



Cognitive Science 46 (2022) e13094

© 2022 The Authors. *Cognitive Science* published by Wiley Periodicals LLC on behalf of Cognitive Science Society (CSS).

ISSN: 1551-6709 online

DOI: 10.1111/cogs.13094

What Children with Developmental Language Disorder Teach Us About Cross-Situational Word Learning

Karla K. McGregor,^{a,b} Erin Smolak,^a Michelle Jones,^c Jacob Oleson,^c
Nichole Eden,^a Timothy Arbisi-Kelm,^a Ronald Pomper^a

^a*Center for Childhood Deafness, Language and Learning, Boys Town National Research Hospital*

^b*Department of Communication Sciences and Disorders, University of Iowa*

^c*Department of Biostatistics, University of Iowa*

Received 5 April 2021; received in revised form 20 December 2021; accepted 27 December 2021

Abstract

Children with developmental language disorder (DLD) served as a test case for determining the role of extant vocabulary knowledge, endogenous attention, and phonological working memory abilities in cross-situational word learning. First-graders ($M_{\text{age}} = 7$ years; 3 months), 44 with typical development (TD) and 28 with DLD, completed a cross-situational word-learning task comprised six cycles, followed by retention tests and independent assessments of attention, memory, and vocabulary. Children with DLD scored lower than those with TD on all measures of learning and retention, a performance gap that emerged in the first cycle of the cross-situational protocol and that we attribute to weaknesses in initial encoding. Over cycles, children with DLD learned words at a similar rate as their TD peers but they were less flexible in their strategy use, demonstrating a propose-but-verify approach but never a statistical aggregation approach. Also, they drew upon different mechanisms to support their learning. Attention played a greater role for the children with DLD, whereas extant vocabulary size played a greater role for the children with TD. Children navigate the problem space of cross-situational learning via varied routes. This conclusion is offered as motivation for theorists to capture all learners, not just the most typical ones.

Keywords: language learning; statistical learning; language disorder

Correspondence should be sent to Karla K. McGregor, Boys Town National Research Hospital, Center for Childhood Deafness, Language and Learning, 425 N. 30th St., Omaha, NE 68131, USA. E-mail: karla.mcgregor@boystown.org

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. Introduction

Humans attempt to determine meaning in the world around them. When exposed to an unfamiliar word for the first time, they use their prior knowledge and the available social, linguistic, and visual contexts to infer its referent, but, in everyday situations, these contexts are not deterministic. For example, a toddler might hear “juice” as Dad is looking for juice in the refrigerator. Does the word *juice* refer to the bottle of sweet orange liquid or a jar of jam, the cold emanating from the refrigerator, or the refrigerator itself? Because Dad did not hold up the bottle of juice and label it, the child cannot determine the referent of the word with any certainty. However, across situations, a given word form tends to be paired with a given referent more frequently than other pairings. The child is more likely to hear “juice” when juice is present than when jam or cold air is present. By using the co-occurrence information across naming situations, the child can determine the meaning of the word.

In the laboratory, cross-situational word learning is evident in infants and toddlers (Smith & Yu, 2008; Vlach & Johnson, 2013; Woodard, Gleitman, & Trueswell, 2016), older children (Suanda, Mugwanya, & Namy, 2014; Vlach & DeBrock, 2017), and adults (Benitez, Yurovsky, & Smith, 2016; Berens, Horst, & Bird, 2018; Kachergis, Yu, & Shiffrin, 2013; MacDonald, Yurovsky, & Frank, 2017; Roembke & McMurray, 2016; Romberg & Yu, 2014; Yu & Smith, 2007; Yurovsky & Frank, 2015). In a typical cross-situational learning task, participants see two (or more) pictures of novel objects and hear one (or more) novel words labeling those objects. In the traditional task, the first time these word-referent pairings occur (i.e., the first cycle of trials), there is no information about the correct mapping between each word and its referent. However, in subsequent cycles, word-referent pairings repeat such that a word is paired with its correct referent 100% of the time but is paired with other referents inconsistently. Thus, participants can learn correct word-referent pairings through across-cycle information. Cross-situational learning opportunities enable not only the linking of words to referents but also the mapping of the precise phonological details of the word forms themselves, and this is true of infants and adults (Escudero, Mulak, & Vlach, 2016a, b). Moreover, by age 4 or 5, children retain words learned from cross-situational opportunities over a 5-min interval (although longer intervals have not been tested) (Vlach & DeBrock, 2019). Given the lack of reliable ostensive cues in real-life environments (Cartmill et al., 2013; Quine, 1960) and the sheer number of words to be learned (Nagy and Anderson, 1982), cross-situational word learning may be necessary to the optimal building of a lexicon.

There are two primary competing accounts of the processes underlying cross-situational learning. These accounts differ primarily in the extent to which implicit statistical learning mechanisms are assumed to contribute to the learning process. Statistical learning can be broadly defined as the gradual (over many exposures), incidental learning of patterns or regularities extracted from the environment (Conway, 2020). Thus, extracting regularities from the co-occurrences of words and their referents across multiple contexts can be viewed as a type of statistical learning. Indeed, one account of cross-situational learning highlights its statistical nature. This account proposes that learning emerges from the gradual aggregation of reliable co-occurrence data that increases the weight of the ultimately correct word-referent associations (Smith & Yu, 2008; Yurovsky, Fricker, Yu, & Smith, 2014). Another account

recognizes that the application of propositional logic leads the learner to hypothesize word-referent associations quickly, at a more or less conscious level. The hypothesis is subsequently strengthened over additional situations unless those situations provide counterevidence, in which case a new hypothesis is proposed and verified (Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell, Medina, Hafri, & Gleitman, 2013; Woodard et al., 2016). One key difference between the two accounts is whether the learner tracks multiple hypotheses or one. Under the aggregation account, learners track the co-occurrence between a word and all possible referents in the context; therefore, even when they incorrectly link a word to a referent on a given trial, they benefit from that experience on subsequent trials because they have retained a memory of the other potential referents. In contrast, under the propose-but-verify account, learners do not encode information about all objects present in the context of a trial. Instead, they commit to a given word-to-referent link and, if subsequent data prove them wrong, they will fall back to the baseline level of performance because they have not tracked other possibilities.

There is growing evidence that gradual aggregation and propose-but-verify strategies likely coexist (Roembke & McMurray, 2016; Stevens, Gleitman, Trueswell, & Yang, 2017; Xu & Tenenbaum, 2007; Yurovsky & Frank, 2015), thus motivating a hybrid model. Under the hybrid account, the learner forms a hypothesis space of probabilistic form-referent links. With each occurrence of the word, the likelihood of each hypothesis being correct changes in proportion to the probability of the hypothesized word meaning being correct given the available data. In other words, learners do propose hypotheses but also keep track of probabilistic co-occurrences. In one hybrid account, the effect of previous accuracy on current accuracy tends to increase over time, suggesting that propose-but-verify strategies may be a product of learning accomplished by gradual aggregation of available data (Roembke & McMurray, 2016). That is, learners may leverage statistical aggregation early in the learning process when word-referent associations are weak. Later in the learning process, learners may make use of strengthening associations to propose hypotheses (or bias attention toward) the correct word-referent association (Bhat, Spencer, & Samuelson, 2021). Hybrid or flexible strategy use would appear to be highly adaptive given the extreme variations in context and task demands that learners likely encounter in everyday life.

Relative reliance on aggregation versus hypothesis-testing mechanisms in cross-situational word learning may vary with the task, context, degree of ambiguity, and the number of unfamiliar words encountered in the task. On the whole, learners in the laboratory depend more heavily upon an aggregation approach when there are fewer words to be learned, many exposures to each word, or no overt response requirements (Dautriche & Chemla, 2014; Roembke & McMurray, 2016; Stevens et al., 2017; Yurovsky & Frank, 2015). Increasing the number of words to be learned and the number of competitors appearing alongside them, reducing the number of exposures, or requiring an overt guess on each trial tends to move learners toward a propose-but-verify strategy, as does structuring the trials in such a way as to permit the application of the principle of mutual exclusivity (Kachergis et al., 2013; Roembke & McMurray, 2016). As for the latter, Roembke and McMurray (2016) allowed objects to appear as targets (when they were named) but also as foils (alongside other objects being named). They found an effect of “last-foil-accuracy” such that learners were more accurate

on a given trial if they had previously identified an object that served as a foil on that trial. As foils repeat, learners can leverage their knowledge of the foils to infer a word-to-referent mapping.

1.1. Learners' access to cross-situational data

Models of cross-situational learning, whether gradual aggregation, propose but verify, or a hybrid of the two, should specify the abilities that the learner must apply to the task, an issue that has received little attention in the literature on cross-situational learning. Is cross-situational learning robust in all learners? If not, which learner characteristics predict better or worse outcomes? Two sets of findings, one from bilingual adults and the other from monolingual children and adults, provide clues.

There is some evidence that bilinguals may outperform monolinguals on cross-situational word learning tasks (Benitez et al., 2016; Escudero, Mulak, Fu, & Singh, 2016) as they do when learning words via ostensive naming (Kaushanskaya and Marian, 2009; Bartolotti and Marian, 2012), although we note that this advantage is not always robust, and may be present only under certain circumstances (Poepsel & Weiss, 2016). When the advantage is found, it is hypothesized to relate to differences in prior metalinguistic knowledge and potential strengths in attention and memory.

Individual differences among monolingual children and adults reinforce this hypothesis. For example, Vlach and DeBrock (2017) found positive correlations between cross-situational word learning outcomes of preschoolers and their extant vocabulary knowledge and memory for paired associates, with memory abilities being the stronger predictor. In a second experiment, they measured children's recognition memory for novel objects, novel words, and the linking of novel words to novel objects following an ostensive teaching paradigm. They then exposed those same children to cross-situational learning opportunities. All three memory measures predicted variance in cross-situational word learning over and beyond the children's chronological age.

Another example comes from Yu and Smith (2011). They grouped infant participants according to their cross-situational word learning outcomes and then examined the eye gaze behaviors recorded during learning. The stronger learners had fewer and longer fixations on the object stimuli during learning than the weaker learners, behaviors taken as a sign of better sustained attention. In subsequent work, Smith and Yu (2013) determined that sustained attention to novelty (externally driven attention) detracted from learning, whereas sustained attention to longer-term regularities, such as reoccurring word-object pairings (endogenously driven attention), was a sign of better learning. Children with stronger endogenous attention not only demonstrated better cross-situational word learning but also had greater extant vocabulary knowledge. Relatedly, Toro, Sinnett, and Soto-Faraco (2005) found that statistical learning of speech segmentation was impaired when participants' attention was diverted from the task. A third and final example comes from Scott and Fisher (2012). They found that, among 2-year-olds, only those with larger vocabularies could learn transitive verbs in a cross-situational paradigm.

