2-20-2020

An investigation of word learning in the presence of gaze: Evidence from school-age children with typical development or Autism Spectrum Disorder

Janet Y. Bang
Stanford University, janet.bang@sjsu.edu

Aparna S. Nadig
McGill University

Recommended Citation

Janet Y. Bang and Aparna S. Nadig. "An investigation of word learning in the presence of gaze: Evidence from school-age children with typical development or Autism Spectrum Disorder" *Cognitive Development* (2020). https://doi.org/10.1016/j.cogdev.2020.100847

An investigation of word learning in the presence of gaze:
Evidence from school-age children with typical development or Autism Spectrum Disorder

Janet Y. Bang PhD* [a,b,c], Aparna Nadig[a,b]
*Corresponding author

Received 4 June 2019; Received in revised form 10 December 2019; Accepted 17 January 2020

Citation

Bang, J. Y., & Nadig, A. (2020). An investigation of word learning in the presence of gaze: evidence from school-age children with typical development or Autism Spectrum Disorder. *Cognitive Development*, *54*, 100847. https://doi.org/10.1016/j.cogdev.2020.100847

Affiliations:

[a] McGill University, School of Communication Sciences and Disorders, 2001 Avenue McGill College Suite 800, Montreal, QC, H3A 1G1, Canada; [b] Centre for Research on Brain, Language & Music, 3460 de la Montagne, Montréal, QC, H3G 2A8, Canada, [c]Department of Psychology, Stanford University, 450 Jane Stanford Way, Stanford, CA, 94305, USA

Funding sources

This research was funded by doctoral awards to the first author from the Autism Research Training Program (Canadian Institutes of Health Research, Sinneave Family Foundation), and the Center for Research on Brain, Language, and Music.

Declaration of Competing Interest

None.

Acknowledgements

We thank Kristine Onishi for valuable input throughout. We would also like to thank Corentin Montiel, Milva Venditti, Verona Soliman, Hida Caliskan, Lisa Bissett, Nicole Khammar, Renuka Giles, and Nowrin Hoque for assistance with testing and data collection.
Thank you to the many organizations that helped with recruitment for the study including: The Montreal Children's Hospital, le Centre int.gr. universitaire de sant. et de services sociaux du Centre-Sud-de-l'.le-de-Montr.al (CIUSSS), Autism Speaks Canada Walk in Montreal, Summit School, Giant Steps, and Autisme Montr.al. Finally, we greatly appreciate the families who participated.

Supplementary data

Supplementary material related to this article can be found, in the online version, at https://doi.org/10.1016/j.cogdev.2020.100847

All data and code are available at https://github.com/janetybang/word_learning_refGaze and all stimuli are available at https://osf.io/dsqmn/

Abstract

Little is understood about how children attend to and learn from gaze when learning new words, and whether gaze confers any benefits beyond word mapping. We examine whether 6- to 11-year-old typically-developing children (n = 43) and children with Autism Spectrum Disorder (n = 25) attend to and learn with gaze differently from another directional cue, an arrow cue. An eye-tracker recorded children's attention to videos while they were taught novel words with a gaze cue or an arrow cue. Videos included objects when they were static or when they were manipulated to demonstrate the object's function. Word learning was measured immediately after videos and one week later. In contrast to an arrow cue, children in both groups looked longer at a gaze cue and had more contingent looks from the gaze cue to the referent. Exploratory analyses demonstrated that across both groups, children with higher versus lower social competence skills recalled more semantic features about the referent in the gaze condition. We discuss how these findings add to our theoretical understanding of how gaze supports word learning.

Keywords: word learning; typical development; autism spectrum disorder; referential gaze; language development; intention understanding

1. Introduction

Seminal studies have evaluated how children learn new words through interaction with an adult (Baldwin, 1993; Carey & Bartlett, 1978), demonstrating that minimal exposure to labels and referents can lead to word learning (fast-mapping). Though there are multiple cues that contribute to learning from live interactions (e.g., head turn, emotion), gaze is a ubiquitous cue that we can use to learn about the world. For example, children can learn new words by hearing a label and following another's gaze, or referential gaze, to the referent.

Yet little is understood about *why* children attend to and learn new words in the presence of gaze. In the field of infant cognition, Csibra and Gergely propose in their Natural Pedagogy (2009) account that gaze is a specific communicative cue evolutionarily adapted for efficient learning. That is, the ostensive nature of gaze conveys to the learner that important generalizable information will immediately follow. On this view, children may treat gaze as an intentional and communicative cue, thereby facilitating efficient learning of a new word. Word learning researchers have also interpreted gaze as an intentional cue (Baldwin, 1993; Moore, Angelopoulos, & Bennett, 1999; Norbury, Griffiths, & Nation, 2010), although less attention has been paid to the benefits of treating gaze as intentional for word learning. Some of the support for gaze as an intentional cue comes from studies of a population with known difficulties in social communication, children with Autism Spectrum Disorder (ASD). Because children with ASD have sometimes demonstrated poorer word learning with gaze in comparison to peers without ASD, this evidence has been interpreted as a lack of intention understanding in individuals with ASD (Baron-Cohen, Baldwin, & Crowson, 1997; Norbury et al., 2010). Yet recent overwhelming evidence of intact word learning in children with ASD in the presence of gaze raises the question of why, and what exactly, they are able to learn in these situations (e.g., Bani Hani, Gonzalez-Barrero, & Nadig, 2012; Luyster & Lord, 2009; Norbury et al., 2010).

We investigate three open questions to better understand how children treat gaze during word learning, while examining differences between a gaze cue versus another directional cue, an arrow. First, do children visually attend to a word learning scene in the presence of a gaze cue any differently from an arrow cue? Second, are there any benefits to learning words with gaze versus an arrow cue? Third, do children with ASD differ from typically-developing children in their visual attention and learning with gaze versus an arrow? We examine differences between groups considering categorical differences in social abilities as measured by an ASD diagnosis, and dimensional differences as measured by a standardized measure of social competence.

1.1 The role of gaze during word learning

Currently, there is little evidence indicating that the way children treat gaze during word learning is truly unique to gaze. Nevertheless, many studies have proposed that children read gaze as an intentional cue, which thereby supports word learning (e.g., Baldwin, 1993; Hollich et al., 2000; Moore et al., 1999)[1]. Though studies have not included a control cue that similarly directs attention, they have compared learning between conditions that manipulated children's focus of attention when hearing a speaker say a novel label. One common paradigm compares a *follow-in condition*, where the speaker labels an object the child is already attending to, with a *discrepant condition*, where the speaker labels an object that is less interesting or not in the child's focus of attention, thereby requiring the child to break his/her own attention to follow the speaker's focus of attention. In both conditions learning is measured immediately after training

[1] These studies used a person's referential gaze, but because they include live interactions, inevitably gaze itself is not the only cue, for example, head movement, body posture. For consistency, we use the term referential gaze.

by asking infants to choose the just labeled object. Prior studies have shown that before the age of 2, children demonstrated learning in the *follow-in* condition relative to chance but had difficulty in the *discrepant* condition. However, from about 2 years onward, children demonstrated learning in the *discrepant* condition, and more contingent looking, or coordinated attention between the speaker and object during *discrepant* versus *follow-in* conditions (Baldwin, 1993; Hollich et al., 2000). Thus, gaze following and contingent looking during a *discrepant* condition have been considered as evidence of a pragmatic understanding of the speaker's naming intentions.

Some early studies tested alternative explanations for word learning in the presence of gaze, including the role of attentional salience by manipulating how interesting the object is (Baldwin, 1993; Baldwin et al., 1996; Moore et al., 1999). Results from these studies have demonstrated that when an object indicated by referential gaze was pitted against a more attentionally salient object, children still learned as well with or better with gaze; these findings add support for an intentional reading of gaze. However, in addition to the role of attentional salience, it is important to address another alternative explanation: how children treat gaze relative to another *directional cue.* We employ an arrow cue as a control for the direction of attention. While this comparison has been used frequently in attention research using spatial cueing paradigms with mixed findings (for reviews see Birmingham, Ristic, & Kingstone, 2012; Nation & Penny, 2008; Rombough, Barrie, & Iarocci, 2012), these results have demonstrated that an arrow cue can be an effective directional cue to compare with gaze. We compare gaze versus an arrow cue to better examine potential additive benefits of gaze, beyond its directional properties, when children use gaze throughout different stages of the word learning process: cue following, contingent looking, fast-mapping, and in-depth word learning.

1.2 In-depth word learning

To delve deeper into the ways that gaze may support word learning, our second research question tests whether there are any benefits of learning with referential gaze in contrast to another directional cue. We expand learning measures to not only include the commonly used measure of word recognition, referent selection, but also a more in-depth learning measure of descriptions (Gladfelter & Goffman, 2017; Norbury et al., 2010). Children can be successful on recognition measures with only partial knowledge of a word: knowing the label associated with the referent. Though a label-referent association is limited and incomplete knowledge of word meaning, this is the beginning of knowledge that is likely slowly built up over time (McMurray, Horst, & Samuelson, 2012; Swingley, 2010). Researchers have proposed that different stages of word learning such as immediate label-referent association, and short-term retention of this association, may emerge from different underlying mechanisms (Hartley, Bird, & Monaghan, 2019; McMurray et al., 2012). Hartley and colleagues (2019) found that while children with ASD had difficulty fast-mapping novel labels relative to typically-developing children, children with ASD demonstrated improved short-term retention in conditions with social feedback (including gaze) versus with non-social feedback (flashing light) or with no feedback; these findings indicate that cues such as gaze may support short-term retention. However, we still know little about what children learn beyond a label-referent association in typically-developing children, and even less so in children with ASD. One reason for the lack of attention to other dimensions of word learning is due to a focus on early processes of word acquisition in infancy and toddlerhood. However, word learning is not a finite process early in development, and children continue to learn words beyond these early stages, making it critical to go beyond label-referent mapping in school-age children.

Given the experimental context that introduces novel objects, we examine in-depth learning via children's recollection of what the object looks like and how it can be used, operationalized as *semantic features*. Only one prior study has examined the relation between learning from referential gaze and object descriptions (Norbury et al., 2010), finding that school-age children recalled semantic features from a fast-mapping paradigm. However, this study did not compare whether descriptions with gaze differed from another directional cue, thus it is not clear whether there are any benefits of learning new words in the presence of gaze. Benefits of learning with gaze versus another directional cue would provide additional support that an intentional reading of gaze generates efficient learning (Csibra & Gergely, 2009).

1.3 Word learning in Autism Spectrum Disorder

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder defined by impairments in social communication, and the presence of restricted, repetitive behaviors and stereotyped interests (American Psychiatric Association, 2013). Reduced spontaneous following of gaze is an early clinical marker (Leekam, Baron-Cohen, Perrett, Milders, & Brown, 1997; Loveland & Landry, 1986), and difficulties in coordinating attention are widely attested (Lord, Rutter, DiLavore, & Risi, 2002; Mundy, Sigman, & Kasari, 1990). Accordingly, much research has examined how individuals with ASD discriminate, attend to, and follow another's gaze (for reviews see (Boraston & Blakemore, 2007; Itier & Batty, 2009; Nation & Penny, 2008).

In early word learning studies, children with ASD demonstrated poorer referent selection after being taught words in the presence of gaze, relative to typically-developing children (Baron-Cohen et al., 1997; Preissler & Carey, 2005). However, participants in these two studies had severe language delays (e.g., mean chronological age of 9 years with mean language age of 2 years). ASD is a notably heterogeneous diagnosis, including children who have severe language delays to those with above average abilities (Kjelgaard & Tager-Flusberg, 2001; Wittke, Mastergeorge, Ozonoff, Rogers, & Naigles, 2017), therefore this initial work represented only a subgroup of children with ASD. Importantly, subsequent work has repeatedly demonstrated that children with ASD who have moderately-delayed to above-average language levels *can* in fact demonstrate correct immediate referent selection when taught new words with gaze (Akechi, Kikuchi, Tojo, & Osanai, 2013; Akechi et al., 2011a; Bani Hani et al., 2012; Field, Lewis, & Allen, 2019; Luyster & Lord, 2009; McDuffie, Yoder, & Stone, 2006; McGregor, Rost, Arenas, Farris-Trimble, & Stiles, 2013; Norbury et al., 2010; Parish-Morris, Hennon, Hirsh-Pasek, Golinkoff, & Tager-Flusberg, 2007; Tenenbaum, Amso, Abar, & Sheinkopf, 2014; Venker, Kover, & Weismer, 2016). In addition, children with ASD have shown intact referent selection after a four-week delay (Norbury et al., 2010). These findings suggest that only those children with ASD with significant language delays have difficulty during laboratory word learning tasks.

