Statistical Word Learning and Non-Social Visual Attention in Children with Autism

By

Courtney E. Venker

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy
(Communication Sciences & Disorders)

At the
UNIVERISITY OF WISCONSIN-MADISON
2013

Date of final oral examination: 12/05/2013

The dissertation is approved by the following members of the Final Oral Committee:
Susan Ellis Weismer, Professor, Communication Sciences & Disorders
Jenny R. Saffran, Professor, Psychology
Margarita Kaushanskaya, Associate Professor, Communication Sciences & Disorders
Jan Edwards, Professor, Communication Sciences & Disorders
Vanessa Simmering, Assistant Professor, Psychology
Dan Bolt, Professor, Educational Psychology

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

CHAPTER 2: METHOD

CHAPTER 3: RESULTS

CHAPTER 4: DISCUSSION

Dedication

To my family. Your love and support mean the world to me.

Acknowledgements

This work was supported by NIH grants F31 DC012451 to Courtney E. Venker, T32 DC005359 to Susan Ellis Weismer, R37 HD037466 to Jenny Saffran, and P30 HD003352 to the Waisman Center. Funding for this study was also provided by the Friends of the Waisman Center and the Wisconsin Speech-Language Pathology and Audiology Association.

I would like to thank my mentor, Dr. Susan Ellis Weismer, for her thoughtful and insightful mentorship and unwavering commitment to my training. My experiences and interactions with her have fostered my development as a scientist. I would like to thank Dr. Jenny Saffran for unhesitatingly welcoming me into her lab, which allowed my research to go in new directions. I also thank the other members of my committee, Dr. Jan Edwards, Dr. Rita Kaushanskaya, Dr. Vanessa Simmering, and Dr. Dan Bolt, for generously sharing their knowledge and experience.

Many thanks to the past and present members of the Language Processes Lab for providing me with a warm and supportive community in which to develop my research interests. In particular, I thank Beth Eernisse, Courtney Karasinski, Sara Kover, Allison Bean, Heidi Sindberg, Corey Ray-Subramanian, Meghan Davidson, and Eileen Haebig. I also thank the LINGO study team—Anna Dorrance, Taryn Stricker, and Kristina Fassbender—for helping me to carry out this project. Thanks to all past and present members of the Infant Learning Lab—including Erica Wojcik, Casey Lew-Williams, Chris Potter, Brianna McMillan, Alexa Romberg, and Jon Willits—for their thoughtful feedback and wise advice. Thanks also to Rob Olson for his patience and technical support, and to the Wednesday morning Barrique's Writing Group—Lisa Vinney, Lindsey Byom, and Peter Meulenbroek—for making writing a social activity.

I offer sincere thanks to my family and friends, who have shared my joys and my challenges. Greg, Elliott, Mom, and Dad: thank you for coming along with me on this journey.

Lastly, I am sincerely appreciative of the time and energy so generously shared by the children and families who participated. This study would not have been possible without them.

Abstract

This study investigated cross-situational word learning and non-social visual attention in 4- to 7-year-old children with autism spectrum disorders (ASD) with heterogeneous language and cognitive abilities (*n* = 27). A comparison group of 2- to 7-year-old typically developing (TD) children matched on receptive vocabulary was also included (*n* = 28). Children participated in a cross-situational word-learning task presented on an eye tracker that required attention to word-object co-occurrences across individually ambiguous trials. Children with ASD who were excluded due to inattention (*n* = 7) had significantly lower vocabulary levels than children who were retained. Contrary to expectations, eye gaze patterns and a behavioral pointing task indicated no significant group differences in cross-situational learning. Eye-gaze and pointing task performance was significantly correlated in the TD group, suggesting convergent validity, but the two measures were not significantly associated in the ASD group. Several child-level abilities were examined to determine the factors underlying successful cross-situational learning. Age, receptive vocabulary level, and simple word learning were significantly associated with cross-situational learning in the TD group; correlations were marginal in the ASD group. There were no significant relationships between cross-situational learning and visual recognition memory, nonverbal IQ, or autism severity.

Children also participated in a non-social visual orienting task. Surprisingly, the TD and ASD groups did not differ in mean latencies to shift and disengage attention. It was hypothesized that children with ASD who had more difficulty disengaging their attention from an ongoing stimulus (i.e., sticky attention) would have more difficulty with the cross-situational task because successful learning required attention to multiple visual stimuli. Consistent with this prediction, the children with ASD who had stickier attention demonstrated worse cross-situational learning, although this relationship was marginal when age was statistically controlled. Sticky attention was also related to familiar word recognition, pointing to a broader impact of visual disengagement on language processing. A post hoc analysis suggested that real-time familiar word recognition mediates the relationship between sticky attention and vocabulary knowledge in children with ASD. Differences in sticky visual attention may help to explain why some children with ASD experience such striking delays in early vocabulary learning.

Chapter 1: Introduction

Investigating Word-Learning Mechanisms in Children with ASD

As a manifestation of their broader social-communication impairments, many children with ASD have difficulty learning words (Ellis Weismer et al., 2011; Ellis Weismer, Lord, & Esler, 2010; Luyster, Kadlec, Carter, & Tager-Flusberg, 2008). In fact, early language delay is one of the first concerns reported by parents of children later diagnosed with ASD (Wetherby et al., 2004). Historically, research on language development in children with ASD has focused on characterizing their language abilities. For example, studies have shown that many children with ASD have early delays in vocabulary and communication skills, and up to 30% of individuals with ASD never develop functional spoken language (Charman et al., 2003a; Charman et al., 2005; Tager-Flusberg & Kasari, 2013). Although descriptive studies provide a foundational understanding of language acquisition in this population, they do not address the underlying factors that allow some children with ASD to learn language more easily than others.

Deficits in social communication are part of the diagnostic criteria for ASD, but structural language abilities are an area of striking heterogeneity. Although the majority of children with ASD demonstrate early language delays, heterogeneity in language skills increases over development (Charman et al., 2003a; Charman et al., 2005; Kjelgaard & Tager-Flusberg, 2001; Tager-Flusberg, Paul, & Lord, 2005). Some children with ASD eventually acquire age-appropriate vocabulary and morphosyntactic skills, whereas others experience persistent structural language deficits (Kjelgaard & Tager-Flusberg, 2001). Research has investigated the developmental course of vocabulary deficits in ASD, but remarkably little is known about the origins of these deficits. Both theory and clinical practice require not only recognition of word-learning deficits, but also understanding of the mechanisms through which they emerge (e.g., atypical learning and attention).

One reason for the limited understanding of heterogeneity in language skills is that researchers have only recently begun to investigate mechanisms of language learning in autism. Another reason is that very few studies of language development in children with ASD have included the diverse participant samples necessary to answer broader questions of heterogeneity. Instead, many studies have focused on only a subset of the broader population—in most cases, 'high functioning' children with ASD whose cognitive and language skills meet or exceed age expectations (Tager-Flusberg & Kasari, 2013). Studying mechanisms of learning in heterogeneous samples of children with ASD may shed light on why language abilities in ASD are so strikingly variable (Eigsti, de Marchena, Schuh, & Kelley, 2011).

Recent studies have begun to investigate mechanisms of language learning in children with ASD, with the goal of better understanding breakdowns in language acquisition (Baron-Cohen, Baldwin, & Crowson, 1997; Bedford, Gliga, & Frame, 2012; de Marchena, Eigsti, Worek, Ono, & Snedeker, 2011; Gliga, Elsabbagh, Hudry, Charman, & Johnson, 2012; Luyster & Lord, 2009; McGregor & Bean, 2012; Preissler & Carey, 2005). Some studies have pointed to deficits in specific word learning mechanisms (McGregor & Bean, 2012; Tek, Jaffery, Rein, & Naigles, 2008), whereas others have identified intact word learning abilities (de Marchena et al., 2011; Luyster & Lord, 2009). The current study continues this line of work by investigating statistical word learning in children with ASD—specifically, cross-situational word learning. It also investigates the skills that may underlie successful cross-situational word learning and examines the relationship between non-social visual attention and word learning. Identifying the mechanisms through which word learning breaks down has the potential to enhance our understanding of ASD as a developmental disorder, to inform theories of language development and word learning, and to identify strategies for language interventions.

How do children learn the meanings of words? One of the biggest challenges facing early language learners is determining the meanings of new words. How do children learn which object a new label describes? In some situations, objects are clearly described (e.g., a father holds up a truck and says 'Look! A truck!'), but in other situations the associations between labels and objects may be less clear (e.g., a father says 'Look! A truck!' while looking around a playroom full of toys). Quine (1960) famously described this problem as the indeterminacy of reference and suggested that input alone can never logically disambiguate the meaning of a word. Although determining the meaning of a new word in an ambiguous context is certainly difficult, a recent line of work (Smith & Yu, 2008; Yu & Smith, 2007) suggests that it is not impossible if learners accumulate information about label-object co-occurrences *across* ambiguous contexts.

Cross-situational word learning. Statistical learning broadly refers to sensitivity to patterns or structure in the environment (Romberg & Saffran, 2010) and has been shown to play a role in various aspects of language acquisition, including word segmentation (Saffran, Aslin, & Newport, 1996), discrimination of phonetic categories (Maye, Werker, & Gerken, 2002), and grammar learning (Gomez & Gerken, 1999). Researchers have recently begun to investigate whether statistical learning mechanisms might also help children learn the meanings of new words. Akhtar and Montague (1999) found that preschoolers learned the meanings of novel adjectives (e.g., *A modi one*) by tracking the elements (e.g., shape or texture) that consistently co-occurred with novel adjectives across ambiguous contexts. This ability to "…pay attention to (and remember) the element that remains constant across different situations in which the same word is used" (Akhtar & Montague, 1999, p. 349) represents a statistical learning mechanism called *cross-situational word learning* (Fisher, Hall, Rakowitz, & Gleitman, 1994; Siskind, 1996; Smith & Yu, 2008).

In a landmark study, Smith and Yu (2008) exposed 12- and 14-month-old typically developing (TD) infants to a series of trials, each of which presented two new labels and two new objects—meaning that there was no information *within* a single trial to indicate the correct label-object pairing. Eye movements revealed that infants learned the names of new objects by accumulating information about label-object co-occurrences across the ambiguous trials; in other words, infants were capable of cross-situational word learning (see Yu and Smith, 2007, for a similar study in adults). It was notable that infants learned new words despite the ambiguity of individual trials, the number of new objects and labels to which they were exposed (6 of each), the limited amount of training provided (less than 4 minutes), their young age, and the lack of explicit instructions.

A continually expanding body of work has extended, replicated, and challenged the original findings by Smith and Yu (2008). Vouloumanos and Werker (2009) found that 18-month-old TD infants were capable of learning new words in a stochastic (i.e., probabilistically distributed) environment, wherein referents were not always present when their label was produced. Infants learned words consistently paired with one referent better than those paired with multiple referents, suggesting that they tracked both high- and low-probability statistics.

Scott and Fisher (2012) found that TD toddlers with higher vocabularies, but not those with lower vocabularies, learned intransitive verbs by tracking cross-situational statistics, suggesting that prior vocabulary knowledge plays a role in cross-situational learning.

There is some debate regarding the mechanisms that underlie cross-situational learning, with some researchers positing that information is accumulated over time through statistical learning (Suanda & Namy, 2012; Yu & Smith, 2007), and others positing a single-trial learning strategy (Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell, Medina, Hafri, & Gleitman, 2013). In other words, as is the case for word learning more generally, some researchers have argued for the existence of implicit, associative, bottom-up mechanisms, whereas others have argued for more explicit, hypothesis-driven, top-down mechanisms.

Yu and Smith (2012) have recently proposed that the debate between hypothesis testing and simple associative models is not well specified, making it difficult to directly test comparisons between the two types of models. Yu and Smith (2012) argued that the associative and hypothesis testing models of word learning may not in fact be fundamentally different, since manipulation of aspects that are seemingly unrelated to the core ideas of each theory can cause the models to resemble one another. Work by Frank, Goodman, and Tenenbaum (2009) using Bayesian computational models to analyze language corpus data has suggested that children may use cues to speakers' intent and statistical information about word-object pairings in concert.

Cross-situational word learning in children with ASD. Despite the rapidly expanding knowledge base on cross-situational word learning in typically developing individuals, almost nothing is known about cross-situational word learning in children with ASD. Studying cross-situational word learning in ASD is important because neural evidence suggests that children with ASD may have impaired statistical learning abilities in other domains, such as using transitional probability statistics to segment words from continuous speech (Scott-Van Zeeland et al., 2010; but see Mayo & Eigsti, 2012). In addition, children with ASD do not generalize new words by shape (i.e., the shape bias), a pattern commonly seen in typical development, which suggests that young children with ASD may not selectively attend to the statistical regularities among the words they have learned (Tek et al., 2008). Although few studies have investigated statistical learning in children with ASD, a study by Evans, Saffran, and Robe-Torres (2009) demonstrated that children with specific language impairment required twice as much exposure to an artificial language than their peers to perform comparably on a word segmentation task.

Understanding cross-situational word learning is also important because impairments in the use of social cues (e.g., joint attention) for disambiguating word meanings may cause children with ASD to experience an increased number of ambiguous learning situations in which the correspondence between a label and an object is not immediately clear—precisely the type of contexts in which cross-situational word learning would be advantageous. Baron-Cohen et al. (1997) found that, contrary to TD children and children with developmental delay, some children with ASD rely on their own direction of gaze to determine the meaning of a new word—a strategy that results in mapping errors when an object is labeled based on a speaker’s direction of gaze (see also Preissler & Carey, 2005; but see Luyster & Lord, 2009). Even when toddlers at risk for ASD follow eye gaze cues to look at a novel object, they may not successfully learn the name of that object (Gliga et al., 2012).

Baron-Cohen and colleagues (1997) hypothesized that, instead of relying on social cues to learn new words, children with ASD might learn words by identifying co-occurrences between the visual and auditory stimuli across different situations—a potentially less efficient, but nonetheless reliable strategy that they termed the ‘Common Feature’ strategy. The Common Feature strategy is the mechanism now typically referred to as cross-situational word learning. Additionally, although some accounts of cross-situational learning have focused on associative learning mechanisms (Yu & Smith, 2011), there is also evidence that social factors, such as sensitivity to a speaker intentions (Frank et al., 2009) and discourse information (Frank, Tenenbaum, & Fernald, 2013) are also relevant for cross-situational learning. Thus, even though some types of lab-based cross-situational learning tasks limit the impact of social interaction, everyday cross-situational learning is likely to incorporate the social aspects that pose difficulty for many children with ASD.

Further, ASD is characterized by atypical patterns of attention (Allen & Courchesne, 2001) that may impact cross-situational word learning. Swettenham and colleagues (1998) found that, in a free play context, toddlers with ASD spent most of their time looking at objects as opposed to people, meaning that they were not spontaneously seeking social interactions about objects that might have led to greater learning opportunities. In addition, the children with ASD made relatively more shifts between objects than between objects and people, or between two people. These same patterns—for example, an increased interest in non-social stimuli—also appear to be present in infants at risk for ASD (Bhat, Galloway, & Landa, 2010; Ibanez, Messinger, Newell, Lambert, & Sheskin, 2008). For example, Bhat et al. (2010) found that infant siblings at risk for ASD demonstrated no difficulties learning an associative relationship but spent more time looking at relevant non-social stimuli than at their caregivers. The fact that children with ASD spend a relatively large proportion of time looking at and shifting attention between objects and non-social stimuli underscores the need to investigate their cross-situational word learning abilities, since cross-situational learning relies in part on children’s attention to objects.

Children with ASD also show reduced exploration of objects in their environment (Pierce & Courchesne, 2001), and infants at risk for ASD show atypical visual exploration of objects as early as 12 months of age (Ozonoff et al., 2008). Children with ASD also demonstrate difficulty disengaging visual attention (Landry & Bryson, 2004), which may have a negative effect on cross-situational learning. Spontaneous allocation of visual attention and attention shifting are particularly relevant because they pertain to play-based contexts in which cross-situational word learning may naturally occur. Intact cross-situational word learning abilities could help children with ASD acquire language, but impaired abilities may result in a continued language-learning disadvantage.

Investigating cross-situational word learning in children with ASD is imperative, but only one empirical study has examined this issue. McGregor, Rost, Arenas, Farris-Trimble, and Stiles (2013) examined the use of probabilistic gaze cues in high-functioning children with ASD and their age-matched peers (mean age 11 years); groups were matched on nonverbal cognition and vocabulary level. The TD and ASD groups performed similarly on tasks of word recognition and word learning incorporating different types of gaze cues (facilitative, neutral, or contradictory). Most relevant to the current study was the finding that performance in a neutral gaze condition, wherein children had to rely on prior cross-situational information from previous trials (i.e., co-occurrences between words and objects), was related to receptive vocabulary level in the ASD group but not the TD group, despite similar variability in both groups. McGregor and colleagues proposed that children with ASD with more limited language abilities may have difficulty using cross-situational statistics to learn new words. Given the exploratory nature of their findings, however, they highlighted the need for additional empirical evidence to support or refute this hypothesis. The current study addressed precisely this issue.

Cognitive and Linguistic Skills Supporting Cross-Situational Word Learning

One approach to understanding deficits in cross-situational learning is to examine the foundational, or prerequisite skills that underlie successful cross-situational learning. The current study adopted this approach by examining the role of visual recognition memory and simple word learning as prerequisite skills for cross-situational word learning. It would stand to reason that if children were unable to recognize previously presented objects, or to associate single labels with their referents, breakdowns in cross-situational word learning may occur. Other child-level variables, including age, IQ, vocabulary level, and autism severity, were also examined, since these factors may play a part in cross-situational language learning or real-time language processing (McGregor et al., 2013; Venker, Eernisse, Saffran, & Ellis Weismer, 2013).

Visual recognition memory. Visual recognition memory is a well-studied phenomenon in infant cognitive development that refers to the ability to remember previously encountered visual stimuli (Colombo, Mitchell, & Horowitz, 1988; Rose, Feldman, & Jankowski, 2004). Since cross-situational word learning requires children to recall the objects they have seen, visual recognition memory is likely an important foundational skill for cross-situational word learning. Visual recognition memory is commonly measured by the visual paired-comparison (VPC) task, which often consists of a familiarization phase, in which two identical images are presented, and a test phase, in which the same stimulus and a novel stimulus are presented (Colombo, Mitchell, Dodd, Coldren, & Horowitz, 1989; Richmond, Colombo, & Hayne, 2007). The dependent variable of interest is typically the amount of time children spend looking at the familiar versus unfamiliar stimulus during test.

Typically, infants who have received a relatively long exposure time—presumably, enough time to form a mental representation of the familiarized stimulus—will demonstrate a looking preference for the novel stimulus (i.e., a novelty preference). This assumption is based on Sokolov's (1963) comparator theory of attention orienting, which suggests that infants should prefer to look at a novel stimulus once their internal representation of a previously viewed stimulus is adequately specified. Indeed, young infants often show a novelty preference in this type of task (Colombo et al., 1988; Fagan, 1990). Interestingly, research has shown that infants' recognition memory for visual stimuli is enhanced when they are presented with two images, either simultaneously or sequentially, as opposed to one (Fagan, 1990; Oakes, Kovack-lesh, & Horst, 2010; Rose et al., 2004), which was why two stimuli were presented in the current study.

Several studies using a variety of different tasks have documented intact recognition memory in children and adolescents with ASD (Lind & Bowler, 2009; Minshew & Goldstein, 2001). For example, Dawson, Osterling, Rinaldi, Carver, and McPartland (2001) found no group differences in a play-based visual recognition task among children with ASD, children with Down syndrome, and typically developing children matched on receptive language. Although these findings suggest that visual recognition memory may not be a specific deficit in ASD, it is nonetheless possible that individual differences in visual recognition memory could impact cross-situational word learning. Researchers have identified individual differences in visual recognition memory that relate to looking patterns and other developmental skills (Colombo et al., 1989; Rose et al., 2004). This study investigated whether visual recognition memory was related to cross-situational word learning. If children were unable to remember the novel images from one trial to the next, this deficit might explain poor cross-situational word learning performance.

Language skills. Although there remains much to be learned about how statistical learning relates to language acquisition (Arciuli & Torkildsen, 2012), cross-situational word learning ability may be supported by general language and cognitive skills. Newman, Ratner, Jusczyk, Jusczyk, and Dow (2006) identified a relationship between word segmentation in infants and expressive language level at 2 years of age. Additionally, children who had successfully segmented words from continuous speech in infancy had higher language abilities, but not general IQ, at 4 to 6 years of age. Akhtar and Montague (1999) found that four-year-olds showed better cross-situational learning than 2-year-olds, suggesting that statistical word learning improves with age and/or language experience. Scott and Fisher (2012) found that TD toddlers with higher vocabularies, but not those with lower vocabularies, learned intransitive verbs by tracking cross-situational statistics, suggesting that vocabulary knowledge plays a role in cross-situational learning. Evans et al. (2009) also found that vocabulary was correlated with statistical learning (i.e., word segmentation) in TD children. McGregor and colleagues (2013) found that some children with ASD used cross-situational statistics to learn the meanings of new words but that this ability is related to children's receptive vocabulary.

In addition to general language abilities, word learning in unambiguous contexts may also be related to cross-situational word learning. As conceptualized in the current study, simple word learning taps the ability to form an association between a word and an object in what is arguably a less complex context than that encountered during cross-situational word learning—exposure to one label-object pair at a time versus exposure to two label-object pairs at a time. It is possible that children with ASD who have general difficulties associating labels and objects have particular difficulty learning label-object correspondences across ambiguous contexts. This question is important because although some studies have indicated no deficits in children with ASD in explicit word learning contexts (Baron-Cohen et al., 1997; Priessler & Carey, 2005), a recent study identified word-learning deficits in children with ASD compared to children with fragile X syndrome (McDuffie, Kover, Hagerman, & Abbeduto, 2012). Although previous studies of cross-situational word learning have not investigated this question, it is important for understanding potential breakdowns in word learning in children with ASD.

Non-Social Visual Attention: Visual Orienting

It has been suggested that, in order to successfully learn from their environments, children must exhibit several separable but inter-related components of attention, including orienting, sustaining, disengaging, and shifting (Dawson et al., 2004; Patten & Watson, 2011; Swettenham et al., 1998). Although exhibiting each of the various components of attention is important, the manner in which children exhibit these features is arguably even more important. Ideally, children's attentional systems should be flexible; should orient to the aspects of the environment most relevant for learning; and should shift easily away from those same stimuli when a learning context demands attention to a new relevant stimulus.

For example, a child who is too 'good' at orienting attention would continually be distracted by potentially irrelevant factors. A child who is too 'good' at sustaining attention might focus in on one aspect of the environment to the exclusion of other important aspects. In other words, possessing the capability to orient, sustain, and shift attention is insufficient. To be successful learners, individuals must flexibly move their attention based on contextual factors. In many ways, then, the attentional system of a successful human learner is highly demanding, insinuating that an individual should look at an object or event, but only when it matters, look away from it when something else matters more, and do so rapidly enough to gain access to the right information at the right time.

Despite the complexity of these processes, many TD children appear to meet these demands quite naturally. Children with ASD, however, demonstrate considerable differences in many aspects of attention that could produce cascading negative effects over the course of development (Amso, Haas, Tenenbaum, Markant, & Sheinkopf, 2013; Elison et al., 2013; Patten & Watson, 2011; Sacrey, Bryson, & Zwaigenbaum, 2013; Sasson & Touchstone, 2013). Posner and colleagues (Posner & Petersen, 1990) have proposed that humans' complex networks of attention might be organized into three systems: a vigilance system comprising alertness and sustained attention; a basic system responsible for moving attention among competing stimuli and selecting stimuli to receive additional processing; and an executive system encompassing voluntary control of attention and goal selection. The current study focused on one aspect of the basic attentional system, namely, visual orienting.

Visual orienting in typical development. Attentional orienting refers to the adjustment of the body toward an auditory or visual stimulus, such as shifting one's eye-gaze or turning one's head (Dawson, Meltzoff, Osterling, Rinaldi, & Brown, 1998; Dawson et al., 2004; Landry & Bryson, 2004). As conceptualized by Landry and Bryson (2004), orienting is comprised of two inherently related but experimentally separable components of attention: disengaging and shifting. Disengaging refers to an individual's ability to release attention from a stimulus that is the current focus of attention. Shifting refers to the actual movement of attention to a second stimulus, usually without a competing stimulus. Although these two components are inherently linked, Landry and Bryson (2004) developed an experimental paradigm—the gap-overlap task—that separates the constructs of shifting and disengaging (see also Posner, 1988). The current study used an adapted version of Landry and Bryson's gap-overlap task to investigate the relationship between visual orienting and cross-situational word learning—a skill inherently intertwined with the ability to flexibly move attention to different stimuli in the environment.

Visual orienting is an early developing skill in typical development; most infants readily disengage and shift their visual attention by 3 – 4 months of age. Johnson, Posner, and Rothbart (1991) found that 4-month-old infants were able to disengage their attention from a salient stimulus and move their attention to a competing stimulus, but that younger infants had difficulty disengaging their visual attention—a phenomenon referred to as 'sticky attention.' In general, infants and adults show longer latencies to disengage their attention from a competing stimulus than to shift their attention without a competing stimulus, and latencies to disengage attention decrease over the course of development (Hood & Atkinson, 1993).

Visual orienting in children with ASD. Despite the early emergence of visual orienting abilities in typical development, infants and children with ASD demonstrate atypical patterns of attentional orienting in response to both auditory and visual stimuli, within both social (Dawson et al., 1998; Dawson et al., 2004) and non-social (Landry & Bryson, 2004) domains. Much research in this area has focused on the diagnostic sensitivity and specificity of orienting deficits and the point of their emergence in infant siblings of children with ASD, who are at a higher genetic risk for ASD (Elsabbagh et al., 2009; Elsabbagh et al., 2013; Sacrey et al., 2013; Zwaigenbaum et al., 2005). These studies have suggested that attentional differences may be one of the earliest-emerging behavioral differences associated with ASD (but see Fischer, Koldewyn, Jiang, & Kanwisher, 2013, for evidence of intact disengagement in high-functioning children with ASD).

In a study of early-emerging risk markers of ASD, Zwaigenbaum et al. (2005) investigated visual orienting in at-risk siblings of children with ASD and found evidence of prolonged disengagement by 12 months of age. All infants later diagnosed with ASD showed longer latencies to disengage visual attention at 12 months than at 6 months, indicating the opposite developmental pattern than expected (i.e., that latencies to disengage would decrease with age; Hood & Atkinson, 1993). Sacrey et al. (2013) conducted a prospective longitudinal study of visual attention in at-risk infants. By 12 months of age, infants later diagnosed with ASD were distinguishable on the basis of their sticky attention while playing with toys. Zwaigenbaum and colleagues (2005) hypothesized than an early-emerging, basic attentional deficit like impaired visual orienting could underlie the later emerging social deficits in ASD.

Elsabbagh et al. (2009) examined shifting and disengaging of visual attention in infant siblings of children with ASD (9-10 months of age). At-risk siblings demonstrated longer latencies to disengage from an overlapping central stimulus—even though this stimulus was static at the time that the peripheral stimulus was presented. In addition, the at-risk infant siblings demonstrated less of a facilitation effect when a gap was inserted between two stimuli, meaning that they were less able to benefit from a temporal cue than controls. Elsabbagh et al. (2013) found that 14-month-old infants at risk for ASD demonstrated longer latencies to disengage from a visual stimulus than TD infants and infants at risk for other developmental delays, though these differences were not apparent at 7 months of age. As in previous work, TD infants showed expected decreases in mean latency over time, but many infants later diagnosed with ASD showed increases in latencies to disengage between 7 and 14 months (see also Elison et al., 2013).