Across these studies of bilingual adults and monolingual children, a pattern emerges: The success of cross-situational word learning varies with the learners' prior knowledge of the language, especially extant vocabulary knowledge, their ability to sustain attention to relevant information, and their memory abilities. Everyday environments provide statistical information that can, theoretically, support word learning, but as Smith, Suanda, and Yu (2014, p. 5) note, "... the data relevant to a statistical learner are not the data in the world but the data in the world filtered through the learners' sensory, attentional and memory systems." A complete account of cross-situational word learning requires understanding how these mechanisms enable or impede success.

The current study constitutes a natural experiment of sorts. Children with developmental language disorder (DLD) are presented as a test case to demonstrate how the approach to and outcomes of cross-situational learning vary with the characteristics of the learner. DLD, also termed specific language impairment, is a neurodevelopmental condition affecting approximately 7% of the population (Norbury et al., 2016). DLD is characterized primarily by deficits in language comprehension and production, especially in morpho-syntax. These children do not have intellectual disabilities, but they have known weaknesses in three relevant domains: extant language knowledge, attention, and memory. The goal of the current study is to compare cross-situational learning performance in children with DLD to typically developing (TD) children. Because these TD children are the same age and thus have had a similar amount of real-world word learning experience, one can draw firmer conclusions about the cognitive mechanisms that enable cross-situational word learning. Specifically, if TD children outperform children with DLD, one might conclude that statistical co-occurrence information is not equally accessible to all learners. Moreover, if patterns of strategy use vary between the groups, regardless of success, we will have evidence of the multiple routes that make word learning viable within a given context. Finally, relationships between outcomes and abilities determined by independent measures of extant vocabulary, attention, and memory will provide supporting evidence for hypothesized bases of cross-situational learning and will serve to clarify the mechanisms that are most important to successful learning.

1.2. Three weaknesses that characterize DLD

1.2.1. Vocabulary knowledge

As a group, people with DLD present with smaller and less rich lexicons than their typical peers (McGregor, Oleson, Bahnsen, & Duff, 2013). Signs of their vocabulary problems include slow (Lahey & Edwards, 1996) or errorful naming (Lahey & Edwards, 1999; McGregor et al., 2002), as well as difficulties parsing and explaining the modifier-head relationships in compound words (McGregor, Rost, Guo, & Sheng, 2010), identifying subordinate word meanings (Norbury, 2005), making lexical decisions (Edwards & Lahey, 1996, but see Haebig, Kaushanskaya, & Weismer, 2015), naming semantic associates (Sheng & McGregor, 2010), and judging the prototypicality of category referents (Alt, Meyers, & Alt, 2013). On receptive single-word vocabulary tests, people with DLD tend to score within the average range but statistically lower than their matched TD peers (Alt et al., 2013; Gray, Plante, Vance,

& Henrichsen, 1999; Haebig et al., 2015; McGregor et al., 2002, 2010, 2012; McGregor et al., 2020; McGregor, Arbisi-Kelm, Eden, & Oleson, 2020).

1.2.2. *Memory*

In contrast to prior research on memory for paired associates (Vlach & DeBrock, 2017), the focus of the current study is on phonological working memory. We chose this as the memory construct of interest because, during cross-situational word learning, the learner must hold an unfamiliar word form in phonological memory while determining its referent. In typical children, phonological working memory as measured by nonword repetition predicts the accuracy of the recognition and production of new words taught directly in the laboratory (Adlof & Patten, 2017) and the accrual of vocabulary over time in the everyday world (Gathercole, 2006). One of the most reliable markers of DLD is low performance on nonword repetition tasks. Gathercole and Baddeley (1990) first reported that children with DLD were less accurate in their repetition of nonwords than their younger TD peers who had similar language development levels as measured by tests of receptive vocabulary and reading. In a meta-analysis of 23 studies, Graf Estes, Evans, and Else-Quest (2007) found that, on nonword repetition tasks, children with DLD averaged 1.27 standard deviations below typical peers matched on age, language ability, or nonverbal cognition.

1.2.3. *Sustained attention*

As a group, people with DLD exhibit poor endogenously controlled sustained attention relative to their TD peers in tasks of auditory or visual-spatial sustained attention (Kapa, Plante, & Doubleday, 2017; Montgomery, Evans, & Gillam, 2009; Noterdaeme, Amorosa, Mildemberger, Sitter, & Minow, 2001; Smolak et al., 2020; Spaulding, Plante, & Vance, 2008). A meta-analysis revealed that children with DLD perform 0.69 standard deviations below their peers in sustained attention tasks (0.82 *SD* in auditory linguistic tasks, 0.61 *SD* in auditory nonlinguistic tasks, and 0.47 *SD* in visual-spatial tasks, Ebert & Kohnert, 2011). Moreover, individual differences in sustained attention abilities among children with DLD are positively related to working memory and language abilities (Kapa et al., 2017; Montgomery et al., 2009; Smolak et al., 2020; Spaulding et al., 2008).

1.3. *Word learning among children with DLD*

Our current understanding of the word-learning abilities of individuals with DLD is based mainly on ostensive naming paradigms in the laboratory or explicit teaching in the clinic. In these contexts, people with DLD require more exposure (Gray, 2003, 2004, 2005; McGregor et al., 2020), more supportive ostensive cues (Kapantzoglou, Restrepo, & Thompson, 2012; Lüke & Ritterfeld, 2014), or more retrieval practice (i.e., naming with feedback, Leonard, Deevy, Karpicke, Christ, & Kueser, 2020; McGregor, Gordon, Eden, Arbisi-Kelm, & Oleson, 2017) to learn as many words as their typical peers. Effect sizes are moderate, on average (Kan & Windsor, 2010). The learning of word forms presents a more significant obstacle than learning word referents, which holds for children (Gray, 2004; Leonard et al., 2019) and adults (McGregor et al., 2020; McGregor, Licandro, et al., 2013) with DLD. For example, while

learning new words via retrieval practice, adults with DLD made more errors in word form production, alone or in combination with errors in linking forms to the correct referents than their typical peers. However, the number of pure link errors did not differ by group. Linking words to the referents they name appears to be a relative strength, at least as measured in ostensive naming contexts (Gray, 2004; Leonard et al., 2019; McGregor et al., 2020) or in contexts that cue that mapping via mutual exclusivity (Dollaghan, 1987; Estis & Beverly, 2015; Lewis, Cristiano, Lake, Kwan, & Frank, 2020). For example, Estis and Beverly (2015) found that, by age 6, children with DLD reliably infer that a novel word names an unfamiliar object, not a familiar object.

To our knowledge, there is only a single study on cross-situational word learning in children with DLD (Ahufinger, Guerra, Ferinu, Andreu, & Sanz-Torrent, 2021). In this study, school-aged TD children and children with DLD were presented with two novel words and two potential referents on each trial. Across trials, a novel word occurred with its referent 100% of the time, a strong competitor 50% of the time, and a weak competitor 25% of the time. Children's eye gaze during learning trials was recorded, and their learning performance was assessed on an alternative forced-choice (AFC) posttest in which the child saw four pictured referents and pointed to the one being named. Although both groups of children performed above chance, TD children significantly outperformed children with DLD. Additionally, children with DLD chose the weak competitor significantly more often than TD children, suggesting that TD children were more adept at extracting co-occurrence frequency information. Finally, eye-tracking data suggested that children with DLD were less confident in their responses than were TD children. Even when they focused on the correct referent, they spent more time looking at the competitor than did their peers. Given results from this study and evidence that children with DLD exhibit problems in using statistical information to support speech segmentation (Evans, Saffran, & Robe-Torres, 2009; Haebig, Saffran, & Ellis Weismer, 2017) and the learning of grammatical constructions (Grunow et al., 2006; Hsu, Tomblin, & Christiansen, 2014; Lum, Conti-Ramsden, Morgan, & Ullman, 2014; Plante et al., 2002; Torkildsen, Dailey, Aguilar, Gómez, & Plante, 2013), grammatical gender marking (Richardson et al., 2006), or stress assignment rules (Plante, Bahl, Vance, & Gerken, 2010), we anticipated a gap in cross-situational word learning between TD children and children with DLD in the present study. The present research expands upon the study by Ahufinger et al. (2021) by (1) instructing children to make an overt response/guess during learning, thereby allowing us to analyze change in performance over time, trial by trial and cycle by cycle, and (2) investigating the relationship between cross-situational learning performance and vocabulary, attention, and phonological working memory.

1.4. The current study

The goal was to determine whether and how children with DLD differ from their typical peers in using cross-situational information in support of word learning. To achieve this goal, we compared the two groups of children on accuracy during learning, strategy use during learning, accuracy after a 5-min retention interval, and the extent to which extant

vocabulary, endogenous attention, and phonological working memory accounted for their performance.

In the cross-situational word learning paradigm used here, a single cycle involved 12 trials, one for each of 12 novel words to be learned. Each trial consisted of two pictured items on a touch screen and a single novel word. Children were instructed to touch the picture being named. The 12 items to be learned appeared as referents (once per cycle) and foils (randomly selected from the remaining 11 items). This manipulation served to structure the hypothesis space for the learners. Specifically, it enabled them to propose word-referent mappings via mutual exclusivity. The cycles repeated for a total of six exposures to each novel word-referent pairing. With this paradigm, we addressed four questions and associated predictions:

- 1) Does performance accuracy vary for the DLD and TD groups within and across learning cycles? We predicted that the DLD group would be less accurate than the TD group (a main effect of diagnostic group) and demonstrate a shallower learning trajectory over the six cycles (a group \times cycle interaction).
- 2) Do learning strategies vary for the DLD and TD groups? Learners could apply three strategies and their combinations: mutual exclusivity, statistical aggregation, and propose-but-verify. Evidence of mutual exclusivity would be above chance performance on cycle 1 before cross-cycle statistical information was available. Additional evidence would be an effect of trial such that performance improved over trials within a cycle as items previously presented as referents were repeated as foils, or vice versa. Evidence of an aggregation strategy would be performance that improved from cycle to cycle regardless of previous item-level accuracy. Evidence in support of a propose-but-verify strategy would be performance that improved over cycles but in a way that depended on the accuracy of the item when previously encountered. If children guessed correctly on certain items in a given cycle, they would continue to guess correctly on the next cycle. However, if they guessed incorrectly, performance would drop back to baseline levels. Evidence in support of a hybrid strategy would be a propose-but-verify strategy in which the dependency of a correct response on a previous correct response grew over time. Alternative evidence supporting a hybrid strategy could be aggregation, which was slower for previously incorrect than correct selections but still above baseline levels. We predicted that, given their age, both groups would apply the mutual exclusivity heuristic (Dollaghan, 1987; Lewis et al., 2020). Given that mutual exclusivity information and overt responses promote a propositional approach (Kachergis et al., 2013; Roembke & McMurray, 2016), we also predicted that both groups would evince a propose-but-verify strategy. However, our paradigm did not preclude the use of statistical aggregation. In fact, two features of our paradigm, few competitors and cycles, tend to promote statistical aggregation (Kachergis et al., 2013; Roembke & McMurray, 2016). Given the statistical learning deficits associated with DLD, we hypothesized that an aggregation strategy would be less evident for the DLD group than the TD group.
- 3) Do learning outcomes vary for the DLD and TD groups as measured after a 5-min retention interval? We assessed children's performance on two 3AFC posttests that

tapped their knowledge of the word forms and the links between the words and their referents. Given predicted differences in accuracy and trajectory across cycles and only six cycles to learn from, we predicted that the DLD group would have poorer learning outcomes than the TD group on both measures. Given that people with DLD tend to find word forms particularly challenging to learn, we predicted an interaction such that the gap between groups on the 3AFC form test would be larger than the gap on the 3AFC link test. Ahufinger et al. (2021) reported lower link recognition among learners with DLD than learners with TD, but they did not measure form recognition.

