Cognition 183 (2019) 181–191

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

Cross-situational and ostensive word learning in children with and without autism spectrum disorder

Courtney E. Venker[†,*]

Waisman Center, University of Wisconsin-Madison, USA

ARTICLE INFO

Keywords:
Cross-situational learning
Word learning
Fast mapping
Statistical learning
Autism spectrum disorder
Language development

ABSTRACT

Numerous experimental studies have shown that infants and children can discover word meanings by using co-occurrences between labels and objects across individually ambiguous contexts—a phenomenon known as cross-situational learning. Like typically developing children, high-functioning school aged children with autism spectrum disorder (ASD) are capable of cross-situational learning. However, it is not yet clear whether cross-situational learning is similarly available to children with ASD who are younger and show a broader range of language and cognitive abilities. Using eye-tracking methodology, the current study provided the first evidence that preschool and early school-aged children with ASD can rely on cross-situational statistics to learn new words. In fact, children with ASD learned as well as typically developing children with similar vocabulary knowledge. In both groups, the children with the highest cross-situational learning accuracy were those who showed the best familiar word processing skills. Surprisingly, children in both groups learned words equally well in the cross-situational task and an ostensive word-learning task, which presented only a single label-object pairing at a time. In combination, these results point to similarities in the word learning abilities available to typically developing children and children with ASD.

1. Introduction

Word learning is a crucial part of language development, but identifying the meanings of new words is not always easy. This is true even in the case of object nouns—arguably one of the most straightforward cases of mapping word (label) to meaning (object). Determining correct label-object mappings can be difficult because there are often many things in the environment that a novel word could describe (Quine, 1960). Although children sometimes have access to adult cues that explicitly identify which object is being labeled (e.g., pointing, gaze), they also encounter many instances in which explicit cues are not available. In the absence of such cues, how might children determine the meanings of object nouns in a busy world filled with many things to see and hear? One possibility is that children discover word meanings by using co-occurrences between labels and objects across individually ambiguous contexts—a phenomenon known as cross-situational learning (Smith & Yu, 2008; Suanda, Mugwanya, & Namy, 2014; Yu & Smith, 2007). Even when the links between objects and labels are not immediately clear, these links may become clear as the same object and label co-occur over time. For example, if the words ‘duck’ and ‘ball’ are first presented in the absence of explicit cues that identify their referents, the learner may not yet know which label describes which object. However, it is possible to determine the correct label-object associations by attending to co-occurrences over time (e.g., that the round object is consistently visible when the word ‘ball’ is produced).

Experimental studies have shown that adults, young children, and even infants are capable of using this type of cross-situational information to learn the meanings of new words (Smith & Yu, 2008; Suanda et al., 2014; Vouloumanos & Werker, 2009; Yu & Smith, 2007). In a landmark study by Smith and Yu (2008), typically developing 12- and 14-month-old infants were exposed to a series of individually ambiguous trials, each of which presented two novel labels (e.g., *bosa, gasser*) and two unfamiliar objects. Within a given trial, no information was available to indicate which label-object associations were correct. The only way for infants to discover correct associations was by picking up on co-occurrence statistics—namely, which object was consistently visible when a given label was presented. Each label-object pairing occurred a total of 10 times during teaching. After less than 4 min, infants’ eye movements provided evidence that they had learned the new words on the basis of cross-situational statistics alone.

Over the past decade, cross-situational learning has continued to

* Address: Dept. of Communicative Sciences and Disorders, 1026 Red Cedar Road, Room 216, East Lansing, MI 48824, USA
E-mail address: cvenker@msu.edu.
† Present address: Michigan State University, USA.

https://doi.org/10.1016/j.cognition.2018.10.025
Received 27 July 2016; Received in revised form 25 September 2018; Accepted 29 October 2018
Available online 20 November 2018

receive a great deal of empirical attention in typically developing infants, children, and adults, primarily in terms of the factors that affect it and the learning mechanisms that underlie it (Medina, Snedeker, Trueswell, & Gleitman, 2011; Suanda & Namy, 2012; Trueswell, Medina, Hafri, & Gleitman, 2013; Vlach & Johnson, 2013; Yu & Smith, 2012). One current topic of considerable debate pertains to precisely how cross-situational learning is accomplished. Do learners gradually accumulate statistical associations between labels and objects (Smith & Yu, 2008; Yu & Smith, 2007; Zettersten, Wojcik, Benitez, & Saffran, 2018), or do they propose and subsequently verify (or abandon) a single word-referent pairing (Smith, Smith, & Blythe, 2009; Trueswell et al., 2013; Woodard, Gleitman, & Trueswell, 2016)? Or, does their learning falls somewhere in between associative learning and hypothesis testing, depending on contextual factors and individual abilities (MacDonald, Yurovsky, & Frank, 2017; Yurovsky & Frank, 2015)? Although this is a critically important issue, particularly in atypical development, the current study was not designed to differentiate among these possibilities and thus we do not discuss them extensively here. However, we return to the issue of gradual associative learning versus hypothesis testing in the Discussion, presenting potential interpretations of the current findings and avenues for future research.

Despite our growing understanding of cross-situational learning in typical development, we know very little about cross-situational learning in children with autism spectrum disorder (ASD)—a neurodevelopmental disorder characterized by deficits in social communication, repetitive behaviors and restricted interests (American Psychiatric Association, 2013), and wide variability in social, language, and cognitive abilities. Because of their unique behavioral profile, children with ASD present an opportunity to better understand the skills that support cross-situational learning, as well as the role that cross-situational learning plays in language development more generally. Investigating cross-situational learning in children with ASD will also shed light on the role of auditory-visual integration in statistical word learning. Numerous studies have identified differences in the way in which individuals with ASD integrate auditory and visual information (Iarocci & McDonald, 2006; but see Grossman, Schneps, & Tager-Flusberg, 2009). For example, visual information may have less of an influence on heard speech individuals with ASD than in individuals without ASD (Irwin, Tornatore, & Brancazio, & Whalen, 2012). In addition, individuals with ASD show "reduced multisensory speech perception for matched audiovisual stimuli" (Woynaroski et al., 2013, p. 2900), as well as a decreased preference for auditory-visual synchrony (Bebko, Weiss, Demark, & Gomez, 2006; Grossman, Steinhard, Mitchell, & McIlvane, 2015). Although it is not typically described in this way, associating auditory labels with visual objects across contexts can be considered a form of auditory-visual integration. Thus, cross-situational learning might be particularly difficult for children with ASD compared to TD children, particularly for children with ASD and co-occurring language delays.

To our knowledge, only one published study has investigated cross-situational learning in children with ASD. McGregor, Rost, Arenas, Farris-Trimble, and Stiles (2013) found that high-functioning 11-year-olds with ASD used cross-situational statistics to learn new words, thereby providing the first evidence that cross-situational learning mechanisms are available to at least some children with ASD. Furthermore, the children who were better cross-situational learners were those who had acquired stronger vocabulary skills outside the lab setting, leading McGregor and colleagues to hypothesize that cross-situational learning is more difficult for children with ASD who have weaker language abilities overall.

The findings of the study by McGregor et al. (2013) advance our understanding of cross-situational learning in several ways. First, the fact that children with ASD learned words based on cross-situational statistics alone demonstrates that cross-situational learning can operate successfully in the absence of strong social communication skills. Second, the correlation with vocabulary skills (but not age) suggests that cross-situational learning may play an important role in vocabulary development for children with ASD. However, the study by McGregor and colleagues included only older children with age-appropriate language and cognitive skills, leaving open the question of whether cross-situational learning is similarly available to children with ASD who are younger and show a broader range of language and cognitive abilities. In addition, it is not yet clear whether the link between cross-situational learning and language is specific, or whether it can be better explained by overall cognitive skills.

One critical issue that remains unexplored across both typical development and ASD is how cross-situational learning compares to word learning in ostensive (unambiguous) contexts that present only one word and one object at a time. From an information processing perspective, it is logical that tracking and integrating multiple co-occurrences across ambiguous contexts may be more difficult than associating a single word and with a single object. However, we are not aware of any published studies that have directly compared cross-situational and ostensive word learning in the same group of children. This is an important issue to investigate because it has implications for understanding how easily children can access particular learning mechanisms, what additional cognitive demands are placed on children in certain contexts, and what word-learning strategies children are most likely to rely on in everyday situations.