It is unclear how children with ASD with less impaired language to above average language abilities learn new words with referential gaze, as well as how this reconciles with typically-developing children who also successfully learn with gaze. It is possible that there are similarities between groups, in that children in both groups may either treat gaze as intentional or children in both groups simply attend to the directional properties of gaze to support word learning. Another possibility is that groups diverge in their learning processes. For example, a common hypothesis is that typically-developing children treat gaze as intentional whereas children with ASD do not. We take two approaches to evaluate how children in both groups attend to and learn from referential gaze. One approach follows traditional analyses, that relies on the categorical, quasi-experimental distinction of diagnosis to compare performance between groups by testing the main effect of group and the interaction of group and cue condition (gaze vs. an arrow). In the case of significant interactions, we conduct follow-up analyses with well-matched groups to better interpret performance as relating to group diagnosis. Our second approach follows recent calls to consider abilities defining atypical development as part of a spectrum of abilities within the general population (Cuthbert & Insel, 2013). We do this by considering individual differences in social competence skills, and examining how this relates to in-depth word learning across both groups. We refer to social competence as a set of developmentally appropriate skills (e.g., from following another's gaze in infancy to going out in social groups in older children and adults) that reflect how individuals engage with their social world. An interaction between social competence skills and cue condition would reveal that learning with gaze versus an arrow cue changes as a function of social competencies, and point to a common underlying process in how children in both groups learn new words with gaze.

1.4 Current study

As outlined above, this study pursues three research questions that address gaps in the word learning literature. First, do children visually attend to a word learning scene differently in the presence of a gaze cue versus a control for the direction of attention, an arrow cue? We account for the salience of the cue and the presence of a person as much as possible by including the person in both conditions and matching the cues themselves on physical properties of size, shape, as well as the direction, and degree of motion. We use eye tracking to examine on-line attention to the cue (referential gaze or arrow) and the target object (referent), as well as contingent looking between these two areas. Second, are there any benefits of in-depth learning in the presence of a gaze cue in comparison to an arrow cue? We measure on-line attention to the referent, latency to look at the referent during the test phase, and referent selection. We also extend these measures to in-depth learning, building on existing word learning paradigms to include information regarding an object's function to thus probe for more information that children could recall about an object's semantic features (what the object looks like and what you can do with the object; Gladfelter & Goffman, 2017; Norbury et al., 2010). All eye tracking and learning measures were planned a priori based on word learning studies with gaze (Akechi et al., 2011a; Baldwin, 1993; Norbury et al., 2010; Preissler & Carey, 2005; Tenenbaum et al., 2014). Finally, do typically-developing children and children with ASD differ in their visual attention and learning in the presence of a gaze cue relative to an arrow cue? We also explore how in-depth word learning with gaze versus an arrow cue varies as a function of the dimensional nature of social competence.

2. Method

2.1 Participants

Ethics approval was obtained by a university institutional review board and the study was conducted according to the ethical principles stated in the Declaration of Helsinki (2004). All parents provided informed consent and children provided informed assent. Forty-seven typically-developing (TD) children were recruited for the study from a university database and flyers. Five children were excluded because 3 participated in an earlier pilot version of the study, 1 had a diagnosis of ADHD, and 1 had a hearing aid. Our final sample included 43, 6- to 10-year-old TD children (*M* = 8.75 years, *SD* = 1.14 years), 32 males and 11 females (male dominant to create matched groups, see below), and children had English (20) or French (23) as their dominant language. Participants did not have developmental, learning or behavioral disorders, nor did they have physical, vison or hearing limitations that would interfere with study procedures. All children had normal or corrected vision; one child had corrected vision but was not included in eye tracking analyses because of prior strabismus. TD children had no first- or second-degree relatives with ASD.

Twenty-seven children with ASD were recruited from multiple sources: a list of previous participants in our lab, organizations serving children with disabilities, community events for families of children with ASD, and a university database. Two children were excluded because 1 child was not able to complete the study protocol, and 1 child did not meet study criteria for ASD. Our final sample included 25, 6- to 11-year-old children with ASD (*M* = 8.93 years, *SD* = 1.34 years), 22 males and 3 females. Children had English (12) or French (13) as their dominant language. Children with ASD had a documented diagnosis obtained prior to study involvement, which was confirmed using the Social Communication Questionnaire (SCQ) – Lifetime form (Rutter, Bailey, & Lord, 2003). The SCQ is a 40-item parent-report questionnaire where caregivers respond to Yes or No questions about their child's social communication skills before age 5. Children did not have any other medical conditions associated with ASD and no physical, vision, or hearing limitations that would interfere with study procedures (e.g., color blindness). Except for 1 child with ASD with strabismus who was not included in eye tracking analyses, all other children had normal or corrected vision by parent report. Eight children with ASD were diagnosed with comorbidities, which included attention-deficit/hyperactivity disorder (ADHD), speech dyspraxia or language impairment.

When there are differences by group in our full sample, we explore group differences within a matched group. Group differences are difficult to interpret with our full sample, where TD children have on average higher nonverbal IQ and some language skills than children with ASD. These differences mean that any group differences may be due to factors other than a diagnosis of ASD (see the section on Testing and Standardized Assessments for more information on assessments). To better ensure that we could interpret findings as related to a group's diagnostic category (e.g., ASD or TD), we matched children with ASD to TD children such that there are minimal differences between groups in their distribution on known characteristics. We matched a subset of TD children to children with ASD to be similar on age and nonverbal IQ using propensity scores (Bang, 2020). One child with ASD was excluded to achieve the best matched group. Critically, matching was determined prior to analyses of any experimental outcome measures, resulting in 24 TD children and 24 children with ASD. Guidelines to evaluate well matched groups include Cohen's *d* close to 0, variance ratios close to 1, and *p* values > .5 (Kover & Atwood, 2013). As seen in Table 1, groups were well matched on age and nonverbal IQ, though they differed significantly on some of their language abilities as measured by the Clinical Evaluation of Language Fundamentals – 4th Edition (CELF-4), Word Classes and Recalling Sentences subtests (Secord, Wiig, Boulianne, Semel, & Labelle, 2009; Semel, Wiig, & Secord, 2003). Groups did not differ significantly on the CELF-4 Word Associations. Consistent with their diagnosis, all children with ASD demonstrated significantly higher scores on the SCQ and significantly lower scores on the Socialization subscale of the Vineland Adaptive Behavior Scales-Second Edition (VABS-II; (Sparrow, Cicchetti, & Balla, 2005). In both groups there were similar proportions of English-speaking vs. French-speaking children, boys vs. girls, and mothers with below a university-level education versus those with a university-level education or higher.

No prior studies have compared within-group word learning in the presence of gaze or an arrow cue. Therefore, we based our target sample size on whether we might expect to find a group difference between TD children and children with ASD on our in-depth word learning measure. Our decision was based on Norbury and colleagues (2010), who found a group difference on the recollection of object features (*d* = 1.73) between TD children and children with ASD with 13 children per group. Using a two-sample *t*-test comparison, analyses with 13 children per group were powered at 99%. However, because we included two language groups, our recruitment target was 13 English- and 13 French-speaking children with ASD, in case we needed to conduct analyses within each language group. Therefore, our total ASD sample recruitment target was 26 children with ASD. The recruitment target of the TD group was doubled to approximately 20 children per language group, for a total of 40 TD children. The larger TD sample was recruited to provide a sample large enough to facilitate group matching. We stopped recruitment at 43 TD children that met our exclusion criteria based on available matches for ASD children, and this was not dependent on any preliminary analyses.

Table 1. Sample Characteristics of TD Children and Children with ASD

	Full group		Matched group		Statistics comparing Matched groups		
	TD (n = 43)	ASD (n = 25)	TD (n = 24)	ASD (n = 24)	*p*	*d*	vr
Age†	8.75 (1.14) 6.5 – 10.67	8.93 (1.34) 6.67 – 11.42	8.70 (1.12) 6.50 – 10.50	8.83 (1.26) 6.67 – 11.33	.713	.11	1.27
Nonverbal IQ†	114.5 (14.18) 83 – 143	107.2 (13.61) 80 – 131	109.50 (13.24) 83 – 131	108.29 (12.65) 87 – 131	.748	-.09	.91
CELF-4 Word Associations†	33.7 (10.28) 16 – 53	30.08 (14.72) 5 – 65	33.29 (11.17) 17 – 53	29.92 (15.01) 5 – 65	.382	-.26	1.80
Gender (M : F)	32 : 11	22 : 3	18 : 6	21 : 3	.461		
Maternal education (below : above university)§	10: 33	13 : 12	6 : 18	12 : 12	.136		
English- and French-dominant speaking children (En : Fr)	20 : 23	12 : 13	10 : 14	11 : 13	1		
CELF-4 Word Classes Total†‡	12.14 (2.96) 6 – 19	9.50 (3.84) 2 – 16	12.08 (3.06) 6 – 19	9.74 (3.74) 2 – 16	.024	-.69	1.49
CELF-4 Recalling Sentences†	10.95 (2.27) 6 – 16	7.84 (4.25) 1 – 14	11.17 (2.18) 7 – 16	8.08 (4.16) 1 – 14	.003	-.93	3.64
Vineland Socialization subscale†	111.65 (15.66) 80 – 160	76.60 (11.45) 61 – 110	110.00 (11.88) 80 – 129	76.83 (11.64) 61 – 110	<.001	-2.82	0.96
Social Communication Questionnaire†	3.88 (2.71) 0 - 11	20.72 (5.76) 12 – 32	4.42 (2.62) 0 – 9	20.88 (5.83) 12 – 32	<.001	3.64	4.95

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

†The values shown are the mean (*SD*) and range.

‡One child with ASD did not complete this measure.

§For one TD child the mother's education was not provided thus the father's education was used instead. Continuous and categorical variables were analyzed using paired sample *t*-tests and Fisher's exact tests, respectively, *d* = Cohen's *d*, vr = variance ratio. Negative values for Cohen's *d* indicate higher values in the TD group.

2.2 Testing and standardized assessments

Testing was conducted over two visits, approximately one week apart. During Visit 1, children participated in standardized testing and the word learning task, where children watched a video teaching a novel word. Each video was immediately followed by off-line word learning assessments by experimenters who were blind to cue condition. Visit 1 was always conducted in our lab testing room because the eye tracker was required. Visit 2 only consisted of only the off-line measures, thus this visit was either at the lab or at the families' homes.[2]

Parents filled out questionnaires regarding their level of education (as a proxy for socioeconomic status), their child's language exposure, and social skills. Both groups of children completed a standardized assessment battery on nonverbal IQ and language abilities. Nonverbal IQ was assessed with the Leiter International Performance Scale, Third Edition (Leiter-3; (Roid & Miller, 2013), with a mean standard score of 100, and SD of 15. Language abilities were measured using subtests of the Clinical Evaluation of Language Fundamentals, Fourth Edition (CELF-4; Semel et al., 2003) or the validated Canadian French version, Évaluation clinique des notions langagières fondamentales – version pour francophones du Canada (Secord et al., 2009). Due to time constraints, we included only three subtests, two that tested semantic knowledge (Word Classes, Word Associations) and were thus related to the experimental measures, and one to identify language impairment (Recalling Sentences). Word Classes and Recalling Sentences are normed (*M* = 10, *SD* = 3). Word Associations is a timed assessment of how many exemplars can be named in a category (animals, food, jobs); this is not normed and scores represent the total exemplars named across all three categories.

[2] Additional word association and naming tasks were administered but are not reported here because of floor and ceiling effects (removed for blinding). A word generalization (i.e., mapping the novel word to other objects of a similar kind) was also administered to a subset of children; more can be seen in our supplemental information.

Children's social competence skills were assessed with the Vineland Adaptive Behavior Scales, Second Edition Parent/Caregiver Rating Form (VABS-II) – Socialization domain. The VABS-II is a parent report of multiple domains of children's adaptive behavior including Communication, Daily Living Skills, Socialization, Motor Skills, and Maladaptive Behavior. We focused on the Socialization domain, which provides a measure of children's everyday social behavior and thus relates to intention understanding. The Socialization domain is comprised of three subsections of Interpersonal Relationships, Play and Leisure Time, and Coping Skills. The three subsections are incorporated into a standard score (*M* = 100, *SD* = 15).