- 4) Which cognitive mechanisms are associated with learning and retention, and do these vary with group? Because cross-situational word learning requires memory for phonological forms, attention to visual referents, and operations that link forms and referents in memory to other lexical entries over time, we predicted that phonological working memory, sustained attention, and vocabulary performance would account for variance in outcomes. Of course, any associations obtained could result from the reverse relationship. In particular, if cross-situational learning is a viable means of word learning in the wild, one would expect a positive relationship between the outcomes on both learning and retention measures (how well they learn words) and extant vocabulary size (how many words they have accrued over time). Given the scant previous literature on cross-situational word learning among people with DLD, we had no firm prediction about how these relationships would vary by diagnostic group.

2. Materials and methods

2.1. Participants

All participants were enrolled in a longitudinal study of changes in word learning over the first 4 years of primary education (McGregor, 2017). The results reported here were collected when the children were in the first grade. The Institutional Review Board at Boys Town National Research Hospital approved all recruitment and study procedures. The participants and their parents gave informed consent to participate.

There were 72 participants, 28 in the DLD group and 44 in the TD group. Table 1 summarizes their demographic characteristics and test scores. To qualify for the DLD group, a child had to score below the 15th percentile on a sentence recall screening task developed by Redmond (2005) and score below a standard score of 92 on the *Test of Narrative Language*, first or second edition (TNL-1, Gillam & Pearson, 2004; TNL-2, Gilliam & Pearson, 2017). The TNL is appropriate for the ages of interest; it taps multiple domains of receptive and expressive language; it is normed nationwide; and it exhibits minimal gender or racial bias (Gillam & Pearson, 2004). The cutoff of 92 allows identification of DLD with a sensitivity and a specificity of 92% (Gilliam & Person, 2017). To qualify for the TD group, a child had to score above the cutoffs on the sentence recall task and the TNL. Extensive case histories were collected to ensure that none of the children in the DLD or TD groups had diagnoses of autism spectrum disorder or intellectual disability and that none in the TD group had a

Table 1

Comparison of demographic characteristics and test scores of the participant groups

Measure	Descriptive statistic	DLD $N = 28$	TD $N = 44$	t	p	Effect size d
Age (months)	mean (SD)	86.68 (6.20)	86.59 (4.63)	-0.07	.95	-0.02
	range	72–96	76–98			
Parental education (years)	mean (SD)	14.21 (2.51)	16.91 (2.17)	4.83	<.001	1.12
	range	10–20	12–22			
TNL (standard score)	mean (SD)	82.29 (6.95)	111.5 (9.04)	14.56	<.001	3.52
	range	61–91	94–127			
Nonverbal IQ (standard score)	mean (SD)	89.32 (10.94)	107.6 (10.50)	7.08	<.001	1.71
	range	71–116	86–130			
NIH Vocab (standard score)	mean (SD)	91.82 (13.71)	111.25 (14.23)	5.73	<.001	1.38
	range	75–116	78–140			
Track-It Attention (proportion correct for correct memory trials, heterogeneous condition)	mean (SD)	0.69 (0.31)	0.86 (0.22)	2.74	.008	0.67
	range	0–1	0–1			
NWR (proportion phonemes correct)	mean (SD)	0.68 (0.13)	0.81 (0.085)	5.04	<.001	1.22
	range	0.29–0.88	0.54–0.96			

Abbreviations: NIH Vocab, the vocabulary subtest from the NIH Toolbox; Nonverbal IQ, the nonverbal composite score from the Wechsler Abbreviated Scale of Intelligence-II Performance Index (WASI-II; Wechsler, 2011) or the KBIT (The Kaufman Brief Intelligence Test 2nd Edition, KBIT-2; Kaufman & Kaufman, 2004); NWR, the Nonword Repetition Task in Dollaghan and Campbell (1998); TNL, Test of Narrative Language.

history of DLD. Because attention deficit hyperactivity disorder (ADHD) is highly comorbid with DLD (Beitchman, Nair, Clegg, Ferguson, & Patel, 1986; Sciberras et al., 2014), the presence of ADHD was free to vary in both groups. Two children in the TD group and three children in the DLD group had a diagnosis of ADHD.

To rule out hearing loss, participants had to pass an audiometric screening (American Speech-Language-Hearing Association, no date) and, to rule out intellectual disability, they had to score 70 or above on a nonverbal IQ measure: either the Matrices and Block Design subtests of the *Wechsler Abbreviated Scale of Intelligence-II Performance Index* (WASI-II; Wechsler, 2011) or, for one child, the *Kaufman Brief Intelligence Test 2nd Edition* (KBIT-2; Kaufman & Kaufman, 2004). All participants were monolingual speakers of English.

With the exception of the matching variable age, the DLD and TD groups differed significantly and with moderate to large effect sizes on all intake measures, including parental education and nonverbal IQ. These differences are characteristics of the population (Gallinat & Spaulding, 2014; Rudolph, 2017). The last three entries in Table 1 are the scores used as predictors of cross-situational outcomes. The Early Childhood Picture Vocabulary Test from the NIH Toolbox (Gershon et al., 2013; Weintraub et al., 2013) assesses auditory single word comprehension in an adaptive procedure. The standard score, corrected for age, served as a proxy for extant vocabulary size. On the Track-It Sustained Attention Task

(Erickson et al., 2015; Fisher et al., 2013, <http://www.psy.cmu.edu/~trackit>), children visually track a target shape as it moves among other foil shapes. After the trial, the child identifies the final position of the target and then demonstrates memory for the target in an AFC array. When the target moves among foils that differ in shapes and color (heterogeneous condition), one can derive a measure of endogenous attention by calculating proportion of trials correct after removing incorrect trials on the memory probe (Fisher, Thiessen, Godwin, Kloos, & Dickerson, 2013). The sustained attention performance of 67 children in the current study is explored in detail in Smolak et al. (2020). The score derived from the nonword repetition task (Dollaghan & Campbell, 1998) was the proportion of phonemes correctly repeated, a proxy for phonological working memory ability.

To anticipate the need to interpret any observed sex differences in learning trajectories or outcomes, note that there were no significant differences between boys and girls in age, parental education, TNL, nonverbal IQ, vocabulary, sustained attention, or nonword repetition (all $ps > .17$).

2.2. Stimuli

The stimuli were eight sets of 12 unfamiliar objects depicted in color photographs and their novel names. The objects were rare (unfamiliar to adult judges) mammals, birds, flowers, and fruits/vegetables. Two familiar words and pictured referents, *dog* and *watermelon*, served as filler stimuli. One of the eight sets was assigned to each child in the year 1 cross-situational learning study reported here (the others were assigned to other learning conditions and years in a manner counterbalanced across children). The novel names, randomly assigned to the objects, were mono- and disyllables created to accord with the phonotactics of English. The range of phonotactic probability and lexical neighborhood density values and their means were matched across the eight sets to ensure comparable learning and recall challenges (Storkel & Lee, 2011). Multiple talkers recorded the stimulus items to facilitate learning (Richtsmeier, Gerken, Goffman, & Hogan, 2009) and recognition (Creel, Aslin, & Tanenhaus, 2008). For example, using multiple talkers has been shown to facilitate cross-situational word learning in children with poor sustained attention skills (Crespo & Kaushanskaya, 2021). Additional details about the novel word stimuli are available in McGregor (2021a,b).

2.3. Procedure

The cross-situational learning paradigm comprised six cycles of 14 trials each, one trial for each of the 12 words to be learned, plus two filler trials in which *watermelon* and *dog* were always paired. Filler trials were meant to prevent the children with DLD from getting discouraged by potentially low performance and to signal to all learners that asking them to identify the same set of referents six times in a row was not necessarily an indication of an error on their part. The order of trials was randomized within cycles across participants.

The number of words and cycles was determined by a review of the literature and piloting. Yu and Smith (2007) gave adult learners 18 words to learn in six cycles, and Trueswell et al. (2013) gave adult learners 12 words to learn in five cycles, but each of these tasks was harder

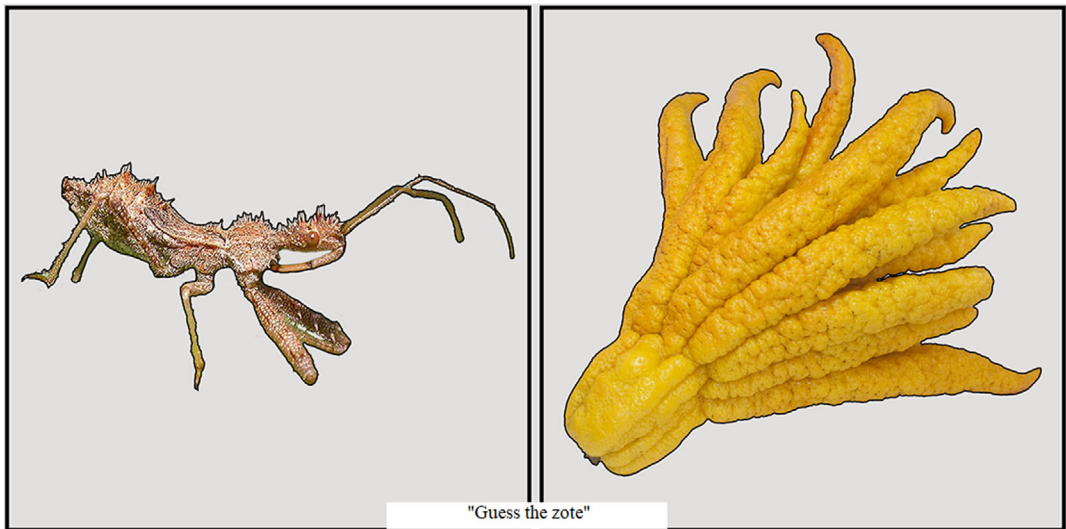


Fig 1. An example of stimulus presentation.

than the one proposed here because each involved more than one word or more than two referents per trial. After piloting five cycles for 12 words, a sixth cycle was added to increase the likelihood that children with DLD would perform better than chance by the end of the training.

