Word-Learning Strategies of Children with Autism Spectrum Disorders:

Understanding of Referential Intent

A Thesis

Submitted to the Faculty

Of

Drexel University

By

Kristina E. Patrick

in partial fulfillment of the

requirements for the degree

of

Doctor of Philosophy

February 2013
Acknowledgments

I am grateful for the support of my thesis committee: Drs. Felicia Hurewitz, Doug Chute, and Amy Booth. I especially want to thank my chair, Dr. Hurewitz for spending numerous hours mentoring me on this project. I also appreciate all the hard work of many research assistants including Amy Krischer, Jennifer Mongiat, Vanessa Dopker, Carolyn Moy, and Sarah Wright. Thank you to Michael Pesch for providing technical support for data organization. Thanks also to Dr. Maria Schultheis, Dr. Jennifer Gallo, and my lab members for providing helpful constructive criticism and guidance on this project.
# Table of Contents

LIST OF TABLES ........................................................................................................... v

LIST OF FIGURES ......................................................................................................... vi

ABSTRACT ..................................................................................................................... vii

1. INTRODUCTION ....................................................................................................... 1
   1.1 Overview ................................................................................................................ 1
   1.2 Gaze Following ...................................................................................................... 1
   1.2 Social Information ............................................................................................... 2
   1.3 Syntactic Information ............................................................................................ 5
   1.4 Present Study ........................................................................................................ 6

2. METHOD ................................................................................................................... 8
   2.1 Participants ............................................................................................................ 8
      2.1.1 Children with ASD ........................................................................................ 8
      2.1.2 Age-Matched Control Group ....................................................................... 8
      2.1.3 Vocabulary-Matched Control Group ............................................................. 9
   2.2 Materials ............................................................................................................... 9
   2.3 Assessments ......................................................................................................... 10
      2.3.1 Autism Diagnostic Observation Schedule .................................................... 10
      2.3.2 Peabody Picture Vocabulary Test, 4th edition ................................................. 11
   2.4 Design .................................................................................................................. 12
   2.5 Procedure ............................................................................................................ 12
      2.5.1 Overview ....................................................................................................... 12
      2.5.2 Familiarization Task ..................................................................................... 13
1. INTRODUCTION .................................................................................................................. 4

2. METHODS ................................................................................................................................ 10

2.1 Participant Selection ........................................................................................................... 10
2.2 Equipment .......................................................................................................................... 11
2.3 Instructions ........................................................................................................................ 11
2.3.1 Vocabulary Instruction .................................................................................................. 12
2.3.2 Procedure ...................................................................................................................... 12
2.4 Coding .................................................................................................................................. 13
2.4.1 Comprehension Testing ................................................................................................. 13
2.4.2 Eye Gaze ....................................................................................................................... 13
2.4.3 Random Factors .............................................................................................................. 14

2.5 RESULTS ............................................................................................................................. 15

2.5.1 Accuracy ......................................................................................................................... 15
2.5.2 Follow-Up Testing ......................................................................................................... 16

2.6 DISCUSSION ........................................................................................................................ 19

2.6.1 Findings and Implications ............................................................................................. 19
2.6.2 Limitations ...................................................................................................................... 20

3. RESULTS .................................................................................................................................. 21

3.1 Word-Mapping Accuracy .................................................................................................... 21
3.2 Children’s Direction of Gaze During Teaching Episodes ................................................... 22
3.3 PPVT-IV and ADOS as Predictors of Word-Mapping Accuracy ......................................... 25

4. DISCUSSION ........................................................................................................................... 28

4.1 Findings and Implications ................................................................................................... 28
4.2 Limitations .......................................................................................................................... 29
4.3 Future Research .................................................................................................................. 31

LIST OF REFERENCES ................................................................................................................... 33
List of Tables

1. Accuracy by Group ................................................................. 18

2. Accuracy by Group and Age Group ........................................... 20

3. Accuracy at Follow-Up .......................................................... 28
List of Figures

1. Full Set of Experimental Objects.................................................................10
2. Word-Mapping Accuracy by Group and Condition ........................................18
3. Proportion of Children Accurately Mapping by Age Group and Condition ........21
4. Proportion of Time Spent Looking at Experimenter and Target by Condition ....23
5. Proportion of Time Spent Looking at Experimenter and Target ........................24
6. Number of Trials Correct Predicted by Nouns and Verbs for ASD Children .......26
7. Trials Correct Predicted by Nouns and Verbs for Vocabulary-Matched Children....26
8. Initial and Follow Up Performance....................................................................29
Abstract

Word-Learning Strategies of Children with Autism Spectrum Disorders: Understanding of Referential Intent
Kristina E. Patrick

As typically developing children begin to understand joint attention, they become avid word learners (Baldwin, 1991). In contrast, children diagnosed with autism spectrum disorders (ASD) often struggle with joint attention and word learning (Baron-Cohen, Baldwin, & Crowson, 1997), although more linguistically advanced syntactic abilities are often intact (Tager-Flusberg et al, 1990). We sought to explore the roles of attention and social understanding in the abilities of children with and without ASD to map novel nouns to their intended referents. After their receptive vocabulary skills were assessed, children with ASD (ages 2 to 10) and matched controls were taught novel words while a speaker gazed at, or gazed at and pointed to, touched, or manipulated a referent object. Sessions were coded for word-mapping accuracy and children’s direction of gaze during teaching episodes. Controls of all ages, and ASD children over 6.5 years, mapped correctly in all four conditions, but ASD children 6.5 years and younger performed above chance only in conditions that included a manual gesture. Children with ASD looked at the experimenter less often during teaching episodes than controls, and time spent looking at the experimenter was predictive of word learning for all children. Looks to the target during teaching did not differ by group and were not predictive of performance. Knowledge of nouns, but not verbs, was predictive of performance for controls while knowledge of verbs, but not nouns, predicted performance in children with ASD. Results suggest not only a word-learning delay, but
qualitatively different developmental trajectories, for children with autism spectrum disorders compared to typically developing children.
1. INTRODUCTION

1.1 Overview

Although typically developing children learn words quickly and efficiently, children with autism spectrum disorders (ASD) are often delayed in their ability to receptively and expressively use language. While it is clear that typically developing children are better word learners than children with ASD, it is not yet clear how these groups differ in their approach to word-learning. Furthermore, although a language deficit is a core feature of ASD, many individuals on the autism spectrum do eventually acquire an extensive vocabulary, but it is not yet known whether they do so using the same mechanisms as typically developing children.

1.2 Gaze Following

Numerous studies have indicated that typically developing children as young as 17 months follow the gaze of a speaker to learn new words (Baldwin, 1991; Baldwin, 1993a; Baldwin, 1993b). However, children do not merely use associative processes to pair objects and word labels. Instead, they use direction of speaker gaze to interpret a relationship between a speaker and an object (Woodward, 2003). In the absence of clear social indicators of the speaker’s intended reference, typically developing children do not pair novel objects and labels (Baldwin et al, 1996; Waxman & Gelman, 2009). Gaze following appears to be a particularly important strategy for inferring speaker intent.

Children with ASD often exhibit delays in gaze-following abilities (Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997), but many of them do develop this skill eventually, either spontaneously (Kylliainen & Heietanen, 2004; Luyster & Lord, 2009) or through explicit training (Klein, MacDonald, Vaillancourt, Ahearn, & Dube, 2009;
Even for those who do develop this skill, applying it to the problem of locating objects of attention in space (Klin, Jones, Schultz, Volkmar, & Cohen, 2002a) and for identifying the referents of new words (Akechi, Senju, Kikuchi, Tojo, Osanai, & Hasegawa, 2011; Baron-Cohen, Baldwin, & Crowson, 1997; Gliga, Elsabbagh, Johnson, Hudry, & Charman, 2012; Priessler & Carey, 2005) remains problematic. Problems following the gaze of a speaker are further compounded when distractors are present (Luyster & Lord, 2009; Parish-Morris, Hennon, Hirsh-Pasek, Golinkoff, & Tager-Flusberg, 2007), possibly due to greater distractibility (Vaidya et al, 2011), information processing difficulties (Baldwin & Moses, 2001), and the reduced saliency of social cues for children with ASD compared to children with typical development (Vaidya et al, 2011). Consequently, even when children with ASD reliably follow speaker gaze, they rarely interpret the referential intent of that gaze.