2.3 Apparatus

A remote faceLAB eye tracker (version 4.5.1) recorded eye movements at a rate of 60Hz. Video stimuli were presented using GazeTracker presentation software (version 8.0.3156.1000) on a 43-inch TV monitor approximately 93 cm from the child. Videos were shown at a resolution of 1280 x 720 units. One fixation was defined by consistent looking within 40 pixels on the screen (spatial parameter) for a minimum of 100 ms (temporal parameter).

2.4 Experimental design

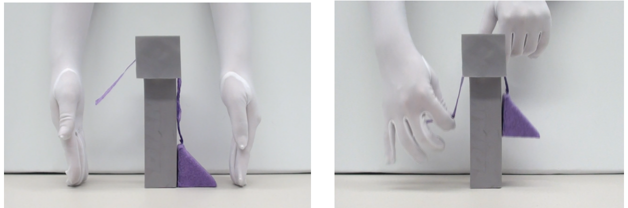
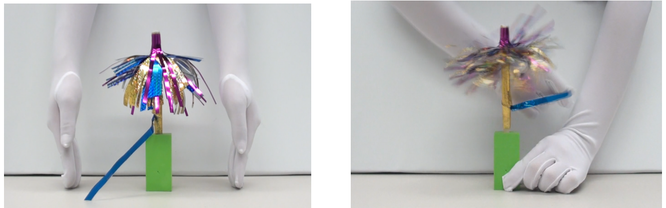
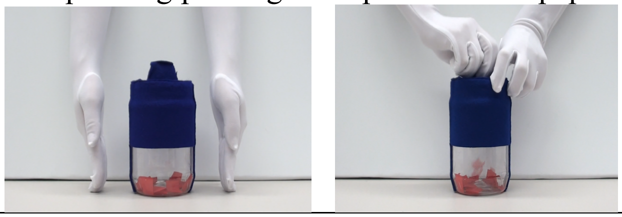
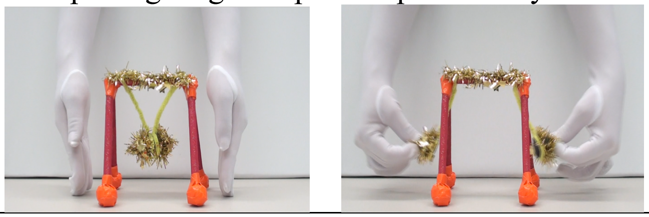
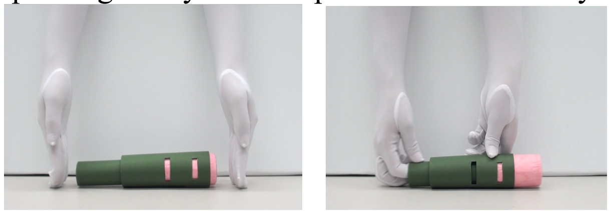
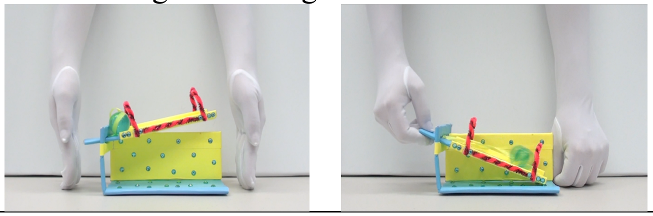
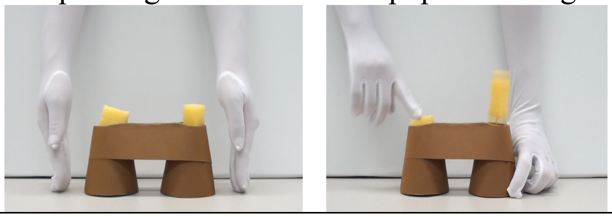
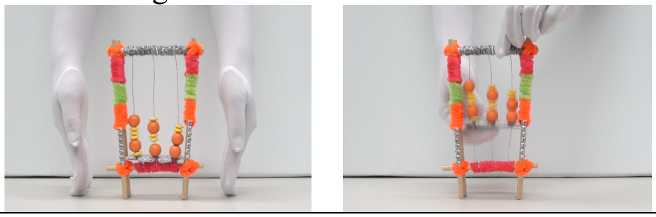
Children were presented with word learning videos in two blocked cue conditions (gaze or arrow), counterbalanced across children. An arrow cue is commonly employed as a control for gaze (e.g., Birmingham et al., 2012) for multiple reasons other than direction, including a level of familiarity and conventional use as a directional cue. Other critical features matched between cue conditions included the presence of the actor, and the approximate size of cue, shape of cue, and the duration and degree of cue motion.

2.5 Novel labels and objects

A subset of four novel labels were selected from a prior norming study in our lab: *fopam, mimole, nalip, pagoune*. Novel words and phrases (e.g., *Where is the pagoune*, *Now point to the pagoune*) were recorded in English and French by a bilingual English and Quebec-French speaker in a sound proof booth using a Marantz PMD660 recorder.

Four targets (referents) and four distractors were designed for this study. Targets were designed to be visually less interesting than distractors: two colors per target with simple shapes and materials (e.g., triangle, paper). We followed these restrictions such that the objects would be simple to describe for children. Distractors were designed to be visually more interesting and perceptually salient, thus ensuring a stronger test of children following the cue to learn the label rather than simply labeling what they were already interested in (Hollich et al., 2000; Parish-Morris et al., 2007). As seen in Table 2, targets and distractors each had a unique cause and effect function and were similar in size (see Table 2). All children were taught four target-distractor pairings. Our online stimuli manual includes more information about stimuli creation and adult pilot testing.

Table 2. Images of Target and Distractor Objects and Their Functions

Target Object and its Function	Distractor Object and its Function
fopam: pulling the string to lift the triangle block up 	**distractor:** pulling the string to twirl the foil strips 
mimole: squeezing/pushing the top to blow the paper inside 	**distractor:** pulling magnetic pieces apart so they reconnect 
nalip: pushing the cylinder to push out the inside cylinder 	**distractor:** moving the rectangle slide down to make the ball roll 
pagoune: pushing the left button to pop out the right button 	**distractor:** moving a bar under the beads to move them up 

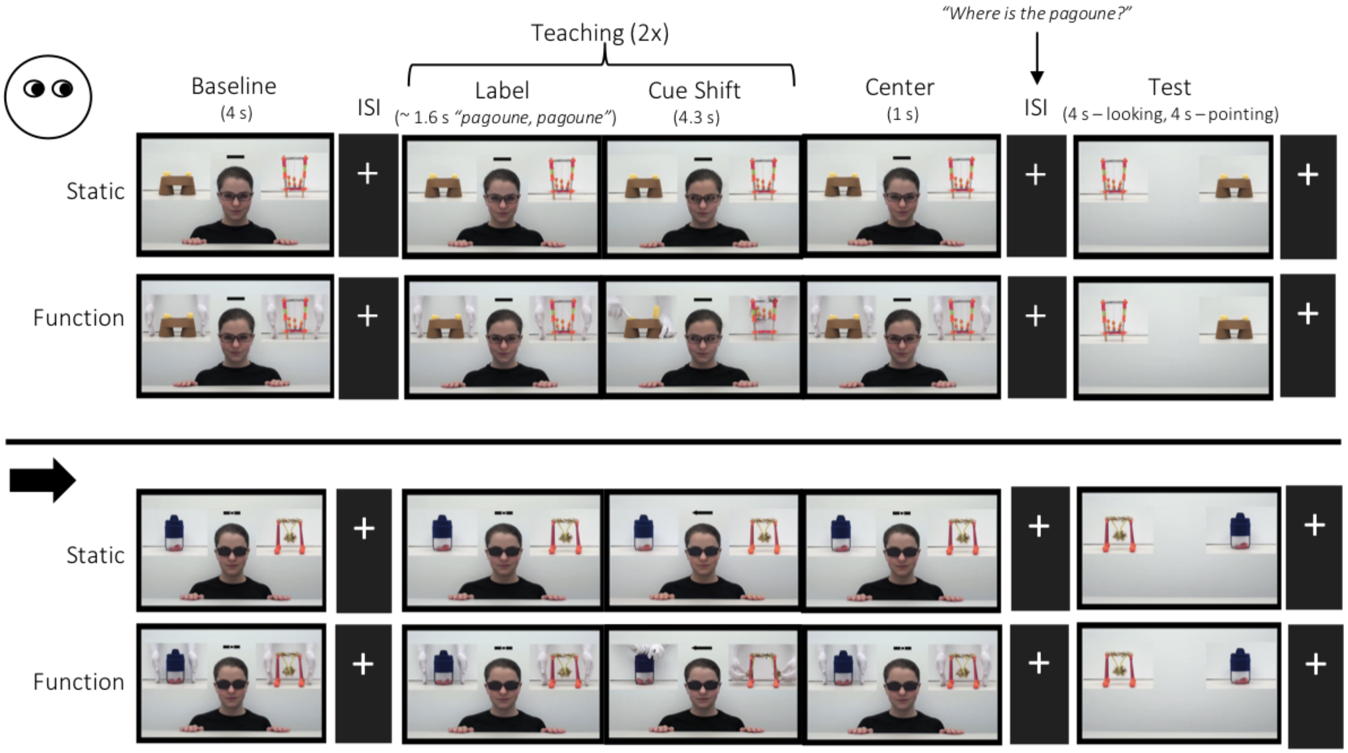
Each of the 4 target objects in the study appeared with the corresponding distractor object, pictured above. Each function had a cause and effect that started and ended with the hands and objects in the same stationary rest position: hands on either side of the object. The hands began moving from rest at the start of the cue portion in the teaching phase (from the start of the cue shift to the target object) and came back to the same rest position before the end of the cue portion (by the end of the cue shift back to direct gaze/static circle).

2.6 Videos

Videos were created using Final Cut Pro software (version 6.0.6). Experimenters provided motivational phrases such as "*Great!*", but no explicit feedback was provided about responses. In each cue condition, one practice trial was conducted prior to each of two experimental trials, for a total of two practice and four experimental trials. The practice trial consisted of similar videos but with known objects instead of novel objects. These objects were chosen to be familiar to children in this age range (i.e., *hammer, cup, scissors, glasses*; Fenson et al., 2007). After watching practice videos, children participated in off-line measures similar to what they would receive with experiment videos. The purpose of practice trials was to keep children motivated in case the experiment videos were too difficult, as well as gauge their understanding of the learning measures. If children did not demonstrate a clear understanding of measures with known objects, then they were not included in corresponding measures for experiment videos (e.g., if the child could not describe even a familiar object, their descriptions of experimental objects were not included). This affected the object description measure, where 3 children with ASD were unable to provide valid descriptions of known objects, and thus were excluded from analyses of object descriptions.

Before watching experiment videos, children were told, "*Now you'll learn some new words for things that you haven't seen before. Make sure you pay close attention because I'm going to ask you questions about what you learn after the video like we did before [with the practice videos].*" Children watched a video sequence first with static objects and then immediately after with hands demonstrating object functions (Figure 1); both static and function videos included naming episodes and the function video was included to provide more information about what children could do with the object. We chose not to counterbalance static and function video presentation in order to optimize our ability to compare performance in our static videos to prior studies that included using similar word learning paradigms.