Although much work has focused on visual attention in early infancy, Landry and Bryson (2004) conducted a study of non-social visual orienting in an older age range: preschool children with ASD (3 to 7 years). Comparison groups of children with Down syndrome and typically developing children matched on nonverbal and verbal mental age were also included to determine whether impairments in visual orienting were specific to ASD. Landry and Bryson adapted Posner's (1988) gap-overlap task to differentiate shifting (gap condition) and disengaging (overlap condition) of visual attention. The critical manipulation was whether the central stimulus was extinguished prior to the presentation of the peripheral stimulus. In gap trials, the central stimulus was extinguished 250ms before the peripheral stimulus appeared. In the overlap trials, the central stimulus continued for the entire trial, requiring the child to disengage attention before looking to the peripheral stimulus. The visual orienting task used in the current study was based on the task developed by Landry and Bryson.

Results of the study by Landry and Bryson (2004) revealed no group differences in the overall mean length of shifts, but the children with ASD had significantly fewer fast shifts than the other groups. Most notably, the children with ASD showed striking deficits in disengaging attention, supporting the premise of overly focused or "spotlight" attention in this population (Bryson et al., 1990). The mean latency in the overlap condition was 2164 ms in the ASD group, compared to a mean of 506 ms in the Down Syndrome group and a mean of 1073 ms in the TD group, and there was almost no overlap between the mean disengagement latencies of children with ASD and children in the other two groups. In 18% of trials, children with ASD never disengaged from the central stimulus for the full 8-second trial, and 80% of children with ASD never disengaged their attention on at least one trial, meaning that the results were not driven by outliers.

The type of "sticky attention" demonstrated by the children with ASD in the study by Landry and Bryson (2004) was characteristic of the visual patterns shown by typically developing 2-month-old infants (Hood & Atkinson, 1993; Johnson et al., 1991). Deficits of this type would be certain to negatively affect development, particularly if they emerged early in development, because visual orienting is one important way in which infants explore their world (Elsabbagh et al., 2009). Although Landry and Bryson (2004) found that neither shifting nor disengaging was related to cognitive or language scores in children with ASD, the authors speculated that such relationships might be evident earlier in life. It is also possible that the broad cognitive and language measures were not adequately sensitive to capture the relationship between these two constructs—which led to the theoretically motivated correlational analyses between visual orienting and cross-situational word learning in the current study.

Why understanding visual orienting is important in ASD. Researchers have long speculated that orienting abilities may underlie a number of key developmental skills, including theory of mind development (Leekam & Moore, 2001), joint attention (Dawson et al., 2004), emotional regulation (Rothbart, Posner, & Rosicky, 1994), and social cognition (Elsabbagh et al., 2013). Unfortunately, many of these hypotheses are underspecified and not well supported by empirical evidence. For example, Swettenham et al. (1998) speculated that, “Because attentional processing is an important component in the execution of complex cognitive operations, an early-occurring attentional deficit in autism might therefore impede the development of higher-order cognitive and social skills” (p. 747). Although these links and hypotheses are likely true, they leave unmentioned the precise mechanisms that link attentional orienting with specific developmental skills.

One broad idea that has been proposed by a number of researchers is that difficulties with visual orienting would negatively affect children’s ability to identify and focus on the aspects of the environment that are most relevant at specific points in time, thus limiting their ability to learn informative contingent relationships across experiences (Elsabbagh et al., 2009; Elsabbagh et al., 2013; Klinger, Klinger, & Pohlig, 2007; Renner, Klinger, & Klinger, 2006). Children with ASD who do not flexibly allocate their attention based on relevant environmental factors are at risk for “…‘locking’ onto certain irrelevant aspects of the environmental input. Similarly, a limitation in the ability to use environmental events to predict and prepare for a shift in attention (anticipation) may prevent typical levels of foraging from visual scenes resulting in reduced foveation of relevant information” (Elsabbagh et al., 2009, p.640). If problems with visual orienting prevent flexible shifting of attention, then children may not flexibly direct their attention based on incoming stimuli (e.g., adult language input). One context in which the detection and extraction of contingent relationships may play out is word learning, which led to the prediction in the current study that visual disengagement would be negatively associated with cross-situational word learning.

Visual attention patterns during cross-situational word learning. Yu and Smith (2011) conducted a replication of their 2008 study with a group of 14-month-old infants to investigate whether an associative learning mechanism using selective attention could explain the behavioral performance. In this study, infants' eye gaze patterns during the training phase were recorded by an automatic eye-tracker and were subsequently analyzed to determine the patterns of eye-gaze that were characteristic of more successful learning. As expected, the behavioral results were similar to Smith and Yu (2008). Analysis of the eye-gaze patterns during training revealed that although both groups of children spent similar amounts of time looking at the screen during training, strong learners made longer fixations and fewer shifts between images per trial.

Yu and Smith (2011) interpreted this finding as evidence that "strong learners have more stable eye movement patterns of sustained attention characterized by fewer but longer fixations" (p. 170). Interestingly, the authors mention that this finding differs from findings for infant visual recognition memory, wherein shorter looks and more shifts are interpreted as faster processing and subsequently yield superior recognition memory. However, as Yu and Smith also pointed out, remembering visual images and learning new labels for visual images are very different tasks, with the latter posing a greater challenge and thus requiring more selective attention to successfully map labels to objects. Yu and Smith (2011) further demonstrated that the differences in length of fixations and number of shifts between the strong and weak learners changed over the course of the experiment.

Finally, Yu and Smith (2011) found that a computational model with a simple associative learning mechanism and selective attention capabilities explained the infant behavioral findings—both in terms of group differences and individual differences. These consistencies in performance suggest that a hypothesis-driven learning mechanism may not be necessary for cross-situational learning. This type of top-down learning mechanism would require an individual to formulate explicit hypotheses about correct label-object pairings and subsequently evaluate and reformulate these hypotheses. Instead, Yu and Smith concluded that a simple associative learning mechanism with real-time selective attention may be sufficient, though this continues to be a topic of much debate.

The Current Study

To better understand word learning in the broader population of children with ASD, the current study examined individual differences in word learning and non-social attention in a participant sample with diverse cognitive, language, and attentional abilities. As stated by Eigsti and colleagues (2011), this approach "…takes the perspective that including a very wide range of abilities for a given skill provides us with the opportunity to investigate the precursors, predictors, and correlates of that skill" (p. 688). Participants were a group of 4- to 7-year-olds with ASD with a wide range of ability levels and a group of 2- to 7-year-old TD children matched on receptive vocabulary level. Matching the groups on receptive vocabulary level provided a conservative test of the hypothesis that children with ASD have cross-situational learning deficits because if the groups had differed on vocabulary level, differences in extant word knowledge could have explained any differences in cross-situational word learning.

This study adopted eye-gaze methodology to answer questions about word learning and attention in children with ASD. In recent years, a number of studies (Goodwin, Fein, & Naigles, 2012; Kelly, Walker, & Norbury, 2012; Naigles, Kelty, Jaffery, & Fein, 2011; Norbury et al., 2009; Norbury, Griffiths, & Nation, 2010; Swensen, Kelley, Fein, & Naigles, 2007; Venker et al., 2013) have successfully used eye-gaze measures to capture information about language comprehension, language processing, and attention in children with ASD (Boraston & Blakemore, 2007; Falck-Ytter, Bölte, & Gredebäck, 2013). These measures are appealing because they measure comprehension and attention in real time and because they incorporate relatively limited task demands (i.e., sitting and looking at a screen).

Some researchers have made a direct call for the use of eye-tracking measures in language learning studies with this population because they provide sensitive information about the time course of learning (Mayo & Eigsti, 2012). Although they were designed to answer different questions, the experimental tasks in this study all examined children's eye movements as a primary outcome measure. The word-learning tasks used automatic eye tracking, and the visual recognition and visual orienting tasks used manual coding of eye gaze based on the looking-while-listening paradigm (Fernald, Zangl, Portillo, & Marchman, 2008).

Research question 1: Do children with ASD show deficits in cross-situational word learning? The first aim was to investigate cross-situational word learning, a mechanism that allows children to learn new words by tracking label-object contingencies across ambiguous learning contexts (Smith & Yu, 2008). Children participated in a cross-situational word-learning task presented on an eye-tracker that assessed children's ability to learn words by tracking label-object pairings across time. Typically developing infants can learn up to four words in a cross-situational task (Smith & Yu, 2008), but many children with ASD demonstrate word-learning deficits. It was predicted that the children with ASD would not perform as well as the group of TD children and that there would be considerable heterogeneity in the ASD group.

Research question 2: What cognitive and linguistic skills support cross-situational word learning in children with ASD? Successful cross-situational learning may rely on a number of skills, including remembering recurring visual stimuli and associating labels and objects in a simple word-learning task. Correlational analyses were conducted to explore the relationship between cross-situational learning and several potential underlying skills, including vocabulary knowledge, IQ, and age, visual recognition memory, and simple word learning. It was predicted that vocabulary, age, and simple word learning were skills most likely to be related to cross-situational learning.

Research question 3: How does non-social visual attention impact cross-situational word learning in children with ASD? The third aim was to investigate the relationship between cross-situational word learning and visual attention. A visual orienting task captured information about children's latency to shift and disengage visual attention. It was predicted that a prolonged latency to disengage visual attention (i.e., sticky attention) would be negatively associated with cross-situational word learning because children need to attend to all available visual stimuli in order to successfully learn across ambiguous trials. Finally, this study examined children's patterns of eye-gaze during the cross-situational training phase to determine whether patterns of visual attention during learning were related to performance at test.

Chapter 2: Method

Participants

Participants with ASD were recruited from a recently completed longitudinal study of language development in children with ASD (Ellis Weismer, PI), local intervention clinics, a developmental disabilities research registry at the Waisman Center, and fliers posted in the community. TD participants were recruited from the Infant Learning Lab (Saffran, PI) and fliers posted in the community. All children in the ASD group had a community diagnosis of ASD, and no children in the TD group had received an ASD diagnosis. ASD diagnoses were confirmed through administration of the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al., 2012). Exclusionary criteria for both groups included uncorrected vision or hearing impairments, known chromosomal abnormalities (e.g., fragile X syndrome or Down syndrome), cerebral palsy, and considerable exposure to a language other than English in the home. Children in the TD group were excluded if they were reported to have any type of developmental delay.

In total, 28 TD children and 29 children with ASD participated in the study. One participant was excluded from the ASD group because he did not meet criteria for a classification of Autism or Autism Spectrum on the ADOS-2, even though he had previously been diagnosed with ASD. One participant with ASD was excluded because he received pilot versions of the experimental tasks. The full sample of children thus included a total of 55 children, 28 children in the TD group and 27 children in the ASD group. However, not all children contributed data for the cross-situational word-learning task, which was vital for inclusion in the analyses. Following data cleaning, 27/28 children in the TD group (96%) and

20/27 children in the ASD group (74%) contributed adequate data for the cross-situational word-learning task (see Data Cleaning procedures below). Because all research questions focused on cross-situational learning, this subset of 27 TD children and 20 children with ASD was used in all analyses.

Participant demographics for this sample of 47 children are presented in Table 1. There were 19 males and 8 females in the TD group. There were 19 males and 1 female in the ASD group. Children in the TD group were between 2 and 7 years of age (24 - 95 months) and children in the ASD group were between 4 and 7 years of age (48 - 95 months). The majority of participants in both groups were White and non-Hispanic. Maternal education ranged from 14 to 25 years in the TD group and 12 to 23 years in the ASD group. Paternal education ranged from 14 to 25 years in the TD group and 12 to 24 years in the ASD group.

Information regarding autism variables is presented in Table 2. In the ASD group, 80% of children met criteria for a classification for Autism and 15% met criteria for a classification of Autism Spectrum on the ADOS-2 (see full description below). One child did not receive the ADOS-2 due to behavioral challenges. ADOS-2 calibrated severity scores (CSS; Gotham, Pickles, & Lord, 2009) ranged from 5 to 10, indicating variability in the severity of autism symptomatology. The full scale ranges from 1 – 10 and indicates the level of autism symptomatology demonstrated during the ADOS-2 administration: High (8 – 10), moderate (5 – 7), low (3 – 4), or minimal to no evidence (1 – 2). Per parent report, 80% of children in the ASD group had ever received intensive autism therapy (i.e., more than 20 hours per week), and 75% of children in the ASD group had ever received language intervention.

General Procedure

Participation in the full study involved a one-time visit lasting two hours (for the TD group) to three hours (for the ASD group). Visits were longer for children in the ASD group because they received the ADOS-2. Children were seen in a child-friendly testing suite at the Waisman Center on the University of Wisconsin-Madison campus. Parents were familiarized with study procedures and provided written consent for their children's participation. All procedures were prospectively approved by the University of Wisconsin-Madison Social and Behavioral Science Institutional Review Board.

Activities typically took place in the following order: 1) the cross-situational task and the visual orienting task; 2) the PPVT; 3) the Leiter; 4) the ADOS-2 (ASD group only); 5) the simple word-learning task and the visual paired comparison (VPC) task. In rare cases of scheduling constraints, the order of activities was altered slightly (e.g., the Leiter was finished after the second set of eye-gaze tasks). The first two eye-gaze tasks always occurred at the start of the visit, and the second two eye-gaze tasks took place at the end of the visit. This step ensured that a considerable length of time had lapsed (typically 1.5 to 2.5 hours), limiting the potential for interference across the two word-learning tasks. Play breaks and snacks were given as needed. Because many children with ASD have difficulty transitioning between activities, a visual schedule (see Appendix 1) was used with all participants to ease transitions between activities. The schedule consisted of pictures representing each activity, as well as a small box where children could put a sticker when each activity was finished.

Standardized Assessments

Peabody Picture Vocabulary Test. The Peabody Picture Vocabulary Test, Fourth Edition (PPVT; Dunn & Dunn, 2006) assessed vocabulary comprehension. This activity was introduced by telling children that it was time to point to some pictures. For each item, children were presented with four pictures on a page and instructed to point to one of the pictures (e.g., *Point to canoe. Show me dilapidated*). The PPVT produced a raw score, growth scale value (GSV), age equivalent, and standard score. Administration typically lasted 15 – 30 minutes. GSV scores were used for group matching and for the majority of analyses because they measured children's raw receptive vocabulary skills on an equal-interval scale.

Leiter International Performance Scale-Revised. The Leiter International Performance Scale-Revised (Leiter; Roid & Miller, 2002) assessed nonverbal cognitive abilities. This activity was introduced by telling children that it was time to play some 'quiet games.' They were told that the examiner would not be talking, but that she would show them with her hands and eyes what they should do. Four subtests from the Visualization and Reasoning Battery were administered: Figure Ground ("The Find It Game"), Form Completion ("The Put Together Game"), Sequential Order "The What's Next Game"), and Repeated Patterns ("The Over and Over Game"). Depending on the first item, children were told that their job was either to point to pictures on the page or to put cards into the slots. Each subtest yielded a raw score, growth scale value (GSV), age equivalent, and scaled score. Compilation of the subtests yielded a Brief IQ, with an associated raw total, GSV, age equivalent, and percentile rank. Administration typically lasted 30 – 45 minutes.

Vineland Adaptive Behavior Scales. Parents completed the Vineland Parent/Caregiver Rating Form (Sparrow, Cicchetti, & Balla, 2005) to provide information about children's adaptive behaviors in everyday situations. The Vineland is divided into four domains, each of which contains subdomains: Communication (Receptive, Expressive, & Written), Daily Living Skills (Personal, Domestic, & Community), Socialization (Interpersonal Relationships, Play and Leisure Time, & Coping Skills), and Motor Skills (Gross Motor & Fine Motor). Each domain yielded a standard score and a percentile rank. Each subdomain yielded a raw score, *v*-Scale Score, and age equivalent. The sum of domain standard scores was converted to an overall adaptive behavior composite (ABC).

Social Communication Questionnaire. The Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003) is a 40-item parent questionnaire used as a screening measure for ASD. Although the SCQ manual indicates that it is appropriate for children 4 years of age and older (with a mental age of at least 2 years), it also states that it may be appropriate for children as young as 2 years of age (with a mental age of 2.0 or greater). Parents completed the SCQ Lifetime form. Half of the items on the Lifetime Form ask whether children had exhibited particular behaviors at any point during their lives, and the other items focus on the time between children's 4th and 5th birthdays. If children were younger than five, parents were instructed to focus on behaviors in the past 12 months. The maximum score was 39. The SCQ manual recommends that children with scores greater than 15 receive further evaluation for an ASD; all children in the TD group scored below this cutoff point. Although lowering the cutoff may increase sensitivity and specificity (Allen, Silove, Williams, & Hutchins, 2007; Wiggins, Bakeman, Adamson, & Robins, 2007), the original cutoff was used because the primary purpose in using the SCQ was simply to confirm by parent report that children in the TD group did not show signs of ASD.

ADOS-2. The ADOS-2 is a semi-structured, behavioral diagnostic measure for autism. It typically lasts 30 – 50 minutes and involves a sequence of activities and social presses that elicit social communication and other behaviors associated with ASD. The ADOS-2 was administered only to children in the ASD group; administration was not completed for two participants because of extremely challenging behaviors. Following standard procedures, participants received one of three ADOS-2 modules, depending on their level of spoken language: Module 1 (for children with limited spoken language); Module 2 (for children with phrase speech); or Module 3 (for children with fluent speech). The ADOS-2 confirmed ASD diagnosis and yielded an overall calibrated autism severity score (CSS; Gotham et al., 2009).

Group Matching

Receptive vocabulary was selected as the primary matching variable. PPVT growth scale values (GSV) were used because they provide a measure of children's absolute receptive vocabulary ability on an equal-interval scale—a more appealing psychometric property than the unequal interval scales characteristic of raw scores or age equivalents. An independent samples t-test indicated that mean PPVT GSV was not significantly different between the TD group, n = 27, M = 137.67, SD = 23.47 and the ASD group, n = 20, M = 133.90, SD = 28.29, $t(45) = .50$, $p = .62$. Not surprisingly, PPVT standard scores, $t(43) = 5.75$, $p < .001$, and Brief IQ, $t(44) = 5.85$, $p < .001$, were significantly higher in the TD group than in the ASD group, and mean chronological age was higher in the ASD group, $t(45) = -3.512$, $p = .001$.

Following recommendations from Kover and Atwood (2013), the effect size and variance ratio were examined in addition to the significance test. Use of effect sizes and variance ratios is advantageous because they provide an interpretable, descriptive measure of the magnitude of group differences while limiting the influence of sample size. Smaller effect sizes are associated with more closely matched samples; a Cohen's *d* of 0 indicates well-matched groups, whereas a Cohen's *d* of 1 indicates more poorly matched groups (see Cohen, 1988). Cohen's *d* for the group difference in PPVT GSV was .15, indicating that the mean difference between the groups was relatively small. A variance ratio of 1 indicates no differences in group variances, whereas a variance ratio close to 2 indicates substantial differences. The variance ratio for PPVT GSV was 1.21, meaning that group variances were relatively similar.

Experimental Tasks

Cross-situational word-learning task.

The cross-situational task was modeled after the task used by Smith and Yu (2008). Children were taught new words and tested on novel word comprehension using an eye-tracking paradigm. Successful learning required children to track co-occurrences between labels and objects across ambiguous trials.

Procedure. At the beginning of each visit, children were introduced to the eye-gaze tasks through a social story presented in a ring-bound book with laminated pages (see Appendix 2). The story informed children that it was now time to watch a short movie; that they would go into a room that would be a little bit dark, but that it was no big deal; that their job would be to sit quietly and watch and listen; and that when the movie was finished, they would get a sticker to put on their visual schedule.

The cross-situational word-learning task was conducted on an eye tracker in a soundproof booth to limit external distraction. In most cases, children sat independently in a chair in front of the screen; in a few cases, children sat on a parent's lap. When children were able to sit in the chair independently, a trained research assistant (RA) was present in the sound booth for the duration of the task. The RA ensured that the child remained sitting straight up, in a good position for the recording of eye movements (e.g., by putting her hands lightly on the child's shoulders to help them remain in place). The RA also reminded the child to "watch the movie" in rare cases of drastic inattention (e.g., the child attempted to get out of the chair). When parents were in the room, they wore opaque glasses so that they could not see the experimental stimuli; they were instructed to remain silent and not to interact with their child. The location of children's gaze was logged during the full task. Children's faces were also video-recorded in the event that hand coding of eye gaze was later deemed necessary.

Equipment. The task was presented on a Tobii T60XL Eye Tracker (Tobii). Visual stimuli were presented on a 24-inch wide-screen monitor with a screen resolution of 1920 x 1200 pixels. Auditory stimuli were presented at a level of 65dB through built-in speakers on the eye tracker. The sampling rate was 60 Hz, meaning that gaze points were sampled 60 times per second (i.e., every 16.7 ms). The Tobii determined eye-gaze location by creating reflections on the cornea and pupil through near infrared illumination. Two image sensors embedded in the lower panel of the Tobii monitor record the images of children's eyes and the patterns of cornea and pupil reflection; standard internal processing algorithms estimated the position of the eye and its location on the screen at each point in time (Tobii T60 XL User Manual). Experimental stimuli were presented on the eye tracker using E-Prime 2.0 software; E-Prime Extensions for Tobii integrated use of the two programs.

Calibration. All participants completed the 5-point infant calibration in Tobii Studio immediately prior to the eye-tracking task. During calibration, the Tobii sampled characteristics about the participant's eye gaze and integrated this information to enhance tracking capabilities. The infant calibration option was used because it was particularly engaging—it presented an animated, moving stimulus at each calibration point along with a corresponding sound (in this case, a shaking rattle with a short musical tune). The infant calibration procedure also allowed the examiner to present an intervening stimulus (i.e., a different moving stimulus with a different sound) between the presentations of the five calibration points, which in many cases brought children's attention back to the screen if they had become distracted. The examiner ensured that children's eye gaze fell generally within the boundaries of the five calibration points; calibration points were re-administered as necessary.

Pointing Task. Immediately following the cross-situational eye-tracking task, children completed a behavioral pointing task as an alternative measure of learning. After exiting the soundproof booth, children were told that they should sit on the red couch and that the examiner was going to ask them a few questions. Children were shown a laminated sheet of paper depicting the four novel images they had just been taught. Children were asked identify each of the four target words (e.g., *point to*___ or *where's the* ___"), following a set sequence. The examiner recorded the color of the object to which the child pointed for each novel word on a data sheet (see Appendix 3); responses were later determined to be correct or incorrect based on the key for each stimulus order. Responses were scored as the number of items correct out of four. Because children were presented with four possible responses on the laminated sheet, the probability of identifying one object correctly by chance was 25%.

***Stimuli*.**

Visual stimuli. The cross-situational word-learning task included images of three familiar objects (*ball, cup, shoe*) and four novel objects (see Figure 1). Images of prototypical familiar objects were obtained thorough an online image search. Images of novel objects were created in Microsoft PowerPoint. A total of twenty-eight candidate novel images were pilot tested with 10 adults blind to the hypotheses of the study to ensure that the objects did not closely resemble real objects. Each adult provided written answers to two questions about each object: *First, does it look like any objects that you've encountered in your everyday life? Second, if you were to give the object a name, what would it be? S*everal images were eliminated because they were commonly named by more than one adult (e.g., all 10 adults labeled one object as *star)*. The objects that were used were given the same label by fewer than four people (e.g., protractor, corner joist, bracket, shelf). All images were cropped, placed on a 375 x 825 pixel gray square, and presented on a black background to enhance salience (see Figure 2).

Auditory stimuli. The task included three familiar labels (*ball, cup, shoe*) and four novel labels (*bosa, coro, manu*, *peri*). The familiar words were selected because they were early-emerging words known by many young children with ASD (Toddler Talk Project, unpublished data). The novel words consisted of two syllables and followed the rules of English phonotactics; they were selected because they had been used successfully in prior studies of word learning. Auditory stimuli were recorded by an adult female using child-directed speech and were edited using Praat software (Boersma & Weenink, 2013). Object labels were recorded in isolation for the training phase and within a question frame for the test phase (e.g., *Where's the coro?*). In test trials, a tag question (e.g., *Do you see it?*) was added to the end of each sound file. At most, the time between the end of the target word and the start of the tag question was 400ms. See Appendix 4 for additional information on the auditory stimuli.

Task design.

The cross-situational word-learning task consisted of three phases: familiarization, training, and test (see Table 3). In the familiarization and training phases, the image on the left was labeled first half of the time, which ensured that the temporal placement of the word labels did not provide a consistent cue to label-object association. All stimuli were counterbalanced for side and order of presentation. Two orders (see Appendix 5 & Appendix 6) were created with trials counterbalanced on a number of factors (e.g., label-object associations were switched and trial order was altered) to ensure that children's performance was not driven by specific aspects of the design. Of the children who contributed data for the task (n = 47), 21 received Order A and 26 received Order B. The two orders were collapsed because there were no significant differences in cross-situational eye-gaze accuracy between Order A, M = 55.35, SD = 22.43, and Order B, M = 60.39, SD = 14.90, $t(45)$ = .92, p = .36. Attention-getter movies were interspersed every 4 – 5 trials. The full task lasted approximately 4 minutes.

Familiarization phase. The familiarization phase consisted of four 4-second trials with an inter-stimulus interval (ISI) of 250 ms. The familiarization phase exposed children to the task design. Each trial presented children with two familiar images and two labels (e.g., *Ball. Shoe*). The first word was presented 500 ms into the trial and the second word was presented 2000 ms into the trial. The images remained on the screen for the full 4000 ms. Each word was labeled 2 – 3 times. The familiarization phase lasted approximately 20 seconds.

Training phase. The training phase consisted of 20 4-second trials with an ISI of 250 ms. The training phase exposed children to the four novel words and four novel labels. Each trial presented two novel images and two labels (e.g., *Manu. Coro.*) with no information about which label-object associations were correct. As in the familiarization phase, the first word was presented 500 ms into the trial and the second word was presented 2000 ms into the trial. The images remained on the screen for the full 4000 ms. Each of the four novel words was presented 10 times. The training phase lasted approximately 1.4 minutes.

Test phase. The test phase consisted of 25 5-second trials with an ISI of 500 ms. Each familiar word was tested 3 times, for a total of 9 trials. Each novel word was tested 4 times, for a total of 16 trials. Familiar and novel trials were interspersed. All objects served as both target and distracter. In familiar test trials, children saw two of the familiar objects introduced during the familiarization phase and were asked about one of them (e.g., *Where's the ball? Do you like it?*). In novel test trials, children saw two of the novel objects taught during the training phase and were asked about one of them (e.g., "*Where's the peri? Do you see it?*"). Test trials began with 1000 ms of silence, followed by the test question; the target noun occurred 2000 ms into the trial. The tag question started at 3000 ms, and each trial ended with 1000 ms of silence (see Figure 3). The images remained on the screen for the full 5000 ms. The test phase lasted approximately 2.3 minutes.

Data processing. Data processing was completed in conjunction with a programming consultant. Matlab scripts integrated timing information from E-Prime with gaze data information from the Tobii. Areas of interest (AOIs) were defined by the outer edges of the 375 x 825 pixel grey boxes containing the target images, plus 10 pixels to allow for slight deviations outside the boxes (see Figure 4). Files of continuous time data during the test phase were generated for each participant. Each 16.7 ms frame of time was assigned a code based on the location of children's eye gaze. A 1 indicated a look to the target image, 0 indicated a look to the distracter image, and '-' was recorded if the child was not fixating either picture (e.g., was shifting between images or looking away from the screen). A '-' was also recorded if there was missing data due to Tobii mistracks. An unknown Tobii error produced empty gaze data files for one child in the TD group and 2 children in the ASD group; these sessions were hand-coded from video by trained coders using i-coder and processed using Datawiz. Coders had met lab requirements for reliability (Venker et al., 2013).

Data cleaning.