1.1. The current study

Research on language development in children with ASD has historically focused on characterizing these children's existing language skills—namely, describing the delays they experience in certain domains and at certain points in development. The current study is part of a new line of work that emphasizes not what language skills children with ASD have acquired, but *how* they acquire these skills (Arunachalam & Luyster, 2016; Mayo & Eigsti, 2012; McGregor, et al., 2013; Naigles, Kelty, Jaffery, & Fein, 2011; Venker, Kover, & Ellis Weismer, 2016). In the current study, we focus not only on the words that children know, but also on the word-learning abilities these children can access when exposed to words they have never heard before. Adopting this type of learning-based approach is beneficial because it allows us to answer new questions about language development in children with ASD, including: what learning mechanisms are intact, and where are the points of breakdown? In addition, focusing on a population with considerable variation in language and cognitive skills will shed light on the mechanisms that support word learning more generally, thereby complementing investigations of TD children.

The current study investigated cross-situational word learning in children with ASD (4–7 years old) and TD children (2–7 years old), matched on vocabulary knowledge. Children took part in a cross-situational eye-tracking task modeled after the task developed by Smith and Yu (2008). Each teaching trial presented two novel labels and two objects, with no information available within a single trial regarding which label described which object. Children also completed an ostensive word-learning task that presented label-object links explicitly (i.e., one label and one object at a time) and therefore did not require children to track multiple co-occurrences to determine word meaning.

We had three goals. The first was to determine whether young children with ASD were capable of cross-situational learning, and whether their learning differed from TD children matched on vocabulary knowledge. Given the general difficulties children with ASD have learning words (Charman, Drew, Baird, & Baird, 2003; Luyster, Lopez, & Lord, 2007; McDuffie, Kover, Hagerman, & Abbeduto, 2012) and integrating auditory and visual information (e.g., Foss-Feig et al., 2010; Iarocci & McDonald, 2006), as well as the fact that our sample included children with language delays, we hypothesized that children with ASD would learn words in the cross-situational task but that they would show deficits relative to TD children. Our second goal was to determine whether individual differences in cross-situational learning were related to children's language skills in either group. Given previous evidence of a relationship between cross-situational learning and language (McGregor et al., 2013; Smith & Yu, 2013), we hypothesized that children with stronger language skills would be better cross-situational learners—regardless of diagnostic group. Such a finding would point to similarities in the skills that may support cross-situational learning across different populations. Our third goal was to determine whether cross-situational word learning was more difficult than ostensive word learning for TD children or children with ASD. We hypothesized that children in both groups would learn words more easily in the ostensive word-learning task than in the cross-situational task, since the ostensive task presented only a single label and object at a time (whereas the cross-situational task presented two labels and objects in each trial).

2. Method

2.1. Participants

The full sample included 24 TD children and 27 children with ASD. Three TD children were excluded: one due to eye tracker error and two because their performance on the cross-situational task was over 2 standard deviations below the mean, indicating potential outliers. Nine children with ASD were excluded: 2 due to eye tracker error, 6 due to excessive missing data,[1] and one because his performance on the ostensive word-learning task was over 2 standard deviations below the mean. The final participant sample included 21 TD children (15 males) and 18 children with ASD (all male). A two-proportions z-test indicated that the proportion of males in the ASD group (18/18) differed significantly from the proportion of males in the TD group (15/21; $z = -2.47$, $p = .01$); this was unsurprising, given the increased prevalence of ASD in males (Christensen et al., 2016). The TD and ASD groups were matched on vocabulary knowledge (growth scale values from the Peabody Picture Vocabulary Test, Fourth Edition [PPVT[2]; Dunn & Dunn, 2006], $t(37) = 0.53$, $p = .60$, $d = 0.17$). Participant characteristics are presented in Table 1.

There were no inclusionary or exclusionary criteria regarding language or cognitive abilities for children in the ASD group. The children with ASD showed a range of spoken language abilities, which were categorized as part of the ADOS-2: no spoken words ($n = 1$); some spoken words ($n = 2$); 'phrase speech' (i.e., flexible 3-word phrases; $n = 5$); and 'fluent speech' (i.e., utterances containing multiple independent clauses; $n = 10$). Three children in the ASD group were reported to have experienced a loss of skills during early childhood (i.e., developmental regression). Fourteen children in the ASD group had ever received speech-language intervention.

2.2. General procedure

Children took part in a single laboratory visit lasting 2–3 h. Parents provided written consent for their child's participation. All procedures were prospectively approved by the university Institutional Review Board. The cross-situational task took place at the beginning of the visit, followed by the standardized assessments and the ostensive word-learning task. Two additional experimental tasks (a visual recognition memory task and a visual orienting task) were also administered but are not reported here.

Table 1
Participant characteristics.

	TD Group (n = 21) *M* (*SD*) range	ASD Group (n = 18) *M* (*SD*) range	Group Difference
Age in months	58 (21) 31–95	76 (17) 48–95	p = .007
PPVT Growth Scale Values	140 (23) 106–175	135 (27) 78–174	p = .603
PPVT Age Equivalent Scores	76 (25) 42–115	70 (26) 22–113	p = .453
PPVT Standard Scores	120 (10) 99–143	94 (20) 62–122	p < .001
Brief IQ	120 (13) 93–145	95 (19) 60–133	p < .001
Autism severity	–	7 (2) 5–10	–
SCQ Total Score	5 (3) 0–14	–	–

Note. PPVT = Peabody Picture Vocabulary Test, 4th Edition. Brief IQ was measured by the Leiter International Performance Scale-Revised. Autism severity was measured by comparison scores on the Autism Diagnostic Observation Schedule, 2nd Edition. SCQ = Social Communication Questionnaire. TD = typically developing. ASD = autism spectrum disorder.

The PPVT assessed vocabulary knowledge. The PPVT produced a raw score, growth scale value, age equivalent, and standard score. Growth scale values were used for group matching and in the analyses. As described in the PPVT manual, growth scale values are PPVT raw scores that are transformed to be on an equal-interval scale. Unlike standard scores, which reflect how a child's vocabulary compares to that of their peers, PPVT growth scale values represent absolute level of receptive vocabulary ability. The Leiter International Performance Scale-Revised (Leiter; Roid & Miller, 2002) assessed nonverbal cognitive abilities. Four subtests from the Visualization and Reasoning Battery were administered: Figure Ground, Form Completion, Sequential Order, and Repeated Patterns. Compilation of these subtest scores yielded a Brief IQ.

Parents of children in the TD group completed the Social Communication Questionnaire Lifetime form to screen for ASD (Rutter, Bailey, & Lord, 2003). The Social Communication Questionnaire manual recommends that children with scores > 15 receive further evaluation for ASD; all children in the TD group scored below this cutoff. All children in the ASD group were reported by their parents to have received a diagnosis of ASD, Autism, Asperger's Disorder, or Pervasive Developmental Disorder-Not Otherwise Specified, from a qualified medical professional. The Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2; Lord et al., 2012), a semi-structured, behavioral diagnostic measure for ASD, was administered to children in the ASD group to confirm ASD diagnosis and measure autism severity (Gotham, Pickles, & Lord, 2009). It was not possible to complete the ADOS-2 for two children due to challenging behaviors, but these children were retained because they had existing ASD diagnoses and showed behaviors consistent with ASD during the evaluation.

2.3. Experimental tasks

2.3.1. Procedure

Children participated in two experimental tasks: a cross-situational word-learning task and an ostensive word-learning task. The two tasks were closely aligned but differed in one crucial way: whether label-object pairings were presented ambiguously (the cross-situational learning task) or explicitly (the ostensive word-learning task). Both tasks taught 4 novel labels and 4 novel objects. Each label-object

[1] The 6 children with ASD excluded due to excessive missing data failed to contribute at least 2 valid trials in one or both of the experimental tasks. These 6 children had lower PPVT growth scale values than the 18 children with ASD who were retained, $t(22) = 2.10$, $p = .047$, indicating that the children who showed the most extreme inattention to these language-based tasks were those with the weakest language skills. Interestingly, children who were excluded did not differ from children who were retained in age, $t(22) = 0.27$, $p = .79$, or Brief IQ, $t(22) = 1.64$, $p = .115$.