1.3 Social Information

In addition to gaze following, typically developing children use other social cues to learn new words. Even before children are able to interpret gaze, they can gather social information with the aid of more salient cues. For instance, by 12 months old, infants understand the referential connection between a speaker and an object that she is touching (Woodward & Guajardo, 2002) or pointing to (Woodward, 2003) even though the ability to interpret referential intent through gaze alone is still developing (Baldwin, 1993a). Furthermore, 16 month olds are able to use gaze direction alone to map labels to referents when their attention is already directed toward the object of referential intent (Baldwin, 1991). By the time children reach their second birthday, they have acquired
not only the ability to utilize gaze direction to interpret communicative intent, but also can use more sophisticated social cues hinging on the novelty of objects to the speaker (Akhtar, Carpenter, & Tomasello, 1996) and the distinction between purposeful and accidental actions (Diesendruck, Markson, Akhtar, & Reudor, 2004). In addition, attention-grabbing features such as enhanced object saliency (Moore, Angelopoulos, & Bennet, 1999), presentation in novel contexts (Samuelson & Smith, 1998), and touching or moving referential objects (Booth, McGregor, & Rohlfing, 2008; Woodward & Guajardo, 2002) increase infants’ abilities to pair objects and labels.

As typically developing children get older, they are able to extract social information necessary for word learning with fewer direct cues. However, it is not entirely clear what amount of cuing is necessary throughout development for children to appropriately attend to a speaker and gather information about the relationship between words and referents. Booth and colleagues (2008) sought to examine whether enhancing children’s attention toward objects by pairing presentation with increasing levels of gestural support led to improved word-mapping in two and a half year-old children. Eighty 28 to 31-month-old toddlers were randomly assigned to one of five groups in which an experimenter presented two objects and labeled one of them while either looking at the center of the table (baseline), gazing at the object, gazing at and pointing to the object, gazing at and touching the object, or gazing at and manipulating the object. Children were tested on their abilities to map the word to the object of referential intent with a forced-choice comprehension task. Researchers also recorded the direction of the children’s gaze throughout the teaching episodes. Results indicated that toddlers were able to use cues from all non-baseline conditions to learn new words, and word learning
improved up the gestural hierarchy. However, the only significant difference in performance between consecutive levels was in the comparison between gaze and point conditions. During word labeling, children spent an equal amount of time looking at the target in the gaze and point conditions. However, time spent looking at the target significantly increased from the point to the touch conditions and again from the touch to the manipulate conditions. Proportion of trials in which children looked at the target object while it was being labeled was significantly positively correlated with performance on the comprehension task. However, gestural condition remained a significant contributor to word learning even when controlling for the effect of attention. This research suggests that as long as they are attending to the task, two and a half year-old children can gather social information to infer the intention of a speaker without the speaker directly guiding their attention toward the object. In fact, a social event as subtle as intentional gaze-direction of a speaker appears to be enough to elicit young children’s attention. Typically developing children appear motivated to gather social information, even in the absence of major attention-grabbing cues.

An inability to effectively interpret social information, or perhaps a lack of motivation to do so (Chevallier, Kohls, Toiani, Brodkin, & Schultz, 2012), may partially account for the word-learning delay common in children with ASD. In addition to difficulty inferring referential intent from speaker gaze, children with ASD struggle with interpretation of a variety of social cues. Compared to typically developing children, children with ASD show impaired abilities to orient towards social stimuli (Dawson, Meltzoff, Osterling, Rinaldi, & Brown, 1998), imitate and infer the intention of other’s actions (Vivanti, Nadig, Ozonoff, & Rogers, 2008; Vivanti et al, 2011), and recognize
strategies implemented by expert models (McGregor & Bean, 2012). However, some evidence suggests that children with ASD can use some cues, such as novelty, as well as children with typical development to learn new words (Priessler & Carey, 2005). Although children with ASD do appear able to extract and interpret some social cues, they may be more likely than typically developing children to pair words based on object salience (Akechi et al, 2011) and associative learning (Baron-Cohen et al, 1997; Preissler & Carey, 2005). As a result, children with ASD may fail to pair objects with labels when they must rely solely on social information to do so (Hennon, 2003). However, the amount of guidance required for children with ASD to attend to and subsequently interpret social information to learn new words remains unclear.

1.4 Syntactic Information

Children amass their initial vocabulary of concrete words through word-to-world pairings, supported by many of the social cues discussed above. The acquisition of more abstract words, however, appears to hinge on more complex linguistic information. According to the theory of syntactic bootstrapping (Fisher, 1996; Gleitman, 1990; Gleitman Cassidy, Nappa, Papafragou, & Trueswell, 2005; Landau & Gleitman, 1985; Trueswell & Gleitman, 2004), children’s accumulated linguistic experience ultimately allows children to infer the meaning of new words based on the structure of the language in which they occur. Specifically, children can use knowledge of previously acquired words in combination with clues from the syntactic structure of spoken sentences to quickly infer the meanings of more abstract words through “structure-to-world” mapping. In this way, word learning is further “bootstrapped” by previously acquired linguistic knowledge. In sum, observational strategies like gaze following are initially used to
acquire a pool of concrete words, which are then used as a scaffold to aid in interpretation of syntactic structure for more efficient word learning and acquisition of abstract words.

Although children with ASD exhibit social delays that may interfere with their ability to efficiently map words to observed referents, the current literature indicates that their syntactic abilities are largely intact (Paul, Fischer, & Cohen, 1988; Swensen et al, 2007; Tager-Flusberg et al, 1990). A recent study by Naigles and colleagues (2011) found that children with ASD engage in syntactic bootstrapping in a manner similar to typically developing children. The language deficit associated with ASD appears to be primarily associated with social and pragmatic skills, rather than grammatical abilities (Kelley, Paul, Feing, & Naigles, 2006). If this is the case, then these children may be able to implement more mature syntactic strategies to learn new words, while continuing to have difficulty with developmentally prior socio-pragmatic strategies.

1.5 Present Study

We adapted the procedure used by Booth and colleagues (2008) to decipher the level of gestural support necessary to direct the attention of children with ASD toward objects of referential intent. Children with ASD, 2-10 years old, were matched with typically developing children on age and gender (age-match group) and receptive vocabulary scores (vocabulary-match group). In a within-subjects design procedure, experimenters presented four sets of two objects to each child and then labeled one object in each set while either gazing at, gazing at and pointing to, gazing at and touching, or gazing at and manipulating the target object. Children were then asked to identify the target object. Sessions were also recorded and participants’ eye gaze patterns were coded. The level of gestural support necessary for children with ASD to correctly map
words above chance will indicate the amount of guidance that children with ASD need to extract social information to learn new words. Analyses of participants’ eye gaze will provide evidence of the types of word-learning strategies that children are implementing. If children who accurately map words to their referents spend more time than inaccurate children looking at the target object, it is likely that word learning is primarily occurring through associative processes. However, if time spent looking at the experimenter is associated with word learning, then identifying relevant social information must be important. Furthermore, if joint attention skills and initiation of social gestures, both measured by a standardized autism assessment, predict word learning on our task, it will indicate that social communication skills are important for early word learning of children with ASD. Although it is clear that typically developing children implement social strategies when learning new words, this study will help contribute to our understanding of whether children with ASD are able to use such strategies.

