (2006). Perception of intonation. In D. B. Pisoni, & Remez, R.E. (Ed.), *The handbook of speech perception* (pp. 236-261). Oxford: Blackwell.

Wallace, S., Sebastian, C., Pellicano, E., Parr, J., & Bailey A. (2010). Face processing abilities in relatives of individuals with ASD. *Autism Research, 3,* 345-349.

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.

Wang, A. T., Lee, S.S., Sigman, M., & Dapretto, M. (2007). Reading affect in the face and voice. *Archives of General Psychiatry, 64,* 698-708.

Welch-Ross, M. K. (1997). Mother-child participation in conversation about the past: Relationships conflicting mental representation predicts suggestibility. *Developmental Psychology, 33*, 618-629.

Wellman, H. M., & Woolley, J.D. (1990). From simple desires to ordinary beliefs: The early development of everyday psychology. *Cognition, 35*, 245-275.

Wellman, H. M., & Lagattuta, K.H. (2000). Developing understandings of mind. In S. Baron-Cohen, Tager-Flusberg, H. & Cohen, D.J. (Ed.), *Understanding other minds: Perspectives from developmental cognitive neuroscience* (2nd edition ed.). Oxford, U.K.: Oxford University Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory of mind development: The truth about false belief. *Child Development, 72*(3), 655-684.

Williams, J. H. G. (2008). Self-other relations in social development and autism: Multiple roles for mirror neurons and other brain bases *Autism Research, 1*, 73-90.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128.

Winner, E., Brownell, H., Happe, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficit and discourse interpretation in right hemisphere brain damage patients. *Brain and Language, 62*, 89-106.

Young, E., Diehl, J., Morris, D., Hyman, S., & Bennetto, L. (2005). The use of two language tests to identify pragmatic language problems in children with autism spectrum disorder. *Language Speech and Hearing Services in Schools, 36*(62-72).

Young, G. S., Merin, N., Rogers, S.J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science, 12*(5), 798-814.