[2] PPVT growth scale values were used because they represent children's vocabulary knowledge on an equal-interval scale and are thus more appropriate than raw scores for statistical analyses.

Table 2
Design of word-learning tasks.

	Cross-Situational Word-Learning Task	Ostensive Word-Learning Task
Familiarization Phase	4 trials were presented Each trial lasted 4 s and presented 2 objects and 2 labels 3 familiar label-object pairs were presented: *ball, cup, shoe*	4 trials were presented Each trial lasted 2 s and presented 1 object and 1 label 3 familiar label-object pairs were presented: *dog, car, book*
Teaching Phase	20 trials were presented Each trial lasted 4 s and presented 2 objects and 2 labels Total of 4 novel objects Total of 4 novel labels: *toma, subo, deepu, modi* Each label-object pair was presented 10 times	40 trials were presented Each trial lasted 2 s and presented 1 object and 1 label Total of 4 novel objects Total of 4 novel labels: *bosa, coro, manu, peri* Each label-object pair was presented 10 times
Test Phase	25 trials were presented: 16 novel, 9 familiar	25 trials were presented: 16 novel, 9 familiar

pairing was presented 10 times during familiarization (see Table 2 for additional details). The cross-situational task presented 4-second trials with 2 objects and 2 labels, whereas the ostensive task presented 2-second trials with 1 object and 1 label.

Children sat in a chair in front of an eye tracker in a soundproof booth. A trained research assistant remained in the booth during the task to provide reassurance and, if necessary, to help the child remain within the tracking range (e.g., by placing hands lightly on the child's shoulders). The research assistant directed her eyes at the floor to ensure the eye tracker did not capture her eye gaze. In a few cases, children sat on their parent's lap to increase their comfort level. In such cases, the parent was instructed to remain silent and, if possible, not to interact with the child. Parents wore opaque glasses to prevent them from viewing the screen and inadvertently influencing their child's behavior.

2.3.2. Equipment and calibration

Both word-learning tasks were conducted on a Tobii T60 XL Eye Tracker (Tobii). Experimental stimuli were presented using E-Prime 2.0 software with E-Prime Extensions for Tobii. Visual stimuli were presented on a 24-inch wide-screen monitor with a screen resolution of 1920 x 1200 pixels. Auditory stimuli were presented at a level of 65 dB through built-in speakers on the eye tracker. The sampling rate was 60 Hz, meaning that gaze location was sampled 60 times per second (i.e., every 16.7 ms). The Tobii determined gaze location by creating reflections on the cornea and pupil using near infrared illumination. Two image sensors embedded in the lower panel of the Tobii monitor recorded images of children's eyes and the patterns of cornea and pupil reflection. Standard internal processing algorithms estimated the position of the eye and its location on the screen. All participants completed a 5-point infant calibration in Tobii Studio immediately before each experimental task. The infant calibration option was used because it engaged attention automatically by presenting an animated, moving stimulus at each calibration point, along with a corresponding sound (a shaking rattle with a brief musical tune). The infant calibration procedure allowed the examiner to present an intervening stimulus between calibration points, which brought children's attention back to the screen if they had become distracted. Before starting a task, the examiner ensured that children's gaze fell generally within the boundaries of the calibration points. Individual calibration points were re-administered as needed.

2.3.3. Cross-situational word-learning task

The cross-situational task consisted of three phases: familiarization, teaching, and test (see Table 2 for additional details about the experimental design). Familiarization trials were included to familiarize participants with the task. Each familiarization trial presented 2 familiar objects and 2 familiar labels (e.g., "*Ball. Shoe.*"). Teaching trials presented 2 unfamiliar objects and 2 novel labels (e.g., "*Toma. Subo.*"). No carrier phrases were provided during teaching. Children were exposed to each label and object a total of 10 times during teaching. Each label-object pair occurred either 3 or 4 times with every other label-object pair. In both the familiarization and teaching phases, the first label was presented 500 ms into the trial, and the second label was presented 2000 ms into the trial. There was no information available within a single familiarization or teaching trial to indicate which label described which object (e.g., gaze cues).

Test trials tested either a familiar word or a novel word. In familiar test trials, children saw two of the familiar objects introduced during the familiarization phase and were asked about one of them (e.g., *Where's the ball? Do you like it?*). In novel test trials, children saw two of the novel objects that had been taught during the teaching phase and were asked about one of them (e.g., "*Where's the coro? Do you see it?*"). All test trials used the same carrier phrase (*Where's the_?*). Each target image appeared with all novel foils at least once during the test phase; see Appendix A for additional details). Test trials began with 1000 ms of silence, followed by the auditory stimulus; the object label was presented 2000 ms into the trial. Images remained on the screen for the full 5000 ms trial.

Stimuli in all phases were counterbalanced for side and order of presentation. The full task lasted approximately 4 min. To maximize children's attention, short movies (animated musical clips of nature scenes) were interspersed every 4 – 5 trials. To ensure that children's performance was not driven by specific aspects of the experimental design, two versions of the task were created that differed in certain ways (e.g., label-object associations were switched, trial order was altered). Because there were no significant differences in accuracy across the two versions, $t(37) = -0.52$, $p = .603$, data were collapsed in the subsequent analyses.

Images of familiar objects were obtained through an online image search. Images of novel objects were modeled after those used by Smith and Yu (2008) and created in Microsoft PowerPoint (see Fig. 1). To ensure that the images did not resemble namable real objects, candidate images were pilot tested with 10 adults blind to the study hypotheses. Objects were eliminated if they were consistently named (e.g., one image was consistently labeled as *star)*. Final images were cropped, placed on a 375 x 825 pixel gray square, and presented on a black background (see Fig. 1). Auditory stimuli were recorded by an adult female using child-directed speech.

2.3.4. Ostensive word-learning task

The ostensive word-learning task also consisted of a familiarization, teaching, and test phase (see Table 2). This task mirrored the cross-situational learning task with one critical difference: label-object links were presented explicitly. Familiarization trials presented 1 familiar object and 1 familiar label (e.g., "*Dog.*"). Teaching trials presented 1 unfamiliar object and 1 novel label (e.g., "*Bosa.*"). In both the familiarization and teaching phases, the object label was presented 500 ms into the trial. The design of the test phase was identical to that in the cross-situational task. Stimuli in all phases were counterbalanced for side and order of presentation, and attention-getter movies were interspersed every 4 – 5 trials. The full task lasted approximately 4 min (the same length as the cross-situational word-learning task). Two versions of the task were created. Because there were no significant differences in accuracy across the two versions, $t(36) = 1.18$, $p = .245$, data were collapsed in the subsequent analyses. One additional child

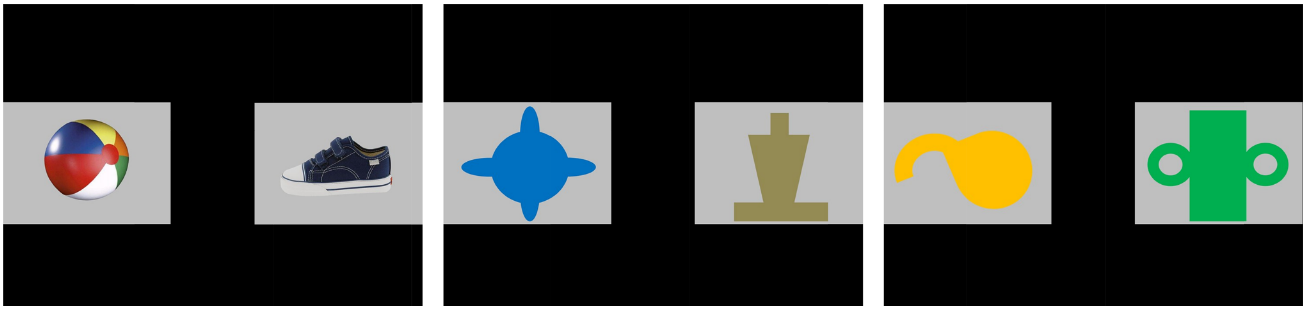

Fig. 1. Sample Visual Stimuli. Sample visual stimuli from a familiar word test trial, a cross-situational word-learning test trial, and an ostensive word-learning test trial.

with ASD was excluded from the ostensive word-learning task analyses due to excessive missing data, leaving $n = 17$ children in the ASD group. As in the cross-situational task, familiar images were obtained through an online image search, and novel objects were piloted with naïve adults and created in Microsoft PowerPoint. Final images were cropped, placed on a 375 x 825 pixel gray square, and presented on a black background. Auditory stimuli were recorded by an adult female using child-directed speech. Mean phonotactic probability of the novel labels did not differ between the cross-situational and ostensive word-learning tasks, $t(6) = 1.86$, $p = .112$ (Vitevitch & Luce, 2004).