In addition to evaluating the use and effectiveness of social strategies for word learning in children with ASD, we also sought to shed light on the role previous lexical knowledge, as indicated by a standardized receptive vocabulary assessment, plays in predicting children’s abilities to understand and use social gaze to learn new nouns. Because we used a noun-learning task, it seems logical that previous acquisition of nouns would predict performance. If, however, acquisition of verbs, rather than nouns, is predictive of performance, it may indicate that implementation of social gaze strategies for noun learning can only occur after children with ASD have acquired sophisticated strategies required for abstract word learning. If this is the case, then ASD children,
unlike typically developing children, might implement social gaze strategies later in development, rather than as an early word-learning tool.

2. METHOD

2.1 Participants

This study included 67 participants, ages 1 year, 7 months to 10 years, 0 months ($M = 5.44$, $SD = 2.30$), recruited from the northern and western suburbs of Chicago.

2.1.1 Children with ASD

Participants with ASD included 25 children, ages 2 years, 11 months to 9 years, 10 months ($M = 6.06$ years, $SD = 1.97$), who were diagnosed with an ASD by a psychologist or medical doctor. Mean language level computed using the Peabody Picture Vocabulary Test, fourth edition (PPVT-IV) was 4.60 years ($SD = 2.81$). Three additional children initially participated as part of the ASD group but were excluded because they did not come from a primarily English-speaking home ($n = 1$) or failed to meet the diagnostic cutoff on the Autism Diagnostic Observation Schedule ($n = 2$). Although inclusion criteria initially included children with ASD ages 2-10, halfway through recruitment, we began to strategically target recruitment of children 6.5 years and younger in order to examine developmental effects of word learning.

2.1.2 Age-Matched Control Group

Participants in this group included 25 typically developing children each matched on gender and within 2 months of the chronological age of a participant with ASD. The age range of this group was very similar to that of the ASD group ($range = 2$ years, 10 months – 10 years, 0 months, $M = 6.07$ years, $SD = 2.02$) but language level was higher ($M = 7.51$ years, $SD = 2.68$).
2.1.3 Vocabulary-Matched Control Group

Participants in this group included 25 children each matched within 2 months of PPVT language equivalency with a participant with ASD. The age range of this group was lower than that of the ASD group (range = 1 year, 7 months – 9 years, 0 months, $M = 4.22$, $SD = 2.44$) but language level was similar ($M = 4.55$, $SD = 2.87$). Eight children participated as an age match for one child with ASD and a vocabulary match for another child.

2.2 Materials

Materials included two sets of familiar stimuli used for a training task at the beginning of the experiment. One set included a miniature shoe, miniature brush, and a toy airplane; the other set included a block, a cup, and a ring from a ring stacker. The objects were intended to be common objects that young children would have seen and heard labels for.

Stimuli for the experimental sections (see Figure 1) consisted of four sets of three novel objects. Each set included a target object, a foil object presented during teaching episodes, and a novel foil object presented during free play. The purpose of the novel foil object was to reduce chance responding to 33%. The experimental objects were unfamiliar objects that have names that are not commonly known by children.
1. Full set of experimental objects. Each of the four sets of objects included a target object, a distractor object (foil), and an additional distractor presented during free play (novel foil).

2.3 Assessments

2.3.1 Autism Diagnostic Observation Schedule (ADOS; Lord et al, 2000)

The ADOS is a 30-45 minute, semi-structured assessment, with 4 modules based on child language level, that includes presentation of toys and a series of presses designed to elicit social, communicative, and perseverative behaviors often seen in typically developing children or children with ASD. Behaviors are scored on a 0-3 scale with higher numbers indicating behavior more typical of children with ASD. Scores are placed in an algorithm that indicates a classification that the child is likely to fall into based on social and communication domains: autism, ASD, or non-spectrum. Module 1, 2, or 3 of the ADOS was administered to children in the experimental group by an evaluator who met competency and reliability standards for research (with the exception of two participants for whom the examiner met clinical, but not research, standards).

Individual items of relevance to this study include measures of Gesture, Pointing, Initiation of Joint Attention, and Response to Joint Attention. Gesture, scored on all three modules used, assesses the frequency and quality of participants’ presentation of gestures. Pointing, assessed only in modules 1 and 2, is scored according to frequency
and quality of participants’ initiation of a social point. Initiation and Response to Joint Attention, assessed only in modules 1 and 2, are scored according to quality of participants’ initiation of at least one joint attention episode (gazing from an object to the administrator and then back to the object) and ability to follow the administrator’s direction of gaze when prompted.

2.3.2 *Peabody Picture Vocabulary Test, 4th Edition (PPVT-IV, Dunn & Dunn, 2007)*

The PPVT-IV consists of presentation of groups of pictures to the child who must indicate the picture labeled by the examiner. It allows for calculation of a raw score, standard score, growth scale curve, and age equivalency. In addition, the manual allows for coding of correctly identified words based on word type: noun, verb, or attribute. For this study, we coded the proportion of nouns and verbs that children correctly identified. The proportion of correct attributes was not calculated because the PPVT-IV requires identification of a very small number of attributes early in the assessment. Consequently, young or language-delayed children, like our participants, are not given very many opportunities to identify attributes.

The PPVT-IV is validated with high internal and external validity and reliability for use with individuals 2.5 years and above. For this study, six children younger than age 2.5 participated because of the difficulty in finding typically developing children with low enough PPVT-IV scores to be matched with young children with ASD. For these children, standard scores could not be computed but raw scores and language age equivalencies were calculated and used to match participants. The PPVT-IV was not administered to one child with ASD based on clinical judgment that the child could not
comprehend the instructions. The vocabulary-matched participant for this child was a typically developing child who did not meet basal set on the PPVT-IV.

2.4 Design

The mixed quasi-experimental design for this study included three groups (ASD, age-match, and vocabulary-match) that each received all four levels of gestural guidance: gaze, gaze and point, gaze and touch, and gaze and manipulation. The order of presentation of gestures was randomized across children but the order of object set presentation was fixed. All participants saw the same stimuli and received the same number of exposures to object labels and gestures.

2.5 Procedure

2.5.1 Overview

Families of children diagnosed with an ASD were recruited through advertisements in Autism-focused newsletters and websites, flyers distributed by diagnostic and intervention specialists, and personal and professional contacts. Children were tested in the Northwestern University Early Learning Lab, local daycares, and participants’ homes. For control participants under seven years old and all ASD participants, parents signed informed consent forms and children gave verbal assent to participate in the study. For control participants over seven years old, parents signed informed consent forms and participants signed informed assent forms. Participants and parents were told that they could change their mind about participating at any time. All participants and accompanying siblings were given a small toy or book after each day of participation. In addition, parents of participants received $10.
The procedure was adapted from Booth and colleagues (2008). Children were allowed several minutes to interact with the experimenter in the playroom in order to familiarize them with the setting and optimize their comfort level. Then, the experimenter brought the child and parent into the study room where the child sat in front of the parent and across from the experimenter. During the experiment, the parent filled out a short questionnaire about the child’s developmental history. Parents were asked not to interact with the child or prompt him or her during the procedure.

**2.5.2 Familiarization Task**

The familiarization task was intended to acquaint the child with the procedure and ensure that the child was comfortable following instructions from the experimenter. In addition, completion of this task demonstrated the child’s ability to understand and respond to a comprehension task similar to the experimental task. The experimenter presented three familiar objects and allowed the child to play with them. After approximately 30 seconds, the experimenter placed the objects in a line and asked the child to give one of the objects by name. When the child gave the correct object, the experimenter praised the child. When the child gave an incorrect object, the experimenter directed the child to the correct object and repeated the comprehension test. The experimenter then repeated the procedure with another set of familiar stimuli. All participants correctly identified the target object on either the first or second trial, indicating that they understood the task.