Figure 1. Word Learning Video Sequence



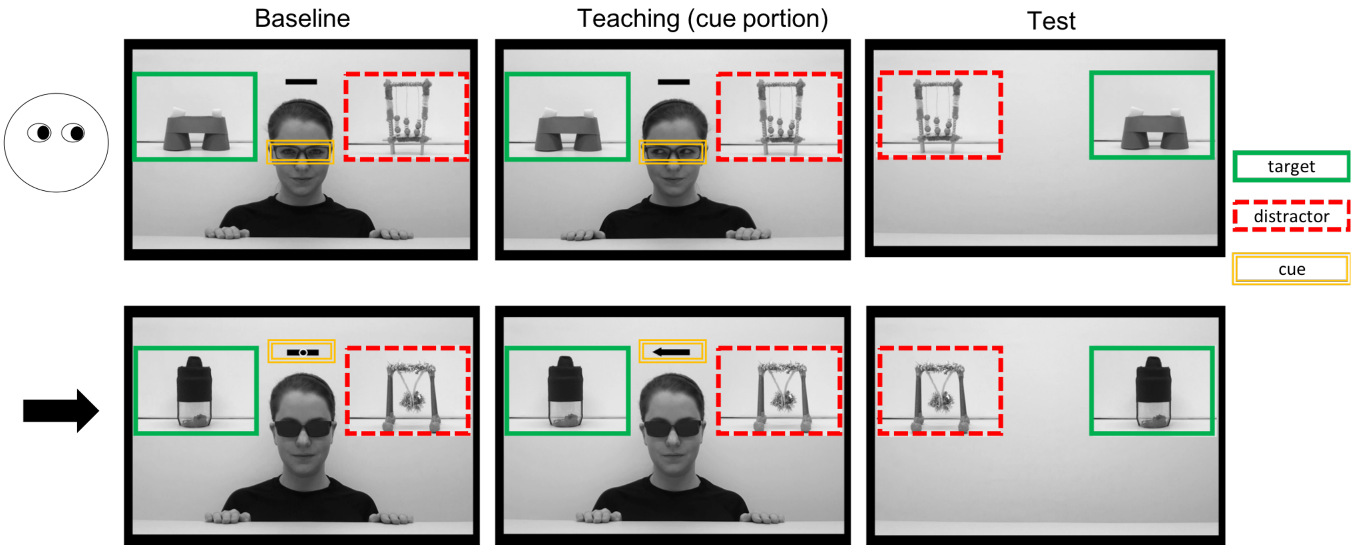
These frames depict an example word learning video sequence in the gaze condition (above the black line) and in the arrow condition (below the black line). Children viewed static videos, which were immediately followed by function videos in both cue conditions. An interstimulus interval with a black screen of a white cross was presented after the test phase during the static video and before the start of the function videos. The duration of the video for the combined static and function videos was approximately 1 min per target object. Children had two experimental trials per cue condition, with each trial including different target and distractor objects.

1) Baseline phase: the child could view the scene for 4s. **2) Teaching phase:** in the *label* portion of the teaching phase, the label was spoken twice, "*pagoune, pagoune*" (~ 1.6 s) The actor's mouth did not move in the videos. After the label portion, in the *cue shift* portion, the cue indicated the referent for 3.6 s, then returned to center. The total duration of the cue shift portion was 4.3 s. After two teaching phases, the video was shown for one extra second with the cue back to the same position as seen in the baseline and label phase; this was included because otherwise the video felt too rushed to move into the test phase. **3) Test phase:** prior to seeing the objects during test, children heard the prompt, *"Where is the pagoune?"* during the ISI. Objects appeared at the offset of the prompt and referent side was switched at test. After 4 s of viewing the objects, children heard the prompt, "*Now point to the pagoune,*" and had 4 s to point; this provided an explicit measure of their referent selection.

2.7 Areas of interest for eyetracking measures

The three video phases included in analyses were 1) baseline, 2) teaching during the cue shift portion only, and 3) test. Looking time during the label portion was not included because looking time to objects and cues were similar between the baseline phase and the label portion. As seen in Figure 2, there were three areas of interest: 1) target 2) distractor 3) cue area (not included during the test phase). The areas around the targets and distractors were the same size (width 440 pixels and height 306 pixels), and the areas around the gaze and the arrow were the same size (width 184 pixels and height 54 pixels).

Figure 2. Examples of the Areas of Interest in the Static Videos



This figure depicts the three areas of interest (AOI) used in the static (pictured) and function videos analyzed in these studies. These AOIs include the area around the target object, the distractor object, and the cue area. The top row depicts the gaze condition and the bottom row depicts the arrow condition.

2.8 Learning Measures

All scoring was completed by coders blind to cue condition and group diagnosis.

2.8.1 Referent selection

A referent selection task measured children's explicit receptive knowledge. During the test phase, children were asked, "*Now point to the [target label]*". Children received 1 point for the target and 0 points for the distractor. Collapsing 2 points over 2 objects in each cue condition, this resulted in for 4 possible points in static videos and 4 possible points in function videos.

2.8.2 Object description.

After watching each video, a description task measured the semantic features recalled from the word learning episode (Gladfelter & Goffman, 2017; Norbury et al., 2010). We asked, "*Now I want you to describe a pagoune for my friend. Remember you can tell me about the size,* *color, shape, what you can do with it, and what kind of object it is. Can you tell me three things about a pagoune?*" Children were not required to provide three things; this was meant to provide a tangible number in the event that saying *"anything"* would be abstract and potentially difficult. Experimenters transcribed descriptions from a video and coding was completed in two passes.

In the first pass, descriptions were verified as descriptions of the target (*valid*) versus those that did not describe the target with certainty (*invalid*). This first pass ensured that semantic features elicited for the objects actually reflected learning of the targets. Table 3 presents examples of the minimum criteria needed for a valid description and Table 4 presents examples of descriptions by children. In the second pass, descriptions were coded for semantic features. We defined *semantic features* as those describing *physical attributes* and *intended function*. *Physical attributes* included features regarding the color, shape, number, spatial location of physical properties (e.g., *top, bottom, center, attached*), and length or weight of object parts. Length or weight was only counted if children demonstrated an understanding of object parts, such as, *the buttons are small* (simply saying *it is small* was not counted since children did not really know novel object sizes). *Intended functions* included features that described the specific use or purpose of the object as seen in the video (e.g., *press a pagoune*). Table 5 details the measures administered at each visit. The full decision tree and coding protocol are available upon request.

Table 3. Four Different Types of Minimum Criteria to Identify a Description as Valid

Minimum Criteria	Example for a pagoune
1. Reference to shape of the whole target object that distinguished it from the distractor object	*looks like binoculars*
2. References to both colors of the target object	*brown and yellow*
3. Reference to both the cause and effect of the target object's function	*you push one side and the other side pops up*
4. Reference to at least one color and one function (cause or effect) of the target object	*brown, pop*

This table provides the four different types of minimum criteria to identify a valid description of a target object. To be classified as a valid description, only one of these criteria needed to be met although the clearest description would meet both criteria (2) and (3). The examples above are from children's descriptions for the target object labeled as a *pagoune*.

Table 4. Examples of Valid and Invalid Descriptions of the Pagoune

Object Identification by Coder	Excerpts of Descriptions (from different children)
Correct Target (Valid)	***"It's brown, it's small, and you can push down one side then the other side comes up…there's two buttons that are yellow. It looks like it's made out of paper and that's it."***
Distractor (Invalid)	*"on the edge, on the side there are feathers…it's made out of wood, there's like strings and there's like balls on the strings that go up and down…"*
General (Invalid)	*"It's for making sound effects, it's for people to play with and that means it's a toy."*
Other – Igloo (Invalid)	*"you run in it, it's a nice house…it's made by blocks of ice, chunks of really hard ice…"*

This table provides examples of different descriptions when different children were asked to describe a pagoune (see Table 1). In the left-hand column are the coder's identification of the object. The first description was identified as a correct target (in bold), thus identified as a valid description; this description is an example of a clear description that met both criteria (2) and (3) as noted in Table 3.

Table 5. A List of the Measures Collected During Visit 1 and Visit 2

Type of Measure	Measure	Visit
Visual attention:	Proportion of looking time to the cue area	1
Cue area	Contingent looking difference	1
Visual attention:	Target advantage	1
Target Object	Latency to the target at test	1
Learning	Referent selection	1
	Latency	1
In-depth learning	Number of valid descriptions	1 and 2
	Number of semantic features	1 and 2

This table details the measures collected in this study, the order in which they were collected, and when they were collected. Visit 2 occurred one week after Visit 1.

2.9 Reliability

A subset of 20% of children's descriptions were independently double-coded. Kappas (object identification) and intraclass correlations (ICCs; semantic features) were .90 and above for the coding groups analyzed in this study, demonstrating a strong level of agreement: object identification (κ = .92, [.86, .98]), intended function (ICC = .93, [.91, .95]), physical attributes (ICC = .95, [.94, .96]).

2.10 Analysis plan

We evaluated continuous dependent variables using linear mixed effects models. All models included the fixed effects of *cue condition* (gaze, arrow) and *group* (TD, ASD), and a random intercept of participant. Depending on the measure, additional fixed effects included the *motion* type of word learning video (static or function), *video phase* (e.g., baseline, teaching - cue shift, and/or test phases), or *visit* (visit 1, visit 2). A random slope of cue condition, our main fixed effect of interest, was included when it significantly improved model fit by comparing AIC values and did not result in high correlation terms (defined as > ± .90; (Barr, Levy, Scheepers, & Tily, 2013). The two categorical dependent variables of *referent selection* and the *number of* *valid target object descriptions* were evaluated using logistic models to examine potential group differences. We did not examine cue condition differences for these variables given the few observations per level (2 observations per cue condition). Based on prior planned analyses and an effort to compare across measures selected for their use in prior studies, we have opted to report full models rather than simplified models of only significant effects (Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Details on model building are provided in our supplementary information.

As noted above in our participants section, group differences in the full sample are not meaningful to interpret given the multiple differences between groups in addition to a diagnosis of ASD. When there were interactions with group and other factors, we conducted the same model for the matched group of TD children and children with ASD to determine if differences could in fact be attributed to group membership rather than other co-occurring differences (e.g., IQ). Interactions were found for the full sample between group and *proportion of looking time to the cue area* and group, as well as *target advantage*. However, analyses with matched groups did not reveal significant main effects or interactions on either measure. These interactions are noted below, but we do not explore them further here; please refer to our supplementary information for additional figures depicting these interactions.

For final linear mixed models, assumptions of normally distributed errors and homoscedasticity were satisfied by visual inspection of quantile-quantile (q-q) plots and histograms. If homoscedasticity appeared to be violated, the dependent variable was transformed. Square root transformations were used with positively skewed data that included many zero values. Log transformations were used for positively skewed data when values were greater than zero. Post hoc tests were conducted on effects when $p < .05$, and multiple comparisons were corrected with the Tukey method and a family-wise error of alpha = .05. Estimated marginal means (EMM) are provided, which are adjusted for missing observations. For all transformed data, means were back-transformed and thus are interpretable in their original units. Standardized effect sizes are provided by Cohen's *d*, using a pooled standard deviation with the unadjusted data. Although standardized effects are helpful to compare across study measures, they do not represent the within-subjects design and missing observations accounted for in the models. Therefore, in addition to the standardized effect size, we report the simple effect size (differences between the EMMs, or the *EMM difference*) and 95% confidence intervals, which are robust measures of meaningful differences using measurement units of the study (Baguley, 2009). The factors of language and block order were not included because visual analysis indicated similar directions of the effects within each factor (e.g., English- and French-speaking children looked more at gaze versus the arrow cue).

2.11 Eyetracking diagnostics

Children were calibrated with a 9-point calibration process. Two TD children and 3 children with ASD were not included in eye tracking analyses because of an inability to calibrate (1 TD, 2 ASD) and a history of strabismus (1 TD, 1 ASD). This resulted in 41 TD children and 22 children with ASD included in eye tracking analyses. Our supplementary information and available code (Bang, 2020) provides more detail on data cleaning.

3. Results

3.1 Visual attention: cue area

3.1.1 Proportion of looking time to the cue area

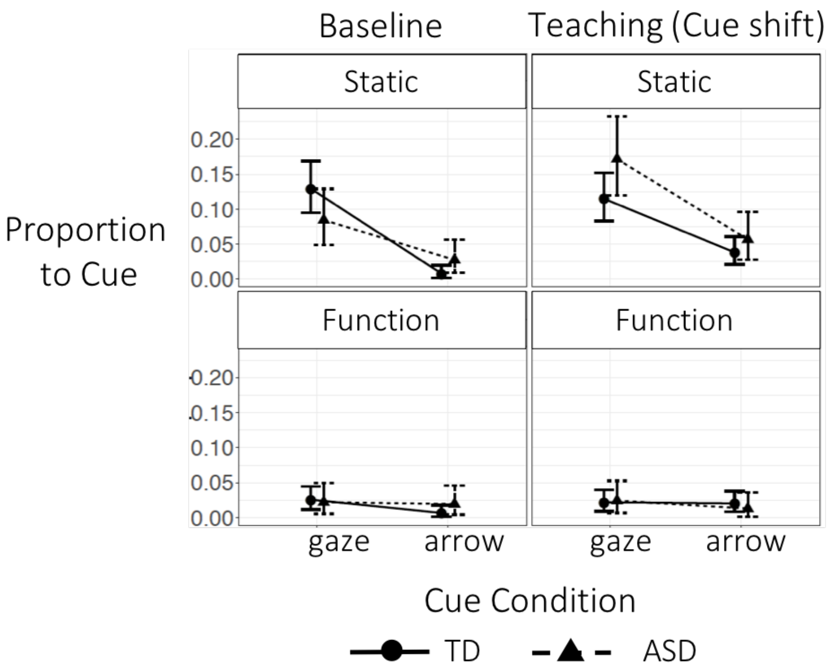
Both baseline and teaching phases were included for this measure, thus examining attention to gaze before and during the cue shift; a visual depiction of this contrast can be seen in

Figure 2. For baseline and teaching phases, we calculated the *proportion of looking time to the cue area,* dividing the duration to the cue area by the total duration of the scene in the respective phase; proportions were square root transformed. The model included a random intercept of participant and fixed effects of cue condition, motion, video phase, group, and all interactions.