2.3.5. *Processing and cleaning eye-gaze data*

Data processing for the two experimental tasks was identical. Areas of interest (AOIs) for the test trials were defined by the outer edges of the grey boxes containing the images, plus 10 pixels (see Fig. 1). Each time frame during the experimental trial was assigned a code based on whether the child was looking at the named image, looking at the unnamed image, or looking at neither image (classified as missing data). Following prior work (e.g., Mahr, McMillan, Saffran, Ellis Weismer, & Edwards, 2015; Wass, Smith, & Johnson, 2013), segments of missing data up to 150 ms were interpolated if the child was looking at the same AOI before and after the missing data occurred. The analysis window was 200–1800 ms after noun onset during the test phase. In the cross-situational task, 2.33% of time frames were interpolated during this analysis window for both the TD group and the ASD group. In the ostensive task, 1.69% of time frames were interpolated for the TD group, and 2.52% of time frames were interpolated during the analysis window in the ASD group.

Following interpolation, test trials with less than 50% looking time to the images during the analysis window were eliminated because they were considered to contain too little data to provide a valid measure of performance.[3] The familiar test trials in each task were combined, increasing the maximum number of familiar trials to 18. On average, TD children contributed 12 familiar test trials ($SD = 3$), 9 cross-situational test trials ($SD = 4$), and 9 ostensive test trials ($SD = 4$). On average, children with ASD contributed 11 familiar test trials ($SD = 4$), 8 cross-situational test trials ($SD = 5$), and 8 ostensive test trials ($SD = 4$). The number of test trials contributed did not differ significantly by group for familiar trials, $t(37) = 0.52$, $p = .607$, cross-situational trials, $t(37) = 0.90$, $p = .372$, or ostensive trials, $t(36) = 0.58$, $p = .567$. Neither age nor PPVT growth scale value was significantly correlated with the number of familiar trials, cross-situational trials, or ostensive trials contributed in either group (all $ps > 0.270$).

We also examined the proportion of data contributed by each group during the analysis window of the test trials across both tasks. The ASD group looked at the images 79.56% of the time in the cross-situational task ($SD = 11.14\%$) and 84.23% of the time in the ostensive task ($SD = 7.70\%$). The TD group looked at the images 87.25% of the time in the cross-situational task ($SD = 4.05\%$) and 85.81% of the time in the ostensive task ($SD = 6.20\%$). On average, the TD group looked at the images significantly more than the ASD group, regardless of condition ($p < .01$); there was no significant main effect of condition and no condition x group interaction ($ps > 0.07$).

2.3.6. *Defining accuracy*

We first examined looking behaviors during the first 2 s of the test trials, before the object label was presented. As expected, children spent approximately the same amount of time looking at each image during baseline in familiar trials, $t(38) = 1.37$, $p = .180$, cross-situational trials, $t(38) = -0.98$, $p = .333$, and ostensive trials, $t(37) = 1.11$, $p = .274$. Because there were no significant baseline effects, the analyses focused on accuracy during the analysis window. Like many prior eye-gaze studies of word learning and word recognition (Tenenbaum, Amso, Righi, & Sheinkopf, 2017; Vouloumanos & Werker, 2009; Yu & Smith, 2011), our dependent variable of interest focused on the extent to which children looked at the object that was named in each test trial, as opposed to the object that was not named. Following standard procedures (Fernald, Zangl, Portillo, & Marchman, 2008), accuracy was defined as the proportion of looks to the named image during the analysis window, divided by looks to both images. Looks away from the images were not incorporated into the denominator to ensure that a decrease in accuracy could not be attributed to a general tendency towards off-task behavior.

3. Results

Our first goal was determine whether young children with ASD could learn words by tracking cross-situational statistics, and whether their learning differed from TD children matched on vocabulary knowledge. We hypothesized that children with ASD would learn words in the cross-situational task, but that they would show poorer learning than TD children, given the general difficulties children with ASD have learning words (Charman et al., 2003; Luyster et al., 2007; McDuffie et al., 2012) and integrating auditory and visual information (e.g., Foss-Feig et al., 2010; Iarocci & McDonald, 2006). As a first step, we examined children's comprehension of the familiar words to ensure that the eye-tracking task was working as intended. Because each test trial displayed two images on the screen at a time, the likelihood of looking at the named image by chance was 0.50. Mean accuracy in familiar word trials was 0.77 ($SD = 0.13$, range = 0.56–0.95) in the TD group and 0.75 ($SD = 0.11$, range = 0.55 – 0.94) in the ASD group. As expected, mean accuracy was significantly higher than chance in both the TD group, $t(20) = 9.58$, $p < .001$, $d = 2.08$, and the ASD group, $t(17) = 9.56$, $p < .001$, $d = 2.27$, and did not differ between the two groups, $t(37) = 0.55$, $p = .585$, $d = 0.17$.

[3] There are currently no standardized criteria for trial-level cleaning, and criteria vary widely across studies. In selecting the 50% criterion, our goal was to maximize trial retention while ensuring that the retained trials included adequate data to represent the construct of interest. In addition, we did not want to overly penalize children for occasional looks away from the experimental stimuli.

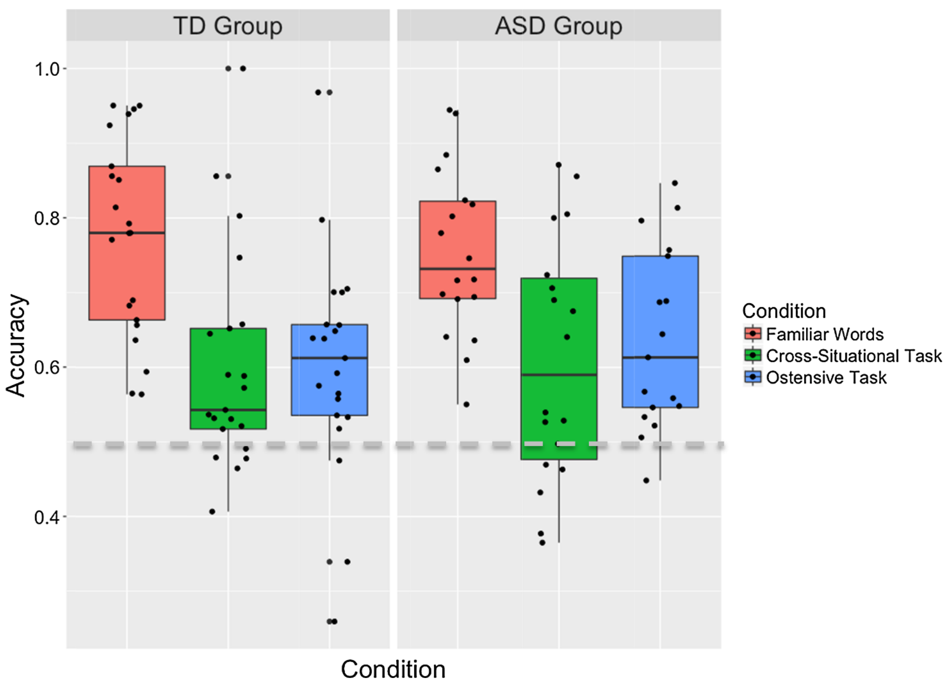

Fig. 2. Experimental task performance. Accuracy was the proportion of looking to the named image. TD = typically developing. ASD = autism spectrum disorder. The dark horizontal bars indicate the median. The upper and lower hinges indicate the first and third quartiles (i.e., 25th and 75th percentile). The upper and lower whiskers indicate the values within 1.5 IQR of the upper and lower hinges, respectively. (IQR = interquartile range, the distance between the first and third quartiles). The grey dotted line indicates chance (0.5).