**2.5.3 Teaching**

After the familiarization task, the experimenter placed two objects in front of her and out of the child’s reach. The target object was placed on the right and the foil on the
left. After ensuring that the child was looking toward the experimenter, objects, or the
table, the experimenter gazed at the target object alone or accompanied her gaze with a
point, touch, or manipulation of the object while using the novel object label three times
within a set of tag lines (e.g., “Look, it’s a mip! What a cool mip! I like that mip!”).
The same set of objects was always associated with the mip, and the mip set was always
presented first, but the order of the accompanying gesture was randomized. In the gaze
condition, the experimenter turned her head slightly and look at the object during labeling
while keeping her hands at her sides. In the point condition, the experimenter maintained
the same position as in the gaze condition but also pointed her right index finger at the
target object, keeping the finger approximately 3-5 inches away from the object. In the
touch condition, the experimenter maintained the same position as in the point condition
with one exception: she lightly touched the object with her index finger. In the
manipulate condition, the experimenter gazed at the object and while holding the object
approximately 1-3 inches off the table, moved it back and forth slightly.

After the first part of the teaching episode, the experimenter pushed both objects
toward the child and allowed the child to play with them for approximately 30 seconds.
The experimenter retrieved the objects and again placed the target object in front of her
and to the right and the foil object in front of her and to the left. She then repeated the
teaching procedure previously described, using similar taglines (e.g., “Remember, this is
a mip. What a cool mip! I like that mip.”).

2.5.4 Comprehension Testing

Upon completion of the teaching episodes, the child was given the target object
and foil, along with a novel foil, to play with. The free play sections were intended to
reduce novelty effects by allowing children time to interact with the objects before choosing one.

After about 30 seconds or when the child stopped interacting with the toys, the experimenter placed the objects in a line, randomizing the location of the target object. The experimenter then asked, “Can you give me the mip?”. If the child indicated an object by pointing to, touching, or handing it to the experimenter, the experimenter took the object and responded with an encouraging remark (e.g., “Great” or “Thank you”). The experimenter gave the same response regardless of whether the child gave the correct object. When a child did not indicate an object or indicated multiple objects, the experimenter picked up all the objects, placed them back in the line in the same order, and said, “Let’s try again. Can you give me the mip?”. When a child still did not indicate an object, the experimenter gave a neutral response (e.g., “Okay”) and cleared the objects off the table.

2.5.5 Repetition of Procedure for Remaining Objects and Gestures

Teaching, production testing, and comprehension testing were repeated for the remaining sets of objects with a different, randomly assigned gesture for each target. All phrases used with the first set of objects (e.g., “Look! It’s a mip. What a cool mip! I like that mip!”) were repeated with the new object labels. Target location during teaching alternated across objects with objects one (mip) and three (dax) being on the right side and objects two (koob) and four (teeg) being on the left side. Target location during comprehension testing was randomized.
2.5.6 Follow-Up Testing

Participants were scheduled for a brief follow-up session between 5 and 7 days after the initial session. During the follow-up, participants were presented with the familiarization and comprehension tasks as described above. Participants in the ASD group were also administered an ADOS assessment during the follow-up session.

2.5.7 Coding

Sessions were video recorded to allow for post-experiment coding.

2.5.7.1 Comprehension Testing

Participants’ choices during the comprehension task were coded for word-mapping accuracy.

2.5.7.2 Eye Gaze

Digitized sessions were coded using Noldus Observer software. For teaching episodes, coders indicated where the child was looking during each frame (every 33 ms). If the child was looking anywhere on the experimenter’s face, the coder indicated “experimenter”; if the child was looking at the target object or one or more of the foil objects, the coder indicated “target” or “foil” respectively; if the child was looking at both the target and the foil, the coder indicated “multiple objects”; if the child was looking anywhere not already mentioned, the coder indicated “away”; and if the coder could not decipher where the child was looking, she indicated “unclear.” Approximately 20% of videos were independently coded by a second coder, and reliability was very high (mean \( \kappa = .83 \)).
2.5.7.3 Random Factors

Several extraneous variables were coded for inclusion as random factors in analyses. Variables coded included the order of the gesture presentation and the position of the target objects during comprehension testing.

3. RESULTS

3.1 Word-Mapping Accuracy

Of the four total trials, children with ASD correctly mapped labels to their referents on a mean of 2.80 ($SD = 1.32$) trials, while vocabulary-matched children correctly mapped on a mean of 3.40 ($SD = 1.08$) trials. This difference is trending towards significance with a medium effect size, $t(46.15) = 1.76, p = .09, d = .50$ (Levene’s correction applied). All children in the age-matched group mapped correctly on all four trials. Parametric tests were not used for comparison of children in the age-matched group to other children because of lack of variance. Figure 2 displays word-mapping accuracy for each group in each condition. Chi Square Tests of Goodness of Fit revealed accuracy significantly above chance (all $p s < .01$) for children in all groups in all conditions. Although performance was above chance for all groups, Chi Square Tests of Independence revealed that children with ASD were significantly less accurate than age-matched controls in all conditions. A Bonferroni correction was applied by dividing the alpha level of .05 by four (the number of conditions) and all comparisons remained significant except for the point condition. When compared to vocabulary-matched controls, children with ASD did not statistically differ in any condition after a Bonferroni correction was applied. See Table 1 for full set of frequency counts and Chi Square comparisons.
2. *Number of children who correctly mapped words to their referents in each condition (n = 25). Performance is above chance (indicated by the dashed line) in all conditions.*

1. **Accuracy by group.** Frequency of correct and incorrect word-mappings in each condition by group and Chi Square Test of Independence comparisons between experimental group and matched controls. * indicates comparisons that are statistically significant using an alpha level of .0125 (Bonferroni corrected alpha).