Accuracy. An initial target window was set from 300 ms – 2000 ms after target word onset (see Venker et al., 2013). Note that the tag question started 1000 ms after target word onset. A plot of the grand mean confirmed that children's looks to the target began to increase approximately 300 ms after target word onset and began to drop off after approximately 2000 ms, supporting selection of this target window. Trials were eliminated when children looked away from the screen more than half of the time during the target window (i.e., did not accumulate 850ms looking time to the images). Participants were eliminated if they contributed fewer than 2 valid trials.

Following standard procedures (Fernald et al., 2008), accuracy was defined as the proportion of time children spent looking at the target image during the target window, divided by the time they spent looking at the target and distracter images combined. Cross-situational eye-gaze accuracy data were contributed by 27/28 children in the TD group (96.43%) and by 20/27 children in the ASD group (74.07%). The number of cross-situational accuracy trials contributed by each participant ranged from 2 – 15 in the TD group, *M* = 8.41, *SD* = 4.22, and from 2 – 15 in the ASD group, *M* = 7.80, *SD* = 4.82. The number of cross-situational accuracy trials contributed did not significantly differ between the TD and ASD groups, *t*(45) = .46, *p* = .65. Cross-situational eye-gaze accuracy was not significantly related to the number of trials children contributed in the full sample or within either group, *p*s > .10. The proportion of looking to target at baseline did not differ significantly from chance, *p* = .29. Thus, cross-situational eye-gaze accuracy analyses compared children's proportion of looks to target against chance probability of looking to the target image (.5).

Given that approximately 25% of the sample of children with ASD did not contribute adequate data for analysis, it was important to determine whether exclusion from the task was systematically related to any child-level characteristics (see Table 4). The 7 children who were excluded from the cross-situational eye-gaze accuracy analyses did not significantly differ from the 20 children who contributed data in age, *t(25)* = -.08, *p* = .94, autism severity, *t(25)* = .88, *p* = .39; or Brief IQ, *t(24)* = -1.72, *p* = .10. Children who were excluded had significantly lower PPVT growth scale values (GSV), *t(25)* = -2.10, *p* = .046, PPVT age equivalents (AE), *t*(25) = -2.20, *p* = .04, and PPVT standard scores (SS), *t(25)* = -2.51, *p* = .02, than children who contributed data, pointing to the impact of language ability on data loss. Children with higher language skills showed greater visual attention to the images during this language-based task.

Children with lower language abilities demonstrated considerable inattention that resulted in their exclusion from the task.

Latency. Following standard procedures (Fernald et al., 2008), latency was defined as the amount of time between word onset and the child's initiation of a shift in gaze away from the distracter image. Thus, only trials in which the child happened to be looking at the distracter image at word onset were included in the latency analyses. Latency values of less than 300 ms or more than 2000 ms were eliminated since they fell outside the target window. Applying the criterion of at least 2 valid trials, 7 children in the ASD group and 15 TD children contributed latency data. Children in the TD group contributed between 2 and 4 trials, *M* = 2.533, *SD* = .640. Children in the ASD group contributed between 2 and 6 trials, *M* = 4.286, *SD* = 1.800.

The large number of trials and participants that were excluded calls into question the validity of the latency data; thus, latency analyses were conducted but were not examined further as part of the overall results. Mean latency in the TD group was 1103.74, *SD* = 342.44, range = 600 ms – 1767 ms. Mean latency in the ASD group was 1009.13, *SD* = 310.10, range = 675 ms – 1511 ms. Based on this limited number of participants, mean RT did not significantly differ between the two groups, *p* = .54. However, given the questionable validity of these data, the latency analyses are not considered further (see also Venker et al., 2013).

Visual paired comparison task.

Procedure. The visual paired comparison (VPC) task assessed visual recognition memory. The task was presented on a 55-inch wall-mounted screen in a soundproof booth to limit external distraction. Children sat on a chair approximately 50 inches in front of the screen. Stimuli were presented using WISP software (Olson, 2011) designed for the Infant Learning Lab

(Saffran, PI). WISP uses a MatLab platform to present image, sound, and movie files. Children's faces were video-recorded and eye-gaze was hand coded offline (see details below).

Visual stimuli. Sixteen novel object images were included in the task. Objects were created in Microsoft PowerPoint and were similar in salience to those used in the two word-learning tasks. None of the images was used in any other task. Two images per trial were presented in grey boxes in the lower left and right corners of a black screen. Because this task assessed visual memory, all trials were silent.

Task design. The VPC task contained 16 trials: 8 familiarization trials and 8 test trials. Each familiarization trial was immediately followed by a test trial, such that familiarization and test trials were 'paired.' Familiarization trials presented children with two identical images, one on each side of the screen. Test trials presented the image that had just been familiarized on one side of the screen and a completely novel image on the other side (see Figure 5). Children were expected to look longer at the novel image (i.e., to show a novelty preference) if they remembered the previous image. Novel images appeared on the left side in half of the test trials. Children never saw the same image in different pairs of trials. Four-second attention getters (animated nature scenes with musical soundtracks) were interspersed every 1 to 2 trials. Two counterbalanced orders were created; half of the objects served as familiarization stimuli and half served as test stimuli in one order; roles were reversed in the other order. For an unknown reason, test accuracy for Order A, $M = 67.54$, $SD = 20.53$, was significantly higher than test accuracy for Order B, $M = 59.66$, $SD = 24.79$, $p < .001$.

A looking contingency was incorporated to ensure that children accumulated 2 seconds of looking time to the screen during each trial. The looking contingency was controlled in real time by the examiner, who was able to see the child's face on a video screen from outside the booth. The examiner pressed a button when the child was looking at the screen and released it if the child looked away. The next trial was presented after the button had been depressed for a total of 2 seconds; if more than 15 seconds elapsed before the child accumulated 2 seconds of looking time, the current trial was aborted and the experiment moved on to the next training trial. This contingency accommodated differences in attentional engagement. The complete VPC task typically lasted around 1.5 minutes (minimum 56 sec, maximum 4.4 minutes).

Manual eye-gaze coding. Children's eye gaze was coded offline by two trained coders, who determined whether children were looking left, right, away from the screen, or were shifting between the images (Fernald et al., 2008). A subset of the VPC task data (4 sessions from each group) was coded independently by both coders. Agreement was calculated by comparing the total proportion of frames on which the two coders agreed. Frame agreement was 94% for the ASD group and 97% for the TD group, indicating good inter-coder agreement. Although inter-coder agreement was high, it was slightly lower than agreement for standard LWL tasks used with children with ASD (Venker et al., 2013). This likely occurred because trials contained no speech describing the images or directing children where to look. Children's eye gaze often flitted away from the images for short periods, making coding more challenging. The VPC task was also the last activity; children's attention may have been limited because at this point they had participated in three similar tasks.

Data cleaning. Eye gaze was examined during the full trial window. Training trials without paired test trials were excluded. The dependent variable was novelty preference, operationalized as the amount of time the child spent fixating the novel image during test trials, divided by the amount of time spent fixating either image. Novelty preference was not affected by variable trial lengths because it considered only looking time to the images; the amount of time spent looking away from the screen was not taken into account. The mean number of trials in the TD group (n = 27) was 7.70, *SD* = 0.61, range = 6 – 8. The mean number of trials in the ASD group (n = 17) was 7.76, *SD* = 0.56, range = 6 – 8. The number of trials contributed by each group did not significantly differ, $t(42) = -0.33$, $p = 74$.

As a preliminary step, children's eye gaze during training trials was examined; looking times to the two identical images were not expected to differ. The target side during familiarization was arbitrarily defined as the image on the same side as the upcoming novel target image. As expected, the amount of time children spent fixating each of the two identical pictures during familiarization did not differ significantly from chance for children in the TD group, M = 47.51%, $t(26) = -1.81$, $p = .08$, or children in the ASD group, M = 51.36%, $t(16) = 0.77$, $p = .46$.

Simple word-learning task.

Procedure. Eye-tracking procedures, equipment, and calibration were identical to those for the cross-situational task. As with the cross-situational task, children completed a pointing comprehension task immediately following the eye-tracking task. Children were presented with a sheet of paper depicting the four novel images they had just been taught and were asked identify each of the four target words by pointing.

Stimuli.

Visual stimuli. The simple word-learning task included images of three familiar objects (*dog, car, book*) and four novel objects (see Figure 6). Familiar images were obtained through an online image search and novel objects were created in PowerPoint. All images were cropped, placed on a 375 x 825 pixel gray square, and presented on a black background.

Auditory stimuli. The simple word-learning task included three familiar labels (*dog, car, book*) and four novel labels (*toma, subo, deepu, modi).* The non-words consisted of two syllables and followed the rules of English phonotactics; additionally, they matched the non-words in the cross-situational task on final vowel sound. Auditory stimuli were recorded by an adult female using child-directed speech and were edited using Praat software (Boersma & Weenink, 2013). Object labels were recorded in isolation for the training phase and within a question frame for the test phase (e.g., *Where's the subo?*). In test trials, a tag question (e.g., *Do you like it?*) was added to the end of each sound file. See Appendix 7 for additional information on the auditory stimuli.

Task design. The simple word-learning task mirrored the cross-situational word-learning task with one critical difference: training trials were unambiguous. Each training trial presented only one image and one label, meaning that children did not have to rely on cross-situational co-occurrences to learn label-object associations. As in the cross-situational task, children heard each label 10 times. The simple word-learning task consisted of three phases: familiarization, training, and test (see Table 3). All stimuli were counterbalanced for side and order of presentation.

Two orders (See Appendix 8 & Appendix 9) were created with trials counterbalanced on a number of factors (e.g., label-object associations were switched and trial order was altered) to ensure that children's performance was not driven by specific aspects of the design. The two orders were collapsed because there were no significant differences in cross-situational eye-gaze accuracy between Order A, *M* = 61.73, *SD* = 4.60, and Order B, *M* = 60.69, *SD* = 2.37, *t*(46) = -

.22, *p* = .83. Attention-getter movies were interspersed every 4 – 5 trials. The complete task lasted approximately 4.28 minutes.

Familiarization phase. The familiarization phase consisted of four 2-second trials with an ISI of 250 ms. The familiarization phase exposed children to the task design. Each trial presented children with one familiar image and one label (e.g., *Book*). The word was presented 500 ms into the trial and was followed by 500 ms of silence. The image remained on the screen for the full 2000 ms. Each familiar word was presented 1 or 2 times. The familiarization phase lasted approximately 13 seconds.

Training phase. The training phase consisted of 40 2-second trials with an ISI of 250 ms. The training phase exposed children to four novel words and four novel labels. Each trial presented one novel image and one label (e.g., *Toma*). The word was presented 500 ms into the trial. The images remained on the screen for the full 2000 ms. As in the cross-situational task, each of the four novel words was presented 10 times. The full training phase lasted approximately 1.8 minutes.

Test phase. The test phase design was identical to that in the cross-situational task. The 3 familiar words were tested 3 times each and the 4 novel words were tested 4 times each for a total of 25 test trials. Test trials were 5 seconds long and had an ISI of 500 ms; familiar and novel trials were interspersed. Trials began with 1000 ms of silence; the target noun occurred 2000 ms into the trial. The tag question started at 3000 ms, and each trial ended with 1000 ms of silence. The test phase lasted approximately 2.3 minutes.

Data processing. Data processing steps were identical to those described in Chapter 2. Due to Tobii malfunction, one child in the TD group had an empty gazedata file, meaning that

27/28 TD kids had eye-tracking data recorded. The simple word-learning task was not administered to 2 children in the ASD group due to behavioral interference, meaning that 25/27 ASD kids had eye-tracking data recorded.

Data cleaning. The target window was set at 300 ms – 2000 ms after target word onset. Trials were eliminated when children looked away from the screen more than half of the time during the target window. Participants were eliminated if they contributed fewer than 2 valid trials. Accuracy was defined as the proportion of time children spent looking at the target image out of the time they spent looking at the target and distracter images combined during the target window (Fernald et al., 2008).

Simple word learning eye-gaze accuracy data were contributed by 27/27 children in the TD group and by 17/20 children in the ASD group. The number of simple word learning eye-gaze accuracy trials contributed by each participant ranged from 2 – 14 in the TD group, *M* = 8.30, *SD* = 3.93, and from 3 – 14 in the ASD group, *M* = 7.94, *SD* = 3.65. The number of simple word-learning accuracy trials contributed did not significantly differ between the TD and ASD groups, *t*(42) = .30, *p* = .77. Simple word learning eye-gaze accuracy was not significantly related to the number of trials children contributed in the full sample or within either group, *p*s > .36. The proportion of looking to target at baseline did not differ significantly from chance, *p* = .83. Simple word-learning eye-gaze accuracy analyses compared children's proportion of looks to the target against chance (.5). Given that the limited number of trials precluded examination of latency in the cross-situational word-learning task, latency was not examined for the simple word-learning task.

Familiar word comprehension. Although it was not a primary research question, data from familiar word trials were analyzed to confirm that children were able to demonstrate comprehension of known words in the eye-gaze paradigm. All children were reported by their parents to comprehend and/or produce each of the familiar words. (One participant in the ASD group was not reported to understand 'car,' but he was reported to say it spontaneously; accuracy was 98% and 100% for each of his *car* trials). The cross-situational task and simple word-learning task each incorporated 3 familiar words. The 6 familiar words were tested 3 times each, yielding a maximum of 18 trials. Data for all familiar word trials were collapsed because mean familiar word accuracy did not differ between the two tasks, $p = .69$.

Of the children who had contributed cross-situational eye-gaze accuracy data, 27/27 children in the TD group and 19/20 children in the ASD group also contributed familiar accuracy data. The number of familiar accuracy trials contributed by each participant ranged from 3 – 17 in the TD group, $M = 11.11$, $SD = 3.60$, and from 4 – 17 in the ASD group, $M = 10.79$, $SD = 3.79$. The number of familiar trials contributed did not significantly differ between the TD and ASD groups, $t(42) = .30$, $p = .77$. The number of trials contributed was not significantly correlated with familiar word recognition accuracy in the full sample or within each group, ps > .55. As with the cross-situational and simple word-learning tasks, the proportion of looking to target at baseline did not differ significantly from chance for familiar trials, $p = .20$.

The proportions of familiar trials, cross-situational trials, and simple word-learning trials retained across the full sample were examined to determine whether data loss was significantly different across the three conditions. The mean proportion of trials retained in the full sample was 60.99% for familiar words; 50.93% for cross-situational word learning; and 50.99% for simple word learning. Paired-samples *t*-tests revealed that on average, a significantly higher proportion of trials was retained for familiar words than for novel words in both the cross-situational task and the simple word-learning task, *p*s < .01.

Gap-overlap task.

Procedure. The visual orienting task was modeled after Landry and Bryson's (2004) gap-overlap task. Stimuli were presented on a 55-inch central screen and two 19-inch peripheral side monitors. This setup, typically used for infant head turn studies, was selected because it incorporated a looking contingency that allowed the examiner to control the timing and presentation of stimuli in real time. Additionally, the relatively large screen increased engagement and use of the peripheral side monitors increased the potential for sticky attention, since the stimuli appeared in the periphery. Children sat on a chair approximately 50 inches in front of the center screen. The side monitors were approximately 40 inches to the left and right of the chair, which required children to turn their heads approximately 65 to 70 degrees to view the side stimuli (see Figure 7).

Several options were considered prior to selecting the setup described above. One possibility was an eye tracker, with stimuli appearing in the center and on the edges of the screen. The advantage would have been automatic tracking of eye gaze. The primary limitation would have been the limited distance between the central and peripheral stimuli (i.e., shifting attention would require gaze to move only a few inches). Additionally, incorporating an automatic gaze contingency was deemed unnecessarily risky. Presenting both central and peripheral stimuli on a large screen was another possibility, but pilot testing suggested that identifying eye movements from the center to the sides of the screen in real time was challenging.

Visual stimuli. Visual stimuli were movie clips of colorful, dynamic shape patterns sampled from *Animated Classical Music for Babies, Vol.1*. Sound was removed because the task focused solely on visual attention. To control for salience, three pairs of clips with identical movement patterns in different colors were yoked within trials, such that one served as the center stimulus and one served as the side stimulus. The attention getter movies were outer space-themed clips (e.g., stars, silver coils, constellations).

Task design. The task included 10 gap trials and 10 overlap trials presented in a semi-randomized order. No more than two trials of one type were presented consecutively. A looking contingency was incorporated so that stimulus presentation was based on children's gaze behaviors. The looking contingency was controlled in real time by the examiner, who was able to see the child's face on a video screen from outside the booth. The examiner pressed a button when the child was looking at the screen and released it if the child looked away. Trials in which a child never looked to the center for 1 second were repeated at the end of the experiment. However, these trials were discarded because so few trials were repeated and including them did not decrease the number of children excluded from the analyses.

In both conditions, attention-getter trials lasting 5 – 11 seconds were interspersed every 2-3 trials on one of the three screens. Pilot testing demonstrated that this manipulation helped prevent children from attempting to anticipate where and when the peripheral stimuli would appear. Two counterbalanced orders with a different sequence of test trials were created to ensure that children's performance was not affected by specific aspects of the experimental design. A *t*-test indicated that mean latency did not significantly differ between the two orders for the gap condition, $p = .57$, or for the overlap condition, $p = .15$. Thus, the two orders were collapsed. A *t*-test also indicated that mean latency did not significantly differ for trials with the peripheral stimulus on the left versus the right side for the gap condition, $p = .34$, or for the overlap condition, $p = .84$.

See Figure 8 for a visual depiction of the gap condition. The gap condition measured shifting of attention; trials contained a short gap between the central and side stimuli. Gap trials proceeded as follows: the center stimulus was presented; the child had up to 12 seconds to accumulate 1 second of looking time to the center; after the child looked at the center for 1 second, the center stimulus was extinguished and the side stimulus appeared. The next trial was presented after the child accumulated 1 second of looking to the side stimulus or when 8 seconds had elapsed. The gap condition was designed so that the side stimulus would appear immediately after the center stimulus was extinguished; previous studies have incorporated a gap of 200 ms (Elsabbagh et al., 2009) or 250 ms (Landry & Bryson, 2004). However, it was later determined that the stimulus presentation software required more time than expected to load the side video, which produced a gap of between 800 and 866ms.

See Figure 8 for a visual depiction of the overlap condition. The overlap condition measured disengagement of attention; trials presented overlapping central and side stimuli. Overlap trials proceeded as follows: the center stimulus was presented; the child had up to 12 seconds to accumulate 1 second of looking time to the center; after the child looked at the center for 1 second, the side stimulus appeared *while the center video played continuously*; the next trial was presented after the child accumulated 1 second of looking to the side stimulus, or when 8 seconds had elapsed. For most children, the complete task lasted approximately 3.5 minutes.

Manual eye-gaze coding. Children's eye gaze was coded offline by two trained coders. Coding took place in two steps. In Step 1 trials were categorized as valid or invalid. Only valid trials were retained for the analyses. In a valid trial, the child was fixating the center stimulus when the side stimulus appeared, and the child either looked at the center stimulus for the entire trial (i.e., a timeout trial) or made their first shift in gaze toward the 'correct' side (i.e., the side where the stimulus appeared). Invalid trials were eliminated because they did not provide an accurate measure of latency for one of several reasons, including: the child was not looking at the center stimulus when the side video appeared; the child first looked at the 'incorrect' (blank) side screen; or the child was blinking or already shifting when the side video appeared (see also Kelly et al., 2012; Landry & Bryson, 2004). In Step 2, coders recorded the moment at which the child first initiated a shift in gaze away from the center stimulus toward the side stimulus. Latency to shift attention (gap condition) or disengage attention (overlap condition) was operationalized as the amount of time that lapsed between the presentation of the side stimulus and the initiation of the child's shift in eye gaze away from the center stimulus.

A subset of the data (4 sessions from each group) was coded independently by two trained coders. Agreement was conducted separately for each of the two coding steps. Percentage of agreement for categorizing valid vs. invalid trials was 96.25% for the ASD group and for the TD group. Coders discussed any discrepancies and reached a consensus. Shift agreement compared frames in which the child's eye gaze was shifting between the images. Shift agreement was 100% for the ASD group and 97.92% for the TD group, which indicates excellent agreement.

Data cleaning. The first step of data cleaning was identification and removal of invalid trials. Ten trials were eliminated due to examiner error (inaccurate chair placement). Trials shorter than 100 ms were removed (*n* = 7) because they were likely to have been planned prior to the appearance of the peripheral stimulus (Elsabbagh et al., 2013). Participants were excluded from a condition if they contributed fewer than 2 valid trials. Of the 20 children with ASD who contributed cross-situational eye-gaze accuracy data, 4 were excluded from the gap condition because of inadequate data, and none was excluded from the overlap condition. Of the 27 TD children who contributed cross-situational eye-gaze accuracy data, 3 were excluded from the gap condition because of inadequate data, and none was excluded from the overlap condition. The gap condition contained more invalid trials because the inadvertent delay in presentation of the side video caused many children to shift their eyes away from the center stimulus before the side stimulus appeared. Additionally, one child in the TD group was excluded from both conditions because he showed no evidence of understanding that he was allowed to look at the side monitors.

Children in the TD group contributed 3 – 9 gap trials, *M* = 4.65, *SD* = 1.80, and 5 – 10 overlap trials, *M* = 8.58, *SD* = 1.33. Children in the ASD group contributed 2 – 9 gap trials, *M* = 4.69, *SD* = 2.12, and 3 – 10 overlap trials, *M* = 7.40, *SD* =1.76. The mean number of gap trials contributed did not significantly differ between the TD and ASD groups, *t*(37) = -.06, *p* = .96. The mean number of overlap trials contributed was significantly higher in the TD group than the ASD group, *t*(44) = 2.58, *p* = .013.

Timeout trials (i.e., trials in which children did not shift their attention to the side stimulus in the 8-second trial) were rare in the gap condition, likely because the center stimulus did not compete with the side stimulus. The mean number of timeout trials in the gap condition was *M* = .04, *SD* = .21 in the TD group, and *M* = .06, *SD* = .25 in the ASD group. The number of gap timeout trials did not significantly differ between the TD and ASD group, *t*(37) = -.258, *p* = .80. Timeout trials were more frequent in the overlap condition, likely because the overlapping stimuli created competition. The mean number of timeout trials in the overlap condition was *M* = 1.73, *SD* = 1.61 in the TD group, and *M* = .75, *SD* = .91 in the ASD group. Contrary to expectations, the number of overlap timeout trials was significantly greater in the TD group than in the ASD group, *t*(44) = 2.43, *p* = .02.

Based on previous work (Landry & Bryson, 2004), timeout trials were assigned latency values equal to the full trial length (8000 ms). Raw latency data for the gap and overlap conditions were skewed; thus, these variables were transformed into natural logarithms in an attempt to more closely approximate the normal distribution (Elsabbagh et al., 2009; Landry & Bryson, 2004). Examination of the distributions indicated that the log-transformed variables more closely approximated the normal distribution.

Eye-gaze patterns during cross-situational learning. Based on work by Yu and Smith (2011), children's eye gaze was examined during the training phase of the cross-situational learning task. A variable was created to record the mean amount of time per trial each child spent fixating the images during the training phase. This variable provided a measure of children's visual attention to the images during training. All training trials were included, even those in which the child's eye gaze was not recorded, because each trial represented an opportunity for the child to look at the novel objects.

Next, variables were created to characterize eye-gaze patterns—number of shifts and length of longest look—during the training phase. Following the cleaning guidelines instated for the test phase (i.e., 50% of looking during the window of interest), any training trials with less than 2000 ms of looking time to the images were excluded. This criterion ensured that children had adequate time to switch gaze between the two images. Two children in the TD group were excluded because they had fewer than 2 training trials after data cleaning. Gaze duration had to be at least 100 ms to be counted as a fixation. If a child made multiple fixations within an image AOI (i.e., the grey boxes plus 10 pixels outside the boxes), these fixations were summed. If children fixated within the AOI of only one image during a training trial, they had 0 attention shifts. If children looked at one image AOI and then the other, they had 1 attention switch.

The length of longest fixation during training was determined by recording the longest amount of time the child spent looking at one of image AOI (minimum 100 ms). Using the definition proposed by Yu and Smith (2011), strong cross-situational learners were those who looked absolutely longer at the target image during test (i.e., cross-situational eye-gaze accuracy > .50), and weak learners who looked at the images equally or looked longer at the distracter image. Overall, 16 children were categorized as weak learners (9 TD children and 7 children with ASD) and 31 children were categorized as strong learners (18 TD children and 13 children with ASD).

Table 1
Participant Characteristics

	TD Group ($n = 27$)		ASD Group ($n = 20$)		
	n (%)		n (%)		
Sex					
Female	8 (30%)		1 (5%)		
Male	19 (70%)		19 (95%)		
Race					
White	25(93%)		17(85%)		
Black	2 (7%)		2 (10%)		
Asian			1 (5%)		
Ethnicity					
Non-Hispanic	27(100%)		19 (95%)		
Hispanic	0 (0%)		1 (5%)		
	Mean (SD)	Range	Mean (SD)	Range	t-test of group differences
Chronological Age (months)	55.78 (20.42)	24 – 95	75.25 (16.32)	48 – 95	$p = .001$**
Maternal Education (years)	19.33 (2.63)	14 – 25	16.40 (3.05)	12 – 23	$p = .001$*
Paternal Education (years)	18.58 (3.23)	14 – 25	16.41 (3.21)	12 – 24	$p = .04$*
PPVT SS	119.68 (10.79)	99 – 143	93.60 (19.26)	62 – 122	$p < .001$**
PPVT AE	73.67 (24.62)	40 – 115	69.00 (25.16)	22 – 113	$p = .53$
PPVT GSV	137.67 (23.47)	103 – 175	133.90 (28.28)	62 – 174	$p = .62$
Leiter Brief IQ	122.22 (12.68)	93 – 145	95.26 (18.60)	60 – 133	$p < .001$**
Vineland ABC	109.56 (11.77)	82 – 130	80.05 (12.90)	66 – 117	$p < .001$**

Note. The TD group included typically developing children and the ASD group included children with autism spectrum disorders. PPVT indicates the Peabody Picture Vocabulary Test. SS are standard scores, AE are age equivalents, and GSV are growth scale values (raw scores on an equal interval scale). Leiter indicates Leiter International Performance Scale-Revised. Vineland ABC is the adaptive behavior composite from the Vineland Adaptive Behavior Scale.

Table 2

Autism and Intervention: Descriptive Information

	ASD Group (*n* = 20)	
	n (%)	
Existing Diagnosis		
Autism	7 (35%)	
PDD-NOS/Autism	12 (60%)	
Spectrum	1 (5%)	
Asperger Syndrome		
Intensive Autism Intervention		
Yes	16 (80%)	
No	4 (20%)	
Language Intervention		
Yes	15 (75%)	
No	5 (25%)	
ADOS-2 Module		
Module 1 No Words	1 (5%)	
Module 1 Some Words	2 (10%)	
Module 2 Younger than 5	2 (10%)	
Module 2 Age 5 or Up	4 (20%)	
Module 3	10 (50%)	
ADOS-2 not completed	1 (5%)	
ADOS-2 Classification		
Autism	16 (80%)	
Autism Spectrum	3 (15%)	
ADOS-2 not completed	1 (5%)	
	Mean (SD)	Range
ADOS-2 CSS	7.42 (1.71)	5 – 10

Note. PDD-NOS represents pervasive developmental disorder, not otherwise specified. Intensive autism intervention was defined as at least 20 hours per week. ADOS-2 represents the Autism Diagnostic Observation Schedule, 2nd Edition. CSS represent ADOS-2 calibrated severity scores, which are comparable across age and language levels.