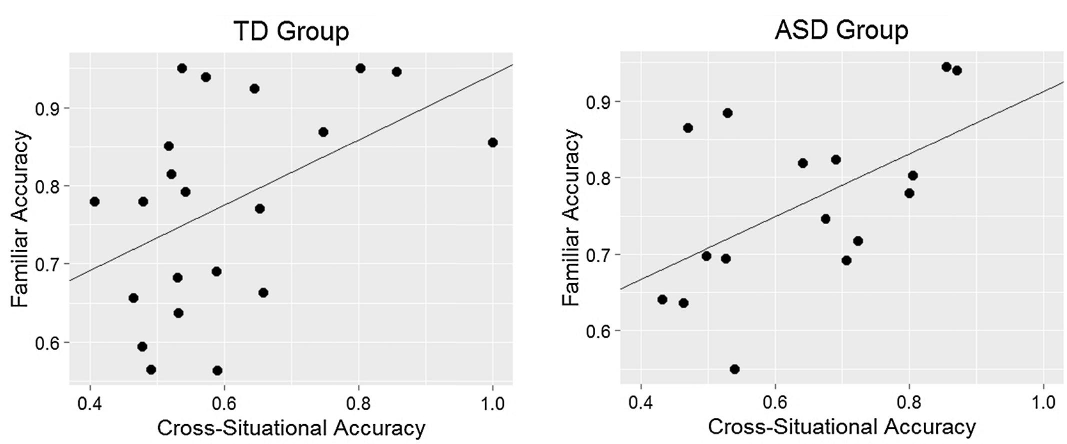

Fig. 3. Relationship between familiar word processing accuracy and cross-situational learning accuracy. Accuracy was the proportion of looking to the named image. TD = typically developing. ASD = autism spectrum disorder.

Mean accuracy in the cross-situational task was 0.60 in the TD group ($SD = 0.15$, range = 0.41–1.0) and 0.61 in the ASD group ($SD = 0.16$, range = 0.37–0.87; see Fig. 2). To assess cross-situational learning, we tested whether accuracy differed from chance (0.50). Accuracy was significantly higher than chance in the TD group, $t(20) = 3.16$, $p = .005$, $d = 0.67$, and in the ASD group, $t(17) = 2.83$, $p = .012$, $d = 0.69$. Contrary to predictions, there were no significant group differences in cross-situational learning, $t(37) = -0.18$, $p = .570$, $d = 0.06$.

Our second goal was to determine whether individual differences in cross-situational learning were related to children's language skills. We hypothesized that children with stronger language skills would be better cross-situational learners—regardless of diagnostic group. To address this question, we tested correlations between cross-situational learning and two different measures of language: vocabulary knowledge (PPVT growth scale values) and familiar word processing accuracy. Vocabulary knowledge was not significantly correlated with cross-situational learning in the TD group, $r(19) = 0.37$, $p = .103$, or in the ASD group, $r(16) = 0.46$, $p = .057$, though these correlations were marginal.[4] Familiar word processing accuracy was significantly correlated with cross-situational learning in the TD group, $r(19) = 0.46$, $p = .034$, and in the ASD group, $r(16) = 0.60$, $p = .009$, indicating that the children who were most adept at processing familiar words also showed the highest cross-situational learning accuracy (see Fig. 3). Because neither language measure accounted for age, we also tested the relationship between age and cross-situational learning. Age was significantly correlated with cross-situational learning in the TD group, $r(19) = 0.71$, $p < .001$; the correlation in the ASD group was marginal, $r(16) = 0.39$, $p = .112$. Cross-situational learning in the ASD group was not significantly correlated with Brief IQ, $r(16) = 0.25$, $p = .314$, or comparison scores from the ADOS-2 (which measured autism severity),

[4] The fact that 6 children with ASD with weak vocabulary knowledge were excluded limited the variability in PPVT scores in this group. This limited variability may have been a contributing factor to the non-significant (but marginal) correlation between cross-situational learning accuracy and vocabulary knowledge ($p = .057$).

$r(16) = -0.20$, $p = .434$. The correlation between Brief IQ and cross-situational learning in the TD group was marginal and negative, $r(19) = -0.35$, $p = .118$.

Our third goal was to determine whether cross-situational word learning was more difficult than ostensive word learning for TD children or children with ASD. We hypothesized that children in both groups would learn words more easily in the ostensive word-learning task than in the cross-situational task. We first examined children's performance in the ostensive word-learning task to ensure that children had learned the new words. Mean accuracy in the ostensive task was 0.60 in the TD group ($SD = 0.15$, range = 0.26–0.97) and 0.64 in the ASD group ($SD = 0.12$, range = 0.45–0.85; see Fig. 2). As expected, accuracy was significantly higher than chance in the TD group, $t(20) = 3.21$, $p = .004$, $d = 0.67$, and in the ASD group, $t(16) = 4.63$, $p < .001$, $d = 1.17$. As in the cross-situational task, word learning accuracy did not significantly differ between the TD and ASD groups, $t(36) = -0.74$, $p = .769$, $d = 0.29$. Next, we tested whether either group showed better cross-situational word learning than ostensive word learning. Contrary to predictions, accuracy across the two word-learning tasks did not differ in either the TD group, $t(20) = 0.08$, $p = .470$, $d = 0$, or the ASD group, $t(16) = 0.33$, $p = .374$, $d = 0.14$.

The ostensive task took place at the end of all sessions. Thus, it was possible that children attended less to the objects during this task than during the cross-situational task. To investigate this possibility, we examined the proportion of time children looked at the images across both teaching phases. In the cross-situational teaching phase, children in the TD group looked at the images 61.03% of the time ($SD = 13.30\%$), and children in the ASD group looked at the images 56.80% of the time ($SD = 18.14\%$). In the ostensive teaching phase, children in the TD group looked at the images 68.16% of the time ($SD = 20.17\%$), and children in the ASD group looked at the images 57.71% of the time ($SD = 21.46\%$). In the ASD group, the amount of time spent looking at the images did not significantly differ between the cross-situational teaching phase or the ostensive teaching phase, $p = .753$. In the TD group, children looked significantly more at the images in the ostensive teaching phase than in the cross-situational teaching phase, $p = .039$. The amount of time looking at the images did not significantly differ between the groups in either task, $ps > 0.234$.

4. Discussion

This study provides the first evidence that preschool and early school-aged children with ASD, like older children with ASD (McGregor et al., 2013), are capable of cross-situational learning. In fact, the children with ASD in this study showed cross-situational learning abilities on par with TD children who, though younger, had similar levels of vocabulary knowledge. This is exciting because it confirms the availability of a type of word learning—cross-situational learning—that young children with ASD can use to determine word meanings without relying on social cues. Like McGregor and colleagues' findings in older, high-functioning children, our findings provide evidence that cross-situational learning can operate effectively even in children with social impairments. These findings add to growing empirical evidence that although many children with ASD have delayed language development, they have access to word-learning abilities that are qualitatively similar to those seen in typical language development—including sensitivity to patterns in speech, use of social cues, and syntactic bootstrapping (Arunachalam & Luyster, 2016; de Marchena, Eigsti, Worek, Ono, & Snedeker, 2011; Luyster & Lord, 2009; Mayo & Eigsti, 2012; McGregor et al., 2013; Naigles et al., 2011; but see Tek, Jaffery, Fein, & Naigles, 2008).