On each trial, the participant saw two pictured items on the computer screen, the referent and a foil, and 500 ms later heard “guess the X,” with X being the name of the referent (Fig. 1). In contrast to Ahufinger et al. (2021), where children were exposed to word-referent pairings during training but did not provide a response, in the current study, participants answered by touching one of the two items on the screen. For any given trial within a cycle, the foil that accompanied the referent was a random selection from the remaining 11 items. In other words, a pictured item appeared once per cycle as a referent, but it could also appear more than once as a foil. In this way, a participant could perform better than chance without tracking across-cycle information.

This possibility is most easily understood by consulting the example in Table 2. In this example, the participant guessed the referent for the target word *zote* (paired with the *zote* and *pargle* referents) correctly in cycle 1-trial 1. The *zote* referent then appeared as a foil for cycle 1-trial 6 *klotig* (paired with the *klotig* and *zote* referents) and cycle 1-trial 14 *brep* (paired with the *brep* and *zote* referents). This participant also correctly identified the referents for *klotig* and *brep*. Armed with the principle of mutual exclusivity and a hypothesis about the meaning of *zote* (or something less conscious, like the strength of the previous *zote*–*zote* association), the participant could infer that the other pictured item was the named referent. Note that even when the hypothesized meaning is wrong, the reappearance of foils could aid the learner. Consider cycle 4-trial 1, where the participant incorrectly selected the foil *garp* as the referent for the target word *naskin*. When the *garp* reappeared as a foil on trial 8, he

Table 2
An example of one participant's responses by trial, cycle, and accuracy

Trial	Cycle 1			Cycle 2			Cycle 3			Cycle 4			Cycle 5			Cycle 6			
	Target	Foil	Acc	Target	Foil	Acc	Target	Foil	Acc	Target	Foil	Acc	Target	Foil	Acc	Target	Foil	Acc	
1	zote	pargle	1	brep	zote	1	seepie	modge	1	naskin	garp	0	brep	fetchik	1	chawg	seepie	1	
2	modge	duver	0	naskin	chawg	0	garp	klotig	1	koost	pargle	0	modge	naskin	1	zote	chawg	1	
3	fetchik	modge	1	klotig	naskin	1	naskin	fetchik	0	duver	koost	0	koost	zote	0	brep	naskin	1	
4	dog	melon	1	modge	pargle	1	pargle	klotig	1	brep	pargle	1	chawg	seepie	1	klotig	brep	1	
5	chawg	modge	0	garp	fetchik	1	dog	melon	1	seepie	zote	1	seepie	naskin	1	duver	chawg	1	
6	klotig	zote	1	chawg	zote	1	koost	seepie	1	zote	fetchik	1	zote	pargle	1	koost	chawg	1	
7	melon	dog	1	duver	klotig	1	klotig	duver	0	klotig	brep	1	fetchik	naskin	1	seepie	brep	1	
8	duver	garp	0	fetchik	brep	0	chawg	koost	1	chawg	garp	1	duver	koost	1	dog	melon	1	
9	koost	chawg	1	melon	dog	1	fetchik	duver	1	dog	melon	1	dog	melon	1	naskin	brep	0	
10	pargle	garp	1	koost	duver	1	zote	koost	1	garp	seepie	1	pargle	naskin	1	garp	brep	1	
11	naskin	fetchik	0	seepie	modge	1	brep	garp	1	fetchik	koost	0	garp	seepie	1	fetchik	duver	1	
12	seepie	fetchik	1	pargle	klotig	0	modge	duver	1	pargle	koost	1	melon	dog	1	modge	garp	1	
13	garp	chawg	0	zote	klotig	1	melon	dog	1	melon	dog	1	klotig	modge	1	pargle	seepie	1	
14	brep	zote	1	dog	melon	1	duver	modge	1	modge	klotig	1	naskin	zote	0	melon	dog	1	
Mean			58%			75%			83%			67%			83%				92%

Abbreviations: Acc, accuracy: 1, yes; 0, no; Mean, number of novel items correct divided by total novel items.

correctly guessed the *chawg*. It does not matter that he thinks *garps* are called *naskins*; he can determine that they are not called *chawgs*.

Across cycles, there was additional information available to support learning. Specifically, the co-occurrence frequency of a word with its target referent was 100% (six of six cycles), and the co-occurrence of a word with a given pictured foil was 17% (one of six cycles) for four foils and 33% (two of six) for one foil. Thus, the consistent pairing of word and referent across all six cycles together with the inconsistent pairing of word to foil enabled a correct outcome under either propose-but-verify or statistical aggregation strategies. Of course, both strategies depend on the ability to attend to and remember information across cycles, so although perfect performance is theoretically possible by the end of the sixth cycle, it is unlikely.

Following the cross-situational learning task was a 5-min retention interval during which the child conversed with the examiner to minimize the chance of rehearsal. The 3AFC form recognition test and the 3AFC word-to-referent link recognition test were then administered, in that order so that the test that involved the most cueing was last. On each trial of the form recognition task, the child heard a trained word (e.g., *zote*) along with two untrained words that differed from the trained word on a single medial or final consonant (e.g., *zoke*, *zofe*). A dot appeared on the screen as each word played. The child was instructed to touch the dot that matched the word they had heard in training. To ensure that the child listened to all three choices, the touch function was activated after all three dots were on the screen. The order of target and foils varied across trials. On each trial of the word-to-referent link recognition test, the child saw three pictured objects from the training set on the computer screen, heard a word that labeled one of them, and touched the picture that matched the label. Each trained referent served once as the target and, via random selection, twice as a foil. For both tests, each trial ended with the child's touch or after 3 s, whichever came first.

2.4. Data analysis

We conducted a Bayesian hierarchical logistic regression transition analysis because it is well suited for modeling complex scientific phenomena in which prior knowledge influences performance, it does not require large samples, and, given the data and a prior distribution, it yields a posterior probability distribution of the effect size rather than a single point estimate (Oleson, Brown, & McCreery, 2019; Smit, Milne, Dean, & Weidemann, 2019; van de Schoot & Depaoli, 2014). The outcome variable was whether the individual correctly identified the referent upon hearing the word on each cycle, which was assumed to follow a Bernoulli distribution with a probability of success π . A logit transformation was then taken on π to set up a logistic regression for the probability of successfully identifying each word correctly. The terms in the logistic regression model included a lag effect to capture the fact that the probability of a correct answer may depend upon accuracy in the previous cycle. Other covariates in the model included diagnostic group (DLD, reference = TD), trial (1,...,12), cycle (reference = 1, 2,...,6), and two-way interactions between diagnostic group and lag, trial, and cycle. Nonverbal IQ was not included as a covariate because it captured no variance beyond that captured by diagnostic group (see details at https://osf.io/u78pr/?view_only=1bb11610c10a415d90b938e12e737fac). Sex was

not included as a covariate because a preliminary two-sample Welch's *t*-test to compare the mean accuracy across all cycles for boys (54%) versus girls (49%) was not significant, $t(60.8)$, $p = .322$. There was also a random effect for subject to allow for correlation between repeated observations per subject and a random effect for item to account for some items being potentially harder than others.

A vague normal prior distribution with a mean of 0 and a variance of 100 was placed on the coefficients representing diagnosis, trial, lag, cycle, and the interactions. The vague normal prior distribution essentially placed a noninformative prior on beliefs about these estimates, thus allowing the data to drive the model. For the subject and item effects, a vague normal prior distribution was used $N(0, \sigma^2)$ with the variance parameters for both subject and item receiving an inverse gamma prior (0.01, 0.01).

The analysis was performed in R (R Core Team, 2020) and run in OpenBUGS using the R2OpenBugs package. There were three chains with different starting values, where each chain ran for 50,000 iterations after a burn-in of 20,000 iterations to reach convergence. All Gelman–Rubin convergence diagnostics equaled 1.0, which indicates that convergence was met for all model parameters.

To test outcomes as measured after a retention interval, proportion correct scores per participant were computed across the 12 items per 3AFC form and 3AFC link tasks. Those proportion correct scores served as the dependent variable in a Bayesian linear mixed model. The fixed effects were word component (form, reference = link) and diagnostic group (DLD, reference = TD) as well as the interactions between word component and diagnostic group. A random intercept for subject was included because the word component was a within-subject variable. Analyses were conducted in the BRugs package in R. There were three chains with different starting values, where each chain ran for 10,000 iterations after a burn-in of 5000 iterations to reach convergence.

To determine predictors of learning and retention, zero-order correlations were examined, and then three relative importance analyses were conducted to determine the relative contribution of endogenous attention, phonological working memory, and extant vocabulary to performance on (a) the final cycle of learning and after a 5-min retention interval as determined by (b) the 3AFC form task and (c) the 3AFC link task. In relative importance analyses (Johnson, 2000), orthogonal variables are created to determine how much unique variance in the dependent measure is explained by each predictor. This procedure offers advantages over multiple regression when predictors are correlated and is a more accurate method of partitioning out variance explained by each predictor variable. The analysis was conducted in R (R Core Team, 2020) using the script in Tonidandel and LeBreton (2015). After computing orthogonal raw correlations and 95% confidence intervals for vocabulary, attention, and phonological working memory, weights were rescaled to represent the proportion of total variance accounted for by each predictor for the sample as a whole and individually for the DLD and TD groups. The significance of each weight was determined via a bootstrap procedure of 10,000 iterations with replacement (Tonidandel, LeBreton, & Johnson, 2009). Finally, whether the weights for attention, phonological working memory, and vocabulary were statistically different between the two groups was determined.

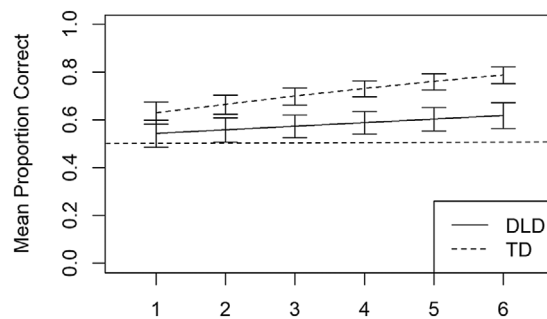


Fig 2. Mean proportion of correct referent selections during cross-situational learning by diagnostic group (developmental language disorder and typical development) and cycles (1–6). Note that chance is 0.50.