Table 3

Cross-Situational and Simple Word-Learning Task Design

	Cross-Situational Word-Learning Task		Simple Word-Learning Task	
	Content	Approximate Length	Content	Approximate Length
Familiarization	4 trials	20 seconds	4 trials	13 seconds
Training phase	20 trials	1.4 minutes	40 trials	1.78 minutes
Test phase	25 trials: 16 novel, 9 familiar	2.3 minutes	25 trials: 16 novel, 9 familiar	2.3 minutes
		4 minutes		4.3 minutes

Note. The cross-situational task presented two label-object pairings per trial. The simple word-learning task presented one label-object pairing per trial. Attention-getters were interspersed throughout the trials in all phases. Inter-stimulus intervals were 250 ms.

Table 4

Characteristics of Children with ASD Excluded from the Study

	Excluded (*n* = 7)		Retained (*n* = 20)	
	Mean (SD)	Range	Mean (SD)	Range
Chronological Age (in months)	74.71 (9.93)	57 – 89	75.25 (16.32)	48 – 95
Brief IQ	81.29 (17.90)	54 – 107	95.26 (18.60)	60 – 133
Autism Severity	8.17 (2.14)	4 – 10	7.42 (1.71)	5 – 10
PPVT GSV*	108.86 (23.37)	80 – 146	133.90 (28.29)	62 – 174
PPVT AE*	46.14 (18.03)	27 – 78	69.00 (25.16)	22 – 113
PPVT SS*	72.14 (20.05)	47 – 109	93.60 (19.26)	62 – 122

Note. **p* < .05. Children with inadequate attention during the test phase of the cross-situational eye-gaze task (i.e., fewer than 2 trials) were excluded. Brief IQ was measured by the Leiter International Performance Scale-Revised. Autism severity was measured by ADOS-2 calibrated severity scores. PPVT represents the Peabody Picture Vocabulary Test. GSV indicates growth scale values. AE indicates age equivalents. SS indicates standard scores.

Figure 1. Familiar images (top panel) and novel images (bottom panel) in the cross-situational task.

Figure 2. Example of visual layout for trials in the cross-situational task.

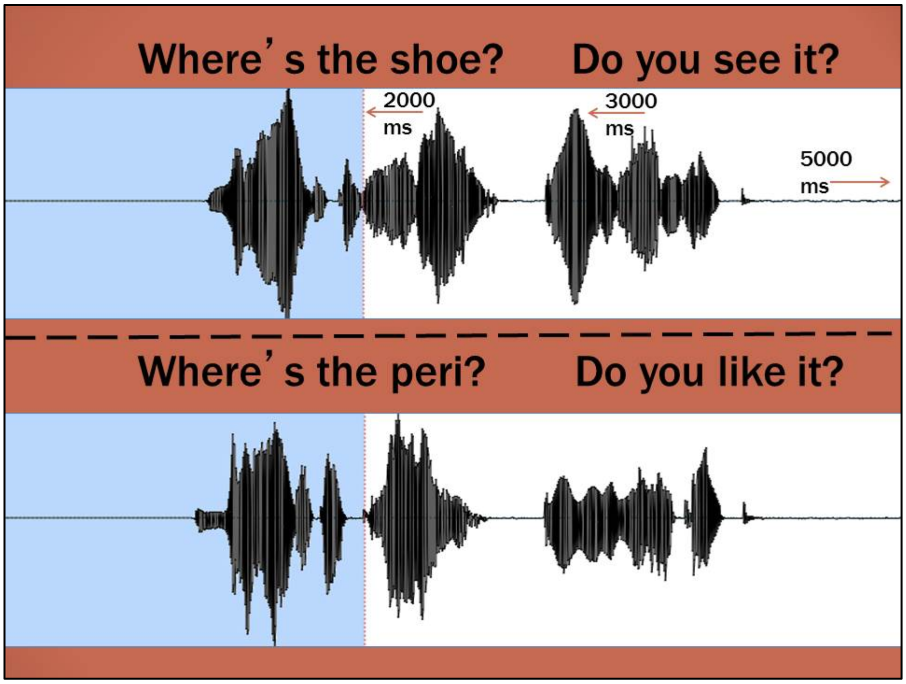

Figure 3. Time course of auditory stimuli for cross-situational and simple word-learning test trials. The target noun was time-locked to occur 2 seconds into the trial. Each test trial lasted 5 seconds.

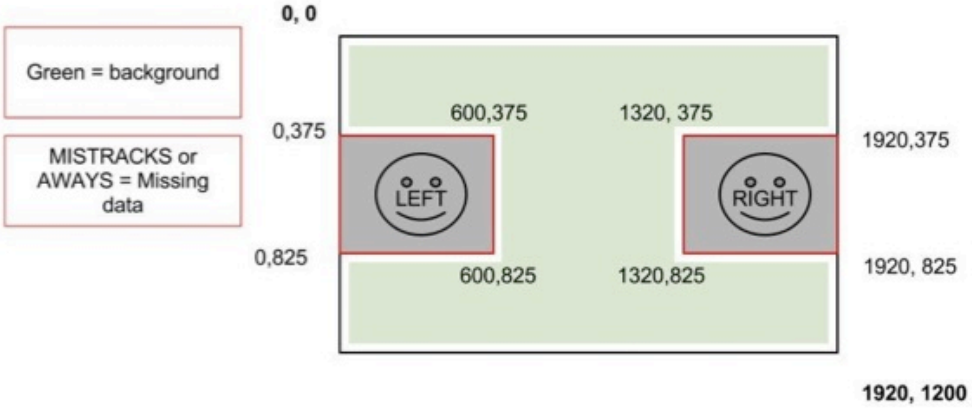

Figure 4. Tobii T60 XL eye tracker screen dimensions and areas of interest.

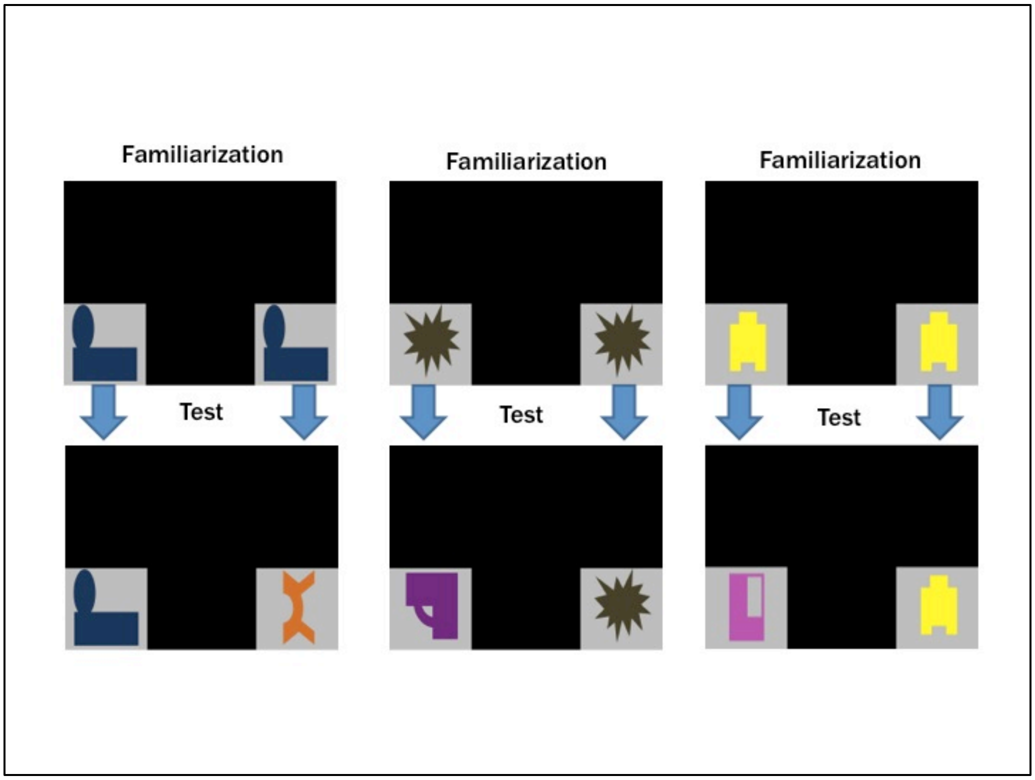

Figure 5. Visual paired comparison task design. Each familiarization trial was immediately followed by a test trial. Examples of three training-test progressions are presented.

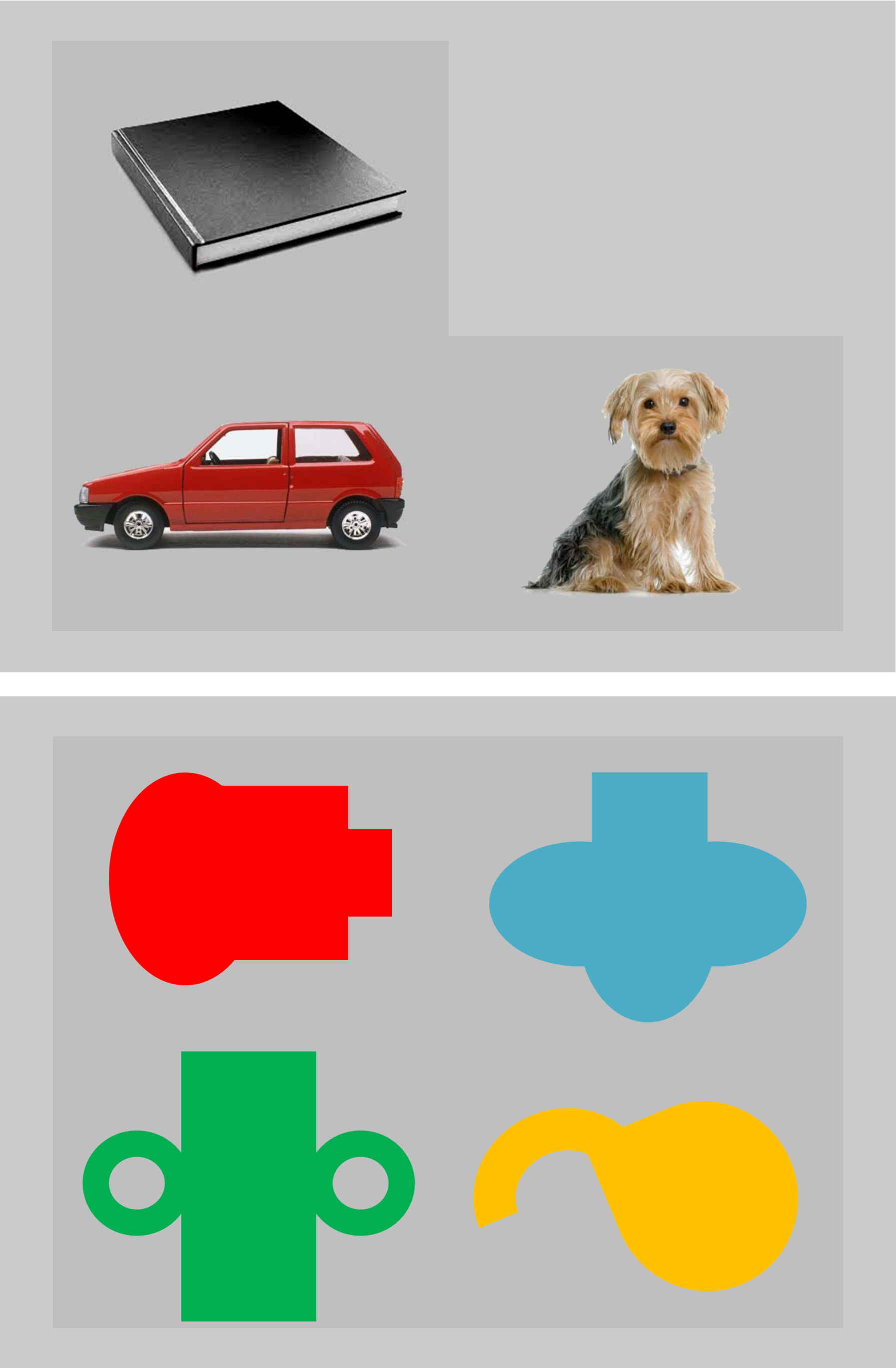

Figure 6. Familiar images (top panel) and novel images (bottom panel) in the simple word-learning task.

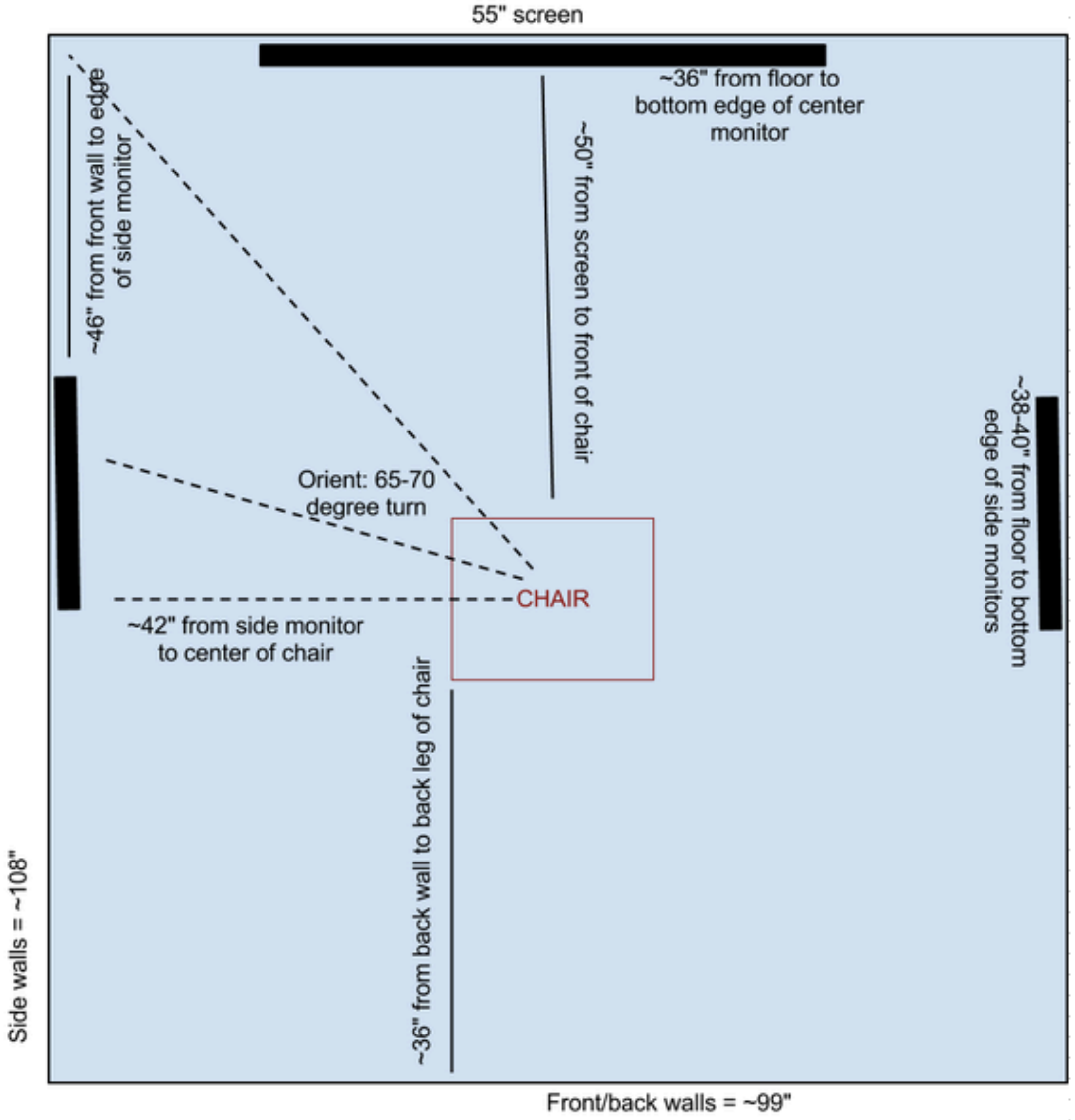

Figure 7. Visual depiction of the booth setup in the gap-overlap task.

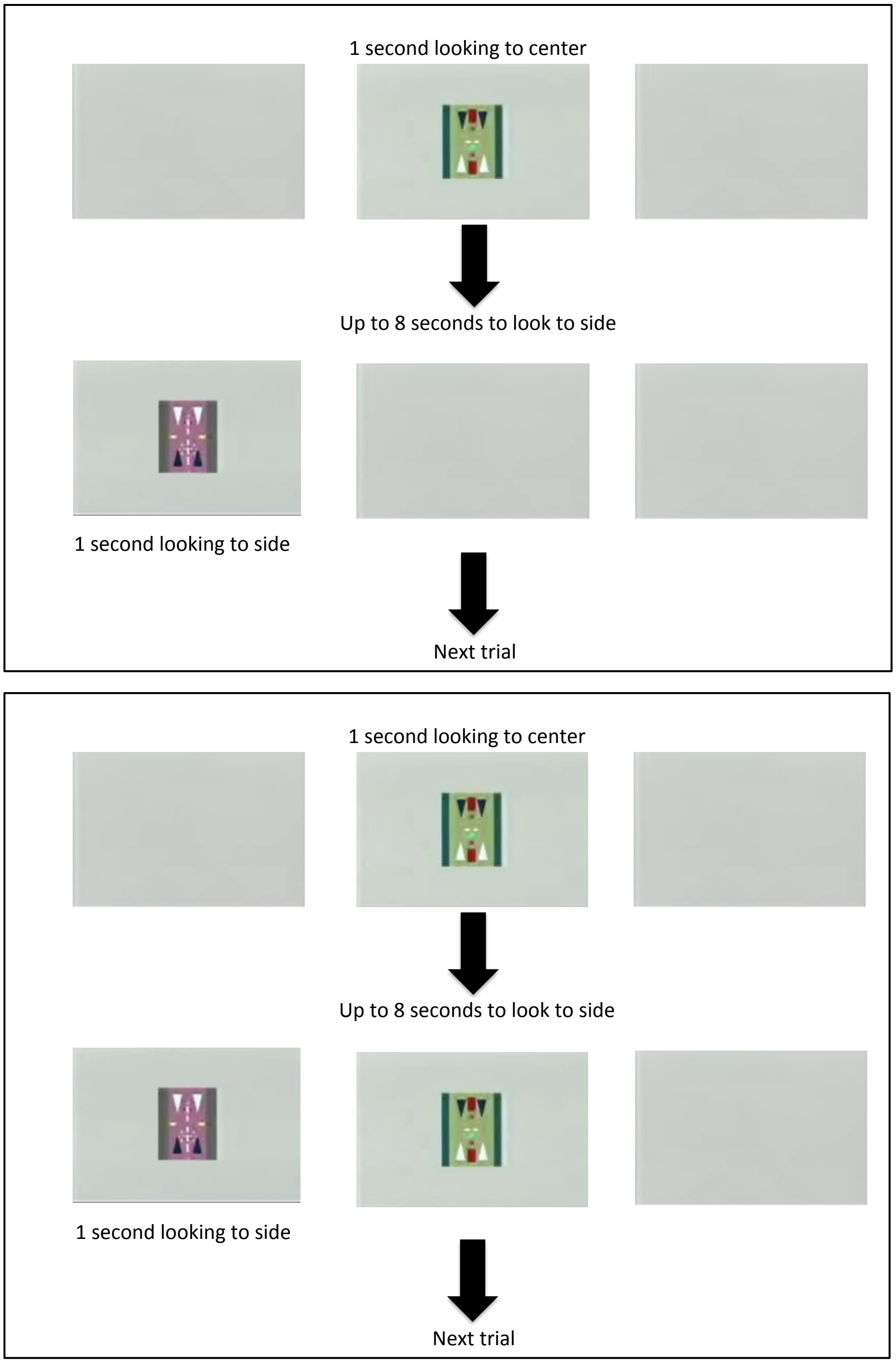

Figure 8. Gap-overlap task design. The gap condition (top panel) measured shifting of attention and the overlap condition (bottom panel) measured disengaging of attention.

Chapter 3: Results

Do Children with ASD Show Deficits in Cross-Situational Learning?

The first aim was to determine whether children with ASD demonstrated cross-situational word-learning deficits as measured by eye-gaze and by a behavioral pointing task. Performance of children with ASD was compared to performance in typically developing (TD) children matched on receptive vocabulary. In addition to group comparisons of cross-situational learning, analyses were also conducted to determine whether learning differed from chance levels of responding—in other words, whether the TD and ASD groups, on average, showed significant evidence of cross-situational learning. Because two different measures of learning were recorded, analyses also examined agreement between the eye-gaze and pointing measures of cross-situational learning. As a first step, children's familiar word recognition was examined to ensure comprehension of known words as measured by eye gaze.

Familiar word recognition. Familiar word trials were interspersed among the novel test trials in both the cross-situational and simple word-learning tasks; all familiar trials were combined because there were no significant differences in familiar word recognition accuracy between the two tasks. Familiar word recognition was examined to confirm that children's eye gaze provided evidence that they comprehended known words. Recall that accuracy was defined as the proportion of looking to the target image, out of the total looking time to the target and distracter images combined. All 27 children in the TD group and 19/20 children in the ASD group contributed adequate trials for familiar word recognition. Time course data for familiar word recognition are presented in Figure 9. Mean familiar word recognition accuracy in the TD group was 78.42%, *SD* = 14.13%, range = 50.50% to 97.58%. Mean accuracy in the ASD group was 75.81%, *SD* = 12.55%, range = 45.67% to 96.48%. Mean familiar word recognition accuracy did not significantly differ between the two groups, *t*(44) = .65, *p* = .52, *d* = .20. Familiar accuracy was significantly greater than chance in the TD group, *t*(26) = 10.45, *p* <. 001, and the ASD group, *t*(18) = 8.96, *p* < .001, demonstrating that, on average, both groups comprehended the familiar words. In summary, findings from the familiar word trials indicated that the TD and ASD groups did not significantly differ in their mean accuracy for familiar word recognition. Despite considerable variability in mean accuracy across children, on average children's eye-gaze provided evidence of their comprehension of known words in both groups.

Cross-situational word learning: group comparisons. Next, analyses were conducted to determine whether cross-situational learning significantly differed between the TD and ASD groups. It was hypothesized that the TD group would outperform the ASD group. Accuracy for the eye-gaze measure was defined as the proportion of time looking to the target image, out of the total looking time to the target and distracter images combined. Time course data are presented in Figure 10. Mean eye-gaze accuracy in the TD group was 55.88%, *SD* = 20.10, range = 0% to 100%. Mean accuracy in the ASD group was 61.19%, *SD* = 16.35, range = 33.46% to 86.52%. Contrary to expectations, mean cross-situational eye-gaze accuracy did not significantly differ between the two groups, *t*(45) = -0.97, *p* = .34, *d* = .29. The group difference remained non-significant when the one child in the TD group with a score of 0 was removed.

Next, analyses were conducted to determine whether mean accuracy in the cross-situational behavioral pointing task differed between the TD and ASD groups. Each of the four novel words was tested once, yielding a maximum score of 4. Mean pointing accuracy in the TD group was 2.26, *SD* = 1.40, range = 0 – 4. Mean pointing accuracy in the ASD group was 1.95, *SD* = 1.64, range = 0 – 4 (see Figure 11). Because the pointing task included only 5 possible responses (0 – 4), the dependent variable was considered ordinal and did not meet the assumption of normality required for parametric tests; for this reason, non-parametric tests were used. First, a Mann Whitney U test—the non-parametric equivalent of an independent samples *t*-test—tested whether median pointing accuracy differed between the TD group and ASD group. The assumptions for this test were met because the dependent variable (pointing accuracy) was ordinal; the independent variable (group) was categorical; the observations were independent; and the distributions of the two groups had the same shape. The mean rank in the TD group was 25.19 (median = 2) and the mean rank in the ASD group was 22.40 (median = 2). Cross-situational pointing accuracy did not significantly differ between the two groups, $U = 238$, $p = .47$.

As another approach to characterizing the cross-situational pointing data, a chi square analysis was conducted to determine whether the proportions of children who performed above versus below chance differed between the two groups. In the TD group, 70.40% of children performed above chance and 29.60% of children performed below chance. In the ASD group, 60.00% of children performed above chance and 40.00% performed below chance. A chi square analysis indicated that the proportions of children in the TD and ASD groups who performed above chance did not significantly differ, $\chi^2 = .55$, $p = .54$. Contrary to predictions, there were no significant differences in cross-situational word learning between the vocabulary-matched TD and ASD groups in both the eye-gaze and pointing measures.

Cross-situational word learning: who learned? Children’s performance in the eye-gaze task and pointing task was compared to chance levels of responding to determine whether children showed significant evidence of learning at the group level. In the eye-gaze measure, accuracy significantly exceeded chance in the ASD group, $t(19) = 3.06$, $p = .006$, providing evidence that, on average, children the ASD group understood the novel words. However, mean cross-situational eye-gaze accuracy did not differ significantly from chance in the TD group, $t(26) = 1.52$, $p = .14$, meaning that eye-gaze patterns provided no evidence that, as a group, TD children had learned the novel words. When the child with a score of 0 (who was 4½ years of age) was removed from the TD group, performance at the group level was significantly greater than chance, $t(25) = 2.40$, $p = .02$.

Next, a one-sample Wilcoxon test was conducted to determine whether children in each group performed significantly above chance in the pointing task (i.e., whether the median value differed significantly from 1). The median value was significantly greater than 1 in the TD group, $p < .001$, and in the ASD group, $p = .026$, indicating that, on average, children in both groups demonstrated above-chance comprehension on the pointing task. This finding stands in contrast to the group-level eye-gaze results, which indicated that only the children with ASD performed above chance. In summary, both the TD and ASD groups on average showed evidence of cross-situational learning in the pointing task, but only the ASD group showed evidence of learning in the eye-gaze measure. However, it should be noted that performance in the TD group was significantly greater than chance when the child in the TD group with a cross-situational eye-gaze accuracy score of 0 was removed.

How did eye-gaze and pointing measures compare? Next, performance in the eye-gaze measure and the pointing measure were compared within each group using non-parametric Spearman correlations to determine whether they provided convergent information. The correlation between cross-situational pointing accuracy and cross-situational eye-gaze accuracy in the TD group was $\rho = .43$, $p = .026$, meaning that the TD children who looked relatively longer at the target image in the eye-gaze task also performed relatively better in the pointing task. The correlation between cross-situational pointing accuracy and cross-situational eye-gaze accuracy in the ASD group was ρ = .05, *p* = .83, meaning that there was not a significant relationship between performance in the eye-gaze task and the pointing task for the children with ASD. Performance on the eye-gaze and pointing tasks was correlated in the TD group, but not in the ASD group, indicating convergent validity between the two measures only for the TD group.

What cognitive and linguistic skills support cross-situational word learning?

Next, the relationship between cross-situational learning and child-level variables was examined. Child characteristics—age, vocabulary level, IQ, and autism severity—were examined first, followed by specific abilities hypothesized to underlie cross-situational word learning—visual recognition memory and simple word learning.

Child-level characteristics.

Table 5 presents correlations between cross-situational eye-gaze accuracy and age, vocabulary level (PPVT growth scale value), and IQ within each group; autism severity level was also examined for the ASD group. Given that vocabulary and age were hypothesized to have a positive association with cross-situational learning, one-tailed *p* values were used for these variables. In the ASD group, accuracy on the eye-gaze measure was not significantly associated with Brief IQ or autism severity, but it was marginally associated with vocabulary level, *r* = .34, *p* = .07 (one-tailed), and with age, *r* = .33, *p* = .08 (one-tailed). In the TD group, cross-situational eye-gaze accuracy was significantly and positively associated with age, *r* = .38, *p* = .03 (one-tailed), and significantly and negatively associated with Brief IQ. The negative correlation between Brief IQ and age in the TD sample suggested a ceiling effect of the Leiter; indeed, the correlation between Brief IQ and cross-situational eye-gaze accuracy in the TD group was no longer significant, *p* = .24, when age was statistically controlled through a partial correlation.