We found main effects of cue condition and motion, as well as a significant interaction of cue condition by motion (*p* < .001). As seen in Figure 3, this interaction revealed cue condition differences during static videos, where children looked longer at the area of the cue in the gaze versus the arrow condition (gaze EMM = .12, [.10, .15], arrow EMM = .03, [.02, .04]), whereas during function videos there was no significant difference between cue conditions in looking to the cue area (gaze EMM = .02, [.01, .04], arrow EMM = .01, [.01, .02]). The lack of attention to the cue area during function videos is evident by children looking significantly longer at the area of the cue during static than function videos in both the gaze condition, *t*(856.86) = 11.32, *p* < .001, and the arrow condition, *t*(855.65) = 2.96, *p* = .022.

In addition to the interaction with cue condition and motion, there was also a significant interaction of video phase and motion, *F*(1, 851.13), (*p* =.011). Children looked longer at the cue area during teaching versus baseline only during static videos (baseline EMM = .05 [.04, .07], teaching EMM = .09 [.07, .11]), whereas attention to the area of the cue was not significantly different between baseline and teaching phases during function videos (baseline EMM = .02, [.01, .03], teaching EMM = .02 [.01, .03]). Similar to the findings above, children looked longer at the area of the cue during static than function videos in both the baseline phase, *t*(853.46) = 5.34, *p* < .001, and teaching phase, *t*(853.46) = 8.91, *p* < .001. There was one significant 3-way interaction between cue condition, video phase, and group, *F*(1, 851.13) = 8.13, *p* = .004. When examining the same model in a matched group, this interaction did not reach significance, *F*(1, 602.14) = 3.81, *p* = .052, indicating that this difference is attenuated when groups are similar on variables of nonverbal IQ and/or age; thus the difference in the full group may not be due to ASD per se. Our supplementary information includes more information on this finding.

Figure 3. Proportion to Cue: Gaze versus Arrow Differences



The proportion to cue variable has been back-transformed and data are interpretable in their original units. Points represent EMM and error bars are 95% confidence intervals. Children looked longer at the gaze cue versus the arrow cue during baseline and teaching phases, but only during the static videos. Children also looked at both cues longer during the teaching versus the baseline phase, demonstrating that they noticed the cues during teaching. By the function videos, children were not looking long at the cue area in either condition

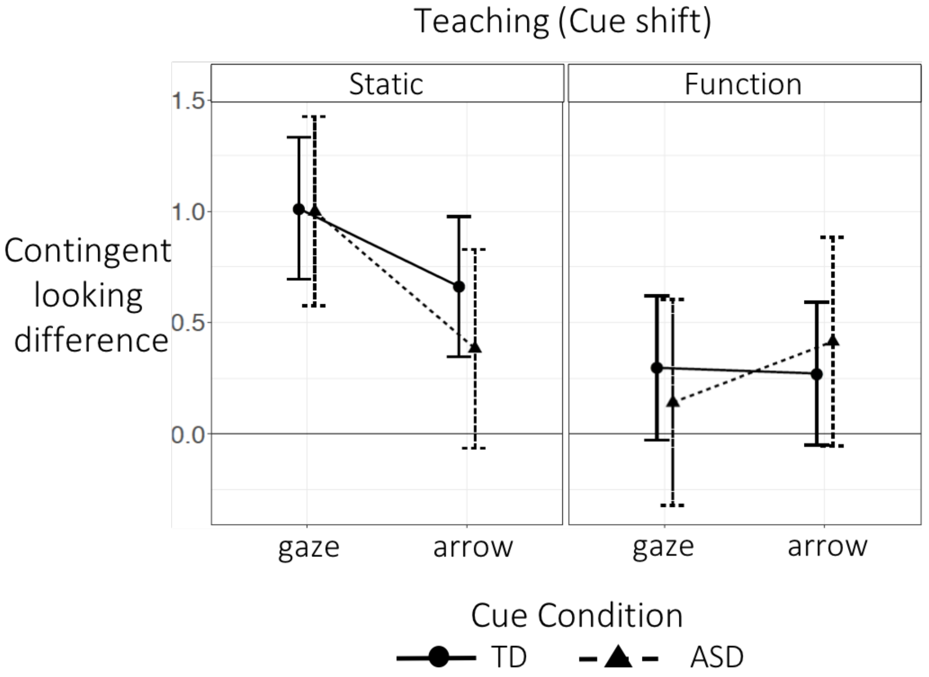
3.1.2 Contingent looking difference

We examined contingent looking, or how often children looked back and forth from the target to the cue area, relative to their contingent looking to the distractor. We now move to focus on the teaching phase alone, when the cue shifts to the target. One contingent look was defined by a fixation on the cue area followed immediately by a fixation on the target or distractor object area, or the reverse direction. We calculated a *contingent looking difference score*, subtracting the total number of contingent looks to the distractor from the total number of contingent looks to the target.[3] The model included a random intercept of participant and fixed effects of cue condition and group, and their interactions.

We found a main effect of motion, as well as a significant interaction of cue condition by motion (*p* = .014). This interaction revealed that during static videos, children had greater contingent looking difference scores in the gaze versus the arrow condition (gaze EMM = 1.00, [.74, 1.27], arrow EMM = .52, [.25, .80]), whereas during function videos, there was no significant cue condition difference (gaze EMM = .22, [-.06, .50], arrow EMM = .34, [.06, .63]). As seen in Figure 4, the cue condition difference was driven by more contingent looking to the target during static versus function videos in the gaze condition, *t*(401.76) = 4.58, *p* < .001, whereas contingent looking difference was not significantly different between static and function videos during the arrow condition, *t*(400.36) = 1.03, *p* = .634.

[3] We did not calculate a standardized difference score because there were many instances of a difference score of 0, which was either because children did not provide any contingent looks to either object, or there were equal contingent looks to both objects.

Figure 4. Contingent looking difference: Gaze and Arrow Condition Differences



Points represent EMM and error bars are 95% confidence intervals. The horizontal line of 0 indicates equal contingent looking to the target and the distractor, with values over 0 indicating more contingent looking at the target and values below 0 indicating more contingent looking at the distractor. This figure demonstrates that during static videos, children in both groups provided more contingent looks to the target (versus the distractor) with a gaze cue than with an arrow cue.

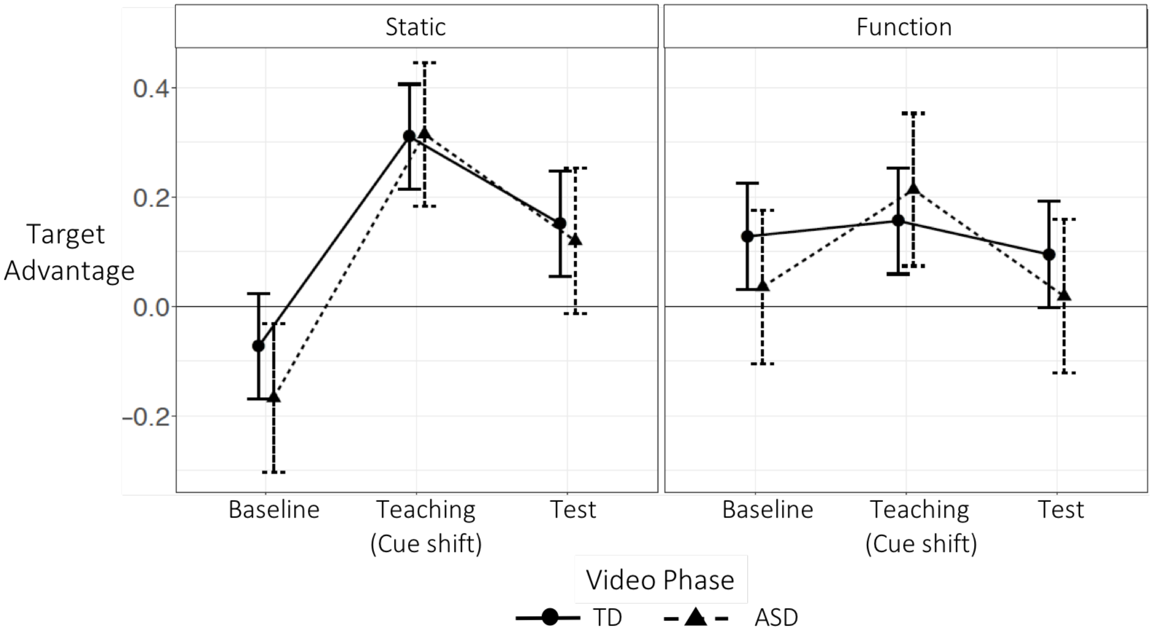
3.2 Visual attention: target object and on-line learning

3.2.1 Target advantage

Ability to attend to the target at baseline, follow the cue to the target during teaching (cue shift portion only) and locate the target during test was measured using the duration to the target relative to the distractor. This was calculated using a standardized difference score to create a *target advantage score* during baseline, teaching, and test phases (Akechi et al., 2013; Akechi et al., 2011b). For each phase, target advantage scores were calculated by subtracting the duration to the distractor (d) from the duration to the target (t), then dividing by the total duration to both the target and distractor, i.e, (t – d) / (t + d). Our primary analysis examined whether target advantage scores differed between cue conditions, and if there were any interactions of this factor with video phase or group. We also examined comparisons of target advantage scores during teaching and test phases relative to baseline to investigate children's attention to the object over the course of static and function videos. When children had greater target advantage scores during teaching versus baseline, this suggests that children followed the cue shift to look at the target, and when children had greater target advantage scores during test versus baseline, this suggests that children learned the target object. The model included a random intercept of participant and a random slope of cue condition, as well as fixed effects of cue condition, video phase (baseline, teaching, test), group, and their interactions.

There were no significant main effects of cue condition, motion, or group, nor any two-way interactions with cue condition. There was one significant 3-way interaction of cue condition, motion, and group, *F*(1, 1258.38) = 4.62, *p* = .032. However, this was not significant with the matched groups, *F*(1, 884.91) = 2.57, *p* = .110. The lack of a finding with matched groups again demonstrates that the full group difference is attenuated when groups are similar on nonverbal IQ and/or age, suggesting that this difference may be less likely due to a group difference of diagnosis. The only other significant effects were that of a main effect of video phase and an interaction between video phase and motion, with the interaction revealing reduced target advantage scores during function videos. As seen in Figure 5, post-hoc comparisons between video phases indicated significant differences between video phases only during static, but not function videos (*p*s > .07). During static videos, there were greater target advantage scores during teaching than baseline, *p* < .001 (baseline EMM = -.12, [-.20, -.04]; teaching EMM = .31, [.23, .39]), indicating that children followed the cues. Moreover, there were greater target advantage scores during test (EMM = .14, [.05, .22]) versus baseline, *p* < .001, indicating that children learned the label-referent association. There were also greater target advantage scores during teaching versus test, *p* = .002, which was expected since during teaching a cue directed attention to an object. Negative target advantage scores during baseline indicated an initial preference in both groups for the distractor regardless of cue. Comparisons between static and function videos demonstrated that attention to the object during function videos was influenced by the preceding static videos, because greater target advantage scores were seen during the baseline phase of function versus static videos, *t*(1236.07) = -4.07, *p* < .001. No differences were seen between static and function videos for teaching or test phases (*p*s > .06).

Figure 5. Attention to Target Versus Distractor Objects During Static and Function Videos



Points represent EMM and error bars are 95% confidence intervals. The horizontal line of 0 indicates equal looking to the target and the distractor, with values over 0 indicating more looking at the target and values below 0 indicating more looking at the distractor. This figure demonstrates that during static videos, children were looking more to the distractor versus the target object during baseline, but then followed the cue to look more at the target versus the distractor during teaching. During test, children continued to look at the target more than the distractor, which was significantly different from their attention to the target during baseline, suggesting that they learned the target object. During function videos, EMM values above zero indicate that children looked more at the target versus the distractor, but this did not differ between video phases.