All raw data and analysis codes are available in Open Science Framework at https://osf.io/u78pr/?view_only=1bb11610c10a415d90b938e12e737fac

3. Results

3.1. Accuracy and improvement over cycles

Fig. 2 displays the smoothed raw data as the mean proportion of items correct per cycle for the two diagnostic groups. On cycle 1, the TD group performed above chance, $t(43) = 5.01$, $p < .001$, whereas the DLD group was at chance, $t(27) = 0.58$, $p = .57$. By the final cycle, the DLD group averaged 62% correct ($SD = 18\%$) and the TD group averaged 77% correct ($SD = 14\%$), a numeric difference of 15%.

The outcomes of the Bayesian logistic regression transition analysis applied to the accuracy data appear in Table 3; details appear in (see details at https://osf.io/u78pr/?view_only=1bb11610c10a415d90b938e12e737fac). Reliable effects were those with credible intervals that did not include zero. There was a reliable effect of diagnostic group, favoring the TD group. The odds were 83% higher that a person in the TD group would answer correctly at cycle 1 relative to a person in the DLD group from cycle 1 to 3; and reliable lag effects at lags 3 to 4, 4 to 5, and 5 to 6. None of the interactions were reliable.

To better understand the contribution of trials and cycles to learning, consider that for every 1 unit increase in trial, there was a multiplicative effect of 1.019 on the odds that a TD individual would answer correctly; a 1.9% increase in odds for every trial accrued from 1 to 12. The effect was numerically larger for the DLD group with an odds ratio estimate of $(\exp(0.019+0.020) = 1.040)$, a 4% increase in odds per trial. The cycle effects represent the size of the deviation from cycle 1 for the TD group (the referent group) after controlling for variance related to subject, item, trial, and lag. The only reliable cycle effect was cycle 3, where the main effect has a 95% credible interval that does not include 0. The odds of a correct answer were 51% higher in cycle 3 than in cycle 1 for the TD group. For the DLD group, the odds of a correct answer were 16% higher in cycle 3 than cycle 1 ($\exp(0.41 - 0.263) = 1.158$). Note that the odds for the TD and DLD group did not differ reliably (the 95% credible interval for the cycle 3 \times diagnostic group interaction includes zero).

Table 3
Summary statistics for the Bayesian logistic regression

Model parameter	Posterior** mean	Posterior SD	95% Credible interval	Odds ratio	95% Odds ratio credible interval	Inverted odds ratio***
Dx (TD is baseline)	-0.606	0.216	(-1.031, -0.185)*	0.545	(0.357, 0.831)	1.834 (1.203, 2.804)
Trial	0.019	0.018	(-0.004, 0.042)*	1.019	(0.999, 1.043)	
Dx × Trial	0.020	0.018	(-0.016, 0.055)	1.020	(0.984, 1.057)	
Cycle 1	0.424	0.140	(0.151, 0.698)*	1.528	(1.163, 2.011)	
Cycle 2	0.101	0.178	(-0.245, 0.453)	1.106	(0.782, 1.573)	
Cycle 3	0.410	0.187	(0.047, 0.779)*	1.507	(1.048, 2.180)	
Cycle 4	-0.078	0.191	(-0.450, 0.298)	0.925	(0.638, 1.347)	
Cycle 5	0.235	0.195	(-0.145, 0.621)	1.265	(0.865, 1.861)	
Cycle 6	0.321	0.217	(-0.097, 0.753)	1.379	(0.907, 2.123)	
Cycle 2 × Dx	0.203	0.266	(-0.318, 0.720)	1.224	(0.727, 2.054)	
Cycle 3 × Dx	-0.263	0.276	(-0.802, 0.275)	0.769	(0.448, 1.317)	
Cycle 4 × Dx	0.172	0.284	(-0.388, 0.726)	1.188	(0.678, 2.067)	
Cycle 5 × Dx	-0.118	0.282	(-0.673, 0.436)	0.889	(0.510, 1.546)	
Cycle 6 × Dx	-0.312	0.300	(-0.902, 0.275)	0.732	(0.406, 1.316)	
Lag1 (1–2)	-0.064	0.195	(-0.446, 0.314)	0.938	(0.640, 1.369)	
Lag2 (2–3)	-0.077	0.205	(-0.482, 0.322)	0.926	(0.618, 1.380)	
Lag3 (3–4)	0.677	0.210	(0.266, 1.086)*	1.969	(1.305, 2.962)	
Lag4 (4–5)	0.770	0.222	(0.334, 1.204)*	2.159	(1.397, 3.333)	
Lag5 (5–6)	0.539	0.235	(0.078, 0.997)*	1.714	(1.082, 2.711)	
Lag1 × Dx	-0.125	0.300	(-0.714, 0.462)	0.882	(0.490, 1.588)	
Lag 2 × Dx	0.472	0.311	(-0.136, 1.087)	1.603	(0.873, 2.965)	
Lag 3 × Dx	-0.443	0.314	(-1.057, 0.167)	0.642	(0.347, 1.181)	
Lag 4 × Dx	-0.446	0.321	(-1.068, 0.181)	0.640	(0.344, 1.198)	
Lag 5 × Dx	0.246	0.338	(-0.414, 0.908)	1.279	(0.661, 2.480)	

Abbreviations: Dx, diagnostic group; SD, standard deviation; TD, typically developing.

*The credible interval does not include 0.

**The posterior mean indicates the log odds of the probability of a correct answer.

***For ease of interpretation, the negative odds ratio was inverted from cycle 1 to 3; and reliable lag effects at lags 3–4, 4–5, and 5–6.

Recall that lag effects indicate the extent to which accuracy at one cycle depended upon accuracy at the just prior cycle. This dependence emerged only in the later cycles. Consider lag 3 for example: for a participant with TD in cycle 4, the odds of answering correctly on an item that had been correct in cycle 3 were 1.97 times the odds of answering correctly on an item that had been incorrect (97% higher). For a participant with DLD, this value was 1.26 ($\exp(0.677 - 0.443) = 1.264$; 26% higher). For lag 4, the odds that a participant with TD would be correct on a previously correct item versus a previously incorrect item were 2.16 (116% higher) and for lag 5, the value was 1.74 (74% higher). For participants with DLD, the lag 4 odds were 38% higher ($\exp(0.77 - 0.446) = 1.38$) and for lag 5, they were 119% ($\exp(0.539 + 0.246) = 2.19$). Again, the credible interval for the diagnostic group × lag interaction terms included 0, indicating that the lag effects for TD and DLD were not reliably different from each other.

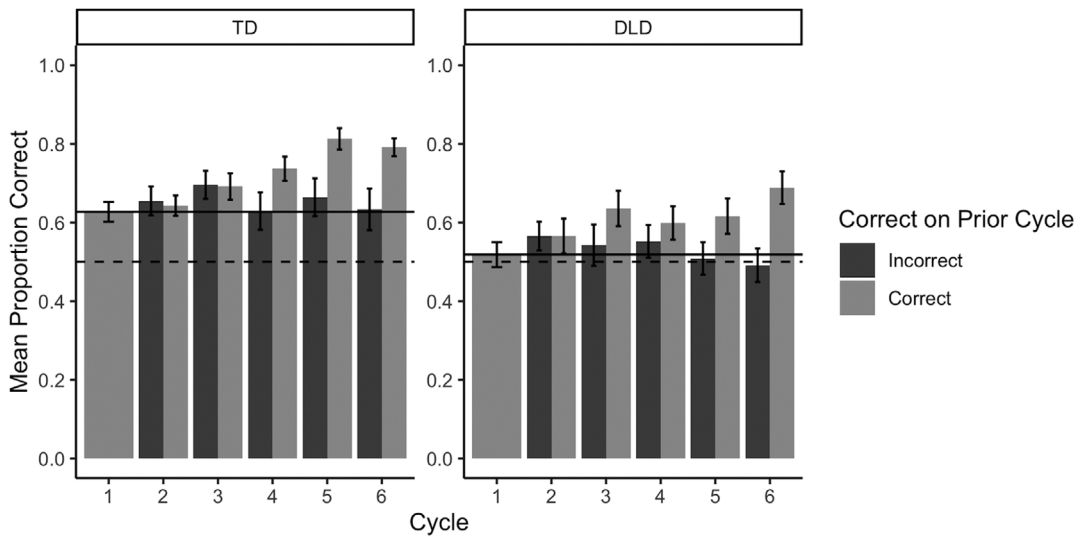


Fig 3. Mean proportion of correct referent selections made during cross-situational learning by cycles (1–6). The dotted line denotes chance; the solid line denotes mean accuracy in cycle 1.

At this point, we have answered the first research question, which pertained to accuracy. As predicted, the DLD group performed less accurately than the TD group during learning. Contrary to prediction, the lack of a group \times cycle interaction suggests that the trajectory of change in accuracy did not vary reliably by group.

We also have partial answers to the second research question, which pertained to strategy use. A variety of strategies were evident. The trial effect suggests the use of mutual exclusivity to propose word-referent mappings and the lack of a group \times trial interaction suggests that both groups made use of mutual exclusivity, although it is clear that the DLD group did not do so with success in cycle 1. Early in the course of learning, from cycles 1 to 3, there was growth in accuracy, but that growth was not dependent upon prior accuracy; in other words, there was a reliable cycle effect but no lag effect. This pattern suggests a statistical aggregation approach. In cycles 4–6, the significant lag effects demonstrate that accuracy becomes dependent on prior accuracy in a manner consistent with propose-but-verify.

Of particular interest is whether the shift from statistical aggregation to propose-but-verify characterized both groups of learners. To test the robustness of these patterns, we examined the posterior mean probabilities from the Bayesian hierarchical logistic regression transition analysis; we compared the probabilities at each cycle to cycle 1. The raw data appear in Fig. 3 in the manner of Trueswell et al. (2013) and the results of this analysis are presented in Table 4 (see details at https://osf.io/u78pr/?view_only=1bb11610c10a415d90b938e12e737fac for plots of the estimates of this model). Note that, in cycle 3, participants with TD demonstrated an increase over the probability of a correct answer in cycle 1, regardless of accuracy in cycle 2. Clearly, they were retaining information about the foils that they had seen previously.