Although both TD children and children with ASD were capable of cross-situational learning at the group level, our results revealed considerable individual differences across children. Some children learned quite well, whereas others showed no clear evidence of learning. Furthermore, cross-situational learning accuracy was significantly associated with familiar word processing in both groups, suggesting that the individual differences we observed reflected not random noise, but meaningful variation. In contrast, cross-situational learning was not significantly associated with autism severity in the ASD group or with nonverbal IQ in either group. Although it is important not to over-interpret null findings, these results suggest that there may be a specific link between cross-situational word learning and language. The presence of a link between cross-situational learning and broader language skills is generally consistent with previous work in both ASD and typical development. For example, McGregor et al. (2013) identified a correlation between cross-situational learning and vocabulary skills in adolescents with ASD. In a study of 12- and 14-month-olds with typical development, Smith and Yu (2013) found that the infants who learned in the cross-situational task had larger vocabularies (per parent report) than children who did not learn.

What does it mean that cross-situational learning was associated with familiar word processing in both groups of children? The fact that children who looked longer at named novel words also looked longer at named familiar words underscores the role of visual attention—specifically, children's tendency to align what they hear with what they see. In addition to reflecting stronger language processing skills, this ability to align auditory and visual input may facilitate language development by helping children strengthen their representations of familiar words and learn new words (Kucker, McMurray, & Samuelson, 2015). Though attending to named objects may not always lead to cross-situational learning (Smith & Yu, 2013), attentional differences may nonetheless prevent some children with ASD from aligning related visual and auditory input as efficiently as TD children, which could disrupt language development (Keehn, Muller, & Townsend, 2013; Tenenbaum et al., 2017; Venker, 2017). Another factor capable of disrupting language development is visual *in*attention. In this study, the 6 children with ASD who looked away from the images in the cross-situational task most often (and as a result were excluded from the analyses) also had the weakest language skills. As a result of this extreme inattention, these children missed out on the learning opportunities their peers received—a serious concern if this type of avoidance is also occurring in their everyday lives.

One important remaining question—both for the results of the current study, and for cross-situational learning in general—is precisely how children discover word meanings when presented with ambiguous learning contexts. On one hand, it has been proposed that learners do so by tracking statistical co-occurrences over time, gradually accumulating multiple label-object associations (Smith & Yu, 2008; Yu & Smith, 2007). On the other hand, it has been proposed that learners hypothesize a single correct label-object association, subsequently considering the correctness of that hypothesized association in light of new exposures (Smith et al., 2009; Trueswell et al., 2013; Woodard et al., 2016). Current empirical evidence suggests that associative learning and hypothesis testing are not entirely separable constructs, but instead exist on opposite ends of a continuum (MacDonald et al., 2017; Trueswell et al., 2013; Yurovsky & Frank, 2015). Where learners fall on that continuum depends on numerous factors, including the degree of referential uncertainty a learner experiences, the structure of the learning exposures (e.g., massed versus interspersed), and the presence and strength of other cues to word meaning (e.g., gaze cues; MacDonald et al., 2017; Trueswell et al., 2013; Yurovsky & Frank, 2015).

Given that we are still in the early stages of understanding word-learning in children with ASD (Arunachalam & Luyster, 2016), our experimental task was not designed to distinguish the extent to which children's learning relied on gradual associative learning versus hypothesis testing. However, we speculate that at least some children gradually accumulated co-occurrence statistics across trials, in part based on the design of our experimental task (MacDonald et al., 2017; Trueswell et al., 2013; Yurovsky & Frank, 2015). Although adult learners tend towards storing only a single representation if there are many possible label-object pairings, they store more information if the number of possibilities is more limited (Yurovsky & Frank, 2015). Our task presented only two label-object pairings per trial, and only four pairings total, suggesting children may have retained more than just a single association for each word. Furthermore, our eye-tracking task was passive (i.e., children did not need to indicate any purposeful response), and children received no explicit instructions about what they were to learn. Of course, it remains possible that at least some children used a hypothesis-testing strategy. In fact, the strategies used by individual children in the current study likely varied a great deal, given vast differences in their age, language and cognitive skills, and social abilities. To better understand this issue in a population of children as diverse as ASD, future studies should include larger participant samples that permit researchers to analyze underlying mechanisms in more homogeneous subgroups of children.

Contrary to our predictions, both groups of children learned equally well in the cross-situational and ostensive word-learning tasks, suggesting that children may rely on either (or both) types of word learning during natural language learning. Interestingly, this was the case even though the children with ASD looked at the images in both teaching phases for approximately the same amount of time, and the TD children spent significantly more time looking at the images during the ostensive teaching phase than the cross-situational teaching phase. From an information processing perspective, this similarity in performance was somewhat unexpected, given that the cross-situational task presented multiple words and objects across ambiguous learning contexts, whereas the ostensive task only involved presentation of only a single word and object at a time. One potential explanation for the poor performance in the ostensive task is that this task offered limited time for encoding the auditory and visual stimuli. As a result of balancing exposure to the novel stimuli across the two tasks, the ostensive trials lasted only two seconds, whereas the cross-situational trials lasted four seconds. Future research using longer exposure times and more repetitions of words (e.g., McDuffie et al., 2012) would be beneficial for determining the contribution of limited encoding time. Additional work is needed to determine whether children rely more heavily on cross-situational or ostensive learning in their natural environments.

The similarity in performance across the cross-situational and ostensive tasks is also interesting to consider in the context of auditory-visual integration deficits in individuals with ASD—including those with strong language and cognitive skills (Grossman et al., 2015; Woynaroski et al., 2013). Why did children with ASD perform equally well in both learning contexts, despite the more complex auditory-visual integration involved in the cross-situational task? Some insight may be gained by considering the stimuli used in this task. The auditory stimuli were words and phrases, which previous work suggests may support auditory-visual integration in individuals with ASD (Grossman et al., 2009). The visual stimuli were recognizable familiar objects, which differed considerably from the simple, low-level stimuli (e.g., flashes of light; Foss-Feig et al., 2010; Stevenson et al., 2014) and the dynamic, speaking faces (e.g., Grossman et al., 2015; Irwin, et al., 2012; Woynaroski et al., 2013) used in many previous studies of auditory-visual integration in ASD. Both of these features of our design may have maximized children's ability to integrate auditory and visual information. It will be useful for future studies to explicitly examine the role of auditory-visual integration in early word learning, with specific attention to the time windows across which auditory and visual information are bound together to represent words and their meanings (Woynaroski et al., 2013).

The results of the current study also suggest that effective teaching contexts need not focus solely on one word and one meaning at a time. Intervention techniques based on cross-situational principles facilitate vocabulary development in children with language delay (Alt, Meyers, Oglivie, Nicholas, & Arizmendi, 2014), and this approach may also be effective for children with ASD. However, our results indicate that children with stronger language skills may be better equipped to utilize cross-situational learning (also see McGregor et al., 2013), which has implications for individualizing treatment strategies. In addition to helping clinicians understand how to support language learning in children with ASD, intervention studies also present a unique opportunity to better understand the role of cross-situational learning in language development more generally. For example, it may be possible to develop interventions that facilitate cross-situational learning in children with ASD. If improvements in cross-situational learning lead to improvements in overall language skills, this would provide strong evidence that cross-situational learning supports language development. In this way, studying children with ASD may inform our understanding of language-learning mechanisms in ways that studying TD children cannot.

Another important question to address in future research is how cross-situational learning interacts with social factors in children with ASD. Although the current study focused on cross-situational learning in the absence of social information, there is evidence that social factors, such as sensitivity to a speaker intentions (Frank, Goodman, & Tenenbaum, 2009) and discourse information (Frank, Tenenbaum, & Fernald, 2013), are also relevant to cross-situational learning. Relying on cross-situational statistics alone without integrating social information may result in learning that takes longer than expected and is more fraught with error. Thus, some children with ASD may remain at a relative disadvantage in natural language-learning environments because they are less likely to capitalize on all available cues for word learning.