Table 6 presents correlations between child-level variables and cross-situational pointing task accuracy. In the ASD group, cross-situational pointing task accuracy was significantly associated with age, ρ = .48, *p* = .02 (one-tailed) but not with vocabulary, IQ, or autism severity. In the TD group, cross-situational pointing task accuracy was significantly associated with age, ρ = .42, *p* = .01 (one-tailed) and vocabulary, ρ = .38, *p* = .03 (one-tailed).

To summarize, in the ASD group, age and vocabulary level were marginally associated with cross-situational eye-gaze accuracy; only age was significantly correlated with pointing task performance. In the TD group, age was significantly correlated with cross-situational eye-gaze accuracy; age and vocabulary were significantly associated with pointing task performance. These results suggest a potential role of age and vocabulary in cross-situational learning, but specific effects may depend on how learning is measured.

Visual recognition memory. Children's visual recognition memory was assessed by the visual paired comparison (VPC) task. Analyses were first conducted to characterize children's visual recognition memory; next, correlations between recognition memory and cross-situational word learning were conducted. All 27 children in the TD group and 17/20 children in the ASD group contributed data for the VPC task. Visual recognition memory was measured by children's novelty preference during test (i.e., whether children looked longer at the new image than the one they had just seen). Mean novelty preference during test was 64.15%, *SD* = 11.24, in the TD group, and 61.31%, *SD* = 8.23, in the ASD group. Mean novelty preference did not significantly differ between the two groups, *t*(42) = .90, *p* = .37, *d* = .29. The proportion of looking to the novel image during test was significantly greater than chance looking in both the TD group, *t*(26) = 6.54, *p* < .001, and the ASD group, *t*(16) = 5.67, *p* < .001, providing evidence of visual recognition memory in both groups.

It was hypothesized that children who were not able to track recurring visual stimuli may perform more poorly on the cross-situational task, since successful performance requires memory of visual images. Novelty preference, however, was not significantly correlated with cross-situational eye-gaze accuracy or with cross-situational pointing accuracy in either group, all *p*s > .74. Similarly, children with higher novelty preferences (based on a median split) did not demonstrate higher cross-situational eye-gaze accuracy in the TD group, *t*(25) = -1.37, *p* = .19, or in the ASD group, *t*(15) = .93, *p* = .37. These results indicated that, as hypothesized, both the TD and ASD group demonstrated evidence of visual recognition memory at the group level, and the groups did not significantly differ from each other in mean novelty preference. There was no evidence of a significant relationship between visual recognition memory and cross-situational learning.

Simple word learning. The simple word-learning task was administered to determine whether performance in a word-learning task that taught one label-object pairing at a time was associated with performance in the cross-situational task. Prior to examining this relationship, however, children's performance in the simple word-learning task was characterized in terms of group differences and comparisons against chance.

Simple word learning: group comparisons. All 27 children in the TD group and 17/20 children in the ASD group contributed data for the simple word-learning eye-gaze accuracy measure. Time course data are presented in Figure 12. Mean simple word-learning eye-gaze accuracy in the TD group was 59.90%, *SD* = 16.24, range = 15.49 – 89.18. Mean simple word-learning eye-gaze accuracy in the ASD group was 61.39%, *SD* = 12.21, range = 41.18 – 82.95. Mean simple word-learning eye-gaze accuracy did not significantly differ between the two groups, *t*(42) = -.32, *p* = .75, *d* = .10.

All 27 children in the TD group and 19/20 children in the ASD group contributed data for the simple word-learning behavioral pointing task. Each of the four novel words was tested once, yielding a maximum score of 4. Mean simple word-learning pointing accuracy in the TD group was 2.19, *SD* = 1.33, range = 0 – 4 (see Figure 13). Mean simple pointing accuracy in the ASD group was 2.42, *SD* = 1.50, range = 0 – 4. Because the pointing task included only 5 possible responses (0 – 4), the dependent variable was considered ordinal and did not meet the assumption of normality required for parametric tests; for this reason, non-parametric tests were used.

First, a Mann Whitney U test—the non-parametric equivalent of an independent samples *t*-test—tested whether median simple word-learning pointing accuracy differed between the TD group and ASD group. The mean rank in the TD group was 22.70 (median = 2) and the mean rank in the ASD group was 24.63 (median = 2). The groups did not significantly differ in simple word-learning pointing accuracy, $U = 235$, $p = .62$. As another approach to characterizing the simple word-learning pointing data, a chi square analysis was conducted to determine whether the proportions of children who performed above versus below chance differed between the two groups. In the TD group, 66.70% of children performed above chance and 33.30% of children performed below chance. In the ASD group, 68.42% of children performed above chance and 31.58% performed below chance. A chi square analysis indicated that the proportions of children in the TD and ASD groups who performed above chance did not significantly differ, $\chi^2 = .02$, $p = 1.0$. These results indicated that simple word learning did not significantly differ between the TD and ASD groups as measured by eye gaze or by the behavioral pointing task.

Simple word learning: who learned? Eye-gaze data were next examined for evidence of learning at the group level. Simple word-learning accuracy was significantly greater than chance in the TD group, $t(26) = 3.17$, $p = .004$, and the ASD group, $t(16) = 3.84$, $p = .001$, meaning that eye-gaze patterns provided evidence of comprehension in both groups. Next, a one-sample Wilcoxon test was conducted to determine whether children in each group performed significantly above chance in the pointing task (i.e., whether the median value differed significantly from 1). The median value was significantly greater than 1 in the TD group, $p < .001$, and in the ASD group, $p = .002$, indicating that, on average, both groups demonstrated above-chance comprehension on the simple word-learning pointing task. These results indicated that both the eye-gaze measure and the pointing task provided evidence of simple word learning in both groups.

How did eye-gaze and pointing measures compare? Next, performance in the simple word-learning eye-gaze measure and the simple pointing measure was compared using non-parametric Spearman correlations within each group to determine whether they provided convergent information. The correlation between simple word-learning pointing accuracy and simple word-learning eye-gaze accuracy in the TD group was $\rho = .15$, $p = .45$. The correlation between simple word-learning pointing accuracy and simple word-learning eye-gaze accuracy in the ASD group was $\rho = -.32$, $p = .22$. Recall that for the cross-situational task, the two measures provided convergent information for the TD group but not for the ASD group. The results for the simple word-learning task did not indicate convergent validity in either group.

Associations between cross-situational and simple word learning. If children who are poor cross-situational learners also have difficulty associating labels and objects more generally, correlations should emerge between performance on the cross-situational and simple word-learning tasks. Within-group correlations were conducted to determine whether simple word learning and cross-situational word learning were associated within each group. Because there were strong a priori hypotheses that cross-situational word learning and simple word learning would be positively correlated, one-tailed *p* values were used.

Cross-situational and simple word-learning eye-gaze accuracy were correlated in the TD group, $r = .49$, $p < .01$ (one-tailed), and marginally related in the ASD group, $r = .38$, $p = .07$ (one-tailed). There was also a significant correlation between cross-situational word learning and simple word learning in the full sample, $r = .46$, $p = .002$. Spearman correlations revealed that cross-situational and simple word-learning accuracy in the behavioral pointing task were not significantly correlated in either group, *p*s > .19 (one-tailed). These results revealed that the cross-situational and simple word-learning tasks as measured by eye-gaze accuracy were significantly correlated in the TD group and marginally correlated in the ASD group. Performance on cross-situational and simple word-learning pointing tasks was not significantly associated in either group.

Cross-situational and simple word learning: which was easier? It is possible that that learning from a cross-situational context that contains two images per training trial is more difficult than learning from a deterministic context that contains one image per trial. However, cross-situational eye-gaze accuracy and simple word-learning eye-gaze accuracy, did not significantly differ in the TD group, $t(26) = -1.12$, $p = .27$. Similarly, cross-situational accuracy and simple word-learning accuracy on the eye gaze measure did not significantly differ in the ASD group, $t(16) = .44$, $p = .67$. As expected, familiar word recognition accuracy was significantly higher than accuracy for either word-learning task in both groups, all *p*s < .01. See Figure 14 for a graphical representation of eye-gaze results for all three conditions. Wilcoxon matched-pair tests revealed that cross-situational pointing accuracy and simple word-learning pointing accuracy did not significantly differ in the TD group, $p = .89$, or in the ASD group, $p =$

.19. These findings indicated that, for both the TD children and the children with ASD, there were no significant differences in accuracy between the cross-situational task and the simple word-learning task on either the eye-gaze or pointing task. Not surprisingly, familiar word recognition was significantly better than novel word recognition in both groups.

Note on mutual exclusivity. Although it was not a primary research question, cross-situational and simple word-learning pointing accuracy data provided information about children's tendency to assume that each object has only one name—a phenomenon known as mutual exclusivity (Markman, 1990). On the cross-situational pointing task children displayed considerable variability, but only one child (in the ASD group) got three items correct on the pointing task. In the simple word-learning task, no children got three items correct. This means that even when they indicated incorrect label-object correspondences, children were unwilling to give multiple labels to the same object (see Figures 11 & 13).

How does Non-Social Visual Attention Impact Cross-Situational Word Learning?

The third aim was to characterize the relationship between cross-situational learning and non-social visual attention. First, children's performance on an independent measure of non-social visual orienting—an adapted version of the gap-overlap task (Landry & Bryson, 2004)—was examined. Second, analyses were conducted to characterize the relationship between cross-situational learning and children's eye-gaze patterns during the cross-situational training phase.

Gap-overlap task: overall performance. The gap-overlap task measured visual attention orienting in a non-social context. The gap condition measured latency to shift attention and the overlap condition measured latency to disengage attention. The primary question was whether shifting and disengagement were significantly associated with cross-situational word learning. As with the other experimental tasks, children's performance in the gap-overlap task was characterized first to aid in the interpretation of results.

In the TD group, 27/27 children contributed data for the gap and overlap conditions. In the ASD group, 16/20 children contributed data for the gap condition and 20/20 children contributed data to the overlap condition. The visual orienting data were first examined for differences by group and condition. Longer latencies were expected in the overlap condition than the gap condition. The ASD group was expected to show longer latencies than the TD group, and a potential interaction was anticipated demonstrating a greater difference between latencies in the overlap condition compared to the gap condition in the ASD group. The mean latency in the gap condition was 0.52 seconds, *SD* = 0.48, range = 0.19 – 2.28 in the TD group, and .81 seconds, *SD* = 0.95, range = 0.18 – 3.34 in the ASD group. The mean overlap latency in the TD group was 2.27 seconds, *SD* = 1.40, range = 0.31 to 4.89. The mean overlap latency in the ASD group was 1.67 seconds, *SD* = 1.49, range = 0.28 – 5.93. Log-transformed variables for the gap and overlap task were used in the analyses. A generalized linear model (repeated measures ANOVA) was conducted to examine orienting latency, with group (TD vs. ASD) as a between-subjects factor and condition as a within-subjects repeated measures factor.

Consistent with expectations, results revealed a significant main effect of condition, $F(1, 37) = 74.96$, $p < .001$, $\eta^2_p = .67$, indicating that across both groups, mean latency was significantly longer in the overlap condition than in the gap condition. Contrary to expectations, the main effect of group was not significant, $F(1,37) = .003$, $p = .95$, $\eta^2_p = .00$, and the condition x group interaction was not significant, $F(1,37) = 2.45$, $p = .13$, $\eta^2_p = .06$, indicating that there were no significant differences in latencies across groups and that the effect of condition was not significantly different across the groups. These findings are contrary to predictions that children in the ASD group would demonstrate longer latencies than children in the TD group, particularly in the overlap condition.

Because this sample included children with a wide range of age and ability levels, the relationship between gap latency, overlap latency and child-level characteristics were examined within each group (see Table 7). The orienting variables were not significantly related to age, vocabulary (PPVT GSV), or IQ in the TD group. In the ASD group, overlap latency was marginally related to age, $r = -.44$, $p = .05$, and PPVT GSV, $r = -.41$, $p = .07$; gap latency was not significantly correlated with any child characteristics. Because of the potential effect of age on overlap latency in the ASD group, the repeated measures ANOVA was conducted with age as a continuous covariate. Results were generally unchanged from the previous analysis; age was a non-significant covariate $F(1, 36) = .69$, $p = .41$, and the age x condition interaction was non-significant, $F(1, 36) = 2.57$, $p = .12$.

Relationship between cross-situational learning and visual orienting. Although there were no significant differences between the ASD and TD groups in mean latencies to shift or disengage, the primary question of interest was how visual orienting related to cross-situational word learning *within* each group. It was hypothesized that overlap latency, but not gap latency, would relate to cross-situational eye-gaze accuracy in the ASD group, and that cross-situational eye-gaze accuracy would be unrelated to either orienting variable in the TD group. Because there was a strong a priori hypothesis about the directionality of findings in the overlap condition for the children with ASD, a one-tailed p value was used. Table 8 presents correlations among the experimental eye-gaze variables, including the cross-situational task, the simple task, visual recognition, and the gap-overlap task.

As expected, gap latency was not significantly correlated with cross-situational accuracy in the TD group, $r = .22$, $p = .31$, or the ASD group, $r = -.01$, $p = .98$. Consistent with expectations, overlap latency was significantly and negatively correlated with cross-situational eye-gaze accuracy in the ASD group, $r = -.43$, $p = .031$ (one-tailed), but not in the TD group, $r = .17$, $p = .40$. This finding indicates that the length of time it took children with ASD to disengage visual attention in the gap-overlap task was significantly and negatively correlated with their eye-gaze accuracy in the cross-situational word-learning task (see Figure 16).

Overlap latency was not significantly correlated with age, vocabulary (PPVT GSV), or IQ in the TD group, *p*s > .20. In the ASD group, overlap latency was marginally correlated with age and with PPVT GSV; age and PPVT GSV were also correlated, $r = .63$, $p = .003$. Follow-up analyses indicated that the correlation between cross-situational accuracy and overlap latency in the ASD group was marginal, $p = .085$ (one tailed), when age was statistically controlled using a partial correlation. This finding suggests that age may play a role in disengagement of attention in the current sample. The relationship between overlap latency and performance in the cross-situational pointing task was also examined. Cross-situational pointing accuracy and overlap latency were not significantly related in the TD group, $\rho = -.14$, $p = .49$, or the ASD group, $\rho = -.26$, $p = .28$. This finding suggests that the relationship between overlap latency and cross-situational learning is specific to the eye-gaze task.

Eye-gaze patterns during cross-situational learning. As another approach to examining the relationship between word learning and attention (Yu & Smith, 2011), eye-gaze patterns during the training phase of the cross-situational learning task were analyzed. The primary goal was to determine whether looking patterns during training differed on the basis of weak versus strong cross-situational learning. Recall that 16 children were categorized as weak learners (9 TD children and 7 children with ASD) and 31 children were categorized as strong learners (18 TD children and 13 children with ASD). Differences between the TD and ASD groups were also examined.

Children with ASD looked at the images during training an average of 2.18 seconds per trial, *SD* = 0.77. Children in the TD group looked at the images during training an average of 2.32 seconds per trial, *SD* = .79. The mean amount of time looking at the images during training did not differ significantly between the TD group and the ASD group, *t*(44) = .63, *p* = .53. Consistent with the findings of Yu and Smith (2011), the amount of time spent fixating the images during training did not significantly differ between strong cross-situational learners and weak cross-situational learners in the TD group, *t*(25) = -.57, *p* = .56, or in the ASD group, *t*(17) = -.83, *p* = .42 (see Table 9). Additionally, correlations revealed that the amount of time children spent fixating the images during training was not significantly related to their eye-gaze accuracy at test in both groups, *p*s > .40. Overall, these results indicated no significant differences in total time spent fixating the novel images during training based on diagnostic status or cross-situational learning accuracy, ruling out the possibility that poor cross-situational learning performance was based entirely on the amount of visual attention devoted to the visual images during training.

Next, the relationship between cross-situational accuracy and the mean number of attention shifts per trial during training was examined. The mean number of shifts was 2.95, *SD* = .69, in the ASD group and 2.88, *SD* = .69, in the TD group, a difference that was non-significant, *t*(42) = -.32, *p* = .75. The number of attention shifts did not significantly differ between strong and weak cross-situational learners in the TD group, *t*(23) = -.41, *p* = .69, or the ASD group, *t*(17) = -1.12, *p* = .28. Next, the relationship between cross-situational accuracy and the mean length of longest fixation during training was examined. The mean length of longest fixation was 1.46 seconds, *SD* = .32, in the TD group and 1.32, *SD* = .26 in the ASD group, a difference that was non-significant, *t*(42) = 1.61, *p* = .12. The mean length of longest fixation did not significantly differ between strong and weak cross-situational learners in the TD group, *t*(23) = -1.20, *p* = .24, or between strong and weak cross-situational learners in the ASD group, *t*(17) = -1.03, *p* = .32.

Exploring Sticky Attention and Language Processing in Children with ASD

Results pointed to an association between sticky attention (i.e., overlap latency) and cross-situational eye-gaze accuracy in the ASD group. This relationship was not present in the TD group, and it was not present for gap latency (shifting of attention). To better understand the association between sticky attention and cross-situational eye-gaze accuracy, several post hoc analyses were conducted. If children's sticky attention directly affected their cross-situational learning, a relationship between overlap latency and eye-gaze patterns during training would be expected to emerge, such that children with stickier attention would demonstrate fewer shifts and/or longer fixations. However, there was no significant correlation between overlap latency and the number of attention shifts or length of longest look during cross-situational training in the children with ASD, *p*s > .85. These non-significant findings suggest, contrary to expectations, that looking patterns during cross-situational learning did not differ on the basis of sticky attention.

If the relationship between sticky attention and cross-situational learning is specific to the type of visual attention required to track label-object co-occurrences across ambiguous trials, overlap latency would not be expected to relate to simple word-learning accuracy or familiar word recognition accuracy. The association between overlap latency and simple word learning accuracy was not significant in the ASD group, *r* = -.14, *p* = .60, suggesting that the relationship may be specific to cross-situational word learning. However, the relationship between overlap latency and familiar word recognition in the children with ASD was *r* = -.62, *p* = .004, suggesting that children's moment-to-moment attention during the test phase alone may be impacted by relatively long latencies to disengage attention—even for known words (see Figure 17). The partial correlation between familiar word recognition and overlap latency remained significant after age was statistically controlled using a partial correlation, *r* = -.53, *p* = .03. In combination, these findings indicated that children with ASD who had longer latencies to disengage visual attention in a non-linguistic task also had lower levels of eye-gaze accuracy for processing of both familiar words and novel words learned in a cross-situational task.

Given these findings, an alternative eye-gaze variable of learning was examined: length of longest look to the target image. The creation of this variable was motivated by the tendency of children with ASD to 'stick' to a visual stimulus; it was possible that their comprehension was more accurately measured by their tendency to look for the longest amount of time at the target, as opposed to the overall proportion of looking to the target (i.e., the standard accuracy variable used in the previous analyses). A variable was then created to represent the proportion of trials on which children's longest look was to the target as opposed to the distracter image. For the familiar trials, 82.04% of the longest looks in the ASD group were to the target (*SD* = 13.32, range = 63 – 100), versus 81.15% (*SD* = 16.84, range = 50 – 100) in the TD group. The proportion of longest looks to target for the familiar trials did not significantly differ between the two groups, *t*(43) = -.19, *p* = .85. For the simple word-learning task, 67.71% of the longest looks in the ASD group were to the target (*SD* = 16.26, range = 33 – 100) versus 61.88% (*SD* = 21.06, range = 0 – 100) in the TD group. Proportion of longest looks to the target in the simple word-learning task did not significantly differ between the two groups, *t*(42) = -.97, *p* = .34.

In the cross-situational task, 67.50% of the longest looks in the ASD group were to the target (*SD* = 19.51, range = 43 – 100), versus 53.74% (*SD* = 23.64, range = 0 – 100) in the TD group. Contrary to the null group differences based on the standard accuracy measure, children in the ASD group had significantly more longest looks to the target for the cross-situational task than the TD group, *t*(44) = -2.08, *p* = .04, *d* = .63. Interestingly, the proportion of longest looks to target in the cross-situational and simple word-learning tasks was correlated in both the TD group, *r* = .40, *p* = .02 (one-tailed), and in the ASD group, *r* = .52, *p* = .02 (one-tailed), which supported previous findings of a marginal association between the cross-situational and simple word-learning tasks in the ASD group based on the standard accuracy measure. Finally, the proportion of longest looks to the target in the cross-situational task and overlap latency were negatively correlated in the children with ASD, *r* = -.42, *p* = .04 (one-tailed), meaning that the children who took longer to disengage their attention in the visual orienting task had fewer trials in which their longest looks were to the target image. Said differently, the tendency to take longer to disengage attention was associated with a higher proportion of trials wherein children's longest look was to the distracter image in the cross-situational task. Surprisingly, this was not the case for the familiar word trials, *r* = -.13, *p* = .62.

A final set of analyses was conducted to shed light on the relationships among sticky attention, language processing, and vocabulary knowledge in the children with ASD. Specifically, the question was whether familiar language processing accuracy mediates the relationship between sticky attention and children's existing vocabulary knowledge. Following Baron and Kenny (1986), four criteria must be met for evidence of a mediating relationship: 1) the independent variable (overlap latency) must be significantly associated with the proposed mediator variable (familiar word recognition accuracy); 2) the mediator variable must be significantly associated with the dependent variable (vocabulary level, PPVT GSV); 3) the independent variable must be significantly associated with the outcome variable; and 4) the association between the independent variable and the outcome variable must be reduced after accounting for variance explained by the proposed mediator (see also McDuffie, Yoder, & Stone, 2006).

Figure 18 presents a visual depiction of the mediation analysis. Correlational analyses revealed that overlap latency was significantly associated with familiar word recognition accuracy, $r = -.62$, $p = .002$ (one-tailed), meeting the first criterion. Second, familiar word recognition accuracy was significantly associated with vocabulary, $r = .63$, $p = .002$ (one-tailed). Third, overlap latency was significantly associated with vocabulary, $r = -.41$, $p = .04$ (one-tailed). These correlations indicated that the independent, mediator, and outcome variables were all significantly correlated. Next, the partial correlation between overlap latency and vocabulary was tested while controlling for the effect of familiar word recognition accuracy. The correlation between overlap latency and vocabulary knowledge was no longer significant after accounting for familiar word recognition accuracy, $r = -.03$, $p = .92$. In combination, these findings suggest that sticky attention may impact children's vocabulary knowledge through their real-time language processing of known words.

Table 5

Correlations between Child-Level Variables and Cross-Situational Eye-Gaze Accuracy

	Cross-Situational Eye-Gaze Accuracy	Age	PPVT Growth	Brief IQ
Cross-Situational Eye-Gaze Accuracy		$r = .38$ $p = .03^{*+}$	$r = .22$ $p = .14^{+}$	$r = -.39$ $p = .04*$
Age	$r = .33$ $p = .08^{+}$		$r = .75$ $p < .001**$	$r = -.56$ $p < .001**$
PPVT GSV	$r = .34$ $p = .07^{+}$	$r = .63$ $p < .01*$		$r = -.33$ $p = .09$
Brief IQ	$r = .20$ $p = .42$	$r = -.24$ $p = .33$	$r = .36$ $p = .14$	
Autism Severity	$r = -.14$ $p = .57$	$r = .24$ $p = .33$	$r = -.01$ $p = .97$	$r = -.36$ $p = .13$

Note. $*p < .05$ $**\ p < .01$ $^{+}$ Indicates one-tailed p values. Bivariate correlations for the TD group ($n = 27$) are presented above the diagonal. Bivariate correlations for the ASD group ($n = 20$) are presented below the diagonal. PPVT GSV indicates growth scale value from the Peabody Picture Vocabulary Test. Brief IQ was measured by the Leiter International Performance Scale-Revised. Autism severity was measured by ADOS-2 calibrated severity scores.

Table 6

Correlations between Child-Level Variables and Cross-Situational Pointing Accuracy

	ASD Group	TD Group
Age	$\rho = .48$	$\rho = .42$
	$p = .02*^{+}$	$p = .01*^{+}$
PPVT GSV	$\rho = .12$	$\rho = .38$
	$p = .31^{+}$	$p = .03*^{+}$
Brief IQ	$\rho = -.20$	$\rho = -.26$
	$p = .42$	$p = .19$
Autism Severity	$\rho = .13$	--
	$p = .61$	

Note. $*p < .05$ $^{+}$Indicates one-tailed p values. PPVT GSV indicates growth scale value from the Peabody Picture Vocabulary Test. Brief IQ was measured by the Leiter International Performance Scale-Revised. Autism severity was measured by ADOS-2 calibrated severity scores.

Table 7

Correlations between Child-Level Characteristics and Visual Orienting

	TD Group		ASD Group	
	Disengagement Latency	Shifting Latency	Disengagement Latency	Shifting Latency
Age	*r* = -.20 *p* = .34	*r* = .26 *p* = .24	*r* = -.44 *p* = .05	*r* = -.18 *p* = .51
PPVT GSV	*r* = -.26 *p* = .20	*r* = .33 *p* = .13	*r* = -.41 *p* = .07	*r* = -.18 *p* = .51
Brief IQ	*r* = .08 *p* = .72	*r* = -.23 *p* = .29	*r* = -.18 *p* = .47	*r* = -.09 *p* = .75
Autism Severity	--	--	*r* = .19 *p* = .22	*r* = .19 *p* = .50

Note. Shifting latency was measured by the gap condition. Disengagement latency was measured by the overlap condition. PPVT GSV indicates growth scale value from the Peabody Picture Vocabulary Test. Brief IQ was measured by the Leiter International Performance Scale-Revised. Autism severity was measured by ADOS-2 calibrated severity scores.

Table 8

Correlation Table of Experimental Eye-Gaze Variables

	Cross-Situational Eye-Gaze Accuracy	VPC Novelty Preference	Simple Eye-Gaze Accuracy	Gap Latency	Overlap Latency
Cross-Situational Eye-Gaze Accuracy		$r = .02$ $p = .93$	$r = .49$ $p = .01$*	$r = .22$ $p = .31$	$r = .17$ $p = .40$
VPC Novelty Preference	$r = -.09$ $p = .74$		$r = -.09$ $p = .66$	$r = .09$ $p = .69$	$r = .02$ $p = .94$
Simple Eye-Gaze Accuracy	$r = .38$ $p = .14$	$r = .06$ $p = .83$		$r = .19$ $p = .38$	$r = .10$ $p = .65$
Gap Latency	$r = -.01$ $p = .98$	$r = .22$ $p = .48$	$r = -.25$ $p = .41$		$r = .29$ $p = .18$
Overlap Latency	$r = -.43$ $p = .03$*$^{+}$	$r = -.10$ $p = .71$	$r = -.14$ $p = .60$	$r = .29$ $p = .27$	

Note. *$p < .05$. $^{+}$ Indicates a one-tailed p value. Bivariate correlations for the TD group ($n = 27$) are presented above the diagonal. Bivariate correlations for the ASD group ($n = 20$) are presented below the diagonal.

Table 9

Eye-Gaze Patterns During Cross-Situational Learning

	Weak Learners		Strong Learners	
TD Group	*M (SD)*	Range	*M (SD)*	Range
Looking Time to Images	2.20 (0.87)	0.22 – 3.16	2.39 (.76)	0.40 – 3.53
Number of Attention Shifts	2.80 (0.72)	2.17 – 4.29	2.92 (0.70)	1.54 – 4.15
Length of Longest Look	1.35 (0.44)	0.91 – 2.27	1.52 (0.24)	1.11 – 2.11
ASD Group				
Looking Time to Images	1.98 (0.94)	0.44 – 3.46	2.29 (0.67)	1.15 – 3.35
Number of Attention Shifts	2.72 (0.51)	2.00 – 3.36	3.08 (0.76)	1.75 – 4.08
Length of Longest Look	1.24 (0.34)	0.85 – 1.78	1.37 (0.21)	0.97 – 1.69

Note. Variables were derived from looking patterns during the cross-situational training phase. Strong learners were those children with cross-situational eye-gaze accuracy above 50%. Weak learners were those children with cross-situational eye-gaze accuracy at or below 50%. There were no significant differences between strong and weak learners in either group for any of the three variables.