Table 4

Summary statistics for probabilities relative to cycle 1 performance with developmental language disorder (DLD) and typical development (TD)

Cycle	Accuracy	TD			DLD		
		Posterior mean	<i>SD</i>	95% Credible interval	Posterior mean	<i>SD</i>	95% Credible interval
2	0	0.023	0.042	(−0.059, 0.105)	0.075	0.049	(−0.022, 0.170)
3	0	0.092	0.041	(0.011, 0.170)*	0.037	0.051	(−0.063, 0.136)
4	0	−0.019	0.046	(−0.110, 0.070)	0.023	0.053	(−0.079, 0.126)
5	0	0.054	0.044	(−0.034, 0.139)	0.029	0.051	(−0.071, 0.128)
6	0	0.072	0.048	(−0.023, 0.164)	0.002	0.052	(−0.098, 0.104)
2	1	0.009	0.036	(−0.062, 0.078)	0.028	0.048	(−0.066, 0.122)
3	1	0.076	0.035	(0.008, 0.144)*	0.133	0.047	(0.041, 0.225)*
4	1	0.131	0.033	(0.066, 0.195)*	0.081	0.046	(−0.009, 0.172)
5	1	0.201	0.032	(0.139, 0.265)*	0.109	0.047	(0.017, 0.200)*
6	1	0.178	0.032	(0.116, 0.241)*	0.192	0.046	(0.101, 0.282)*

Abbreviation: *SD*, standard deviation.

*The credible interval does not include 0.

Table 5

Performance, relative to chance, on retention measures of word form and the link between the word form and its referent

Word component	<i>N</i>	Dx	Proportion correct	95% Credible interval	Probability > chance (.33)
Form	28	DLD	0.378	(0.300, 0.455)	.886
Form	44	TD	0.559	(0.498, 0.621)	1.00
Link	28	DLD	0.461	(0.383, 0.539)	.999
Link	44	TD	0.582	(0.521, 0.643)	1.00

Abbreviations: DLD, developmental language disorder; Dx, diagnostic group; *N*, number of participants; TD, typical development.

In contrast, in cycles 4–6, they demonstrated an increase over the probability of a correct answer in cycle 1, but only for answers that had been correct on the previous cycle. Together, these results are consistent with an aggregation strategy in early cycles, switching to a propose but-verify strategy later. As predicted, the participants with DLD never evinced statistical aggregation. However, on cycles 3, 5, and 6, they appear to have applied a propose-but-verify strategy with success.

3.2. Learning outcomes

Proportion correct scores per participant were computed across the 12 items in the link and form 3AFC tasks and used as the dependent variable in a Bayesian linear mixed model. From the model, we computed the estimated proportion correct for each group and word component combination. Then, we used the posterior estimate from the Bayesian model to compute the probability that the diagnostic group performed better than chance (33%) on the 3AFC tasks of form and link recognition (Table 5). Note that the DLD group did not perform

Table 6
Linear mixed model of retention performance

Effect	Posterior mean	SD	95% Credible interval	Prob > 0
Intercept	0.58	0.03	(0.52, 0.64)	1.00
Form (link reference)	-0.02	0.04	(-0.09, 0.05)	.260
DLD (TD reference)	-0.12	0.06	(-0.022, -0.02)	.007*
Form × Dx	-0.06	0.06	(-0.17, 0.05)	.139

Abbreviations: DLD, developmental language disorder; *SD*, posterior standard deviation; TD, typically development.

*The credible interval does not include 0.

Table 7
Zero-order correlations of phonological short-term memory, extant vocabulary knowledge, and endogenous sustained attention in predicting learning (cycle 6) and memory (3AFC form and link) outcomes

	TD			DLD		
	Vocabulary	Attention	Memory	Vocabulary	Attention	Memory
Cycle 6	.35*	-.11	.09	.17	.43*	-.08
Form	.43*	.10	.24	.21	-.14	.03
Link	.48*	.16	.09	-.04	.11	.29

Note. Memory was measured as percent phonemes correct on a nonword repetition task; Vocabulary was measured as the standard score on the NIH Toolbox vocabulary test; Attention was measured as the endogenous attention score on the Track-It sustained attention task.

Abbreviations: DLD, developmental language disorder; TD, typically developing.

* $p < .05$.

reliably above chance on form recognition. Consistent with prediction three, the TD group outperformed the DLD group on the retention tasks (Table 6). Participants with DLD scored about 13% lower than those with TD, on average. However, contrary to prediction, there was no interaction between word component and diagnostic group. The estimated Pearson correlation between performance on form and link recognition was 0.40, $p = <.001$.

3.3. Mechanisms

For the TD group, vocabulary was positively and significantly associated with performance on all learning outcome measures; for the DLD group, it was not (Table 7). The performance of the DLD group during cycle 6 was positively and significantly associated with sustained attention but not vocabulary or memory.

3.3.1. Cycle 6

In the first relative importance analysis, accuracy on the sixth and final learning cycle was the dependent variable. Overall, 24% of the variance in accuracy was accounted for by the three predictor variables. Vocabulary was the most important and only significant predictor, accounting for 15% of the variance, and 65% of the total variance accounted for by all

three predictors combined (Table 8). With groups examined separately, these predictors combined accounted for 15% of the TD group variance and 21% of the DLD group variance. As expected, given the pattern of correlations, there was a trend such that vocabulary played a more significant role for the TD group, whereas attention played a more significant role for the DLD group, but these weights were not statistically different between groups.

3.3.2. *Form retention*

Next, the relative importance of attention, phonological working memory, and vocabulary to the outcomes on the 3AFC measure of word form recognition was determined. Together, the three variables accounted for 30% of the total variance (Table 8). Overall, vocabulary was the most important and only significant predictor, accounting for 22% of the variance and 73% of the total variance accounted for by all three variables combined. With groups examined separately, the three variables combined accounted for 21% of the variance in the TD group but only 7% of the variance in the DLD group.

3.3.3. *Link retention*

Together, the three variables accounted for 18% of the total variance in 3AFC link performance (Table 8). Overall, vocabulary was the most important and only significant predictor, accounting for 12% of the variance and 63% of the total variance accounted for by all three variables combined. With groups examined separately, these variables account for 23% of the variance in the TD group but less than 10% of the variance in the DLD group. There was a trend such that vocabulary played a greater role for the TD group, whereas phonological working memory played a greater role for the DLD group. The difference in vocabulary as a predictor for the two groups was statistically significant.

Thus, the fourth prediction, that individual differences in vocabulary knowledge, sustained attention, and working memory would be associated with cross-situational word learning performance, was partially supported. Vocabulary played a role in learning outcomes for TD children, and sustained attention played a role for children with DLD in cycle 6. Relative importance analyses identified vocabulary as the most important contributor to performance overall.

4. Discussion

In this study, learners with DLD comprised a test case to explore the role of extant vocabulary knowledge, endogenous attention, and phonological working memory abilities in the accessibility of cross-situational word learning. On average, children with DLD perform more poorly than typically developing age-mates in these areas (Ebert & Kohnert, 2011; Graf Estes et al., 2007; McGregor et al., 2013), and that was true in the current sample. That said, individual variation within groups of learners with DLD can be considerable (McGregor, Arbisi-Kelm, & Eden, 2017), which was also the case in the current sample. We made group comparisons and mined the individual variation to advance understanding of cross-situational word learning.

Table 8
Relative importance of phonological working memory, extant vocabulary knowledge, and endogenous sustained attention in predicting learning (cycle 6) and memory (3AFC form and link) outcomes

Task	Predictors	Overall relative weights			95% CI		DLD relative weights			TD relative weights			95% CI	
		Raw	Rescaled	Lower bound	Upper bound	Raw	Rescaled	Raw	Rescaled	Raw	Rescaled	Lower bound	Upper bound	
Cycle 6	Vocabulary	0.153	64.604	0.035	0.332*	0.024	11.065	0.129	85.732	0.129	85.732	-0.419	0.085	
	Attention	0.059	25.174	-0.009	0.275	0.184	86.276	0.018	12.088	0.018	12.088	-0.013	0.501	
Form	Memory	0.024	10.222	-0.031	0.155	0.006	2.659	0.003	2.180	0.003	2.180	-0.140	0.164	
	Vocabulary	0.215	72.649	0.058	0.366*	0.045	64.365	0.155	74.893	0.155	74.893	-0.363	0.112	
Link	Attention	0.008	2.779	-0.107	0.027	0.023	32.710	0.010	4.662	0.010	4.662	-0.072	0.177	
	Memory	0.070	24.571	-0.048	0.174	0.002	2.925	0.042	20.445	0.042	20.445	-0.228	0.076	
	Vocabulary	0.116	63.122	0.002	0.279*	0.001	0.886	0.215	90.311	0.215	90.311	-0.477	-0.0003*	
	Attention	0.023	12.403	-0.025	0.106	0.012	12.215	0.018	7.400	0.018	7.400	-0.124	0.090	
	Memory	0.045	24.475	-0.018	0.169	0.084	86.899	0.005	2.289	0.005	2.289	-0.046	0.342	

Note: Memory was measured as proportion of phonemes correct on a nonword repetition task; Vocabulary was measured as the standard score on the NIH Toolbox vocabulary test; Attention was measured as the endogenous attention score on the Track-It sustained attention task.
Abbreviations: DLD, developmental language disorder; TD, typically development.
*All confidence intervals (CI) that do not include zero indicate $p < .05$.

4.1. Word learning trajectories

Although there was a reliable main effect of diagnostic group, there were no interactions between group and cycle or trial. The DLD groups were less accurate than their peers on cycle 1 and remained lower in accuracy throughout the course of learning, but their growth trajectories did not differ reliably. Contrary to our predictions, this would suggest that their poorer outcome was not a matter of statistical learning deficits per se but of their ability to reduce the computational load of the task before cross-situational cues were present.

The *only* information that could support better than chance performance in cycle 1 was the repetition of referents as foils. As the trials progressed, a child's mapping of a previous referent that then reappeared as a foil could allow an inference that the never-before-presented item was being named. Thus, at first glance, we might attribute the poorer outcome to a problem in using a mutual exclusivity heuristic. However, such a problem seems at odds with reports of children with DLD using mutual exclusivity successfully to infer word-to-referent links by age 6 (Estis & Beverly, 2015). Moreover, the lack of diagnostic group \times trial interaction in the current study suggests that the children with DLD did make use of trial-level information. In fact, the effect for the DLD group was numerically larger than the effect for the TD group.

Previous reports of successful application of the mutual exclusivity heuristic by children with DLD are based on referent arrays of one familiar and one novel object. Here, participants had to infer that a previously presented novel object was an unlikely referent when paired with a never presented novel object. The memory demands of novel–novel pairings present a more difficult fast-mapping challenge than familiar–novel pairings even for typical children (Lewis et al., 2020). We hypothesize that the challenge was too great for the children with DLD. With subsequent cycles, the items and their names became more familiar and the children with DLD were then able to make use of repetitions of items across trials to benefit from a mutual exclusivity inference. In other words, we do not view performance gap in cycle 1 as a problem with logical inference but as a problem in building the memory trace needed to support the inference.