<table>
<thead>
<tr>
<th>Group</th>
<th>Correct</th>
<th>Incorrect</th>
<th>X²</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gaze</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>14</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Matches</td>
<td>25</td>
<td>0</td>
<td>14.10</td>
<td>&lt; .001*</td>
</tr>
<tr>
<td>Vocab Matches</td>
<td>20</td>
<td>5</td>
<td>3.31</td>
<td>.06</td>
</tr>
<tr>
<td><strong>Point</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>19</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Matches</td>
<td>25</td>
<td>0</td>
<td>6.82</td>
<td>.01*</td>
</tr>
<tr>
<td>Vocab Matches</td>
<td>20</td>
<td>5</td>
<td>.12</td>
<td>.50</td>
</tr>
<tr>
<td><strong>Touch</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>20</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Matches</td>
<td>25</td>
<td>0</td>
<td>5.56</td>
<td>.03</td>
</tr>
<tr>
<td>Vocab Matches</td>
<td>22</td>
<td>3</td>
<td>.60</td>
<td>.35</td>
</tr>
<tr>
<td><strong>Manipulate</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experimental</td>
<td>17</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age Matches</td>
<td>25</td>
<td>0</td>
<td>9.52</td>
<td>.002*</td>
</tr>
<tr>
<td>Vocab Matches</td>
<td>23</td>
<td>2</td>
<td>4.50</td>
<td>.04</td>
</tr>
</tbody>
</table>
In order to examine developmental effects, children with ASD were stratified into two age groups. The older ASD group \((n = 10)\), which consisted of school-aged children, ages 6 years, 7 months – 9 years, 10 months \((M = 8.01, SD = 1.05)\), performed near ceiling in all four gesture conditions. The younger ASD group \((n = 15)\), which consisted of pre-school aged children, ages 2 years, 11 months – 6 years, 5 months \((M = 4.76, SD = 1.13)\), performed above chance in the manipulate, touch, and point conditions, but exactly at chance in the gaze condition. Vocabulary-matched children in both age groups performed well above chance and equally well in all conditions. See Table 2 for comparisons of accuracy in each condition within each age group to chance, using Chi Square Tests of Goodness of Fit. Age-matched children were not used for examination of developmental effects because accuracy was perfect in all conditions for all children in this group. Figure 3 displays accuracy comparisons by group (ASD and vocabulary-matched controls) for each age group.
2. Accuracy by group and age group in each condition. Number of children who correctly and incorrectly mapped words to their referents in each condition by age group. Performance was significantly above chance in all conditions except the gaze condition for the young ASD group.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Correct</th>
<th>Incorrect</th>
<th>$X^2$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experimental Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older Group ($n = 10$, age range 6;7 – 9;10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaze</td>
<td>9</td>
<td>1</td>
<td>14.70</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Point</td>
<td>9</td>
<td>1</td>
<td>14.70</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Touch</td>
<td>9</td>
<td>1</td>
<td>14.70</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Manipulate</td>
<td>8</td>
<td>2</td>
<td>9.99</td>
<td>.002**</td>
</tr>
<tr>
<td>Younger Group ($n = 15$, age range 2;11 – 6;05)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaze</td>
<td>5</td>
<td>10</td>
<td>0.001</td>
<td>.98</td>
</tr>
<tr>
<td>Point</td>
<td>10</td>
<td>5</td>
<td>7.69</td>
<td>.006**</td>
</tr>
<tr>
<td>Touch</td>
<td>11</td>
<td>4</td>
<td>11.04</td>
<td>.001**</td>
</tr>
<tr>
<td>Manipulate</td>
<td>9</td>
<td>6</td>
<td>4.95</td>
<td>.03*</td>
</tr>
<tr>
<td>Vocabulary-Matched Group</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matched with Older Experimental Group ($n = 10$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaze</td>
<td>10</td>
<td>0</td>
<td>20.30</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Point</td>
<td>9</td>
<td>1</td>
<td>14.70</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Touch</td>
<td>10</td>
<td>0</td>
<td>20.30</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Manipulate</td>
<td>10</td>
<td>0</td>
<td>20.30</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Matched with Younger Experimental Group ($n = 15$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaze</td>
<td>12</td>
<td>3</td>
<td>14.99</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Point</td>
<td>12</td>
<td>3</td>
<td>14.99</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Touch</td>
<td>13</td>
<td>2</td>
<td>19.54</td>
<td>&lt; .001**</td>
</tr>
<tr>
<td>Manipulate</td>
<td>14</td>
<td>1</td>
<td>24.70</td>
<td>&lt; .001**</td>
</tr>
</tbody>
</table>
3. Proportion of children who accurately mapped words to their referents in each condition by diagnostic group and age group. The dashed line denotes chance performance (33%). Young ASD children (ages 2;11 – 6;05) performed exactly at chance in the gaze condition and above chance in all other conditions. Older ASD children (ages 6;06 – 10;00) mapped well above chance in all conditions. Typically developing children in both age groups mapped well above chance in all conditions.

3.2 Children’s Direction of Gaze During Teaching Episodes

In addition to examining developmental effects, we sought to determine how children’s direction of gaze during teaching episodes differed as a function of group and whether these differences predicted performance on our task. We calculated the proportion of time that children spent looking at the experimenter or target by total time in the teaching episode. Children with ASD spent a mean of 17.50% ($SD = 20.20$) of the teaching episode looking at the experimenter and 55.63% of the time looking at the target ($SD = 25.73$). Age matched controls spent a mean of 31.66% ($SD = 21.31$) of the teaching episode looking at the experimenter and 59.97% ($SD = 21.94$) looking at the target. Vocabulary matched controls spent a mean of 30.96% ($SD = 23.16$) of the
teaching episode looking at the experimenter and 54.08% (SD = 24.37) looking at the target. See figure 4 for proportion of time spent looking at the target and experimenter for each group. Differences were analyzed with mixed effects models, using the lme4 package (Bates, Maechler, & Bolker, 2011) in R statistical software (R development core team, 2012). With this model, likelihood ratio tests reveal whether a fixed or random factor significantly contributes to the model’s goodness of fit (Baayen, Davidson, & Bates, 2008). Mixed effects modeling was chosen for analyses of data that included crossed random effects, such as variability across participants, stimuli, and order of presentation of stimuli. By accounting for much of the error resulting from these random effects, this model allows for more powerful analysis and is particularly robust against violations of normality and presence of outliers. Because some children participated as both age matches and vocabulary matches, the two control groups could not be analyzed in the same model without violating the assumption of independence of observations. Therefore, each control group was separately compared to the ASD group. We first tested whether including group as a predictor variable improved the model’s ability to predict time spent looking at the experimenter and target. The mixed effects model indicated that children with ASD spent significantly less time looking at the experimenter than children matched on age ($a = .35, b_{GROUP} = -.14, t = -3.11, p < .01$) and on vocabulary ($a = .24, b_{GROUP} = .14, t = 2.62, p = .01$). However, groups did not

1 The following models were compared to analyze differences in direction of gaze based on group (ASD vs. Age Match and ASD vs. Vocab Match):

lmer(Tar/ExpProp~Age+Gesture+(1|Subject)+(1|Trial)+(1|CompLoc))
lmer(Tar/ExpProp~Age+Gesture+Group+(1|Subject)+(1|Trial)+(1|CompLoc))
significantly differ in time spent looking at the target (ASD and Age, \(a = .48, b_{GROUP} = -.04, t = -.99, p = .32\); ASD and Vocab, \(a = .32, b_{GROUP} = .03, t = .59, p = .55\)).

4. Note. Children were coded as looking at the target, experimenter, foil object(s), or away. Therefore, sum proportions for looks to the target and experimenter do not add up to 100%.

We also tested whether direction of gaze significantly contributed to the prediction of accurate word mapping for children with ASD and vocabulary-matched controls (predictions could not be made for age-matched children due to lack of variance). Figure 5 displays children’s direction of gaze over the course of teaching episodes for children who correctly and incorrectly mapped labels to referents. For both ASD children and vocabulary-matched controls, a mixed effects binomial model

\[ lmer(Choice\sim Age+Gesture+(1|Subject)+(1|Trial)+(1|CompLoc), family = \text{‘binomial’}, REML=F) \]
\[ lmer(Choice\sim Age+Gesture+Exp/Tar+(1|Subject)+(1|Trial)+(1|CompLoc), family = \text{‘binomial’}, REML=F) \]

---

*The following models were compared to analyze whether direction of gaze predicted correct word mapping (for each group):*
indicated that looking at the experimenter during teaching episodes was predictive of word-mapping accuracy (ASD, \(a = -5.67, b_{Exp\text{Looks}} = 5.66, z = 2.24, p = .01\); Vocab matches, \(a = -9.80, b_{Exp\text{Looks}} = 9.58, z = 2.40, p = .01\)), but looking at the target was not (ASD, \(a = -3.99, b_{Tar\text{Looks}} = -1.55, z = -1.06, p = .31\); Vocab matches, \(a = -3.76, b_{Tar\text{Looks}} = 1.01, z = .50, p = .64\)).

5. ASD and vocabulary-matched participants’ direction of gaze over the course of teaching episodes after which they correctly or incorrectly mapped the object label to its referent. The time series (in seconds) is centered around the onset of the first object label (indicated by the dotted line). Children who mapped correctly looked at the experimenter significantly more than those who mapped incorrectly. There was no significant difference in time spent looking at the target for correct and incorrect children.