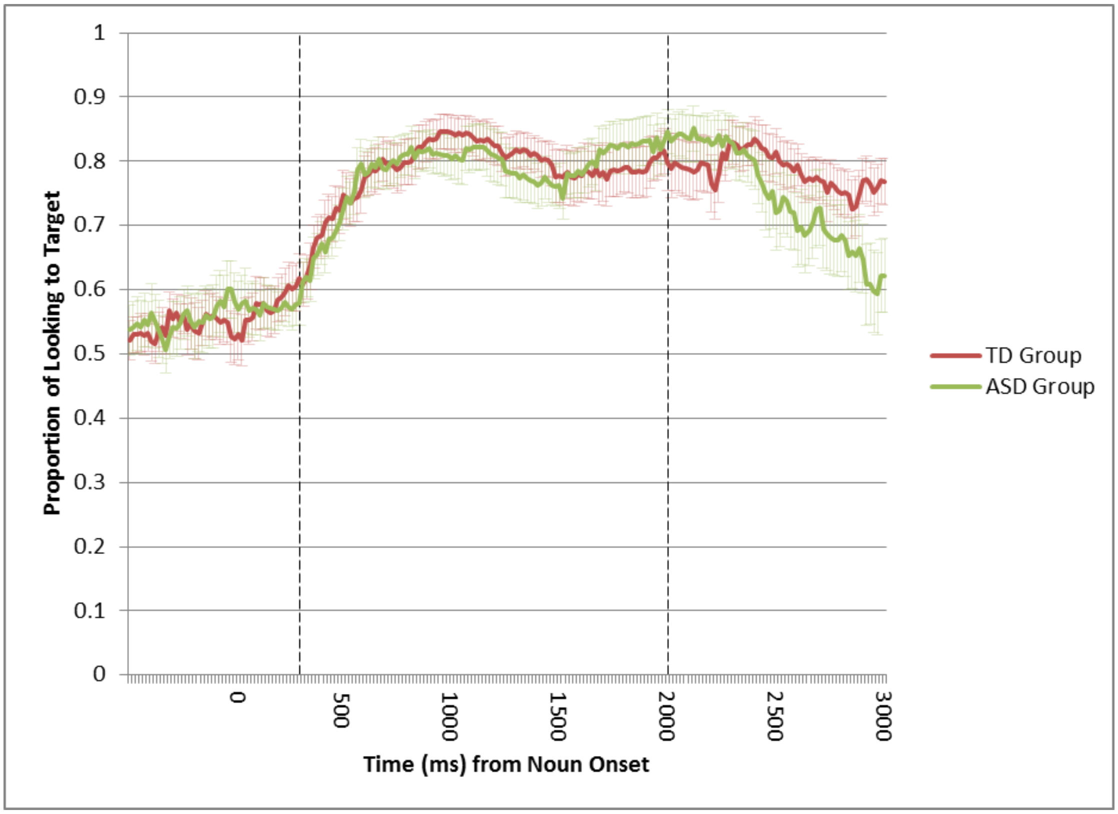

Figure 9. Time course data for familiar word trials. The grand mean for each group is presented with standard error bars. 0 ms indicates target word onset. Dashed lines indicate the target window (300 ms to 2000 ms after word onset). Full trials lasted 5 seconds. Familiar word accuracy (i.e., the proportion of looking to the target) did not significantly differ between the TD group (78%) and the ASD group (76%) during the target window.

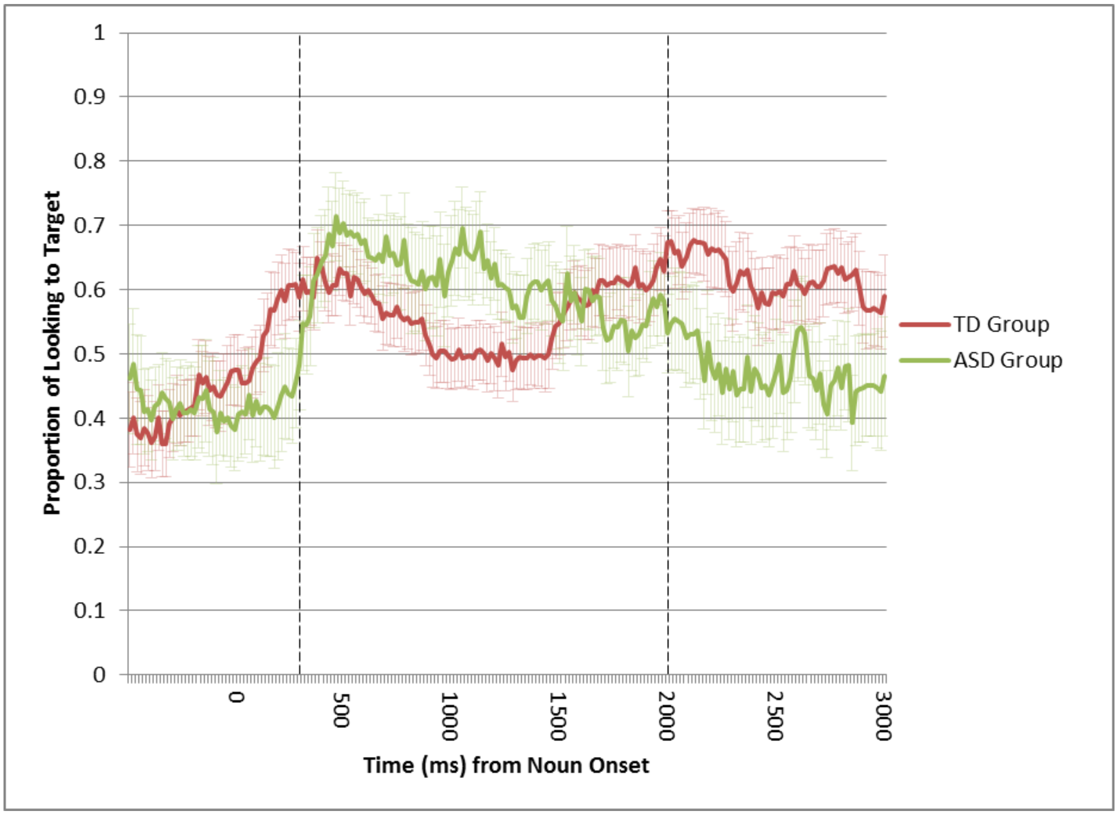

Figure 10. Time course data for the cross-situational word-learning task. The grand mean for each group is presented with standard error bars. 0 ms indicates target noun onset. Dashed lines indicate the target window (300 ms to 2000 ms after word onset). Full trials lasted 5 seconds. Cross-situational eye-gaze accuracy (i.e., the proportion of looking to the target) did not significantly differ between the TD group (56%) and the ASD group (61%) during the target window.

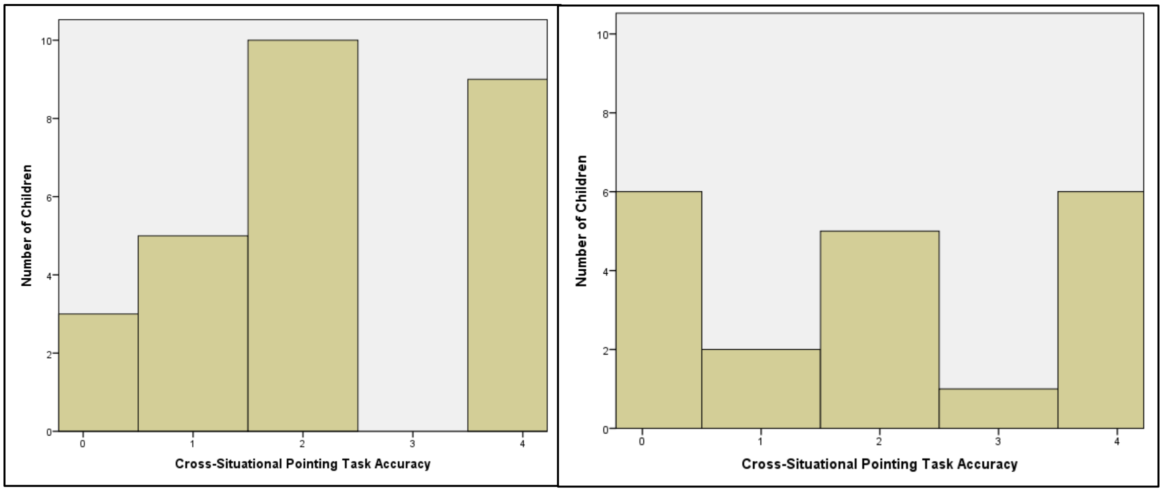

Figure 11. Performance of children in the TD group (left panel) and the ASD group (right panel) on the cross-situational pointing task. Maximum score was 4. Cross-situational pointing task accuracy did not significantly differ between the two groups, nor did the proportions of children performing above chance (25%) in each group.

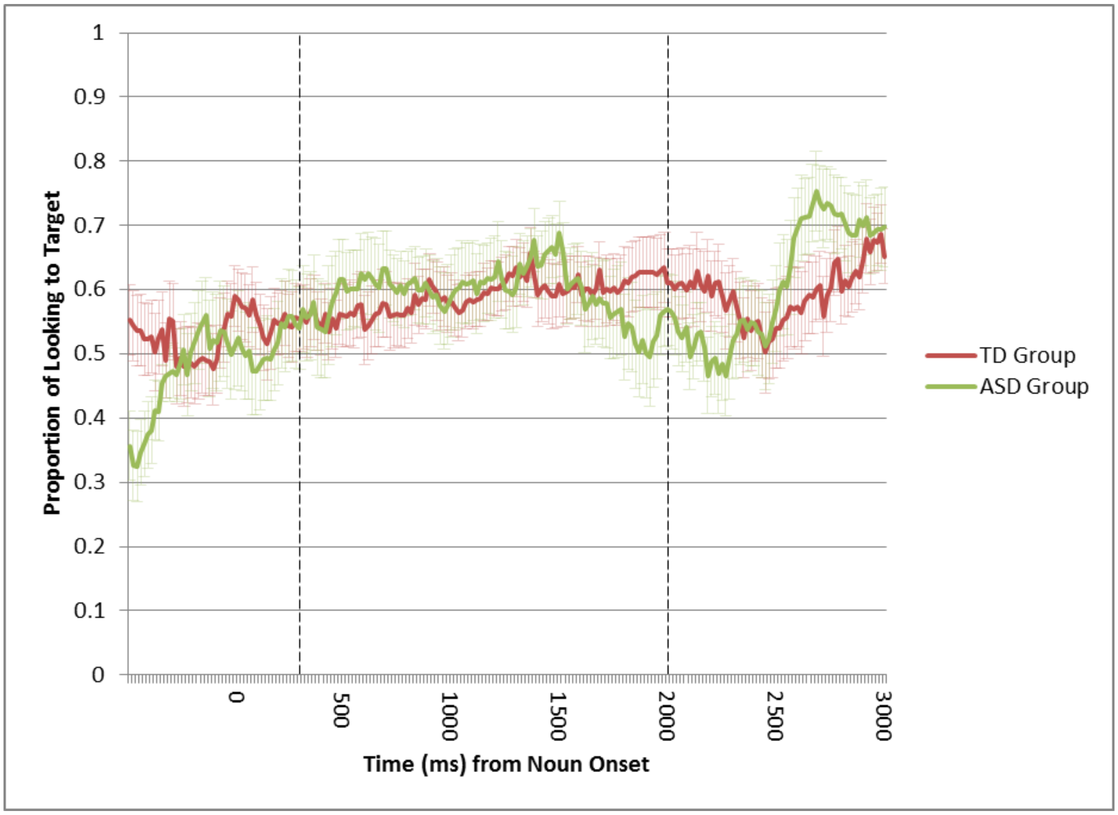

Figure 12. Time course data for the simple word-learning task. The grand mean for each group is presented with standard error bars. 0 ms indicates target noun onset. Dashed lines indicate the target window (300 ms to 2000 ms after word onset). Simple eye-gaze accuracy (i.e., the proportion of looking to the target) did not significantly differ between the TD group (60%) and the ASD group (61%) during the target window.

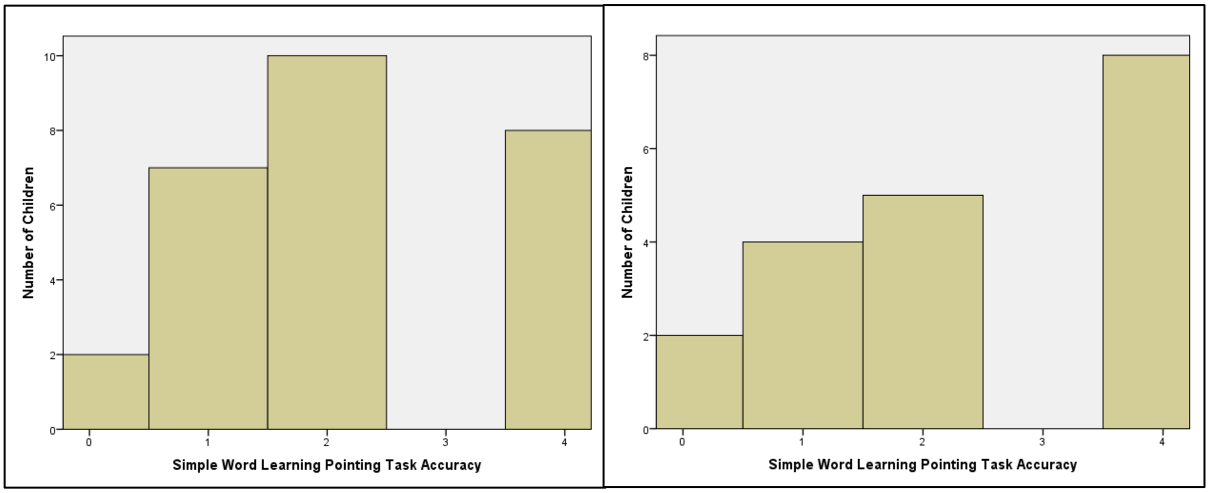

Figure 13. Performance of children in the TD group (left panel) and the ASD group (right panel) on the simple word-learning pointing task. Maximum score was 4. Simple word-learning pointing task accuracy did not significantly differ between the two groups, nor did the proportions of children performing above chance (25%) in each group.

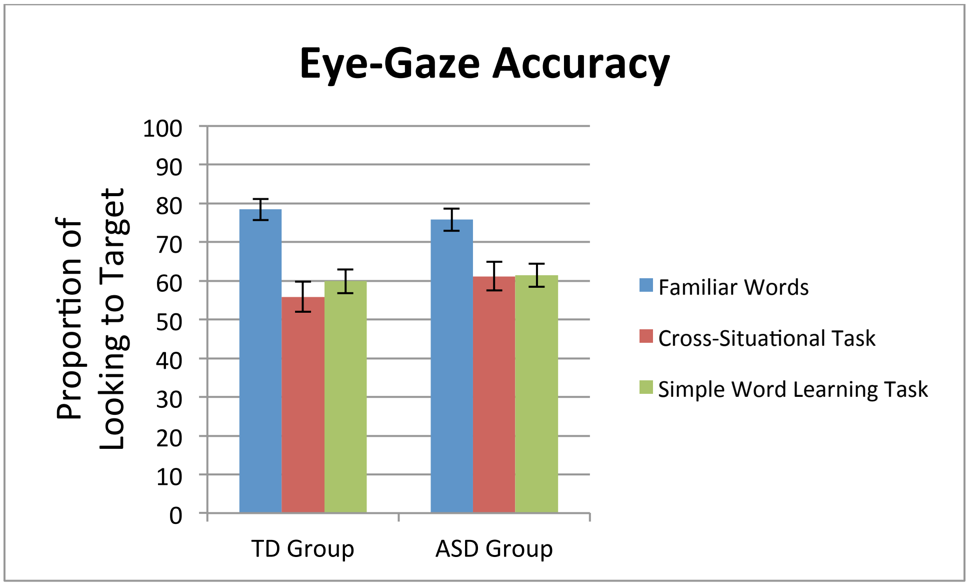

Figure 14. Mean eye-gaze accuracy for the familiar word trials, cross-situational task, and simple word-learning task. Eye-gaze accuracy was the proportion of looking to the target image relative to the time spent looking at either image during the target window (300 ms to 2000 ms after target word onset). Mean eye-gaze accuracy did not significantly differ between the two groups in any of the three conditions.

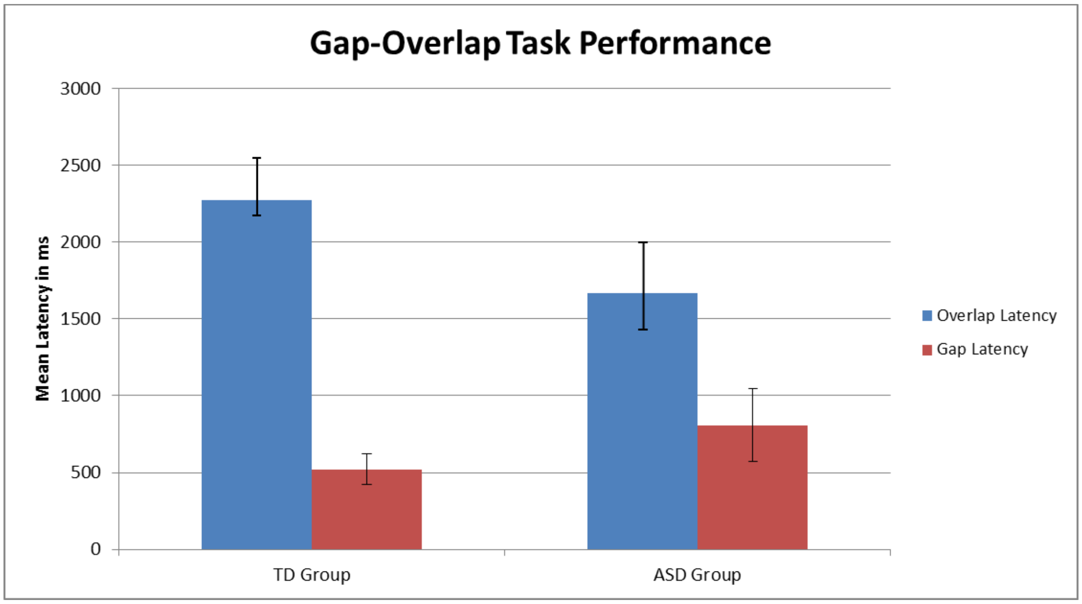

Figure 15. Mean latencies for the gap-overlap task. The gap condition measured shifting of attention and the overlap condition measured disengaging of attention. Standard error bars are presented. Analyses were performed on log transformed latency values. There was a main effect of condition, such that latencies were significantly longer in the overlap condition than the gap condition. There was no significant main effect of group and no significant group x condition interaction.

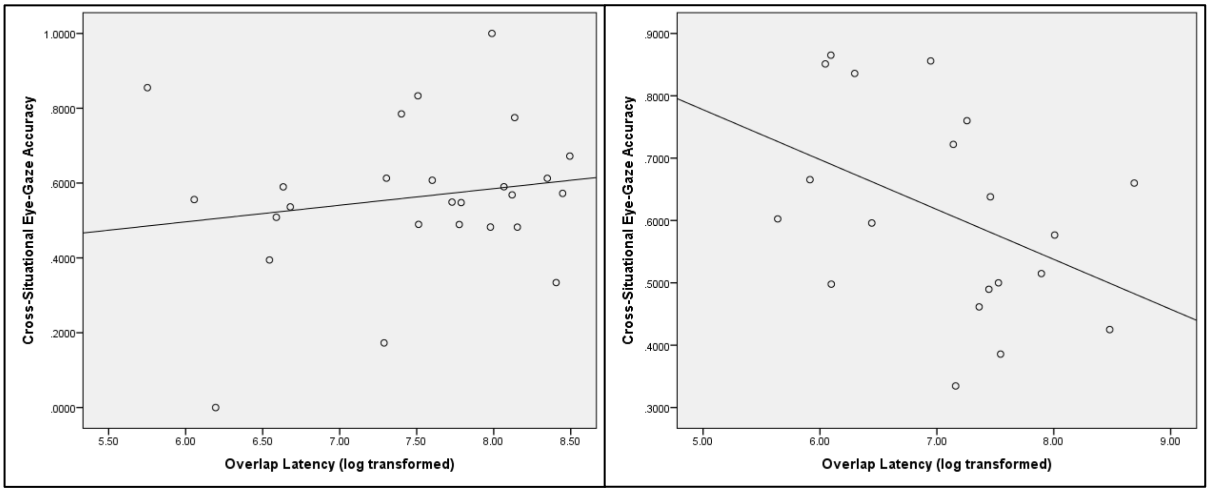

Figure 16. Scatterplots depicting the relationship between overlap latency (log transformed) and cross-situational eye-gaze accuracy for the TD group (left panel) and the ASD group (right panel). Cross-situational eye-gaze accuracy was the relative proportion of looking to the target image. Overlap latency was significantly and negatively correlated with cross-situational eye-gaze accuracy in the ASD group, $r = -.43$, $p = .031$ (one-tailed), but not in the TD group, $r = .17$, $p = .40$.

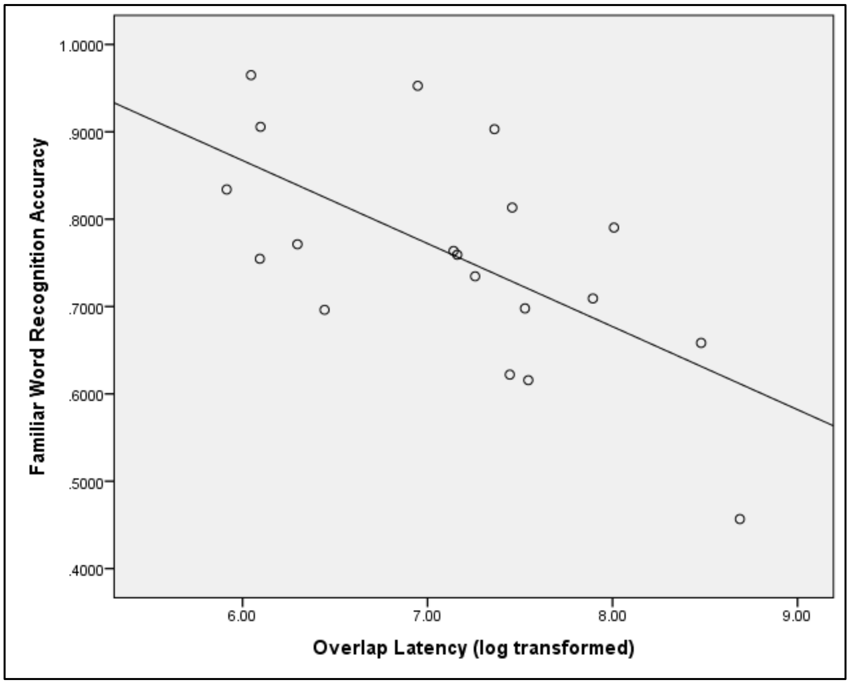

Figure 17. Scatterplot depicting the relationship between overlap latency (log transformed) and familiar word recognition accuracy in the ASD group. Familiar word recognition accuracy was the relative proportion of looking to the target image. The correlation between overlap latency and familiar word recognition in the children with ASD was *r* = -.62, *p* = .004 and remained significant after controlling for age.

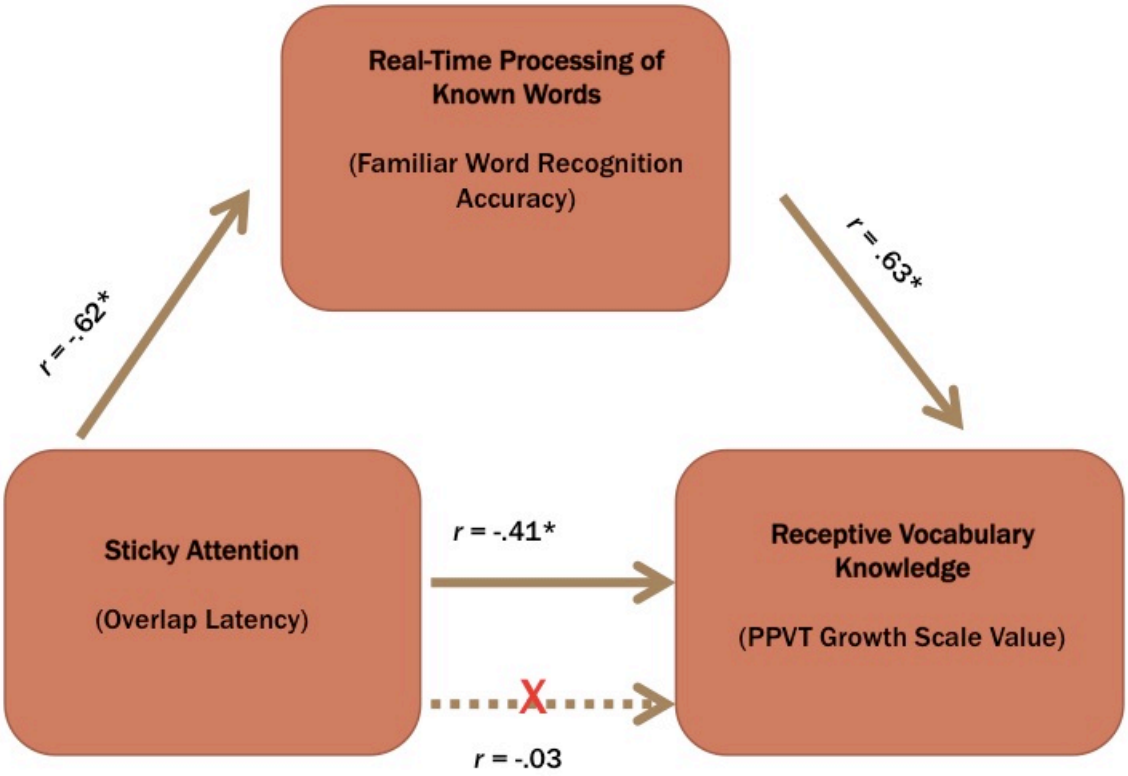

Figure 18. * $p < .05$, one-tailed p values. Visual depiction of the mediation analysis. Sticky attention was the independent variable. Real-time familiar word processing was the proposed mediator variable. Receptive vocabulary knowledge was the outcome variable. The dotted line indicates that the relationship between sticky attention and receptive vocabulary knowledge was no longer significant after familiar word processing was statistically controlled using a partial correlation.

Chapter 4: Discussion

Do Children with ASD Show Deficits in Cross-Situational Word Learning?

In an attempt to better understand language-learning deficits in children with autism spectrum disorders (ASD; Ellis Weismer et al., 2011; Luyster et al., 2008), this study investigated children's abilities to learn new words through a mechanism known as cross-situational word learning (Akhtar & Montague, 1999; Siskind, 1996; Smith & Yu, 2008; Yu & Smith, 2007). Cross-situational word learning requires children to track co-occurrence statistics among labels and objects across a series of individually ambiguous trials. Children participated in a cross-situational eye-tracking task that taught them the names of four new objects. Their comprehension of the newly learned words was then tested in two ways: through their eye gaze and through a behavioral pointing task. Predictions were that a group of younger, typically developing (TD) children matched on receptive vocabulary would demonstrate better learning than the children in the ASD group and that variability would characterize the performance of the children with ASD.