Indirect support for this hypothesis can be found in Bishop and Hsu (2015). In that study, 7-to-11-year-olds with DLD and two groups of typical peers, one matched for age and the other for syntactic ability, learned to associate unfamiliar words with their pictured referents (eight unfamiliar animals) over trials via a computer game. During the first phase of the game, each referent was named twice for the child. In the next phase, the child saw an array of four of the animal pictures and they selected the one being named. Feedback on the accuracy of the selection was provided. This phase was repeated three times for each target word on each of 4 days. Learners in this study had many cues available: ostensive naming on the first two trials, cross-situational information, mutual exclusivity information, and feedback. Although their learning protocol differed from ours, their findings were parallel. The children with DLD (and their younger syntax-matched peers) were poorer than the typical age-mates on the first trial of the second phase, but they improved over trials at a similar rate. Relevant data also appear in Haebig et al. (2019) and Leonard et al. (2019). In these studies, children with DLD and their typical age-mates learned words via repeated spaced retrieval practice. Specifically,

they saw one pictured referent at time, heard it named three times, and then were asked to name the referent when it appeared on subsequent trials. Note that the mutual exclusivity heuristic is not a relevant support for word learning in these two studies at all yet, again, the children with DLD were less accurate than their TD peers on their first naming attempt but they improved in accuracy over trials at a similar rate. Thus, we hypothesize that the difficulties exhibited by the DLD group during cycle 1 had nothing to do with application of the mutual exclusivity heuristic but everything to do with the novel information they were trying to encode. When their memory for the new names and referents was fragile, they fared poorly but with subsequent exposures, they learned at the same rate as their peers.

Note that the deficit in initial encoding—if we are correct in positing such a deficit—does not constitute a problem with cross-situational learning. Instead, it put the children with DLD at a disadvantage upon first receipt of cross-situational information, a gap they never overcame. In fact, given no diagnostic group \times cycle interaction, we have no evidence that children with DLD found cross-situational learning problematic. Nevertheless, cross-situational learning can be achieved by various strategies and the strategies used by the DLD and TD groups appeared to differ. We turn there now.

4.2. *Word learning strategies*

The typical learners demonstrated a mixture of strategies—statistical aggregation from cycle 1 to cycle 3 and propose-but-verify on cycles 4–6. The DLD group demonstrated a pattern consistent with propose-but-verify on cycles 3, 5, and 6, but they never evinced a statistical aggregation strategy. Note that these differences do not relate to the trajectory of growth over cycles: Given no reliable interaction between group and cycle, trajectories did not differ. Instead, the difference concerns the means by which the children managed to grow over trials.

The absence of any apparent statistical accrual strategy suggests that children with DLD have difficulties in computing co-occurrence frequencies during cross-situational word learning, as concluded by Ahufinger et al. (2021). This difference also accords with numerous reports of weak statistical learning in the DLD population as measured by their ability to segment speech (Evans et al., 2009; Haebig et al., 2017) and to learn grammatical constructions (Hsu et al., 2014; Lum et al., 2014; Torkildsen et al., 2013; Grunow et al., 2006; Plante et al., 2002), grammatical gender marking (Richardson et al., 2006), or stress assignment rules (Plante et al., 2010). The children with DLD in the current study demonstrated some success during cross-situational learning but appeared to do so via application of propositional logic rather than statistical aggregation. We cannot rule out the latter completely, but we can conclude that their typical peers presented a reliable pattern of statistical aggregation in the early cycles of learning that was not evident in the DLD group.

4.3. *Word learning outcomes*

Outcomes were measured in two recognition tasks administered 5 min after the final cycle of learning. Averaged across the two recognition tasks, the performance of the DLD group was reliably lower than that of the TD group, but by only a small amount, 13%, a

performance gap similar to that reported by Ahufinger et al. (2021). Recall that the performance gap at the end of cycle 6 was similar, 15%. Therefore, differences on the outcome measures likely reflected differences in learning rather than retention across the 5-min interval (see also McGregor, Arbisi-Kelm, et al., 2017; McGregor, Gordon, et al., 2017). Although we conclude that retention was robust, it is worth noting that the only difference between learning and outcome measures was a 2 versus 3 AFC array; thus, additional learning could have taken place during the test.

The retention of links between the newly learned words and their referents is consistent with Vlach and DeBrock (2019), who found that children of 50–70 months of age (but not younger children) retained word-to-referent pairings trained in a cross-situational learning paradigm over a 5-min interval. The participants in the current study were 72–98 months old.

Unlike Vlach and DeBrock (2019), we also examined the retention of word form information. Although the novel word targets were designed to be readily distinguishable, the retention test required precise knowledge of the word forms. For example, children whose training set included *zote* had to decide whether their newly learned word was *zote*, *zoke*, or *zofe*; those whose training set included *fechik* were tested on the array *fesik*, *fepik*, *fechik*. The TD group performed above chance, which is not surprising given their age. Adults have demonstrated the ability to represent the word forms learned in cross-situational paradigms with fine-grained precision (Escudero et al., 2016a), although it is easier to do so in response to ostensive naming paradigms (Mulak, Vlach, & Escudero, 2019). Infants, too, can learn precise word forms from cross-situational opportunities, although not as robustly as adults (Escudero et al., 2016b).

The DLD groups were not reliably above chance on form recognition; however, form recognition was not disproportionately harder than link recognition for the children with DLD relative to their peers. The positive correlation between performance outcomes on the 3AFC link and form tasks provides additional evidence that form learning was not highly divergent from link learning: Better learners tended to perform well on both tasks and weaker learners tended to perform worse on both tasks. Previous reports of exceptional difficulty with word form on the part of learners with DLD involved ostensive naming paradigms, where the link between the word and its referent is made explicit for the learner (Leonard et al., 2019; McGregor et al., 2017). This extra support for link learning may have amplified the gap between link and form recognition, an amplification that would not pertain here. A direct comparison of word form learning under conditions that vary these supports is underway in our laboratory.

4.4. *Cognitive mechanisms*

DLD was a focus here because, as a group, children with DLD present with weaknesses in sustained attention, phonological working memory, and extant vocabulary, the very mechanisms proposed to support cross-situational word learning. Indeed, in the current sample, there were large between-group differences in vocabulary and phonological working memory and a moderate difference in sustained attention, all favoring the TD group. As predicted, sustained attention, phonological working memory, and extant vocabulary combined to account for variance in learning outcomes: In the relative importance analysis, the combined

scores accounted for 24% of the variance in cycle 6 link recognition, 30% of the variance in form recognition retention, and 18% of the variance in link recognition retention.

Overall, the extant vocabulary score was the most important and, indeed, the only significant predictor of these three learning outcomes. The current findings are consistent with those of Smith and Yu (2013), Vlach and DeBrock (2017), and Scott and Fisher (2012) in that, in each of those studies, extant vocabulary size correlated positively with cross-situational learning outcomes. The robustness of the vocabulary effect likely reflects the reciprocal nature of the relationship. Children with larger vocabularies may have stronger phonotactic knowledge and a more robust lexical network in which to encode new memories. On the other hand, if children build a lexicon, in part, by learning words from information gathered across situations, then a significant relationship between the two is to be expected. Vocabulary knowledge may support cross-situational word learning, and cross-situational learning, in turn, may be a means of acquiring additional vocabulary knowledge.

That said, there were patterns in the data that suggest that cross-situational learning may contribute less to building a lexicon among children with DLD than among children with TD. First, according to the zero-order correlations, vocabulary and performance on all learning outcome measures were positively associated for the TD group but not the DLD group. Second, according to the relative importance analysis, vocabulary accounted for less variance among the DLD participants than the TD participants on all three outcome measures. For the link retention outcome, this difference was significant. There, vocabulary scores accounted for 21.5% of the TD group variance but less than 1% of the variance among the DLD group. If cross-situational learning is a viable means of building a lexicon, one would expect a significant relationship between cross-situational learning outcomes and extant vocabulary. Given the weak relationship in the DLD group, it appears that children with DLD may be more reliant upon ostensive naming contexts for vocabulary learning, a hypothesis consistent with theories that posit intact declarative mechanisms among learners with DLD (Hsu & Bishop, 2010; Ullman & Pierpont, 2005).

Endogenous, sustained attention was positively and significantly correlated with cycle 6 performance in the DLD group, and that group only, a medium effect size. In the relative importance analysis, attention alone was never a significant predictor of performance, although it was the best predictor of variance in cycle 6 performance among DLD learners. It may be that children with DLD who present with stronger attention skills compensate for their weak word learning in a way unavailable to those who have weaker attention. Alternatively, it may be that the reason children with DLD (especially those with weaker attention skills) rely upon propose-but-verify strategies is because they miss valuable co-occurrence information when their attention lapses.

Sustained attention captured little variance within the TD group, which may be surprising given previous work. Smith and Yu (2013) and Yu and Smith (2011) found a relationship between cross-situational learning and endogenous attention measured *during* the cross-situational task, not attention as measured on an independent task, and among babies, not older children. Thus, the difference in outcome could have at least two roots. Perhaps, to be a useful predictor, attention must be measured within or proximal to the task itself; ours was a distal measure, the Track-it, which was independent of the cross-situational learning task.

Also, because suppression-based mechanisms of attentional control emerge in infancy (Fisher & Kloos, 2016), the ability to resist novelty so as to focus attention on relevant information is likely more variable and fragile earlier than later in development. As a result, sustained attention may not be as strong a predictor of word learning in older children who are developing typically, especially those like the first graders studied here, who have experienced the attentional demands of formal instruction. First-graders with frank attention deficits could be an illuminating comparison group.

Although phonological working memory alone was not a significant predictor of learning outcomes among the participants as a whole, it was the best predictor of variance in link retention performance among children with DLD. Vlach and DeBrock (2017) found a significant relationship between memory abilities and cross-situational word learning after controlling for age, vocabulary, and attention. Not only were their participants younger than the current participants, but also, they measured the participants' memory in a paired-associate learning paradigm rather than a nonword repetition task. Given that the task required during learning was to determine the link between a word and its referent, the paired-associate memory is more directly applicable than phonological working memory. A visual working memory capacity that permits the associative binding of phonological strings to objects in the visual environment is likely to be particularly relevant for the cross-situational learning of words that label objects, the focus of the current study (Bhat et al., 2021). Memory is a complex construct that can be measured in multiple ways. Discovering the aspects of memory most pertinent to cross-situational learning is an essential part of the research agenda.

5. Limitations and future opportunities

Given the scant literature on cross-situational learning among children with DLD, the results reported here require both replication and extension. A longer training protocol may reveal abilities on the part of learners with DLD that were masked by the limited cycles of opportunity that characterized the current study. Protocols that remove opportunities to apply the mutual exclusivity heuristic may reduce the tendency toward propose-but-verify strategies and better isolate learning across trials from learning across cycles. Moreover, one might better distinguish between statistical aggregation and propose-but-verify by aligning the foils on the 3AFC link recognition measure according to their co-occurrence with the target during learning. Patterns of errors in which foils that had been paired with the target are selected more often than foils that had not been paired with the target would constitute evidence of statistical aggregation even if the behavioral lens of referent selection is too coarse to identify aggregation during learning.