### 3.3 PPVT-IV and ADOS as Predictors of Word-Mapping Accuracy

Next, we examined whether receptive lexical knowledge as measured by the PPVT-IV predicted accuracy on our task. A bivariate regression revealed that language
age equivalency, computed from PPVT-IV raw scores, was a significant predictor of
number of trials in which children with ASD and vocabulary-matched controls correctly
mapped words to their referents, $F(1, 49) = 14.09, p < .01, R^2_{Adj} = .21, \ a = 2.14, b_{PPVT} = .02, SE(b) < .01$. Using the PPVT-IV manual, we then calculated the proportion of nouns
and verbs that children correctly identified out of the total administered on the PPVT-IV.
Children with ASD correctly identified a mean of 64.04% ($SD = 18.93$) of nouns and
51.97% ($SD = 30.02$) of verbs presented to them on the PPVT-IV and correctly identified
the target item on our task on a mean of 2.80 ($SD = 1.32$) out of 4 trials. Typically
developing children correctly identified a mean of 65.12% ($SD = 10.11$) of nouns and
60.13% ($SD = 26.83$) of verbs and correctly identified our target in a mean of 3.40 ($SD = 1.01$) trials. A multiple linear regression revealed a significant effect of proportion of
correct verbs, but not proportion of correct nouns, on number of trials correct for children
with ASD, $F(2,23) = 4.23, p = .03, R^2_{Adj} = .22, a = 1.39, b_{verbs} = 2.14, SE(b)_{verbs} = .89,
b_{nouns} = .59, SE(b) = 1.41$. For vocabulary-matched children, the opposite was true;
proportion of correct nouns, but not proportion of correct verbs, significantly predicted
trials correct, $F(2, 24) = 6.47, p < .01, R^2_{Adj} = .31, a = -.70, b_{verbs} = .20, SE(b)_{verbs} = .98,
b_{nouns} = 6.10, SE(b) = 2.60$. Scatterplots of trials correct predicted by percentage of nouns
and verbs correctly identified for each group are shown in figures 6 and 7.
6. For children with ASD, proportion of verbs correctly identified significantly predicted the number of trials in which children correctly mapped words to their referents. Proportion of nouns correctly identified was not significantly related to performance.

7. For typically developing children matched on overall vocabulary level with children with ASD, accuracy on our task was high despite variable performance on verb identification. However, ability to correctly identify nouns was significantly predictive of word-mapping accuracy.
For children with ASD, we analyzed whether joint attention abilities and initiation of social gestures, both measured by the ADOS, predicted performance. Children in the experimental group were given an ADOS classification of Autism ($n = 17$) or ASD ($n = 8$). An Independent Samples $t$-test revealed no significant difference in trials correct between children classified as Autism ($M = 2.65, SD = 1.37$) and those classified as ASD ($M = 3.13, SD = 1.25$), $t(23) = .84, p = .41$. Children were scored on each of four measures (Initiation of Joint Attention, Response to Joint Attention, Initiation of Gesture, Initiation of Pointing) on a scale ranging from 0 to 3. A multiple linear regression revealed no significant effect of individual scores on any of the four measures on the ADOS on the number of trials in which children correctly identified the target in our task, $F(4, 18) = .44, p = .78, R^2 = .11$. In an effort to increase statistical power, the four scores were then collapsed into one social gesture score, which ranged from 0 to 12. A bivariate regression revealed no significant effect of the social gesture score on trials correct with a very low effect size, $F(1, 18) = .19, p = .67, R^2 = .01$.

### 3.4 Accuracy at Follow-Up

Twenty-two children with ASD and 24 typically developing children attended the follow-up session. Age-matched and vocabulary-matched children were collapsed into one group for comparisons due to high attrition in the control groups. Word-mapping accuracy for children with ASD who attended both sessions dropped from a mean of 3.05 ($SD = 1.24$) trials correct at the first session to 2.62 ($SD = 1.24$) trials correct at follow-up. For typically developing children who attended both sessions, accuracy dropped from a mean of 3.74 ($SD = .86$) trials correct to a mean of 3.35 ($SD = .93$) trials correct. Performance in each condition is shown in Table 3.
3. Accuracy at Follow-Up. Frequency of children in each group who correctly and incorrectly identified the target object at follow-up. Follow-up data were collected for 22 children with ASD and 24 controls. One control is missing data for the gaze and point conditions and one child with ASD is missing data for the manipulate condition.

<table>
<thead>
<tr>
<th>Condition</th>
<th>ASD Children</th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaze</td>
<td>12</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Point</td>
<td>16</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Touch</td>
<td>15</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Manipulate</td>
<td>13</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Typically Developing Children</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaze</td>
<td>16</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Point</td>
<td>21</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Touch</td>
<td>20</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Manipulate</td>
<td>22</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

Mixed effects binomial models revealed main effects for time, group, and gesture.

Across groups and conditions, performance was better during the initial session than the follow up, \( a = -2.52, b = .94, z = 2.82, p < .01 \). Across groups and time, performance was better in point, touch, and manipulate conditions than the gaze condition, \( a = -2.52, b_{\text{point}} = 1.10, z_{\text{point}} = 2.51, b_{\text{touch}} = 1.13, z_{\text{touch}} = 2.57, b_{\text{manip}} = 1.03, z_{\text{manip}} = 2.35, p = .03 \). Across time and condition, typically developing children performed better than children with ASD, \( a = -2.52, b = 2.07, z = 3.77, p < .01 \).

We examined interaction effects through individual planned comparisons. For typically developing children, mixed effects models revealed significantly better

---

3 lmer(Choice~Age+Gesture+Session+Group+(1|Subject)+(1|Trial), family = ‘binomial’, REML=F) was compared to each of the following:
lmer(Choice~Age+Gesture+Session+(1|Subject)+(1|Trial), family = ‘binomial’, REML=F)
lmer(Choice~Age+Gesture+Group+(1|Subject)+(1|Trial), family = ‘binomial’, REML=F)
lmer(Choice~Age+Session+Group+(1|Subject)+(1|Trial), family = ‘binomial’, REML=F)

4 For each group, the following models were compared with and without “Session” and with and without “Gesturegroup”:
lmer(Choice~Age+Gesturegroup+Session+(1|Subject)+(1|Trial), family = ‘binomial’, REML=F)
performance in the initial session than the follow up \((a = -1.30, b = 1.48, z = 2.28, p = .01)\) and significantly better performance in the manual conditions (collapsed across point, touch, and manipulate) than in the gaze condition \((a = -1.30, b = 1.41, z = 2.35, p = .02)\). However, there was a time by gesture interaction effect such that typically developing children performed equally well across gestures during the initial session \((a = -19.17, b = 1.43, z = .89, p = .40)\) but performed better in the manual conditions than the gaze condition during the follow up \((a = .02, b = 1.66, z = 2.37, p = .02)\)\(^5\). Using the same model comparisons as for typically developing children (see footnotes 4-5), we examined differences in performance based on time (initial session compared to follow up) and gesture (manual gestures compared to gaze) for children with ASD. There was a trend towards a significant effect for time \((a = -1.87, b = .66, z = 1.75, p = .09)\) and children performed significantly better in the manual conditions than in the gaze condition \((a = -1.87, b = .93, z = 2.23, p = .03)\). In contrast with the typically developing children, children with ASD did not show a time by gesture interaction effect on performance. Children with ASD showed a trend towards better performance in manual conditions than the gaze condition at both the initial session \((a = -2.99, b = 1.31, z = 1.81, p = .08)\) and follow up \((a = -.80, b = .83, z = 1.49, p = .16)\). Performance at initial visit and follow up is shown in figure 8.

\(^5\) For each session (initial and follow up), the following models were compared:
- \texttt{lmer(Choice~Age+(1|Subject)+(1|Trial), family = ‘binomial’, REML=F)}
- \texttt{lmer(Choice~Age+Gesturegroup+(1|Subject)+(1|Trial), family = ‘binomial’, REML=F)}
8. Initial and follow up performance. Number of children from each group who correctly mapped labels to their referents at the initial visit and follow up.