Acknowledgements

I thank Susan Ellis Weismer and Jenny Saffran for providing mentorship during the development and execution of this research, and to Rhiannon Luyster and Martin Zettersten for providing comments on a previous version of this manuscript. I also appreciate discussions with Viridiana Benitez and Haley Vlach. Thanks to Anna Dorrance, Taryn Stricker, and Kristina Fassbender for helping with data collection and data entry, and to the members of the Language Processes Lab (Ellis Weismer, PI) and the Infant Learning Lab (Saffran, PI) for feedback and support. I sincerely thank the families and children who partnered with us in this research.

Thanks also to our funding sources. This work was supported by F31DC012451 (Venker, PI), R01DC011750 (Ellis Weismer & Kaushanskaya, PIs), R37HD037466 (Saffran, PI), P30HD003352 (Mailick, PI), the Friends of the Waisman Center, and the Wisconsin Speech-Language Pathology and Audiology Association. These funding sources had no direct involvement in the study design, data collection or analysis, manuscript writing, or decision to submit the article for publication.

Appendix A

Cross Teaching Order A

Left Image	Right Image	Word 1	Word 2	Sequential
Purple flag	Yellow vase	bosa	peri	1
Blue circle	Brown cup	coro	manu	0
Brown cup	Purple flag	coro	bosa	1
Purple flag	Blue circle	manu	bosa	0

Brown cup	Blue circle	manu	coro	0
Movie: 4 s				
Brown cup	Yellow vase	peri	coro	0
Blue circle	Purple flag	manu	bosa	1
Purple flag	Brown cup	bosa	coro	1
Yellow vase	Brown cup	coro	peri	0
Blue circle	Yellow vase	manu	peri	1
Movie: 4 s				
Brown cup	Yellow vase	coro	peri	1
Blue circle	Yellow vase	peri	manu	0
Blue circle	Purple flag	manu	bosa	1
Yellow vase	Brown cup	peri	coro	1
Blue circle	Purple flag	bosa	manu	0
Movie: 4 s				
Purple flag	Yellow vase	peri	bosa	0
Brown cup	Purple flag	bosa	coro	0
Brown cup	Blue circle	coro	manu	1
Yellow vase	Blue circle	peri	manu	1
Yellow vase	Purple flag	bosa	peri	0
Movie: 4 s				

Note. Object-label pairings for Order A were as follows: *bosa* = purple flag; m*anu* = blue circle; *peri* = yellow vase; *coro* = brown cup. Sequential trials were those in which the item on the left was labeled first.

Cross Teaching Order B

Left Image	**Right Image**	**Word 1**	**Word 2**	**Sequential**
Blue circle	Brown cup	peri	coro	0
Purple flag	Yellow vase	manu	bosa	1
Yellow vase	Blue circle	bosa	coro	1
Brown cup	Blue circle	coro	peri	0
Purple flag	Brown cup	manu	peri	1
Movie: 4 s				
Yellow vase	Brown cup	peri	bosa	0
Blue circle	Yellow vase	bosa	coro	0
Blue circle	Purple flag	coro	manu	1
Brown cup	Purple flag	peri	manu	1
Brown cup	Yellow vase	bosa	peri	0
Movie: 4 s				
Yellow vase	Brown cup	bosa	peri	1
Purple flag	Blue circle	coro	manu	0
Blue circle	Yellow vase	coro	bosa	1
Yellow vase	Purple flag	manu	bosa	0
Blue circle	Purple flag	manu	coro	0
Movie: 4 s				
Blue circle	Brown cup	coro	peri	1
Purple flag	Brown cup	peri	manu	0
Purple flag	Yellow vase	manu	bosa	1
Brown cup	Blue circle	peri	coro	1
Purple flag	Yellow vase	bosa	manu	0
Movie: 4 s				

Note. Object-label pairings for Order B were as follows: *bosa* = yellow vase; m*anu* = purple flag; *peri* = brown cup; *coro* = blue circle. Target images for each trial are bolded. Sequential trials were those in which the item on the left was labeled first.

Cross Test Order A

Left Image	Right Image	TestQ	TargetWord	LeftTarget
Cup	Shoe	where_cup_do_see_5	cup	1
Shoe	**Ball**	where_ball_like_5	ball	0
Blue circle	**Brown cup**	where_coro_do_see_5	coro	0
Purple flag	Blue circle	where_bosa_like_5	bosa	1
Yellow vase	Purple flag	where_peri_do_see_5	peri	1
Movie: 4 s				
Cup	**Shoe**	where_shoe_like_5	shoe	0
Blue circle	Yellow vase	where_manu_do_see_5	manu	1
Yellow vase	**Brown cup**	where_coro_like_5	coro	0
Ball	Cup	where_ball_do_see_5	ball	1
Yellow vase	Blue circle	where_peri_like_5	peri	1
Movie: 4 s				
Yellow vase	**Purple flag**	where_bosa_do_see_5	bosa	0
Cup	Ball	where_cup_like_5	cup	1
Purple flag	**Blue circle**	where_manu_like_5	manu	0
Purple flag	**Yellow vase**	where_peri_do_see_5	peri	0
Shoe	Cup	where_shoe_do_see_5	shoe	1
Movie: 4 s				
Brown cup	Purple flag	where_coro_like_5	coro	1
Brown cup	**Blue circle**	where_manu_do_see_5	manu	0

Ball	Shoe	where_ball_do_see_5	ball	1
Brown cup	**Yellow vase**	where_peri_like_5	peri	0
Purple flag	Brown cup	where_bosa_like_5	bosa	1
Movie: 4 s				
Ball	**Cup**	where_cup_do_see_5	cup	0
Blue circle	**Purple flag**	where_bosa_do_see_5	bosa	0
Ball	**Shoe**	where_shoe_like_5	shoe	0
Brown cup	Yellow vase	where_coro_do_see_5	coro	1
Blue circle	Brown cup	where_manu_like_5	manu	1
Movie: 4 s				

Cross Test Order B

Left Image	Right Image	TestQ	TargetWord	LeftTarget
Ball	Cup	where_ball_do_see_5	ball	1
Cup	**Shoe**	where_shoe_like_5	shoe	0
Yellow vase	Blue circle	where_bosa_like_5	bosa	1
Purple flag	Brown cup	where_manu_do_see_5	manu	1
Blue circle	**Brown cup**	where_peri_like_5	peri	0
Movie: 4 s				
Ball	**Cup**	where_cup_do_see_5	cup	0
Brown cup	**Blue circle**	where_coro_like_5	coro	0
Purple flag	Blue circle	where_manu_like_5	manu	1
Ball	Shoe	where_ball_do_see_5	ball	1
Brown cup	**Yellow vase**	where_bosa_do_see_5	bosa	0
Movie: 4 s				
Purple flag	**Blue circle**	where_coro_do_see_5	coro	0
Cup	Ball	where_cup_like_5	cup	1
Purple flag	**Yellow vase**	where_bosa_do_see_5	bosa	0
Brown cup	Purple flag	where_peri_like_5	peri	1
Shoe	Cup	where_shoe_do_see_5	shoe	1
Movie: 4 s				
Yellow vase	**Purple flag**	where_manu_like_5	manu	0
Brown cup	Yellow vase	where_peri_do_see_5	peri	1
Shoe	**Ball**	where_ball_like_5	ball	0
Blue circle	Yellow vase	where_coro_like_5	coro	1
Yellow vase	**Brown cup**	where_peri_do_see_5	peri	0
Movie: 4 s				
Cup	Shoe	where_cup_do_see_5	cup	1
Blue circle	**Purple flag**	where_manu_do_see_5	manu	0
Ball	**Shoe**	where_shoe_like_5	shoe	0
Blue circle	Brown cup	where_coro_do_see_5	coro	1
Yellow vase	Purple flag	where_bosa_like_5	bosa	1
Movie: 4 s				

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2018.10.025.

References

Alt, M., Meyers, C., Oglivie, T., Nicholas, K., & Arizmendi, G. (2014). Cross-situational statistically-based word learning intervention for late-talking toddlers. *Journal of Communication Disorders, 52*, 207–220. https://doi.org/10.1016/j.jcomdis.2014.07.002.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, D.C.: Author.