3.3 Latency

This measure examined how quickly children first fixated (i.e., first time children spent 100 ms) on the target object at test after hearing the prompt during the ISI (e.g., "*Where is the pagoune*?"). Latencies under 200 ms were excluded based on the duration it takes to make a saccade. Latencies over 4000 ms were excluded because this was when children heard the second prompt to point to the target, e.g., "*Now point to the pagoune*." The model included a random intercept of participant and a random slope of cue condition, as well as fixed effects of cue condition, group and their interactions; latencies were log transformed. There were no significant main effects or interactions (*F*s < 2.21, *p*s > .142)

3.4 On-line learning: referent selection task

At the end of the test phase, children were asked to point to the target as an explicit measure of referent selection; this phase did not include the presence of the person or any cue. All children were included in these analyses (TD n = 43, ASD n = 25), and trials from both cues were collapsed for a total of 4 trials per motion video; 4 children with ASD (2 static and 2 function videos) and 1 TD child (1 function video) did not point at all for one trial. There were no significant effects of group or motion, nor an interaction of group and motion (χ^2s < 1.91, *p*s > .167). As seen in Table 6, over 70% of TD children and children with ASD were able to correctly identify the referent at test for three to four trials during static and function videos.

Table 6. Number and Percentages of Children with Correct Referent Selection

Number of correct referents identified	TD	ASD	TD	ASD
	Static		Function	
4	29 (67%)	12 (48%)	28 (65%)	15 (60%)
3	3 (7%)	6 (24%)	7 (16%)	4 (16%)
2	8 (19%)	1 (4%)	6 (14%)	1 (4%)
1	3 (7%)	4 (16%)	2 (5%)	3 (12%)
0	0	2 (8%)	0	2 (8%)

This table indicates how many children were able to correctly point to the target (correct referent) during static and function videos. There were four possible target objects for static and function videos. In the ASD group, 1 child was never able to identify the target during both static and function videos, and 2 different children were never able to identify the target in either the static or the function videos.

3.5 In-depth learning: description task

After watching the video, we assessed children's learning based on their ability to describe the target. Only those children who demonstrated understanding of how to provide descriptions were included in the task, which was based on valid descriptions of familiar objects during practice trials. This resulted in including 43 TD children and 22 children with ASD.

3.6 Number of valid descriptions

Before assessing the number of semantic features, only valid descriptions could be included in these analyses. As seen in Table 7, over 50% of children provided 3 to 4 valid descriptions at visit 1 in both TD and ASD groups. By visit 2, this decreased substantially for both groups. Over a third of children in both groups did not have any valid descriptions at visit 2. There were no significant effects of group nor an interaction of group and visit (χ^2s < .01, *p*s > .923), but there was a significant effect of visit, $\chi^2 = 51.38$, $p < .001$, with more valid descriptions at visit 1 than visit 2.

Table 7. Number and Percentages of Children with Valid Target Object Descriptions

Number of valid	TD	ASD	TD	ASD
descriptions	Visit 1		Visit 2	
4	15 (35%)	9 (41%)	1 (2%)	0
3	11 (26%)	3 (14%)	4 (9%)	3 (14%)
2	6 (14%)	5 (23%)	9 (21%)	7 (32%)
1	8 (19%)	2 (9%)	15 (35%)	2 (9%)
0	3 (7%)	3 (14%)	14 (33%)	10 (46%)

This table indicates the number of valid descriptions at visit 1 and visit 2. There were four possible target objects to describe at both visits. Percentages were rounded up to a whole number.

3.7 Number of semantic features for valid target descriptions

We next examined the number of semantic features recalled by children for their valid descriptions. The model included a random intercept of participant and fixed effects of cue condition, visit, group, and their interactions. There was no significant main effect of cue condition, *F*(1, 190.97) = 3.46, *p* = .064, although children recalled more semantic features in the gaze (EMM = 5.73, [4.86, 6.60]) relative to the arrow condition (EMM = 5.03, [4.17, 5.90]). There were no other main effects or interactions (*F*s < 2.63, *p*s > .106).

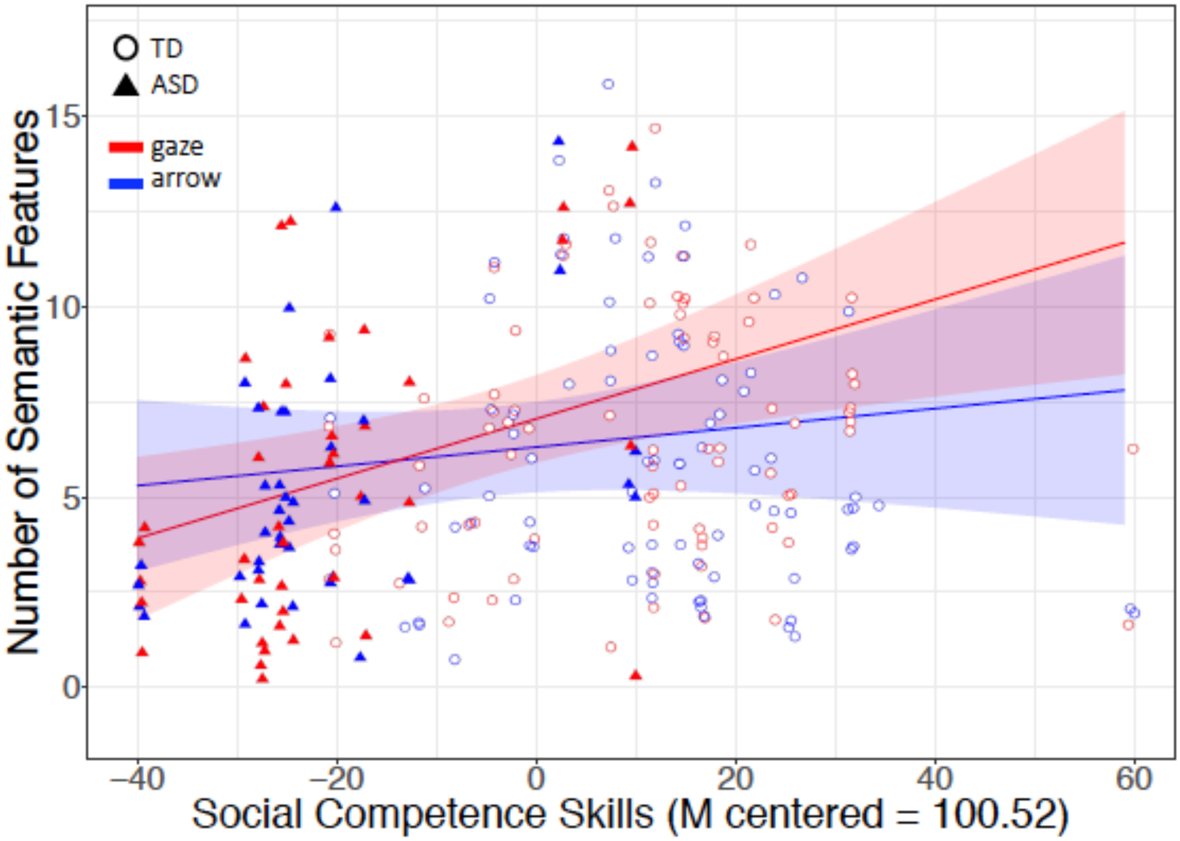
3.8 Exploratory analyses

We explored the relation between cue condition and semantic features by examining whether this could be related to children's parent-reported social competence on the VABS-II. The model included a random intercept of participant and fixed effects of cue condition, visit, group, social competence, and their interactions. There was no main effect of cue condition, *F*(1, 178.22) = 3.52, *p* = .062. However, there was a main effect of social competence, *F*(1, 59.93) = 6.43, *p* = .014, and a significant interaction of cue condition and social competence, *F*(1, 182.64) = 5.55, *p* = .020. Figure 6 depicts these slopes. Post hoc tests of this interaction revealed a positive slope in the gaze condition, slope = .11, [.05, .17], significantly different from the null hypothesis of no relation, *t*(79.25) = 3.44, *p* = .001. The slope in the arrow condition was not significantly different from the null, slope = .04, [-.03, .11], *t*(100.13) = 1.18, *p* = .241. The slope in the gaze condition was significantly greater than that of the slope in the arrow condition, *t*(182.64) = 2.36, *p* =.020. In the TD group, one child had social competence skills over 2.5 SD above the mean (*M* = 100.52, *SD* = 21.80); removing this participant did not change the significance of the interaction between cue condition and social competence. Table 8 summarizes the differences seen between cue conditions on all measures.

In addition to the above interaction, there was an interaction between group and social competence scores, *F*(1, 59.93) = 4.86, *p* = .031. An examination of the data indicated higher social competence skills in children with ASD was related to more semantic features in both cue conditions whereas social competence was not related to semantic feature recall in typically-developing children. However, in the matched group this interaction was not significant, *F*(1, 41.61) = 3.56, *p* = .066, suggesting that if children with ASD do provide more features when they have higher social competence scores, this effect may be weak or due in part to co-occurring factors such as IQ and/or age.

Overall, social competence was weakly and not significantly correlated with other known variables of age: *r* = -.20, nonverbal IQ: *r* = .15, and CELF-4 Word Associations: *r* = .14, but was significantly correlated with CELF-4 Word Classes: *r* = .28, *p* = .028 and CELF-4 Recalling Sentences: *r* = .38, *p* = .003. The same model above with social competence was conducted instead with each of these language measures (CELF-4 Word Classes and with CELF-4 Recalling Sentences). There was a main effect with CELF-4 Word Classes, such that children in both groups with higher scores on the Word Classes subtest were those that provided more semantic features overall, *F*(1, 72.17) = 5.49, *p* = .022. There was no main effect with Recalling Sentences, *F*(1, 77.39) = 1.51, *p* = .223. There were no significant interactions of cue condition and language test scores with either language measure (*p*s > .911).

Figure 6. The Interaction Between Cue Condition and Social Competence on the Number of Semantic Features



Circles (TD) and Triangles (ASD) represent the observed data. Shaded areas represent the 95% confidence interval around the slopes of social competence in the gaze condition (red) and the arrow condition (blue). This figure demonstrates that children with higher social competence in both TD and ASD groups recalled more semantic features in the gaze condition, whereas in the arrow condition higher social competence did not result in more semantic features for those same children.

Table 8. Summary of Comparisons Between Gaze and Arrow Conditions

Measure	Factor	Test statistic	*p*	EMM difference [95% CI]	*d* [95% CI]
Proportion to Cue	**Cue condition (C)**	***F*(1, 860.79) = 71.83**	**< .001**		
	C x M	***F*(1, 856.77) = 33.86**	**<.001**		
	Static: gaze - arrow	***t*(854.09) = 10.48**	**<.001**	**.09 [.07, .11]**	**.77 [.58, .96]**
	Function: gaze - arrow	*t*(862.78) = 1.84	.308	.01 [0, .02]	.13 [-.06, .31]
	C x VP x G n = 63 (full sample)	*F*(1, 851.13) = 8.13	.004		
	C x VP x G n = 46 (matched group)	*F*(1, 602.14) = 3.81	.052		
Contingent Looking Difference	Cue condition	*F*(1, 405.10) = 2.14	.144		
	C x M	***F*(1, 401.50) = 6.12**	**.014**		
	Static: gaze - arrow	***t*(398.39) = 2.88**	**.015**	**.48 [.15, .81]**	**.29 [.04, .55]**
	Function: gaze – arrow	*t*(407.53) = -.69	.830	-.12 [-.48, .23]	-.04 [-.3, .23]
Target Advantage	Cue condition	*F*(1, 62.43) = .06	.803	-.01 [-.08, .06]	-.03 [-.14, .07]
	C x M x G n = 63 (full sample)	*F*(1, 1258.38) = 4.62	.032		
	C x M x G n = 46 (matched group)	*F*(1, 884.91) = 2.57	.110		
Latency	Cue condition	*F*(1, 61.42) = .58	.448	-.05 [-.18, .08]	-.02 [-.22, .17]
Number of Semantic Features	Cue condition	*F*(1, 190.97) = 3.46	.064	.70 [-.03, 1.43]	.15 [-.1, .4]
Number of Semantic Features – model with Vineland	Cue condition	*F*(1, 178.22) = 3.52	.062	1.20 [-.04, 2.46]	.15 [-.1, .4]
	C x Social competence slope: gaze - arrow	***F*(1, 182.64) = 5.455**	**.020**	**.07 [.01, .13]**	

C = cue condition, VP = video phase, M = motion, G = Group. This table summarizes the cue condition comparisons tested in this study for the full sample, unless otherwise indicated. Values in bold indicate measures with significant differences between the gaze and arrow conditions.