4. DISCUSSION

4.1 Findings and Implications

We taught children with ASD and age- and vocabulary-matched controls novel words paired with gaze alone or gaze accompanied by a point, touch, or manipulation of the target object, and then tested word-mapping comprehension immediately and at a follow-up session 5-7 days later. Direction of eye gaze during teaching episodes was coded to provide information regarding word-learning strategies. Previously acquired lexical knowledge was also assessed and used as a predictor for word learning in our task. Surprisingly, as a whole, children with ASD performed well above chance in all conditions and nearly as well as vocabulary-matched controls, suggesting that the majority of participants with ASD had acquired strategies to use social information in word mapping. Interestingly, differences between children with ASD and vocabulary-matched controls, though not significant, were most pronounced in the gaze and manipulate conditions. Group differences in the gaze condition were expected based on our hypothesis that children with ASD would not be able to extract social information as well as typically developing children without additional cues to direct their attention.
However, we expected that the manipulate condition, which contains the most attention-grabbing cues, would facilitate word learning for children with ASD as well as for typically developing children. A possible explanation for the nearly significant difference in performance between children with and without ASD in the manipulate condition is that children with ASD are more likely to orient their attention toward non-social than social stimuli (Dawson et al., 1998). Perhaps the enhanced salience of the target object in the manipulate condition actually led children with ASD to focus too much attention on the object itself thereby ignoring the social context of the word learning task. If children with ASD were distracted by the movement of the object, they may not have attended to the presentation of the object label or social cues, such as the speaker’s direction of gaze, that provided information about the referential intention of the speaker. Therefore, although we hypothesized that object salience would enhance word learning for children with ASD, in this case, it may have distracted children from the social aspects of the task.

To examine developmental effects on word learning, children with ASD were stratified into pre-school and school-aged groups, and we found striking differences in accuracy based on age and gesture. Children with ASD older than 6.5 years were nearly perfect in all conditions, performing similarly to controls. For younger children with ASD, accuracy was lower than for older children but still well above chance in conditions in which the experimenter’s gaze was accompanied by a point, a touch, or object manipulation. However, when labeling was paired only with gaze, accuracy was exactly at chance for the young ASD group. It seems then that children with ASD use gaze following as a strategy for pairing new words with their referents like typically
developing children do, but they begin to do so much later in development. Although this seems to support a hypothesis of delayed, rather than different, development of children with ASD compared to typically developing children, it is not entirely clear from performance data alone whether the addition of pointing enhances performance in young children with ASD because it clarifies the social relationship between the speaker and the referent object or because it simply draws attention toward the referent object. Research from Booth and colleagues (2008) indicates that for typically developing children, the information that pointing provides about the social context is important for word learning even when controlling for the effect of guided attention. However, it is possible that for children with ASD, pointing is effective for its effect on children’s direction of attention rather than for its indication of referential intent.

To provide additional evidence for how gestural guidance contributes to performance, we examined the role of children’s direction of gaze in accurately pairing novel words and targets. Children with ASD are generally less likely than typically developing children to look at faces (Klin, Jones, Schultz, Volkmar, & Cohen, 2002b; Trepagnier, Sebrechts, & Petersen, 2002), so it was not surprising that children with ASD looked at the experimenter less than controls during our teaching episodes. However, the effect that looking at the experimenter had on word-mapping accuracy was enlightening. Looking at the experimenter appeared to be an effective strategy used by children with ASD, as well as typically developing children, for accurately learning words in our task. This indicates that the ability to gather social information from a speaker’s face leads to enhanced word-mapping performance, even with children with ASD and even when other cues, such as manual gestures, are available. In contrast, time spent looking at the target
did not predict word-learning accuracy, indicating that purely associative processes, such as simply looking at the object while hearing its label, were not effective for learning new words. Furthermore, children with ASD looked at the target just as often as typically developing children, indicating that difficulty with word-to-object pairing was not primarily due to failure in attending to the object, but more likely, failure to understand that the object was the target of referential intent. However, children in all groups spent substantially more time during teaching episodes looking at the target than the experimenter. Furthermore, previous research (Akechi et al, 2011; Preissler & Carey, 2005) indicates that attending to the object referenced by the speaker’s gaze is important for learning words for both typically developing and ASD children. Therefore, it is clear that associative processes play some role; however, as long as children are attending to the teaching process, it seems that the ability to gather social information from the speaker’s face is the crucial element in correctly pairing new words to objects. As seen with previous research (Parish-Morris et al, 2007), word learning in children with ASD cannot be explained exclusively by either attentional or intentional processes, but intentional understanding appears to play a special role. The ability to link novel words to objects appears highly dependent on implementation of social gaze strategies, and our younger ASD participants failed to implement this word-learning strategy, which is used with ease by typically developing toddlers. This failure might contribute to their delay in acquiring vocabulary.

In order to gain some understanding of the role that lexical and syntactic knowledge play in the word-learning strategies of children with ASD, we looked at the predictive value of previously-acquired knowledge of specific types of words, namely,
nouns and verbs, on word learning in our task. For typically developing children matched on overall vocabulary with children with ASD, knowledge of nouns, but not knowledge of verbs, was predictive of word-mapping accuracy in our task. This makes sense given that it was a noun-learning task. However, for children with ASD, we saw the opposite effect: knowledge of verbs, not knowledge of nouns, predicted accuracy on the noun-learning task. It is important to note that children were matched on PPVT-IV scores, meaning that there were not differences in general lexical acquisition, only in type of words acquired. For the typically developing children in our sample, acquisition of a large pool of nouns indicated effective strategies for mapping objects to their referents in our task. However, for our participants with ASD, previous acquisition of nouns did not indicate development of effective noun-learning strategies. This seems to indicate that, although implementation of social strategies used by typical word learners may simply occur on a delayed schedule for children with ASD, typical word learners and children with ASD markedly differ in the type of lexical knowledge acquired before they are able to strategically learn words in a novel setting.

One possible explanation of this pattern of results rests on the idea that the ability of children with ASD to take advantage of socio-pragmatic cues (like eye-gaze) in learning novel words emerges over a different developmental trajectory than in typically developing children. Although we can only speculate about the precise nature of this distinctive developmental trajectory, the current data, viewed in the context of the broader literature, suggests one strong possibility. As previously mentioned, typically developing children learn first words through word-to-world pairing, often using social strategies like gaze following. After acquiring a pool of concrete words, along with
considerable linguistic experience, these children are able to learn abstract words through structure-to-world pairing. Concrete nouns acquired through social observations serve as bootstraps for acquisition of abstract words through syntactic understanding. For children with ASD, the process may progress differently. Unable to initially implement social strategies, children with ASD instead may begin by acquiring concrete words through strictly associative processes. In fact, many children with ASD receive early intervention services in Applied Behavior Analysis in which they are taught object labels and other words through repetitive association (Sundberg & Partington, 2010). Although often effective, these processes are certainly less efficient than those supported by an appreciation of social intent, so words are acquired at a relatively slow pace for the children who use them. Nevertheless, children with ASD can eventually acquire a large pool of concrete words through this direct associative learning mechanism. If, as research suggests (Naigles et al, 2011), syntactic understanding is indeed intact for children with ASD, those words might then serve as the scaffold for structure-to-world pairing. That is, using acquired concrete words and linguistic experience, children with ASD may be able to acquire abstract words in a manner similar to typically developing children, even though acquisition of concrete words initially occurred through a different process. If this is the case, word acquisition for children with ASD will be much more efficient once they are able to implement structure-to-world pairing processes, allowing for more room for observations, perhaps even social observations, during linguistic experiences. Based on our data, the ability of children with ASD to use social information such as speaker direction of gaze to learn new nouns occurs after the acquisition of abstract words like verbs. Perhaps for these children, acquisition of abstract words serves as the scaffold for
the ability to make social observations. That is, after achieving a certain level of linguistic sophistication, children with ASD may begin to notice common social communicative situations, such as speakers often looking at objects while talking about them. If this is the case, they are engaging in “backwards bootstrapping” by acquiring words first and using them to understand social information rather than using social information to acquire words.