Arunachalam, S., & Luyster, R. J. (2016). The integrity of lexical acquisition mechanisms in autism spectrum disorders: A research review. *Autism Research, 9*, 810–828. https://doi.org/10.1002/aur.1590.

Bebko, J. M., Weiss, J. A., Demark, J. L., & Gomez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 47*(1), 88–98. https://doi.org/10.1111/j.1469-7610.2005.01443.x.

Charman, T., Drew, A., Baird, C., & Baird, G. (2003). Measuring early language development in preschool children with autism spectrum disorder using the MacArthur Communicative Development Inventory (Infant Form). *Journal of Child Language, 30*(1), 213–236. https://doi.org/10.1017/S0305000902005482.

Christensen, D. L., Baio, J., Braun, K. V. N., Bilder, D., Charles, J., Constantino, J. N., … Yeargin-Allsopp, M. (2016). Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years - Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. Morbidity and Mortality Weekly Report. Surveillance Summaries (Washington, D.C. : 2002), 65(3) (pp. 1–23) 10.15585/mmwr.ss6503a1.

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

Dunn, L. M., & Dunn, D. M. (2006). *Peabody picture vocabulary test* (fourth ed.). Bloomington, MN: NCS Pearson Inc.

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

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism spectrum disorders. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale, 203*(2), 381–389. https://doi.org/10.1007/s00221-010-2240-4.

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

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

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(5), 693–705. https://doi.org/10.1007/s10803-008-0674-3.Standardizing.

Grossman, R. B., Schneps, M. H., & Tager-Flusberg, H. (2009). Slipped lips: Onset asynchrony detection of auditory-visual language in autism. *Journal of Child Psychology* *and Psychiatry and Allied Disciplines, 50*(4), 491–497. https://doi.org/10.1111/j.1469-7610.2008.02002.x.

Grossman, R. B., Steinhard, E., Mitchell, T., & McIlvane, W. (2015). Look who's talking!. *Gaze Patterns for Implicit and Explicit Audio-Visual Speech Synchrony Detection in Children With High-Functioning Autism, 8*(5), 583–592. https://doi.org/10.1002/aur.1474.Replication.

Iarocci, G., & McDonald, J. (2006). Sensory integration and the perceptual experience of persons with autism. *Journal of Autism and Developmental Disorders, 36*(1), 77–90. https://doi.org/10.1007/s10803-005-0044-3.

Irwin, J. R., Tornatore, L.a., Brancazio, L., & Whalen, D. H. (2012). Can children with autism spectrum disorders "hear" a speaking face? *Child Development, 29*(6), 997–1003. https://doi.org/10.1016/j.biotechadv.2011.08.021.Secreted.

Keehn, B., Muller, R.-A., & Townsend, J. (2013). Atypical attentional networks and the emergence of autism. *Neuroscience & Biobehavioral Reviews, 37*(2), 164–183. https://doi.org/10.1016/j.neubiorev.2012.11.014.Atypical.

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

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

Luyster, R., Lopez, K., & Lord, C. (2007). Characterizing communicative development in children referred for autism spectrum disorders using the MacArthur-Bates Communicative Development Inventory (CDI). *Journal of Child Language, 34*(3), 623–654. https://doi.org/10.1017/S0305000907008094.

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.Word.

MacDonald, K., Yurovsky, D., & Frank, M. C. (2017). Social cues modulate the representations underlying cross-situational learning. *Cognitive Psychology, 94*, 67–84. https://doi.org/10.1016/J.COGPSYCH.2017.02.003.

Mahr, T., McMillan, B. T. M., Saffran, J. R., Ellis Weismer, S., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition, 142*, 345–350. https://doi.org/10.1016/j.cognition.2015.05.009.

Mayo, J., & Eigsti, I.-M. (2012). Brief report: A comparison of statistical learning in school-aged children with high functioning autism and typically developing peers. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-012-1493-0.

McDuffie, A., Kover, S. T., Hagerman, R., & Abbeduto, L. (2012). Investigating word learning in fragile X syndrome: A fast-mapping study. *Journal of Aurtism and Developmental Disorders, 43*(7), 1676–1691. https://doi.org/10.1007/s10803-012-1717-3.

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*, 745–753. https://doi.org/10.1111/jcpp.12073.

Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America, 108*(22), 9014–9019. https://doi.org/10.1073/pnas.1105040108.

Naigles, L. R., Kelty, E., Jaffery, R., & Fein, D. (2011). Abstractness and continuity in the syntactic development of young children with autism. *Autism Research, 4*, 422–437. https://doi.org/10.1002/aur.223.

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

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

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

Smith, K., Smith, A. D. M., & Blythe, R. A. (2009). Reconsidering human cross-situational learning capacities : A revision to Yu & Smith ' s (2007) experimental paradigm. *Cognitive Science,* 2711–2716.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition, 106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010.

Smith, L. B., & Yu, C. (2013). Visual attention is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development, 9*, 25–49. https://doi.org/10.1080/15475441.2012.707104.

Stevenson, R. A., Siemann, J. K., Woynaroski, T. G., Schneider, B. C., Eberly, H. E., Camarata, S. M., & Wallace, M. T. (2014). Evidence for diminished multisensory integration in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(12), 3161–3167. https://doi.org/10.1007/s10803-014-2179-6.

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology, 126*, 395–411. https://doi.org/10.1016/j.jecp.2014.06.003.

Suanda, S. H., & Namy, L. L. (2012). Detailed behavioral analysis as a window into cross-situational word learning. *Cognitive Science, 36*, 545–559. https://doi.org/10.1111/j.1551-6709.2011.01218.x.

Tek, S., Jaffery, G., Fein, D., & Naigles, L. R. (2008). Do children with autism spectrum disorders show a shape bias in word learning? *Autism Research, 1*, 208–222. https://doi.org/10.1002/aur.38.Do.

Tenenbaum, E. J., Amso, D., Righi, G., & Sheinkopf, S. J. (2017). Attempting to "Increase Intake from the Input": Attention and Word Learning in Children with Autism. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-017-3098-0.

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology, 66*(1), 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001.

Venker, C. E. (2017). Spoken word recognition in children with autism spectrum disorder: The role of visual disengagement. *Autism: The International Journal of Research and Practice, 2*, 821–829.

Venker, C. E., Kover, S. T., & Ellis Weismer, S. (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.

Vitevitch, M. S., & Luce, P. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, & Computers, 36*, 481–487.

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. Retrieved from *Cognition, 127*(3), 375–382. http://www.sciencedirect.com/science/article/pii/S0010027713000437.

Vouloumanos, A., & Werker, J. F. (2009). Infants' learning of novel words in a stochastic environment. *Developmental Psychology, 45*(6), 1611–1617. https://doi.org/10.1037/a0016134.

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods, 45*(1), 229–250. https://doi.org/10.3758/s13428-012-0245-6.

Woodard, K., Gleitman, L. R., & Trueswell, J. C. (2016). Two- and three-year-olds track a single meaning during word learning: Evidence for propose-but-verify. *Language Learning and Development, 12*(3), 252–261. https://doi.org/10.1080/15475441.2016.1140581.

Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T. (2013). Multisensory speech perception in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43*(12), 2891–2902. https://doi.org/10.1007/s10803-013-1836-5.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. Retrieved from *Psychological Science, 18*(5), 414–420. http://pss.sagepub.com/content/18/5/414.short.

Yu, C., & Smith, L. B. (2011). What you learn is what you see: Using eye movements to study infant cross-situational word learning. *Developmental Science, 14*(2), 165–180. https://doi.org/10.1111/j.1467-7687.2010.00958.x.

Yu, C., & Smith, L. B. (2012). Modeling cross-situational word-referent learning: Prior questions. *Psychological Review, 119*(1), 21–39. https://doi.org/10.1037/a0026182.

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition, 145*, 53–62. https://doi.org/10.1016/J.COGNITION.2015.07.013.

Zettersten, M., Wojcik, E., Benitez, V. L., & Saffran, J. (2018). The company objects keep: Linking referents together during cross-situational word learning. *Journal of Memory and Language, 99*, 62–73. https://doi.org/10.1016/J.JML.2017.11.001.