4. Discussion

This work reveals new evidence regarding how children attend to and learn words from the presence of referential gaze and how this differs from a non-gaze directional cue. This comparison was designed to equate the critical point in a word mapping paradigm when the direction of the cue identifies the referent. We investigated three questions: 1) whether children's visual attention to a word learning scene differed in the presence of gaze versus a directional control cue of an arrow, 2) whether children's in-depth word learning beyond label-referent associations differed in the presence of a gaze cue versus an arrow cue, and 3) whether typically-developing children and children with ASD differed in how they attend to and learn from referential gaze versus another directional cue.

4.1 Children attend to gaze differently than another directional cue

We found that an arrow was as effective at directing attention to a referent as gaze, but gaze was treated differently from an arrow in two key ways. First, children spent overall more time attending to the area of the cue in the gaze versus the arrow condition, and this began at baseline prior to when cues shifted to identify the referent. Yet despite an initial increased attention to gaze during baseline, children still looked longer overall at each cue during teaching relative to baseline and followed each cue to the target, indicating that initial attentional differences during baseline did not affect children's ability to follow the cues to the referent during teaching. The latter point is further confirmed by the lack of differences in target advantages scores between gaze and arrow conditions throughout the video.

Second, we found that during static videos, children provided more contingent looks to the target versus the distractor in the gaze relative to the arrow condition. We note that this effect is small, with an advantage of half of a contingent look to the target in the gaze versus the arrow condition. One explanation could be that increased contingent looking with gaze versus an arrow cue is an artifact of attentional processes, related to the initial salience of gaze, although similar attention to the target in both gaze and arrow conditions suggests that the cue was as effective in directing attention in both conditions. Another explanation is that contingent looking reflects a type of referencing, where children are figuring out more about the link between gaze and the referent in contrast to when the arrow indicates the referent. This referencing may be also related to what children expect when a gaze cue is directing attention versus an arrow cue, which could lead to differences in how information is processed. This possibility is in line with prior work implicating intention reading in conditions that found contingent looking between gaze and a referent (Baldwin & Moses, 2001; Vivanti et al., 2011). Critically, our work demonstrates that even when both gaze and arrow cues similarly direct children's attention to a referent, there is more contingent looking with gaze versus another directional cue. We propose a fourth possible explanation, that is not mutually exclusive with the above: increased contingent looking between gaze and referent may stem from children's differential experience with gaze versus an arrow cue. Due to the nature of gaze embedded within daily social interactions, children may have learned that they should pay close attention to this link. The role of experience with gaze is an overlooked property, which is made clearer when comparing gaze to a well-known, but still less ubiquitous cue of an arrow. Whether this experience sets the stage for, or results from intention understanding is an intriguing and open question, which we continue in our discussion regarding semantic features.

4.2 Gaze may support in-depth word learning

Our measure of in-depth learning examined what children recalled about words they learned, examining how gaze could support word learning beyond a receptive understanding of the label-referent association. While there was no significant difference between the number of features recalled in the gaze versus the arrow condition in our planned comparison, exploratory analyses revealed an interesting interaction between cue condition and children's parent-reported social competence skills. We found a positive slope in the gaze condition such that children with higher versus lower social competence skills provided more semantic features in the gaze condition relative to the arrow condition. The slope of .11 in the gaze condition can be interpreted as if an average increase in 15 points above the mean for social competence skills (approximately 1 standard deviation) results in an increase of 1.65 more semantic features. Though this effect is small, it is still notable given that cue conditions were minimally different from each other, children were not asked to attend to semantic features, and features were collected and coded by individuals blind to cue condition, group, and study hypotheses.

There are multiple reasons for why learning with gaze may support children's recollection of semantic features. As noted earlier, according to Natural Pedagogy (Csibra & Gergeley, 2009), learning with gaze may support *efficient* learning because children perceive gaze as an ostensive and communicative cue reflecting a person's intention to share knowledge that can support generalizable learning. This proposal could explain why in our paradigm children with higher social competence skills, skills in daily life that may reflect an understanding of communicative intent, were able to use gaze, not the arrow cue, in our paradigm to learn semantic features. The semantic features children learned are generalizable, in that they can help children form an initial categorical representation of the object, particularly features of an object's function which remain generalizable if the shape, size, or color were to change. Yet as noted above, the comparison of gaze versus an arrow cue makes it difficult to ignore the role of children's familiarity and experience with gaze in social interactions relative to an arrow. Children with higher versus lower parent-reported social competence skills may be those who have become more sensitive to this cue and are better able to draw from their own experiences with responding to and initiating gaze, thus understanding how to make the most of the actor's gaze. These findings provide preliminary evidence that better in-depth learning with gaze may be a function of children's broader social competence skills. Future work should also examine whether with some initial threshold of familiarity or experience comes a level of intention understanding.

The findings from this study and others (Gladfelter & Goffman, 2017; Norbury et al., 2010) indicate that school-age children are able to recall important semantic features about an object even after a brief teaching episode, though the strength and stability of this representation deserves further attention (Kucker, McMurray, & Samuelson, 2015). It is important to note that while there was no relation between social competence and semantic features in the arrow condition, children still recalled semantic features in the arrow condition; this demonstrates that a non-gaze directional cue was also sufficient for children to learn the word and recall semantic features. Children in both groups with higher scores on the Word Classes subtest also provided more semantic features overall, suggesting that those with a stronger semantic understanding of known words are those who were able to provide more semantic features of novel words. Future work will need to test the direction of this relation. This relation may reflect that those who already have a strong grasp of their semantic knowledge of known words can extract more semantic features from novel words, or it could also be that the ability to extract semantic information about novel words reflects a process of recognizing and piecing together knowledge about words that can lead to stronger semantic knowledge for all words. Additionally, children in both groups described more referents at the first visit than the second visit, which is not surprising given that the word learning episode was short and only presented once. Stronger retention over time may require multiple episodes as well as multiple exemplars (Luyster & Lord, 2009; Vlach & DeBrock, 2019). Prior studies have also demonstrated how ostensive feedback after fast-mapping paradigms supports stronger retention (e.g., gaze, pointing, illumination; Axelsson, Churchley, & Horst, 2012; Hartley et al., 2019).

4.3 Autism Spectrum Disorder and successful word learning

We find few differences between children with ASD and TD children in their attention to the cue, referent, and word learning between gaze and arrow conditions. Measures of proportion of looking time to the cue and target advantage scores resulted in interactions involving cue condition and group, but these did not remain significant when examined in matched groups where non-verbal IQ and other characteristics were accounted for. Importantly, the pattern of attention was similar in the full group versus the matched group, as seen in our supplementary information. This is consistent with many prior studies demonstrating that when children with ASD are similar to TD children on covariates such as nonverbal IQ, gender, parental education, and/or language abilities, few to no differences are seen between groups in their attention to various aspects of the scene or word learning abilities (Akechi et al., 2013; Akechi et al., 2011a; Bani Hani et al., 2012; Gladfelter & Goffman, 2017; Luyster et al., 2009; McGregor et al., 2013). Notably, the evidence that children can follow referential gaze comes from fairly simple controlled paradigms. Others have found that when needing to track multiple pieces of information beyond the direction of speaker's gaze, children with ASD have difficulties using referential gaze to learn new words in real world settings (Jing & Fang, 2014).

Including both diagnostic groups in this work is important to address conflicting evidence surrounding how children use referential gaze to learn new words. Earlier studies in both diagnostic groups based interpretations on assumptions of gaze as an intentional cue (Baldwin, 1993; Baron-Cohen et al., 1997; Norbury et al., 2010). Yet increasing evidence of successful referent selection with referential gaze in both children with ASD or those with typical development (Akechi et al., 2013; Akechi et al., 2011a; Bani Hani et al., 2012; Luyster & Lord, 2009; McDuffie et al., 2006; McGregor et al., 2013; Parish-Morris et al., 2007; Tenenbaum et al., 2014; Venker et al., 2016) raises questions about the extent to which intention understanding plays a role for all children. One difficulty is that intention understanding is not directly measurable or observable (Huang, Heyes, & Charman, 2002; Premack & Woodruff, 1978), thus support for its presence would result from ruling out alternative explanations until intention understanding is the only possibility (Heyes, 2014). Yet in addition to ruling out explanations, it is important to reconcile findings across different populations. Intact social communication has been considered as a prerequisite of intention understanding, but whether this understanding is engaged during word learning is still unclear for both typically-developing children and children with ASD. Baldwin and Moses (2001) noted the importance of needing to consider the variation even with typically-developing children in their social understanding abilities. We extend this call, encouraging researchers to consider variation across a full spectrum of abilities in social communication, bridging across typical and atypical populations. Our findings highlight that one important contributor to children's abilities to successfully use gaze for in-depth learning may be the individual differences in social competence skills, suggesting a common thread reconciling the performance of typically-developing children and children with ASD.

The findings demonstrating that children with ASD can use gaze to learn new words may be counterintuitive to many given their known impairments in social communication. Our findings are not due to our sample of children with ASD having unusually strong social communication abilities relative to typically-developing children. As would be expected, children with ASD were significantly impaired in their everyday social competence relative to their typically-developing peers (Vineland Socialization domain: ASD *M* = 76.60, TD *M* =

111.65, *d* = - 2.82). Though children with ASD are impaired in the development, rate, or style of social communication and interaction (American Psychiatric Association, 2013), the cause of these difficulties is long-debated and controversial. Despite prevailing thought, behaviors that surface as social difficulties may not always have a social cause. In addition to social communication impairments, ASD is characterized by a second symptom domain of restricted and repetitive behaviors (RRBs), which can have downstream effects on social behavior (Uljarević et al., 2017). Researchers have demonstrated that aspects of social communication difficulties in children with ASD may be related to impairments in RRB (Leekam, 2016; Nadig, Lee, Singh, Bosshart, & Ozonoff, 2010; Nadig, Seth, & Sasson, 2015; Vernetti, Senju, Charman, Johnson, & Gliga, 2018). For example, Nadig and colleagues (2010) demonstrated that children who had more severe RRB scores were more likely to monologue, elaborating on their own interests in a one-sided manner, despite being in a conversational setting. Therefore, social communication difficulties may result from multiple, not purely social, origins (Leekam, 2016), demonstrating the complex nature of the disorder.

5. Limitations

Findings from our exploratory analyses need to be confirmed with hypothesis-testing designs and larger sample sizes to better understand how social competence skills relate to word learning. Additionally, because static and function videos were not counterbalanced, future work will need to address whether the effects seen with static videos are due to order effects or object motion. Our sample also includes more males than females, given the increased identification of ASD in males, as well as only children with ASD with normal to above-average intelligence; this limits generalizability to females and the full spectrum of children with ASD. Future work should incorporate a more naturalistic learning scenario with other facial and gestural cues associated with referential gaze (e.g., pointing; Southgate, Van Maanen, & Csibra, 2007). Additionally, different placement of the cue may have affected differential attention to the areas of interest. It is also critical to address that word learning is incremental (Smith, Suanda, & Yu, 2014). Fast-mapping demonstrates one phase of language acquisition (McMurray et al., 2012; Swingley, 2010). Word learning paradigms capture the process at initial exposure, but differs from slow-mapping, which refers to how children slowly build up representations of words in the real world (Kucker et al., 2015; McMurray et al., 2012). It is important to find ways that can capture these incremental representations to help us better understand how fast-mapping relates to the process of word learning. Finally, it is important to investigate different developmental timepoints that span early to later language learning to understand how children learn new words from others.

6. Conclusion

This work explores how school-age children with typical development and children with ASD use referential gaze to support their word learning. We demonstrate that though gaze is a more salient cue than a directional control of an arrow, children in both groups use the directional information provided by gaze as well as an arrow to learn label-referent mappings. Yet, children attended to gaze differently than an arrow cue. First, children looked longer at the area of the cue in the gaze versus the arrow condition. Second, children looked back and forth between gaze and the referent more than they looked between the arrow and referent. Finally, children with higher versus lower parent-reported social communication abilities provided more semantic features with gaze specifically. While some of this evidence could support theoretical accounts of gaze as an intentional cue (Baldwin et al., 1993; Baron-Cohen et al., 1997; Csbira & Gergeley, 2009; Norbury et al., 2010), we note that prior experience with gaze relative to an arrow cue can also explain these differences. Importantly, recognizing this experience with gaze allows us to consider typically-developing individuals and individuals with ASD along a continuum that can provide new insights into children's learning.