Contrary to predictions, there were no significant differences in cross-situational word learning between the TD and ASD groups on either the standard eye-gaze accuracy measure or the behavioral pointing task. Although this finding was somewhat surprising, it is consistent with findings from a recent study in which adolescents with ASD demonstrated similar statistical word segmentation skills to their TD peers matched on age, IQ and vocabulary (Mayo & Eigsti, 2012). Scott Van-Zeeland et al. (2010) also found no differences between statistical word segmentation abilities in TD adolescents and adolescents with ASD using a behavioral task, suggesting that group-level differences in statistical learning may not emerge when groups are well matched on language abilities and that examination of within-group patterns may be a more sensitive measure of underlying processes.

Results also revealed no significant differences between the two groups in the simple word-learning task, which taught only one label-object pairing at a time. These results are consistent with findings that children with ASD, even those with considerable linguistic and cognitive impairments, learn new words quite well when an object within their current focus of attention is labeled (Baron-Cohen et al., 1997). Additionally, children with ASD are able to assign novel labels to unnamed objects in situations where social cues are not required (Preissler & Carey, 2005), suggesting that forming associations between labels and objects may not be specifically impaired in ASD. However, it is possible that children with ASD may have more difficulty retaining information about the meanings of words over time (Gliga et al., 2012; Norbury et al., 2009)—an issue that must be addressed by future studies.

Interpreting null results is challenging, but there are a number of potential explanations for the finding of no significant differences in cross-situational or simple word learning. The groups were well matched on receptive vocabulary level, which provided a conservative test of the hypothesis that children with ASD have cross-situational learning deficits. This strategy was adopted because if the groups had differed on vocabulary level, it would have been impossible to know whether differences in extant word knowledge accounted for differences in cross-situational word learning. Future work may determine whether or not children with ASD are poorer cross-situational learners than their peers, even though their performance is comparable to younger children with similar vocabulary levels.

As expected, there was considerable variability in cross-situational learning in the ASD group (*M* = 61%, *SD* = 16, range = 34 – 87). The great amount of variability in the TD group (*M* = 56%, *SD* = 20, range = 0 – 100), however, was somewhat unexpected and likely contributed to the similar levels of performance across the two groups. It was expected that most children in the TD group would learn the novel words quite easily, given that infants have shown learning in similar tasks (Smith & Yu, 2008; Yu & Smith, 2011). This was not the case. At the group level, the ASD group showed significant evidence of comprehension in both the cross-situational eye-gaze measure and pointing task, but the TD group performed above chance only in the pointing task. Further, some children in the TD group had mean eye-gaze accuracies at or close to 0, suggesting that they may have learned the incorrect label. (It should be noted that when the TD child with a score of 0 was removed, the TD group mean was significantly above chance, demonstrating the impact of this child's mean on the group-level findings).

The fact that the TD children performed significantly above chance in the pointing task provides convincing evidence that they did, on average, learn the new words as a group, leading to the question of why their eye-gaze patterns did not show this same result. The TD children in the study were as old as 7—much older than the infants and toddlers with whom similar eye-gaze methodologies are typically used. It is possible that the looking time measure was simply not sensitive and that asking them to identify the correct referents yielded a more accurate measure of learning. In any case, these findings indicate that the task was certainly not too simple for children in the TD group and captured variability in their learning.

Results indicated that although the cross-situational word-learning task was designed to be relatively easy, with 10 repetitions each of 4 label-object pairs, in actuality the task was quite difficult for many of the children in each group. The task was based on a similar design to that used with infants (Smith & Yu, 2008); it is possible that altering the design to be more engaging, or even to include a greater number of label-object pairs (Yu & Smith, 2007) may have increased performance of children in the TD group. Additionally, Smith and Yu's task presented 6 label-object pairings and tested words somewhat differently; instead of using a question frame (*Where's the* __), the name of the target object was presented repeatedly (e.g., *Coro. Coro. Coro.*).

The heterogeneous abilities of the children with ASD who were recruited for the current study may also have played a role in the finding of no significant differences between groups. Many previous studies have recruited more homogeneous samples of children with ASD with average to above-average cognitive and language skills (see Tager-Flusberg & Kasari, 2013), which precludes characterization of the mechanisms that contributed to variable outcomes. This study took an individual differences approach by specifically recruiting children with ASD with a broad range of abilities for the purpose of examining the underlying skills that contributed to variability in language learning (Eigsti et al., 2011); few exclusionary criteria were applied. Although recruitment of a representative sample of children with ASD increases the generalization of findings, it is likely that the inclusion of such a diverse group of children contributed to the null group differences, since some children demonstrated successful learning but others did not. Future studies may benefit from adopting an approach that involves separately analyzing subgroups of children with ASD with lower and higher language skills in a narrower age band.

The selection of the target window may also have affected the findings. Based on previous work (Venker et al., 2013), the current study adopted a target window of 300 to 2000 ms after target word onset. Examination of the looking time patterns for the cross-situational task suggested that the TD group may have demonstrated a later peak of looking to the target image starting 2000 ms after word onset—looking behaviors that were not examined because they fell outside the window of interest. Additionally, both groups showed an unexpected peak in looking to target at the end of the test trial in the simple word-learning task. Visualization of looking time patterns also suggested that children's looking times may have begun to increase earlier than 300 ms after word onset, suggesting that an earlier time window may also be explored.

Examination of the proportion of longest looks revealed that, during the selected target window, the children with ASD outperformed the children in the TD group. Specifically, they had a significantly higher proportion of longest looks to the target than children in the TD group, highlighting the potential value of examining novel word comprehension through this alternative measure. Although findings could change with the use of a different target window, this finding is consistent with the idea that children with ASD may show strengths in some aspects of word learning (Norbury et al., 2010).

The current study exposed children to only 4 novel labels and objects to facilitate learning; however, it is possible that including even more label-object pairs would have increased learning, at least in some children (Yu & Smith, 2012). Adults in general show better cross-situational learning of words presented with a greater number of distracters (i.e., many weak spurious correlations) than words presented with a small number of distracters (i.e., a few strong spurious correlations; (Kachergis, Yu, & Shiffrin, 2007a; Suanda & Namy, 2012). Indeed, Yu and Smith (2007) found that adults learned more words overall when presented with 18 label-object pairs than with only 9. The advantage of contextual diversity may be the increased difficulty of learning words with strong competitors (a negative impact of high spurious correlations), or it may be due to the decreased difficulty of learning words in more variable contexts (a positive impact of low spurious correlations; see Suanda & Namy, 2012).

Although distinguishing between these two explanations is challenging because they are highly correlated in real-word situations, future work should determine whether children with ASD also show similar sensitivities to variations in statistical structure. Diversity may enhance aspects of word learning in some circumstances (Perry, Samuelson, Malloy, & Schiffer, 2010), but it may limit it in others (Maguire, Hirsh-Pasek, Golinkoff, & Brandone, 2008). Knowing about a child's cross-situational learning abilities may have implications for individualizing intervention strategies. For example, a child who has difficulty tracking multiple label-object contingencies simultaneously might learn more effectively in a less distracting environment with fewer objects. On the other hand, children with a greater ability to focus on the relevant object and remember a greater number of labels and objects may benefit from richer, more diverse word-learning contexts. Interestingly, recent work has suggested that cross-situational learning may increase when learners play an active role in the learning process, namely by selecting the objects they would like to be labeled on individual trials (Kachergis, Yu, & Shiffrin, 2013).

Agreement between Cross-Situational Eye-Gaze and Pointing Measures

Studies of cross-situational learning typically measure performance through either eye-gaze patterns (for infants) *or* forced choice tasks (for adults). Because the children in this study were older than the infants but younger than the adults who have participated in the majority of cross-situational learning studies, comprehension was measured in two ways: a more implicit measure based on eye-gaze patterns and a more explicit measure based on a behavioral pointing task. For the cross-situational task, results indicated significant agreement between the eye-gaze and pointing results in the TD group, but not in the ASD group. This finding indicated convergent validity between cross-situational learning in the two measures for the TD children. In the ASD group, however, the children who performed well on the eye-gaze measure were not the same children who performed well on the pointing task.

What does it mean that performance on the cross-situational eye-gaze measure and pointing measure was significantly correlated in the TD group but not in the ASD group? Broadly, these results suggest convergent validity of the two measures for the TD children but not the children with ASD. Studies of cross-situational learning in adults have suggested anecdotally that participants are unaware of their learning and surprised to discover that they did, in fact, learn new words (Yu & Smith, 2007). Suanda and Namy (2012) found that adults underestimated their learning but that self-ratings and actual performance were significantly correlated. The current findings complement work with adults by showing that in TD children, the extent to which they looked at a named novel object was correlated with their tendency to select that object in a forced-choice task. Even if children were unaware of their learning, some TD children's knowledge of the novel words had reached an adequately high threshold they were able to select the object when presented with 4 options.

Use of the more implicit, non-social eye-gaze measure with the children with ASD was appealing because to some extent it circumvented the difficulties these children may have with examiner-directed tasks (Charman, Drew, Baird, & Baird, 2003b; Kjelgaard & Tager-Flusberg, 2001; Venker et al., 2013). However, the fact that the children with ASD who performed best in the eye-gaze task were not the same children who performed best in the pointing task suggests that measurement of comprehension should receive further attention in future studies of word learning in ASD. Some children indicated understanding of the novel words in the pointing task but did not demonstrate comprehension in the eye-gaze measure. On the other hand, some children indicated understanding of the novel words in the eye-gaze measure but not in the pointing task; these children may have been less likely to indicate knowledge in a socially mediated task. It should be noted that although the eye-gaze and pointing measures were correlated in the TD group for the cross-situational task, this was not the case for the simple word-learning task.

Child-Level Skills Associated with Cross-Situational Learning

The second aim of this study was to determine which child-level skills—including age, vocabulary level, IQ, autism severity, visual recognition memory, or simple word learning—were related to cross-situational word learning in the children with ASD. Despite considerable variability in performance across the sample, some children in both groups appeared to be able to learn words from a cross situational context quite well. What factors contributed to children's success? One possibility is that language ability played a larger role in word learning and word recognition than ASD diagnosis. Matching the groups on PPVT GSV may have equated the groups on exactly the child-level skill that matters the most for statistical word learning: receptive vocabulary. In a study of children with and without language impairment and with and without ASD, Brock et al. (2008) found that sensitivity to contextual linguistic information differed not on the basis of ASD diagnosis, but on the basis of language impairment status. Other studies have identified a relationship between vocabulary knowledge and cross-situational learning in TD children (Akhtar & Montague, 1999; Evans et al., 2009; Scott & Fisher, 2012) and children with ASD (McGregor et al., 2013).

In the current study, results of the correlational analyses were presented separately for each group since relationships between these child-level skills may differ in children with and without ASD diagnoses. Correlational analyses revealed a marginal association between cross-situational eye-gaze accuracy and vocabulary level in the ASD group. Similarly, vocabulary level was significantly associated with cross-situational pointing task accuracy in the TD group. However, it is also important to note that age was marginally or significantly correlated with performance on both cross-situational tasks in both groups. Given that age and vocabulary level were strongly correlated in both groups of children, it is difficult to completely disentangle their separate effects.

Although these results were not presented above, examination of the full sample revealed somewhat similar results, namely that age and vocabulary were associated with performance on both measures of cross-situational learning (adopting one-tailed *p* values) but that the absolute strength of the correlation was larger for age than for vocabulary. Future studies may or may not reveal strong effects of vocabulary level and/or age on cross-situational learning. For example, statistical learning on a word segmentation task was not significantly related to any cognitive, language, or autism severity variables in a study of high-functioning adolescents with ASD (Mayo and Eigsti, 2012).

One factor that appears to enhance cross-situational learning is knowledge about distracter objects. Suanda and Namy (2012) found that adults were more likely to choose the correct target if they had learned names for the foils, which represents an exclusion strategy. Yurovsky, Fricker, Yu, and Smith (2013) found that partial knowledge of competitor words aids in the learning of new words in a cross-situational task—even if the competitor words themselves remain only partially understood. Recent studies have also demonstrated that competition in cross-situational word learning, wherein stronger associations weaken other associations, is present at both local levels (within trials) and global levels (across trials; Yurovsky, Yu, & Smith, 2013). If children had some knowledge about competitor objects, they would likely be better cross-situational learners—creating a cascading effect of vocabulary learning. It is important to remember that in the current study, the distracter objects were not familiar; instead, they were the same novel words that children may or may not have learned. Future research should investigate whether children with ASD can use mutual exclusivity in the service of cross-situational word learning—in other words, whether knowing the names of competitor objects increases learning.

Why might older children have performed better than younger children on the cross-situational learning task? One previously considered answer is that they have superior vocabulary knowledge and additional experience learning new vocabulary words. Another possibility is that although cross-situational learning is presumed to be generally implicit in infants (Smith & Yu, 2008, 2011), older children may adopt more purposeful strategies that enhance their ability to selectively attend to the features that matter for learning new words (Akhtar & Montague, 1999). On the other hand, it is also possible that age had a positive impact on the manner in which children responded during test, and not specifically on different learning strategies. Although the current study did not determine whether successful learning in the children with ASD relied more heavily on associative learning or hypothesis testing, or a combination of the two (Yu & Smith, 2012), future studies should investigate this issue because it speaks to the processes by which successful cross-situational learning occurs. If some children are employing a type of explicit, top-down strategy, it is possible that future studies instructing them about the nature of the task, as in some adult studies (Yu & Smith, 2007), may increase performance (Kachergis, Yu, & Shiffrin, 2007b). Indeed, it has been proposed that children with ASD may compensate for implicit learning deficits with explicit learning strategies (Klinger et al., 2007).

As discussed by Smith and Yu (2012), successful cross-situational learning requires the association of visual objects and auditory labels, and it also requires memory of those associations across some period of time (i.e., a "unified multisensory memory," p. 22). The children who performed poorly in the current study may have had difficulty extracting the co-occurrence statistics that disambiguated the label-object pairings; they may have had difficulty forming a stable association between labels and objects; or they may have had difficulty remembering the associations over a period of a few minutes. A recent call has been made for research investigating the impact of memory on word learning (Wojcik, 2013), and memory also appears to play a key role in cross-situational word learning (Vlach & Johnson, 2013). Vlach and Johnson found that younger infants were capable of cross-situational learning only when pairings were presented in immediate succession, not when pairings were spread out across trials, whereas older infants were capable of learning in both conditions.

The TD and ASD groups both demonstrated evidence of visual recognition memory, and their novelty preferences did not significantly differ. It was not surprising that the children with ASD showed intact recognition memory, given previous studies suggesting unimpaired skills in this area (Dawson et al., 2001; Lind & Bowler, 2009). Contrary to predictions, visual recognition memory did not relate to cross-situational learning in either group. Visual recognition memory may not have been a sensitive predictor of cross-situational learning because most children were able to track visual stimuli across a short period of time. As pointed out by Yu and Smith (2011), there are considerable differences between remembering previously viewed objects in a visual recognition memory task and having to learn the names for those objects.

Although results for the pointing tasks indicated no significant association between simple word learning and cross-situational word learning, the standard measure of eye-gaze accuracy revealed a significant correlation between cross-situational and simple word learning in the TD group and a marginal association in the ASD group. The alternative variable based on the proportion of longest looks to target revealed a significant correlation between cross-situational and simple word learning in both groups. In combination, these findings point to the importance of forming label-object associations as a pre-requisite skill for cross-situational learning. TD children and children with ASD who are better at learning the names for objects in a simpler task are also better at learning the names for objects in a more complex cross-situational task.

The cross-situational word-learning task taught two label-object pairings at a time, and the simple word learning-task taught only one pairing at a time. Although both word-learning tasks provided children with a total of 10 exposures to each pairing, it was possible that children would have an easier time learning in the simpler as opposed to the more complex learning context. Interestingly, there were no significant differences between cross-situational learning accuracy and simple word-learning accuracy on eye-gaze or behavioral pointing measures in either group, meaning that children's performance provided no indication that they had an easier time learning in the simple word-learning condition. Although the reasons for this null finding are unknown, it is possible that differences in the training phases played a role. The cross-situational task presented children with two labels and two objects in 4-second trials, whereas the simple word-learning task presented children with one label and one object in 2-second trials. The rapid rate of presentation in the simple word-learning task may have limited children's learning; they may benefit from a longer exposure to the objects, even if they are presented with more information within that longer time span.

Results of both the cross-situational and simple pointing tasks revealed that, even if their overall learning of novel words was poor, both the typically developing children and the children with ASD were unwilling to give the same novel object multiple names—a principle known as mutual exclusivity (Markman, 1990). This finding supports previous work (Bedford et al., 2012; de Marchena et al., 2011; Parish-Morris, Hennon, Hirsh-Pasek, Golinkoff, & Tager-Flusberg, 2007; Preissler & Carey, 2005) demonstrating intact mutual exclusivity in children with ASD—even in children with considerable language impairment. For example, despite their challenges in using eye-gaze cues to learn new words, children with ASD in the study by Preissler and Carey (2005) mapped new labels onto novel objects, as opposed to familiar objects—meaning that the mutual exclusivity constraint appeared to be intact. Interestingly, informal comments from adult cross-situational learners in the study by Yu and Smith (2007) indicated that they had used a one-word/one-object strategy (see also Ichinco, Frank, & Saxe, 2009).

The plausibility of a cross-situational word learning mechanism has been criticized because it would be computationally overwhelming if humans stored all potential word-label associations to which they were exposed. However, Blythe, Smith, and Smith (2010) used a modeling approach to demonstrate that cross-situational word learning strategies can yield large vocabularies despite the considerable referential ambiguity during learning. The infants and the computational model in the study by Yu and Smith (2011) study only stored some of the potential label-object pairings, not all of them. Instead of remembering massive amounts of associations that in most cases will turn out to be useless, humans may employ selective attention to limit the environmental contingencies to be stored in memory.

Although there is currently no empirical evidence to support this claim, a protracted course of word learning, such as that shown by children with ASD, could be representative of a cross-situational word learning strategy. Vogt and Smith (2005) found that color words could be learned through a cross-situational mechanism, but that the process took a relatively long time to identify the correct mappings. In their discussion of the Common Feature strategy, Baron-Cohen and colleagues (1997) pointed out that, "…although the Common Feature strategy would be effective in eventually ironing out mapping errors, it is clearly not as efficient as the [speaker's direction of gaze] strategy, and so should still lead to some delay in vocabulary development" (p. 55).

Non-Social Visual Attention and Cross-Situational Learning

The third aim of this study was to investigate the relationship between non-social visual attention and cross-situational word learning in children with ASD. Investigating this question is important because orienting impairments, specifically disengagement of attention (i.e., sticky attention), emerge very early in life in infants who later develop ASD (Elsabbagh et al., 2009, 2013; Zwaigenbaum et al., 2005), suggesting that these deficits could impact a number of early developmental skills. Additionally, and most relevant to the current study, researchers have speculated that orienting deficits could lead to difficulties detecting and extracting social and non-social contingent relationships (Elsabbagh et al., 2013; Klinger et al., 2007; Renner et al., 2006).

One example of a contingent relationship is the association between a label and an object—an essential aspect of early word learning. Cross-situational word learning is a particularly important mechanism to investigate because it requires extraction of co-occurrences between labels and objects across ambiguous trials. It was hypothesized that disengagement of visual attention, but not shifting, would be associated with cross-situational word learning since successful learning requires attention to multiple visual stimuli. If children focused attention primarily on one of the visible images, they would be unlikely to learn the correct label-object co-occurrences across the ambiguous trials. Consistent with predictions, findings revealed a relationship between latency to disengage attention in the overlap condition (i.e., sticky attention) and cross-situational eye-gaze accuracy in the children with ASD but not in the TD group. Importantly, age may also play a role in the relationship between sticky attention and cross-situational learning, given that the correlation between sticky attention and cross-situational eye-gaze accuracy was marginal after accounting for age.

Several post hoc analyses were conducted to better understand why sticky attention and cross-situational eye-gaze accuracy were related. The hypothesis was that children with stickier attention would have more difficulty learning because their deficits in disengaging attention would prevent them from allocating adequate visual attention to both of the images presented during the training trials. If this were the case, sticky attention should not be significantly correlated with simple word learning eye-gaze accuracy. In support of this hypothesis, the correlation between sticky attention and simple word learning was not significant in either group, suggesting that sticky attention may have a specific effect on cross-situational learning.

The next step was to determine whether children with sticky attention showed different eye-gaze patterns during the training phase. If sticky attention affected visual attention to the training images, children's number of attention switches and length of longest fixation would be expected to relate to sticky attention. Correlational analyses indicated, however, that disengagement latency (sticky attention) was not significantly correlated with children's eye-gaze patterns during cross-situational training. Additionally, disengagement latency was not significantly correlated with children's performance in the cross-situational pointing task, meaning that sticky attention impacted children's eye gaze accuracy, but not pointing accuracy—during the test phase, but not during training.

As a next step, the correlation between disengagement latency (sticky attention) and familiar word recognition was tested in the children with ASD. Surprisingly, the negative correlation was significant and remained so even after age was statistically controlled. This finding indicated that stickier attention (i.e., longer mean latencies to disengage attention) was associated with lower accuracy levels of familiar word processing, pointing to a broader implication—namely, that deficits in disengaging attention appear to affect children's real-time processing of known words as measured by their visual attention.

One potential explanation for this relationship is that sticky attention may have slowed children's ability to recognize the target word and use that information to rapidly switch their gaze. Children with sticky attention who happened to be looking at the distracter image at the start of the trial may have taken longer to shift their attention to the target image when the other image was named. Unfortunately, the large amount of data loss prevented examination of latency in the cross-situational task. Approaches using multi-level mixed effects modeling may circumvent the issue of data loss and provide information about children's rates of increase in looking to the target image (Barr, 2008; Mirman, Dixon, & Magnuson, 2008). The post hoc mediator analysis shed light on the significant associations between sticky attention, familiar word processing, and children's vocabulary knowledge. The correlation between sticky attention and vocabulary `knowledge was no longer significant after accounting for the variance explained by familiar word recognition accuracy, suggesting that sticky visual attention affects children's processing of words in real time, which in turn affects vocabulary knowledge.

Researchers have speculated that visual orienting deficits, specifically deficits in disengagement that prevent the flexible switching of attention, may decrease children's visual examination of stimuli that are relevant at a particular moment in time. The current findings suggest that this may, in fact, be true in one specific circumstance: processing of newly learned and familiar words. Sticky attention may be driven by a strong, sustained interest in the current focus of attention, a reduced interest in looking at something new, or a combination of these factors. One study has shown that minimally verbal children with ASD demonstrate attentional patterns that are driven by stimulus salience (i.e., bottom-up orienting preferences), suggesting that the nature of the stimulus is particularly important in determining patterns of attention (Amso et al., 2013). Alternatively, it could be the case that superior comprehension of the target words could motivate children to reject the incorrect image once they heard it named. In this way, better word recognition and referent knowledge may override the tendency to fixate to an unnamed object.

The influence of sticky attention on eye-gaze patterns appears to affect performance in contexts besides language processing. In a visual search task, individuals with ASD took longer to launch their saccades and had overall longer fixations than their peers, pointing to a possible role of prolonged disengagement on visual search (Kourkoulou, Kuhn, Findlay, & Leekam, 2013). Interestingly, Kelly and colleagues (2012) found similar latencies in children with ASD and/or language impairment and their peers to 'reflexively' fixate a peripheral target when they were told to move their eyes to the target as quickly as possible, suggesting that explicit instructions may allow children to overcome potential effects of the tendency for sticky attention. In the study by Kelly et al., children with and without ASD who had impaired language demonstrated deficits in the volitional control of eye movements in an anti-saccade task. Fixating a named object may be one situation wherein some aspect of volitional control must come into play, especially when an increased interest in a salient stimulus prevents ready disengagement of attention.

It was somewhat surprising that there were no significant differences between the TD and ASD groups in their mean latency to shift or disengage attention in the gap-overlap task. Studies have identified atypical patterns of disengaging attention in infants at risk for ASD as young as 9 to 14 months of age (Elsabbagh et al., 2009, 2013; Zwaigenbaum et al., 2005), suggesting that these deficits may be apparent primarily early in development. Work by Landry and Bryson (2004), however, has suggested that deficits in atypical disengagement of attention are apparent even in children with ASD between 3 and 7 years of age. The current sample included children with ASD between 4 and 7 years of age, which is a similar age range; however, the mean age of the children with ASD and the TD children in the current sample was nearly one year older than the mean age in the study by Landry and Bryson, which may have affected the findings. In addition, the mean IQ of children in the study by Landry and Bryson was 70, whereas the mean IQ in the current sample was 95. The current findings are consistent with results of a study by Fischer and colleagues (2013) that found no deficits in disengaging visual attention in high-functioning children with ASD matched with a group of TD children on age and IQ.

Additionally, the gap-overlap task used in the current study used peripheral screens to present the side stimuli instead of screens that were adjacent to the center screen; this task adaptation may also have had an effect. Other studies have used an eccentricity of 13 degrees (Elsabbagh et al. 2009) or 15 degrees (Elsabbagh et al., 2013). The increased angle was used in the current study to increase the potential for sticky attention in the children with ASD, but it appears that doing so also elicited longer latencies from children in the TD group. Anecdotally, it was noted that some children, particularly older children in the TD group, seemed to grow tired of looking at the screens after several trials. There was no direct motivation for children to look at the screens, aside from interest in simply looking at the dynamic patterns. Participants were given no explicit instructions and the side stimuli did not present them with any relevant information from which they could learn. It is possible, then, that the lack of group differences was affected by differing levels of engagement in the task.

It is also important to note that age and/or vocabulary level may relate to sticky attention; both factors were marginally related to overlap latency in the ASD group. Boot, Pel, Evenhuis, and van der Steen (2012) found that visual processing time was significantly related to age and IQ in a group of participants with intellectual disability between ages 4 and 14. Landry and Bryson (2004), however, found no correlations between latencies to shift or disengage and any developmental measures, including age, nonverbal mental age, IQ, or language ability.

The current study revealed no significant relationships between cross-situational learning eye-gaze accuracy and eye-gaze patterns during cross-situational training in either group. Specifically, the number of attention shifts and the length of longest look did not significantly differ between strong and weak cross-situational learners in the TD group or the ASD group. This finding stands in contrast to findings by Yu and Smith (2011) that infants who were stronger cross-situational learners had fewer attention shifts and longer lengths of fixations during training. It is important to note that the participants in Yu and Smith's study were TD infants, not older children or children with ASD.

Visual patterns may differ earlier in development; the weak learners in Yu and Smith's study generated nearly 4 attention switches during the 4-second trials, whereas children in the present study typically made approximately 3 switches.

The current study examined eye-gaze patterns across all training trials, and thus did not differentiate patterns of eye-gaze over early versus late phases of training. Patterns of selective attention may vary over the course of cross-situational learning, meaning that future studies may identify differences in attention shifts, length of looks, or entropy (i.e., the stability or instability of looking patterns) based on early versus later points in training (Yu & Smith, 2011). Several studies have identified complex interactions between selective attention during cross-situational learning and learning outcomes (Smith & Yu, 2012; Yu, Zhong, & Fricker, 2012). For example, attention to objects and labels during cross-situational learning may be insufficient for learning in infants, and sensitivity to longer-term regularities over the course of learning as opposed to shorter-term regularities may be indicative of better cross-situational learning (Smith & Yu, 2012). Interestingly, looking at the incorrect as opposed to the correct referent during initial trials of cross-situational learning may have a positive effect on learning in adults (Fitneva & Christiansen, 2011).