This theory is not without limitations. Most importantly, the idea that social observations are bootstrapped on abstract lexical knowledge was a post-hoc theory developed as an explanation for our findings rather than an ad-hoc hypothesis that was tested. Furthermore, we are basing this theory on a limited set of data. A relationship between verb knowledge and gaze following in a specific word-learning task does not necessarily mean that syntactic bootstrapping catapults social observation in children with ASD. However, whether social inferential abilities develop as a result of complex lexical comprehension or not, our data certainly suggest a relationship between the two. Both attention to the speaker’s face and comprehension of verbs predicted word learning for children with ASD, but neither attention to the target nor comprehension of nouns did so. This double association/dissociation provides solid evidence for a relationship between complex lexical comprehension and social inferences in this population. It is also possible that verbs specifically, rather than abstract words in general, are important. Many children with autism have difficulty noticing and interpreting the actions of others (Vivanti et al., 2011). Perhaps once the ability to notice others’ actions develops, children with ASD begin to accumulate verb knowledge more rapidly, which then
jumpstarts the process of syntactic bootstrapping (portions reprinted from Patrick, Hurewitz, & Booth, 2013).

Performance on the joint attention and gesture measures of the ADOS was not predictive of word learning abilities for children with ASD in our sample. This may be due to a lack of power because of such a small sample size, but there did not appear to be a trend toward significant correlations either. The intended use of the ADOS is for diagnostic purposes, not prognosis or assessment of abilities. It seems likely then that the ADOS cannot and should not be used to predict whether children with ASD exhibit the early skills that will support the ability to learn new words. Although it is clear that joint attention and early non-verbal language skills are associated with word learning, the assessment of these abilities on the ADOS may only be sensitive enough to determine whether a child falls on the autism spectrum, not whether a child will be able to learn new words. For instance, the response to joint attention measure on the ADOS requires a child to simply look in the direction of a speaker’s gaze without making any further inferences about the meaning of that gaze. However, Gliga and colleagues (2012) found that the ability to follow a speaker’s gaze was not predictive of word-learning abilities in a population at-risk for autism. Our data support these findings. Furthermore, McGregor and Bean (2012) found no relationship between severity of autism symptoms as measured by the ADOS and the ability to engage in categorical word learning, indicating that either autism severity is not predictive of many core language abilities or the ADOS is not the proper tool for predicting language development of children with ASD.

Following a 5-7 day delay, children were again tested for word-mapping comprehension. We found main effects for group, time, and gesture: typically
developing children performed better across time periods than children with ASD, all children performed better during the first session than the follow up, and across time and children performance was better in conditions that included a manual gesture than the gaze alone condition. Interestingly, typically developing children learned words equally well in all conditions but retained better in conditions in which labeling was accompanied by a manual gesture than gaze alone. However, for children with ASD, there was a trend towards better performance in manual conditions than the gaze condition at both the initial session and follow up, meaning that retention did not differ as a function of gesture condition. Interestingly, the word learning profile that we found in children with ASD is similar to that found by Booth and colleagues (2008) in typically developing two-and-a-half year olds. In that study, toddlers performed well in all conditions but significantly better in point, touch, and manipulate conditions than the gaze alone condition. Performance was similar at follow up. One difference between our children with ASD and the Booth and colleagues’ typically developing toddlers was in ability to retain across time. The typically developing toddlers (Booth et al., 2008) remembered words equally well in the initial and follow up sessions. In our study, children with ASD declined in performance at follow up. Likewise, typically developing children in our study declined in performance at follow up, particularly in the gaze condition. It is possible that the additional attention-grabbing cues associated with the point, touch, and manipulate conditions enhanced retrieval of object-word associations for typically developing children, perhaps because of multiple modalities available during encoding. For children with ASD, social gaze following and interpretation of speaker intention come much later and may develop in conjunction with other more complex word-learning skills. As a
result, once children with ASD are able to interpret social gaze to learn new words, they may already have developed the skills necessary for long-term storage of lexical knowledge. Taking Booth and colleagues (2008) into account, it is possible that young typically developing children perform more poorly in the gaze condition because of difficulty interpreting the social context of gaze (i.e., gaze is not as clearly a “social” act as pointing). As typically developing children get older, our data suggest that they are able to interpret the social context of speaker gaze equally as well as when that gaze is accompanied by point, touch, or manipulation. However, they may not retain words learned in the gaze alone condition as well as other words because the lack of attention-grabbing features make the object less memorable. We may speculate that children with ASD are more likely than typically developing children to learn words by pairing labels with objects in their own attentional focus (Hennon, 2003; Preissler & Carey, 2005), which would suggest that speaker gaze is guiding their attention toward the object more so than providing a cue of social intention. If this is the case, decreased salience of the object in the gaze condition would explain differences in performance between the gaze alone and manual gesture conditions and we would not expect these differences to become more pronounced at follow up. Although highly speculative at this point, this interpretation further supports a hypothesis of qualitatively different word-learning processes of children with and without ASD.

4.2 Limitations

The findings from this study provide some evidence for ways in which word-learning strategies of children with ASD are not simply delayed but actually qualitatively different than word-learning strategies of typically developing children. However, there
are several limitations to consider. First, much of our data violated assumptions of parametric tests. For instance, many of the distributions of data for the ASD group were not normally distributed. As a result, we sometimes had to rely on non-parametric tests with lower power. However, we also used analyses, such as mixed effects modeling, which were highly robust against violations of the assumption of normality. Data analysis was further complicated because performance data from age-matched controls showed no variance, making it difficult to use these data in comparisons. We can, however, conclude that children with ASD are performing below expectations based on age, given the perfect accuracy of the age-matched group.

Additionally, these results should be generalized with caution. Our sample contained primarily middle to upper class children all from the Chicago area. We cannot necessarily generalize these results to children from lower-income families because word-learning environments drastically differ between low and high SES contexts (Hoff, 2006). Furthermore, although the teaching procedure we used was similar to those used in previous research (Baldwin, 1993a; Booth et al, 2008), it may lack ecological validity because naturalistic noun learning generally involves adults saying words in the presence of their referential objects (Gleitman et al, 2005), rather than directing the child to the object while stating its name. However, young children are directly shown novel objects while they are being labeled fairly often, so results should be generalizable.

Results of word-mapping accuracy at follow-up should particularly be considered with caution. Attrition within the typically developing group was very high, particularly for younger children who were recruited from daycares. With our youngest typically developing children differentially excluded, interpretation of follow up results in the
context of development is difficult. However, findings from a similar study that included 28-31-month old children (Booth et al., 2008) allowed for some examination of our follow up results with consideration for performance of young children.

4.3 Future Research

We have only begun to explore the time-based data available from this study. For instance, children’s direction of gaze can be analyzed for joint attention episodes and the possible contribution of response to joint attention as a word-learning tool in this population. Although the correlations between ADOS measures of joint attention and word learning were not significant, more sensitive measures, such as children’s response to joint attention during the teaching episodes in our task (i.e., whether the child looks from the experimenter’s face to the object or from the object to the experimenter’s face) could help predict word-learning abilities.

In addition, analyses of the pattern of children’s direction of gaze during each condition may allow us to see how reliant on social information (i.e. how often children look at the experimenter’s face) children are at different levels of gestural guidance. Specifically, it would be interesting to examine whether children with ASD who correctly map words to their referents are exhibiting the same gaze patterns as typically developing children and whether enhancement in performance provided by the addition of pointing remains when controlling for children’s attention as it does with typically developing children (Booth et al, 2008). Furthermore, for children with ASD who incorrectly map, it is important to determine whether they are exhibiting similar gaze patterns in each condition as typically developing children who err (indicating a developmental delay) or
exhibiting entirely different patterns of attention (indicating qualitative and possibly more substantial differences in language development processes of children with ASD).

Finally, further examination of our data should include analyses of children’s looking patterns during the earliest section of teaching episodes and during labeling. Perhaps, looking at the target or experimenter is only important until the first label is spoken. If this is the case, initial attention (social attention and/or attention directed at the target) may be more important than sustained attention throughout a word-learning episode. If attention directed toward the target is important while the experimenter is labeling, associative pairing may play a stronger role than our initial analyses imply. Further investigation of children’s direction of gaze throughout teaching and the role that gaze plays in correctly mapping labels to referents can provide further information about the strategies that children with and without ASD implement while learning new words.


Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42. http://CRAN.R-project.org/package=lme4