Acknowledgements
We thank Kristine Onishi for valuable input throughout. We would also like to thank Corentin Montiel, Milva Venditti, Verona Soliman, Hida Caliskan, Lisa Bissett, Nicole Khammar, Renuka Giles, and Nowrin Hoque for assistance with testing and data collection. Thank you to the many organizations that helped with recruitment for the study including: The Montreal Children's Hospital, le Centre intégré universitaire de santé et de services sociaux du Centre-Sud-de-l'Île-de-Montréal (CIUSSS), Autism Speaks Canada Walk in Montreal, Summit School, Giant Steps, and Autisme Montréal. Finally, we greatly appreciate the families who participated.

Funding sources
This research was funded by doctoral awards to the first author from the Autism Research Training Program (Canadian Institutes of Health Research, Sinneave Family Foundation), and the Center for Research on Brain, Language, and Music.

References

Akechi, H., Kikuchi, Y., Tojo, Y., & Osanai, H. (2013). Brief report: Pointing cues facilitate word learning in children with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 43*(1), 230–235.

Akechi, H., Senju, A., Kikuchi, Y., Tojo, Y., Osanai, H., & Hasegawa, T. (2011a). Do children with ASD use referential gaze to learn the name of an object? An eyetracking study. *Research in Autism Spectrum Disorders, 5*, 1230–1242.

Akechi, H., Senju, A., Kikuchi, Y., Tojo, Y., Osanai, H., & Hasegawa, T. (2011b). Do children with ASD use referential gaze to learn the name of an object? An eyetracking study. *Research in Autism Spectrum Disorders, 5*, 1230–1242.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders.* Arlington, VA: American Psychiatric Publishing.

Axelsson, E. L., Churchley, K., & Horst, J. S. (2012). The right thing at the right time: Why ostensive naming facilitates word learning. *Frontiers in Psychology, 3*(88), 1–8. https://doi.org/10.3389/fpsyg.2012.00088.

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology, 100*(3), 603–617. https://doi.org/10.1348/000712608X377117.

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language, 20*(2), 395–418.

Baldwin, D. A., & Moses, L. J. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development, 10*(3), 309–329.

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development, 67*(6), 3135–3153.

Bani Hani, H., Gonzalez-Barrero, A. M., & Nadig, A. S. (2012). Children's referential understanding of novel words and parent labeling behaviors: Similarities across children with and without Autism Spectrum Disorders. *Journal of Child Language, 40*(5), 971–1002.

Bang, J. Y. (2017). *The role of intention in reading referential gaze: Implications for learning in typical development and in Autism Spectrum Disorder (Doctoral dissertation)*.

Bang, J.Y., Propensity Scores Open Repository [Internet]. [cited 2020 Feb 10]. Available from: https://github.com/janetybang/propensity_scores.

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., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Birmingham, E., Ristic, J., & Kingstone, A. (2012). Investigating social attention. In J. A. Burack, J. T. Enns, & N. A. Fox (Eds.). *Cognitive neuroscience, development, and psychopathology: Typical and atypical developmental trajectories of attention*. New York, NY: Oxford University Press.

Boraston, Z., & Blakemore, S. J. (2007). The application of eye-tracking technology in the study of autism. *The Journal of Physiology, 581*(3), 893–898. https://doi.org/10.1113/jphysiol.2007.133587.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language and Development, 15*, 17–29.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005.

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine, 11*(126), 1–8.

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates communicative development inventories: User's guide and technical manual (2nd edition)*. Baltimore: Paul H. Brookes.

Field, C., Lewis, C., & Allen, M. L. (2019). Referent selection in children with Autism Spectrum Condition and intellectual disabilities: Do social cues affect word-to object or word-to-location mappings? *Research in Developmental Disabilities, 91*, 103425. https://doi.org/10.1016/j.ridd.2019.05.004.

Gladfelter, A., & Goffman, L. (2017). Semantic richness and word learning in children with Autism Spectrum Disorder. *Developmental Science, 59*(4), e12543. https://doi.org/10.1111/desc.12543.

Hartley, C., Bird, L.-A., & Monaghan, P. (2019). Investigating the relationship between fast mapping, retention, and generalisation of words in children with Autism Spectrum Disorder and typical development. *Cognition, 187,* 126–138. https://doi.org/10.1016/j.cognition.2019.03.001.

Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science, 9*(2), 131–143. https://doi.org/10.1177/1745691613518076.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., ... Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development, 65*(3), 1–123.

Huang, C.-T., Heyes, C., & Charman, T. (2002). Infants' behavioral reenactment of "failed attempts": Exploring the roles of emulation learning, stimulus enhancement, and understanding of intentions. *Developmental Psychology, 38*(5), 840–855. https://doi.org/10.1037//0012-1649.38.5.840.

Itier, R. J., & Batty, M. (2009). Neural bases of eye and gaze processing: The core of social cognition. *Neuroscience & Biobehavioral Reviews, 33*(6), 843–863. https://doi.org/10.1016/j.neubiorev.2009.02.004.

Jing, W., & Fang, J. (2014). Brief Report: Do children with Autism gather information from social contexts to aid their word learning? *Journal of Autism and Developmental Disorders, 44*(6), 1478–1482. https://doi.org/10.1007/s10803-013-1994-5.

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.

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.

Kucker, S. C., McMurray, B., & Samuelson, L. K. (2015). Slowing down fast mapping: Redefining the dynamics of word learning. *Child Development Perspectives, 9*(2), 74–78. https://doi.org/10.1111/cdep.12110.

Leekam, S. (2016). Social cognitive impairment and autism: What are we trying to explain? *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*, 20150082. https://doi.org/10.1098/rstb.2015.0082.

Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-direction detection: A dissociation between geometric and joint attention skills in autism. *The British Journal of Developmental Psychology, 15*(1), 77–95.

Lord, C., Rutter, M., DiLavore, P. S., & Risi, S. (2002). *Autism diagnostic observation schedule: Manual*. Los Angeles: Western Psychological Services.

Loveland, K. A., & Landry, S. H. (1986). Joint attention and language in autism and developmental language delay. *Journal of Autism and Developmental Disorders, 16*(3), 335–349.

Luyster, R., & Lord, C. (2009). Word learning in children with autism spectrum disorders. *Developmental Psychology, 45*(6), 1774–1786. https://doi.org/10.1037/a0016223.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., ... Lord, C. (2009). The autism diagnostic observation schedule—Toddler module: A new module of a standardized diagnostic measure for Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders, 39*(9), 1305–1320.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001.

McDuffie, A. S., Yoder, P. J., & Stone, W. L. (2006). Labels increase attention to novel objects in children with autism and comprehension-matched children with typical development. Autism: *The International Journal of Research and Practice, 10*(3), 288–301. https://doi.org/10.1177/1362361306063287.

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, 54*(7), 745–753. https://doi.org/10.1111/jcpp.12073.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review, 119*(4), 831–877. https://doi.org/10.1037/a0029872.

Moore, C., Angelopoulos, M., & Bennett, P. (1999). Word learning in the context of referential and salience cues. *Developmental Psychology, 35*(1), 60–68.

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders, 20*(1), 115–128.

Nadig, A., Lee, I., Singh, L., Bosshart, K., & Ozonoff, S. (2010). How does the topic of conversation affect verbal exchange and eye gaze? A comparison between typical development and high-functioning autism. *Neuropsychologia, 48*(9), 2730–2739. https://doi.org/10.1016/j.neuropsychologia.2010.05.020.

Nadig, A., Seth, S., & Sasson, M. (2015). Global similarities and multifaceted differences in the production of partner-specific referential pacts by adults with Autism Spectrum Disorders. *Frontiers in Psychology, 6*, 1–14. https://doi.org/10.3389/fpsyg.2015.01888.

Nation, K., & Penny, S. (2008). Sensitivity to eye gaze in autism: Is it normal? Is it automatic? Is it social? *Development and Psychopathology, 20*(01), 79–97. https://doi.org/10.1017/S0954579408000047.

Norbury, C. F., Griffiths, H., & Nation, K. (2010). Sound before meaning: Word learning in autistic disorders. *Neuropsychologia, 48*(14), 4012–4019. https://doi.org/10.1016/j.neuropsychologia.2010.10.015.

Parish-Morris, J., Hennon, E. A., 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), 1265–1287.

Preissler, M. A., & Carey, S. (2005). The role of inferences about referential intent in word learning: Evidence from autism. *Cognition, 97*(1), 13–23. https://doi.org/10.1016/j.cognition.2005.01.008.

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

Roid, G. H., & Miller, L. J. (2013). Leiter international performance scale (third edition). Wood Dale, IL: Stoelting.

Rombough, A., Barrie, J. N., & Iarocci, G. (2012). Social cueing elicits a distinct form of visual spatial orienting. In J. A. Burack, J. T. Enns, & N. A. Fox (Eds.). *Cognitive neuroscience, development, and psychopathology: Typical and atypical developmental trajectories of attention* (pp. 221–250). New York, NY: Oxford University Press.

Rutter, M., Bailey, A., & Lord, C. (2003). Social communication questionnaire (SCQ). Western Psychological Services.

Secord, W., Wiig, E., Boulianne, L., Semel, E., & Labelle, M. (2009). *Évaluation clinique des notions langagières fondamentales—Version pour francophones du Canada*, Toronto: Pearson Canada Assessment.

Semel, E., Wiig, E., & Secord, W. (2003). *Clinical evaluation of language fundamentals (4th edition)*. San Antonio, TX: The Psychological Corporation.

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word-referent learning. *Trends in Cognitive Sciences, 18*(5), 251–258. https://doi.org/10.1016/j.tics.2014.02.007.

Southgate, V., Van Maanen, C., & Csibra, G. (2007). Infant pointing: Communication to cooperate or communication to learn? *Child Development, 78*(3), 735–740. https://doi.org/10.1111/j.1467-8624.2007.01028.x.

Sparrow, S. S., Cicchetti, D., & Balla, D. A. (2005). *Vineland adapative behavior scales (2nd edition)*. Minneapolis, MN: NCS Pearson, Inc.

Swingley, D. (2010). Fast mapping and slow mapping in children's word learning. *Language Learning and Development, 6*(3), 179–183. https://doi.org/10.1080/15475441.2010.484412.

Tenenbaum, E. J., Amso, D., Abar, B., & Sheinkopf, S. J. (2014). Attention and word learning in autistic, language delayed and typically developing children. *Frontiers in Psychology, 5*, 490. https://doi.org/10.3389/fpsyg.2014.00490.

Uljarević, M., Baranek, G., Vivanti, G., Hedley, D., Hudry, K., & Lane, A. (2017). Heterogeneity of sensory features in autism spectrum disorder: Challenges and perspectives for future research. *Autism Research, 10*(5), 703–710. https://doi.org/10.1002/aur.1747.

Venker, C. E., Kover, S. T., & Weismer, S. E. (2016). Brief Report: Fast mapping predicts differences in concurrent and later language abilities among children with ASD. *Journal of Autism and Developmental Disorders, 46*(3), 1118–1123. https://doi.org/10.1007/s10803-015-2644-x.

Vernetti, A., Senju, A., Charman, T., Johnson, M. H., & Gliga, T. (2018). Simulating interaction: Using gaze-contingent eye-tracking to measure the reward value of social signals in toddlers with and without autism. *Developmental Cognitive Neuroscience, 29*, 21–29. https://doi.org/10.1016/j.dcn.2017.08.004.

Vivanti, G., McCormick, C., Young, G. S., Abucayan, F., Hatt, N., Nadig, A., ... Rogers, S. J. (2011). Intact and impaired mechanisms of action understanding in autism. *Developmental Psychology, 47*(3), 841–856. https://doi.org/10.1037/a0023105.

Vlach, H. A., & DeBrock, C. A. (2019). Statistics learned are statistics forgotten: Children's retention and retrieval of cross-situational word learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(4), 700–711. https://doi.org/10.1037/xlm0000611.

Wittke, K., Mastergeorge, A. M., Ozonoff, S., Rogers, S. J., & Naigles, L. R. (2017). Grammatical language impairment in Autism Spectrum disorder: Exploring language phenotypes beyond standardized testing. *Frontiers in Psychology, 8*(320), 594. https://doi.org/10.3389/fpsyg.2017.00532.