Methodological Comments on Eye-Gaze Measures

The eye-gaze paradigms in the current study used either automatic eye tracking or manual coding of eye gaze. These methods both used eye movements to measure learning and language comprehension, but they obtained information about eye movements in different ways. The primary limitation of the LWL method was the time required to code the eye movements; the primary limitation of the eye-tracking method was potential data loss, which occurred when children moved slightly out of range and the eye tracker could no longer capture eye movements. Which method was best for children with ASD? Anecdotal evidence suggested that children were quite tolerant of both paradigms, and neither method appeared to produce unacceptable rates of data loss, although it would be desirable to retain even more than approximately 75% of the children with ASD. Hand coding may offer a benefit since it is not subject to mistracks. It also provided a valuable backup in the cases when the Tobii produced empty gazedata files.

One limitation of the current study was that children's inattention to the images resulted in some data loss at the level of individual trials and individual children. Data cleaning is a complex process because it requires balancing of two factors: 1) retaining as much data as possible to increase statistical power and validity; and 2) excluding data that provides an invalid measure of the construct in question. Strict cleaning criteria increase confidence in the validity of retained trials but limit the extent to which results can be generalized to the population represented by the original participant sample. Liberal cleaning criteria (i.e., retention of a greater number of trials) increase generalizability of results but run the risk of including invalid data.

Some cleaning criteria exclude trials in which the child looks away from the screen for 500 ms, thus including only those trials with very high levels of attention to the screen. On the other hand, some studies require children to devote very little attention for trials to remain in the analyses. For example, Brock et al. (2008) only required that their adolescent participants look at the screen 25% of the time during their window of interest, and Naigles and colleagues have required only 300 ms of looking time (Goodwin et al., 2012; Naigles et al., 2011; Swenson et al., 2007; Tek et al., 2008). The impact of data cleaning on the integrity and interpretation of eye-tracking results has received little attention, but future work should investigate this issue. In an attempt to balance validity and generalizability, and to account for the possibility that children with ASD may demonstrate short periods of inattention during eye-gaze tasks, the current study adopted a criterion of 50% looking time during the test window (see also Venker et al., 2013). This intermediate criterion ensured the exclusion of trials wherein children's gaze was fixated on the images only for a very short time, since these trials were unlikely to capture meaningful eye-gaze patterns. It was liberal enough, however, to include nearly 75% of the sample of children with ASD—a retention rate that compares favorably to the study by Yu and Smith (2011), which excluded 44% of their sample of TD children.

It was notable that language level had an effect on data loss in two different ways. First, the children with ASD who were excluded from the cross-situational analyses due to data loss had significantly lower receptive vocabulary levels than children who contributed data. No other factors, including age, IQ, or autism severity, were significantly different between the groups of children with ASD who did and did not contribute data. This finding suggests that children's language levels are related to their attention to a language-based eye-tracking task, such that children with lower language levels are more likely to look away from the screen and thus be excluded from the analyses. Unfortunately, children with more limited language skills are also the subgroup of children for whom the use of more implicit eye-gaze measures is most appealing, given potential difficulty with verbal, examiner-administered tasks (Charman et al., 2003b; Kjelgaard & Tager-Flusberg, 2001; Tager-Flusberg & Kasari, 2013).

Second, regardless of their diagnostic status, children in this study showed better attention to trials testing familiar words than trials testing novel words. Indeed, approximately 60% of familiar word trials were retained, compared to approximately 50% in the two novel word conditions. This finding suggests that inattention and data loss differ by trial content, such that children are more likely to look away from the screen on trials testing novel words than on trials testing familiar words. Future studies may consider a number of adaptations to address this issue. Including an increased number of novel test trials may increase the number of trials retained, but it would also increase the length of the task. Future studies may also limit data loss by providing older children with more specific instructions about their 'job' in the task and the length of time required. However, the impact of explicit instructions on interpretation of findings would also need to be considered.

Implications and Conclusions

The current findings pointed to a relationship between sticky attention (i.e., visual disengagement) and language processing, which may have implications for intervention. There is much to be learned, however, about how attention impacts language learning in children's everyday lives and within treatment contexts. Patten and Watson (2011) conducted a review and concluded that interventions targeting attention in children with ASD have received little empirical consideration, despite the likelihood that atypical orienting, sustaining, and shifting of attention lead to difficulties in learning. This type of work is important because attentional orienting may moderate the effects of early autism treatments (Rogers et al., 2012), possibly by affecting children's ability to learn from the opportunities with which they are provided.

An important avenue of future research is determining the direct impact of sticky attention on children's attention during naturalistic word learning contexts. Do children continue to demonstrate difficulty disengaging when a clear learning opportunity is provided, or can they override this tendency? How do auditory and visual cues interact to engage children's attention? Sticky attention to objects may increase the likelihood that language input from adults will not describe the visual stimuli to which children are attending. Although adults may show some benefit from looking at the incorrect referent during training (Fitneva & Christiansen, 2011), no studies have investigated this issue in children with ASD, whose representations of new label-object associations may be more fragile. Hearing a non-matching label when looking at an object with a different name may be particularly difficult for children with ASD, supporting treatment strategies wherein adults talk about the child's focus of attention instead of redirecting the child's attention (Haebig, McDuffie, & Ellis Weismer, 2013; McDuffie & Yoder, 2010; Venker, McDuffie, Ellis Weismer, & Abbeduto, 2012).

Work by Koegel, Shirotova, and Koegel (2009) suggested that providing an individualized orienting cue (e.g., providing hugs, kisses, and tickles; combining a motor action with a verbal cue; or giving a high five) prior to an adult prompt for verbal imitation facilitated children's orienting to and subsequent verbal imitation of an adult model. These cues led to increases in children's imitation of verbal models and in parent report of children’s number of spoken words. Additionally, Walton and Ingersoll (2013) found that verbal orienting cues may help children with ASD to learn the names of objects outside their own focus of attention. Children with ASD may have a particularly difficult time appropriately orienting their visual attention during social interactions that also contain preferred objects, indicating difficulty balancing social and non-social stimuli (Sasson & Touchstone, 2013).

Characteristic of many lab-based tasks, the cross-situational task used in the current study incorporated a high level of experimental control. All participants saw the same objects and heard the same labels in the eye-tracking task. Although experimental control is desirable because it eliminates many potential confounds, it also results in limited ecological validity. For example, children are unlikely to experience repetitive and simplified learning contexts like those included in the experimental tasks outside of the lab setting. Despite this acknowledged limitation, recent work by Yurovsky, Smith and Yu (2013) suggests that children’s perspectives in more naturalistic naming events may actually facilitate cross-situational learning in comparison with the more contrived contexts presented in lab-based tasks. Adults were able to learn novel labels for objects when presented with the child's view of the event (recorded with a head mounted camera), but not when presented with a third-person view from a tripod-mounted camera. Given that perspective clearly matters, it will be particularly interesting for future work to determine how repetitive motor behaviors, visual inspection of objects, and sticky attention affect the regularities learned by children with ASD.

As indicated by Arciuli and Torkildsen (2012), more research is needed on statistical learning in ASD—particularly longitudinal research that examines the relationship between statistical learning and language learning over time in the same group of children. Future work should examine the development of cross-situational learning abilities in young children with ASD and how these abilities change over time. Ideally, studies of statistical learning would focus on infants and very young children at the earliest stages of language learning, when cross-situational learning and visual orienting may exert the strongest impact on early vocabulary development.

References

Akhtar, N., & Montague, L. (1999). Early lexical acquisition: the role of cross-situational learning. *First Language*, *19*(57), 347–358.

Allen, C. W., Silove, N., Williams, K., & Hutchins, P. (2007). Validity of the social communication questionnaire in assessing risk of autism in preschool children with developmental problems. *Journal of Autism and Developmental Disorders, 37*(7), 1272-1278.

Allen, G., & Courchesne, E. (2001). Attention function and dysfunction in autism. *Frontiers in Bioscience*, *6*(1), d105–19.

Amso, D., Haas, S., Tenenbaum, E., Markant, J., & Sheinkopf, S. J. (2013). Bottom-up attention orienting in young children with autism. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-013-1925-5

Arciuli, J., & Torkildsen, J. V. K. (2012). Advancing our understanding of the link between statistical learning and language acquisition: The need for longitudinal data. *Frontiers in Psychology*, *3*, 324.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182.

Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development*, *68*(1), 48–57.

Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and LanguageLanguage*, *59*(4), 457–474.

Bedford, R., Gliga, T., & Frame, K. (2012). Failure to learn from feedback underlies word learning difficulties in toddlers at risk for autism. *Journal of Child Language*, *40*(1), 29–46.

Bhat, A. N., Galloway, J. C., & Landa, R. J. (2010). Social and non-social visual attention patterns and associative learning in infants at risk for autism. *Journal of Child Psychology and Psychiatry, 51*(9), 989–97.

Blythe, R. A., Smith, K., & Smith, A. D. M. (2010). Learning times for large lexicons through cross-situational learning. *Cognitive Science*, *34*(4), 620–42.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer [Computer program]. Version 5.3.43, http://www.praat.org/

Boot, F. H., Pel, J. J. M., Evenhuis, H. M., & van der Steen, J. (2012). Factors related to impaired visual orienting behavior in children with intellectual disabilities. *Research in Developmental Disabilities*, *33*(5), 1670–1676.

Boraston, Z., & Blakemore, S.-J. (2007). The application of eye-tracking technology in the study of autism. *The Journal of Physiology*, *581*, 893–8.

Brock, J., Norbury, C., Einav, S., & Nation, K. (2008). Do individuals with autism process words in context? Evidence from language-mediated eye-movements. *Cognition*, *108*(3), 896-904.

Bryson, S. E., Wainwright-Sharp, J. A., & Smith, I. M. (1990). Autism: A developmental spatial neglect syndrome? In J. Enns (Ed.), *The Development of Attention: Research and Theory* (p. 405-427).

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Drew, A., & Cox, A. (2003a). Predicting language outcome in infants with autism and pervasive developmental disorder. *International Journal of Language & Communication Disorders*, *38*(3), 265–285.

Charman, T., Drew, A., Baird, C., & Baird, G. (2003b). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language*, *30*(1), 213–236.

Charman, T., Taylor, E., Drew, A., Cockerill, H., Brown, J. A., & Baird, G. (2005). Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry*, *46*(5), 500–13.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2^{nd} ed). Hillsdale, NJ: Lawrence Earlbaum Associates.

Colombo, J., Mitchell, D. W., & Horowitz, F. D. (1988). Infant visual attention in the paired-comparison paradigm: Test-retest and attention-performance relations. *Child Development*, *59*(5), 1198–210.

Colombo, J., Mitchell, D. W., Dodd, J., Coldren, J. T., & Horowitz, F. D. (1989). Longitudinal correlates of infant attention in the paired-comparison paradigm. *Intelligence*, *42*, 33–42.

Dawson, G, Osterling, J., Rinaldi, J., Carver, L., & McPartland, J. (2001). Brief report: Recognition memory and stimulus-reward associations: indirect support for the role of ventromedial prefrontal dysfunction in autism. *Journal of Autism and Developmental Disorders*, *31*(3), 337–41.

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), 479–85.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental Psychology*, *40*(2), 271–83.

De Marchena, A., Eigsti, I.-M., Worek, A., Ono, K. E., & Snedeker, J. (2011). Mutual exclusivity in autism spectrum disorders: Testing the pragmatic hypothesis. *Cognition*, *119*, 96–113.

Dunn, L. M., & Dunn, D. M. (2006). *Peabody Picture Vocabulary Test, Fourth Edition*. Bloomington, MN: NCS Pearson, Inc.

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

Elison, J. T., Paterson, S. J., Wolff, J. J., Reznick, J. S., Sasson, N. J., Gu, H., … Piven, J. (2013). White matter microstructure and atypical visual orienting in 7-month-olds at risk for autism. *The American Journal of Psychiatry*, *170*(8), 899–908.

Ellis Weismer, S., Gernsbacher, M. A., Stronach, S., Karasinski, C., Eernisse, E. R., Venker, C. E., & Sindberg, H. (2011). Lexical and grammatical skills in toddlers on the autism spectrum compared to late talking toddlers. *Journal of Autism and Developmental Disorders*, *41*(8), 1065–75.

Ellis Weismer, S., Lord, C., & Esler, A. (2010). Early Language Patterns of Toddlers on the Autism Spectrum Compared to Toddlers with Developmental Delay. *Journal of Autism and Developmental Disorders*, *40*(10), 1259–1273.

Elsabbagh, M., Fernandes, J., Jane Webb, S., Dawson, G., Charman, T., & Johnson, M. H. (2013). Disengagement of visual attention in infancy is associated with emerging autism in toddlerhood. *Biological Psychiatry*. doi:10.1016/j.biopsych.2012.11.030

Elsabbagh, M., Volein, A., Holmboe, K., Tucker, L., Csibra, G., Baron-Cohen, S., … Johnson, M. H. (2009). Visual orienting in the early broader autism phenotype:

disengagement and facilitation. *Journal of Child Psychology and Psychiatry*, *50*(5), 637–42.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 321–35.

Fagan, J. F. (1990). The paired-comparison paradigm and infant intelligence. *Annals of the New York Academy of Sciences*, *608*, 337–357.

Falck-Ytter, T., Bölte, S., & Gredebäck, G. (2013). Eye tracking in early autism research. *Journal of Neurodevelopmental Disorders*, *5*(1), 28.

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. . Fernandez, & H. Clahsen (Eds.), *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing* (pp. 97–135). Amsterdam: John Benjamins.

Fischer, J., Koldewyn, K., Jiang, Y. V., & Kanwisher, N. (2013). Unimpaired attentional disengagement and social orienting in children with autism. *Clinical Psychological Science*. doi:10.1177/2167702613496242

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua*, *92*, 333–375.

Fitneva, S. A., & Christiansen, M. H. (2011). Looking in the wrong direction correlates with more accurate word learning. *Cognitive Science*, *35*(2), 367–80.

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science, 20*(5), 578-585.

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, *9*(1), 1–24.

Gliga, T., Elsabbagh, M., Hudry, K., Charman, T., & Johnson, M. H. (2012). Gaze following, gaze reading, and word learning in children at risk for autism. *Child Development*, *83*(3), 926–38.

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*(2), 109–135.

Goodwin, A., Fein, D., & Naigles, L. R. (2012). Comprehension of wh-questions precedes their production in typical development and autism spectrum disorders. *Autism Research*, *5*(2), 109–123.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(5), 693–705.

Haebig, E., McDuffie, A., & Ellis Weismer, S. (2013). Brief report: parent verbal responsiveness and language development in toddlers on the autism spectrum. *Journal of Autism and Developmental Disorders*, *43*(9), 2218–27.

Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behavior and Development*, *16*(4), 405–422.

Ibanez, L. V, Messinger, D. S., Newell, L., Lambert, B., & Sheskin, M. (2008). Visual disengagement in the infant siblings of children with an autism spectrum disorder (ASD). *Autism*, *12*(5), 473–85.

Ichinco, D., Frank, M., & Saxe, R. (2009). Cross-situational word learning respects mutual exclusivity. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, *31*.

Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual orienting in early infancy: contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience*, *3*(4), 335–44.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2007a). Frequency and contextual diversity effects in cross-situational word learning. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society* (Vol. 31).

Kachergis, G., Yu, C., & Shiffrin, R. M. (2007b). Cross-situational statistical learning: Implicit or intentional? *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2013). Actively learning object names across ambiguous situations. *Topics in Cognitive Science*, *5*(1), 200–213.

Kelly, D. J., Walker, R., & Norbury, C. F. (2012). Deficits in volitional oculomotor control align with language status in autism spectrum disorders. *Developmental Science*, *16*(1), 56-66.

Kjelgaard, M. M., & Tager-Flusberg, H. (2001). An investigation of language impairment in autism: implications for genetic subgroups. *Language and Cognitive Processes*, *16*(2-3), 287–308.

Klinger, L., Klinger, M., & Pohlig, R. (2007). Implicit learning impairments in autism spectrum disorders. In J. M. Perez, P. M. Gonzalez, M. L. Comi, & C. Nieto (Eds.), *New Developments in Autism: The Future Is Today*. London: Jessica Kingsley Publishers.

Koegel, R. L., Shirotova, L., & Koegel, L. K. (2009). Brief report: using individualized orienting cues to facilitate first-word acquisition in non-responders with autism. *Journal of Autism and Developmental Disorders*, *39*(11), 1587–92.

Kourkoulou, A., Kuhn, G., Findlay, J. M., & Leekam, S. R. (2013). Eye movement difficulties in autism spectrum disorder: implications for implicit contextual Learning. *Autism Research*. doi:10.1002/aur.1274

Kover, S. T., & Atwood, A. K. (2013). Establishing equivalence: methodological progress in group-matching design and analysis. *American Journal on Intellectual and Developmental Disabilities*, *118*(1), 3–15.

Landry, R., & Bryson, S. E. (2004). Impaired disengagement of attention in young children with autism. *Journal of Child Psychology and Psychiatry*, *45*(6), 1115–22.

Leekam, S. & Moore, C. (2001). The development of attention and joint attention in children with autism. In The development of autism: perspectives from theory and research (ed. J. A. Burack, T. Charman, N. Yirmiya & P. R. Zelazo), pp. 105–129. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lind, S. E., & Bowler, D. M. (2009). Recognition memory, self-other source memory, and theory-of-mind in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *39*(9), 1231–9.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part 1): Modules 1 – 4.* Torrence, CA: Western Psychological Services.

Luyster, R. J., Kadlec, M. B., Carter, A., & Tager-Flusberg, H. (2008). Language assessment and development in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*, 1426–38.

Luyster, R., & Lord, C. (2009). Word learning in children with autism spectrum disorders. *Developmental Psychology*, *45*(6), 1774–1786.

Maguire, M. J., Hirsh-Pasek, K., Golinkoff, R. M., & Brandone, A. C. (2008). Focusing on the relation: fewer exemplars facilitate children's initial verb learning and extension. *Developmental Science*, *11*(4), 628–34.

Markman, E. (1990). Constraints children place on word meanings. *Cognitive Science*.

Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*(3), B101-B111.

Mayo, J., & Eigsti, I.-M. (2012). Brief report: A comparison of statistical learning in school-aged children with high functioning autism and typically developing peers. *Journal of Autism and Developmental Disorders, 42*(11), 2476-2485.

McDuffie, A., Kover, S. T., Hagerman, R., & Abbeduto, L. (2012). Investigating word learning in fragile X syndrome: a fast-mapping study. *Journal of Autism and Developmental Disorders*, *43*(7), 1676-1691.

McDuffie, A., & Yoder, P. (2010). Types of parent verbal responsiveness that predict language in young children with autism spectrum disorder. *Journal of Speech, Language, and Hearing Research : Journal of Speech, Language, and Hearing Research*, *53*(4), 1026–39.

McDuffie, A., Yoder, P., & Stone, W. (2006). Fast-mapping in young children with autism spectrum disorders. *First Language*, *26*(4), 421–438.

McGregor, K. K., & Bean, A. (2012). How children with autism extend new words. *Journal of Speech, Language, and Hearing Research*, *55*(February), 70–84.

McGregor, K. K., Rost, G., Arenas, R., Farris-Trimble, A., & Stiles, D. (2013). Children with ASD can use gaze in support of word recognition and learning. *Journal of Child Psychology and Psychiatry.* doi:10.1111/jcpp.12073

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(22), 9014–9.

Minshew, N. J., & Goldstein, G. (2001). The pattern of intact and impaired memory functions in autism. *Journal of Child Psychology and Psychiatry, 42*(8), 1095–101.

Mirman, D., Dixon, J. & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494.

Naigles, L. R., Kelty, E., Jaffery, R., & Fein, D. (2011). Abstractness and continuity in the syntactic development of young children with autism. *Autism Research, 4*(6), 422-437.

Newman, R., Ratner, N. B., Jusczyk, A. M., Jusczyk, P. W., & Dow, K. A. (2006). Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology*, *42*, 643–655.

Norbury, C. F., Brock, J., Cragg, L., Einav, S., Griffiths, H., & Nation, K. (2009). Eye-movement patterns are associated with communicative competence in autistic spectrum disorders. *Journal of Child Psychology and Psychiatry*, *50*(7), 834–42.

Norbury, C. F., Griffiths, H., & Nation, K. (2010). Sound before meaning: word learning in autistic disorders. *Neuropsychologia*, *48*(14), 4012–9.

Oakes, L. M., Kovack-lesh, K. A., & Horst, J. S. (2010). Two are better than one: Comparison influences infants' visual recognition memory. *Journal of Experimental Child Psychology*, *104*(1), 124–131.

Olson, R. (2011) WISP User's Guide.

Ozonoff, S., Macari, S., Young, G. S., Goldring, S., Thompson, M., & Rogers, S. J. (2008). Atypical object exploration at 12 months of age is associated with autism in a prospective sample. *Autism*, *12*(5), 457–72.

Parish-Morris, J., Hennon, E., Hirsh-Pasek, K., Golinkoff, R. M., & Tager-Flusberg, H. (2007). Children with autism illuminate the role of social intention in word learning. *Child Development*, *78*(4), 1265–87.

Patten, E., & Watson, L. R. (2011). Interventions targeting attention in young children with autism. *American Journal of Speech-Language Pathology*, *20*(1), 60–9.

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally. Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*(12), 1894–902.

Pierce, K., & Courchesne, E. (2001). Evidence for a cerebellar role in reduced exploration and stereotyped behavior in autism. *Biological Psychiatry*, *49*(8), 655–664.

Posner, M.I. (1988). Structures and functions of selective attention. In T. Boll & B. Bryant (Eds.), Clinical neuropsychology and brain function: Research, measurement and practice (pp. 169–202). Washington, DC: American Psychological Association.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Preissler, M. A., & Carey, S. (2005). The role of inferences about referential intent in word learning: evidence from autism. *Cognition*, *97*(1), B13–23.

Quine, W. (1960). *Word and Object*. Cambridge, MA: MIT Press.

Renner, P., Klinger, L., & Klinger, M. R. (2006). Exogenous and endogenous attention orienting in autism spectrum disorders. *Child Neuropsychology*, *12*(4-5), 361–82.

Richmond, J., Colombo, M., & Hayne, H. (2007). Interpreting visual preferences in the visual paired-comparison task. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(5), 823–31.

Rogers, S. J., Estes, A., Lord, C., Vismara, L., Winter, J., Fitzpatrick, A., ... & Dawson, G. (2012). Effects of a brief Early Start Denver Model (ESDM)–based parent intervention on toddlers at risk for autism spectrum disorders: a randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(10), 1052-1065.

Roid, G., & Miller, L. (2002). Leiter International Performance Scale-Revised. Wood Dale, IL: Stoelting Co.

Romberg, A., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 906-914.

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2004). Infant visual recognition memory. *Developmental Review*, *24*(1), 74–100.

Rothbart, M. K., Posner, M. I., & Rosicky, J. (1994). Orienting in normal and pathological development. *Development and Psychopathology*, *6*(04), 635–652.

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

Sacrey, L.-A. R., Bryson, S. E., & Zwaigenbaum, L. (2013). Prospective examination of visual attention during play in infants at high-risk for autism spectrum disorder: A longitudinal study from 6 to 36 months of age. *Behavioural Brain Research*, *256*, 441–450.

Saffran, J. R., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Sasson, N. J., & Touchstone, E. W. (2013). Visual attention to competing social and object images by preschool children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*. doi: 10.1007/s10803-013-1910-z

Scott, R. M., & Fisher, C. (2012). 2.5-Year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163–180.

Scott-Van Zeeland, A. A., McNealy, K., Wang, A. T., Sigman, M., Bookheimer, S. Y., & Dapretto, M. (2010). No neural evidence of statistical learning during exposure to artificial languages in children with autism spectrum disorders. *Biological Psychiatry*, *68*(4), 345–351.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1-2), 39–91.

Smith, L. B., & Yu, C. (2012). Visual attention is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development*, *9*(1), 25-49.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–68.

Sokolov, E. (1963). *Perception and the Conditioned Reflex.* New York: Macmillan.
Sparrow, S.S., Cicchetti, V.D., & Balla, A.D. (2005). Vineland Adaptive Behavior Scales, 2nd Edition. American Guidance Service: Circle Pines, MN.

Suanda, S. H., & Namy, L. L. (2012). Detailed behavioral analysis as a window into cross-situational word learning. *Cognitive Science*, *36*, 545–559.

Swensen, L. D., Kelley, E., Fein, D., & Naigles, L. R. (2007). Processes of language acquisition in children with autism: evidence from preferential looking. *Child Development*, *78*(2), 542–57.

Swettenham, J., Baron-Cohen, S., Charman, T., Cox, a, Baird, G., Drew, A., … Wheelwright, S. (1998). The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and nonautistic developmentally delayed infants. *Journal of Child Psychology and Psychiatry*, *39*(5), 747–53.

Tager-Flusberg, H., & Kasari, C. (2013). Minimally verbal school-aged children with autism spectrum disorder : the neglected end of the spectrum. *Autism Research*. doi:10.1002/aur.1329

Tager-Flusberg, H., Paul, R., & Lord, C. (2005). Language and Communication in Autism. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of Autism and Pervasive Developmental Disorders* (3rd ed., pp. 335–364). Hoboken, NJ: Wiley and Sons.

Tek, S., Jaffery, G., Rein, D., & Naigles, L. R. (2008). Do children with autism spectrum disorders show a shape bias in word learning? *October*, *1*(4), 208–222. Do

Tobii T60 XL Eye Tracker User Manual. (2010). Revision 2. Tobii Technology AB.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156.

Venker, C. E., Eernisse, E. R., Saffran, J. R., & Ellis Weismer, S. (2013). Individual differences in the real-time comprehension of children with ASD. *Autism Research*, *6*(5), 417–32.

Venker, C. E., McDuffie, A., Ellis Weismer, S., & Abbeduto, L. (2012). Increasing verbal responsiveness in parents of children with autism:a pilot study. *Autism*, *16*(6), 568–85.

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition, 127*(3), 375-382.

Vogt, P., & Smith, A. (2005). Learning colour words is slow: A cross-situational learning account. *Brain and Behavior Sciences, 28,* 509-510.

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology*, *45*(6), 1611–7.

Walton, K. M., & Ingersoll, B. R. (2013). Expressive and receptive fast-mapping in children with autism spectrum disorders and typical development: The influence of orienting cues. *Research in Autism Spectrum Disorders*, *7*(6), 687–698.

Wetherby, A. M., Woods, J., Allen, L., Cleary, J., Dickinson, H., & Lord, C. (2004). Early indicators of autism spectrum disorders in the second year of life. *Journal of Autism and Developmental Disorders*, *34*(5), 473–493.

Wiggins, L. D., Bakeman, R., Adamson, L. B., & Robins, D. L. (2007). The utility of the social communication questionnaire in screening for autism in children referred for early intervention. *Focus on Autism and Other Developmental Disabilities, 22*(1), 33-38.

Wojcik, E. H. (2013). Remembering new words: integrating early memory development into word learning. *Frontiers in Psychology*. doi:10.3389/fpsyg.2013.00151

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*(5), 414-420.

Yu, C., & Smith, L. B. (2011). What you learn is what you see: using eye movements to study infant cross-situational word learning. *Developmental Science*, *14*(2), 165–180.

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: prior questions. *Psychological Review*, *119*(1), 21–39.

Yu, C., Zhong, Y., & Fricker, D. (2012). Selective attention in cross-situational statistical learning: evidence from eye tracking. *Frontiers in Psychology*, *3*, 1-16. doi:10.3389/fpsyg.2012.00148

Yurovsky, D., Fricker, D., Yu, C., & Smith, L. (2013). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin and Review*. doi: 10.3758/s13423-013-0443-y

Yurovsky, D., Smith, L., & Yu, C. (2013). Statistical word learning at scale: the baby's view is better. *Developmental Science, 16*(6), 959-966.

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*. doi:10.1111/cogs.12035

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience*, *23*(2-3), 143–52.