The consideration of underlying mechanisms should also be extended. The relationship between endogenous sustained attention and cross-situational learning in the DLD group (but not the TD group) suggests the utility of other test cases, particularly those with ADHD who do not have frank language deficits, in further refining our understanding of the role of attention in cross-situational learning.

We should also consider other potential mechanisms. Given that sustained attention, phonological working memory, and extant vocabulary together accounted for only 18% of the

variance in outcomes as measured by link recognition and 30% as measured by form recognition, there is a large amount of variance left unaccounted for. The relative success that the participants with DLD demonstrated in the use of a propose-but-verify strategy motivates examination of propositional logic and inference-making as mechanistic supports.

6. Conclusions

Typical first graders demonstrated a flexible mixture of strategies for learning words in an ambiguous context. They relied upon mutual exclusivity in cycle 1 when statistical information was not yet available. They demonstrated a statistical aggregation strategy over cycles 1–3 and then switched to rely upon propositional logic via a propose-but-verify strategy in later cycles. Within the TD group, variation in cross-situational learning success was associated with variation in extant vocabulary. Better learners had larger receptive vocabularies.

First graders with DLD presented a different profile. They were less accurate in linking words to their referents than their typical peers by the end of cycle 1, a problem we attribute to weaknesses in initial encoding. They learned at a similar rate as their peers, but they never evinced a statistical accrual strategy. The weighting of mechanisms supporting their learning differed. Unlike their peers, their success during cross-situational learning varied with their attention skills, and their learning outcomes and the size of their extant vocabularies were unrelated.

Optimal language learning depends on multiple routes through the problem space that can be taken flexibly in response to the information that contexts provide. Children with DLD eventually manage this problem space, but without the flexibility or efficiency that characterizes their typical peers. This conclusion is offered as motivation for theorists to capture all learners, not just the most typical ones. The problem to be modeled is not only what information is available to learners, but also what information is accessible to them.

Acknowledgments

We thank the participants and their families. Nancy Ohlmann assisted with data collection and Chris Conway and Katherine Gordon provided useful feedback on an earlier draft. This work was supported by the National Institutes of Health-NIDCD, grant number 2R01DC011742-06.

Funding

The National Institutes of Health-NIDCD, grant number 2R01DC011742-06. The funder had no part in the study design; the collection, analysis, and interpretation of data; the writing of the report; or the decision to submit the article for publication.

References

- Adlof, S. M., & Patten, H. (2017). Nonword repetition and vocabulary knowledge as predictors of children's phonological and semantic word learning. *Journal of Speech, Language, and Hearing Research*, *60*(3), 682–693. https://doi.org/10.1044/2016_JSLHR-L-15-0441
- Alt, M., Meyers, C., & Alt, P. M. (2013). Using ratings to gain insight into conceptual development. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1650–1660. [https://doi.org/10.1044/1092-4388\(2013\)11-0317](https://doi.org/10.1044/1092-4388(2013)11-0317)
- American Speech-Language-Hearing Association (January 9, 2022). Childhood hearing screening. Retrieved from https://www.asha.org/practice-portal/professional-issues/childhood-hearing-screening/#collapse_1
- Ahufinger, N., Guerra, E., Ferinu, L., Andreu, L., & Sanz-Torrent, M. (2021). Cross-situational statistical learning in children with developmental language disorder. *Language, Cognition and Neuroscience*, *36*, 1180–1200. <https://doi.org/10.1080/23273798.2021.1922723>
- Bartolotti, J., & Marian, V. (2012). Language learning and control in monolinguals and bilinguals. *Cognitive Science*, *36*(6), 1129–1147. <https://doi.org/10.1111/j.1551-6709.2012.01243.x>
- Beitchman, J. H., Nair, R., Clegg, M., Ferguson, B., & Patel, P. G. (1986). Prevalence of psychiatric disorders in children with speech and language disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *25*(4), 528–535. [https://doi.org/10.1016/S0002-7138\(10\)60013-1](https://doi.org/10.1016/S0002-7138(10)60013-1)
- Benitez, V. L., Yurovsky, D., & Smith, L. B. (2016). Competition between multiple words for a referent in cross-situational word learning. *Journal of Memory and Language*, *90*, 31–48. <https://doi.org/10.1016/j.jml.2016.03.004>
- Berens, S. C., Horst, J. S., & Bird, C. M. (2018). Cross-situational learning is supported by propose but-verify hypothesis testing. *Current Biology*, *28*(7), 1132–1136.e5. <https://doi.org/10.1016/j.cub.2018.02.042>
- Bhat, A. A., Spencer, J. P., & Samuelson, L. K. (2021). Word-Object Learning via Visual Exploration in Space (WOLVES): A neural process model of cross-situational word learning. *Psychological Review*. <https://doi.org/10.1037/rev0000313>
- Bishop, D. V., & Hsu, H. J. (2015). The declarative system in children with specific language impairment: A comparison of meaningful and meaningless auditory-visual paired associate learning. *BMC Psychology*, *3*(1), 1–12. <https://doi.org/10.1186/s40359-015-0062-7>
- Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283. <https://doi.org/10.1073/pnas.1309518110>
- Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience & Biobehavioral Reviews*, *112*, 279–299. doi: 10.1016/j.neubiorev.2020.01.032
- Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*(2), 633–664. doi: 10.1016/j.cognition.2007.03.013
- Crespo, K., & Kaushanskaya, M. (2021). Is 10 better than 1? The effect of speaker variability on children's cross-situational word learning. *Language Learning and Development*, *17*, 397–410. <https://doi.org/10.1080/15475441.2021.1906680>
- Dautriche, I., & Chemla, E. (2014). Cross-situational word learning in the right situations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(3), 892–903. <https://doi.org/10.1037/a0035657>
- Dollaghan, C. A. (1987). Fast mapping in normal and language-impaired children. *Journal of Speech and Hearing disorders*, *52*(3), 218–222. <https://doi.org/10.1044/jshd.5203.218>
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language and Hearing Research*, *41*(5), 1136–1146. <https://doi.org/10.1044/jslhr.4105.1136>
- Ebert, K. D., & Kohnert, K. (2011). Sustained attention in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, *54*(5), 1372–1384. [https://doi.org/10.1044/1092-4388\(2011\)10-0231](https://doi.org/10.1044/1092-4388(2011)10-0231)

- Edwards, J., & Lahey, M. (1996). Auditory lexical decisions of children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 39(6), 1263–1273. <https://doi.org/10.1044/jshr.3906.1263>
- Erickson, L. C., Thiessen, E. D., Godwin, K. E., Dickerson, J. P., & Fisher, A. V. (2015). Endogenously and exogenously driven selective sustained attention: Contributions to learning in kindergarten children. *Journal of Experimental Child Psychology*, 138, 126–134. <https://doi.org/10.1016/j.jecp.2015.04.011>
- Escudero, P., Mulak, K. E., Fu, C. S., & Singh, L. (2016). More limitations to monolingualism: Bilinguals outperform monolinguals in implicit word learning. *Frontiers in Psychology*, 7, 1218. <https://doi.org/10.3389/fpsyg.2016.01218>
- Escudero, P., Mulak, K. E., & Vlach, H. A. (2016a). Cross-situational learning of minimal word pairs. *Cognitive Science*, 40(2), 455–465. <https://doi.org/10.1111/cogs.12243>
- Escudero, P., Mulak, K. E., & Vlach, H. A. (2016b). Infants encode phonetic detail during cross situational word learning. *Frontiers in Psychology*, 7, 1419. <https://doi.org/10.3389/fpsyg.2016.01419>
- Estis, J. M., & Beverly, B. L. (2015). Children with SLI exhibit delays resolving ambiguous reference. *Journal of Child Language*, 42(1), 180–195. <https://doi.org/10.1017/S0305000914000038>
- Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 52(2), 321–335. [https://doi.org/10.1044/1092-4388\(2009\)07-0189](https://doi.org/10.1044/1092-4388(2009)07-0189)
- Fisher, A., & Kloos, H. (2016). Development of selective sustained attention: The role of executive functions. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research* (pp. 215–237). American Psychological Association.
- Fisher, A., Thiessen, E., Godwin, K., Kloos, H., & Dickerson, J. (2013). Assessing selective sustained attention in 3- to 5-year-old children: Evidence from a new paradigm. *Journal of Experimental Child Psychology*, 114(2), 275–294. <https://doi.org/10.1016/j.jecp.2012.07.006>
- Gallinat, E., & Spaulding, T. J. (2014). Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 57(4), 1363–1382.
- Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27(4), 513–543.
- Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language*, 29(3), 336–360. [https://doi.org/10.1016/0749-596X\(90\)90004-J](https://doi.org/10.1016/0749-596X(90)90004-J)
- Gershon, R. C., Slotkin, J., Manly, J. J., Blitz, D. L., Beaumont, J. L., Schnipke, D., ... Weintraub, S. (2013). NIH Toolbox Cognition Battery (CB): Measuring language (vocabulary comprehension and reading decoding). *Monographs of the Society for Research in Child Development*, 78(4), 49–69. <https://doi.org/10.1111/mono.12034>
- Gillam, R. B., & Pearson, N. A. (2004). *TNL: Test of Narrative Language*. Austin, TX: Pro-ed.
- Gillam, R. B., & Pearson, N. A. (2017). *TNL-2: Test of Narrative Language*. Pro-ed.
- Graf Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 50(1), 177–195. [https://doi.org/10.1044/1092-4388\(2007\)015](https://doi.org/10.1044/1092-4388(2007)015)
- Gray, S. (2003). Word-learning by preschoolers with specific language impairment: What predicts success. *Journal of Speech, Language, and Hearing Research*, 46(1), 56–67. [https://doi.org/10.1044/1092-4388\(2003\)005](https://doi.org/10.1044/1092-4388(2003)005)
- Gray, S. (2004). Word learning by preschoolers with specific language impairment: Predictors and poor learners. *Journal of Speech, Language, and Hearing Research*, 47(5), 1117–1132. [https://doi.org/10.1044/1092-4388\(2004\)083](https://doi.org/10.1044/1092-4388(2004)083)
- Gray, S. (2005). Word learning by preschoolers with specific language impairment: Effect of phonological or semantic cues. *Journal of Speech, Language & Hearing Research*, 48(6), 1452–1467. [https://doi.org/10.1044/1092-4388\(2005\)101](https://doi.org/10.1044/1092-4388(2005)101)

- Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools*, 30(2), 196–206. <https://doi.org/10.1044/0161-1461.3002.196>
- Grunow, H., Spaulding, T. J., Gómez, R. L., & Plante, E. (2006). The effects of variation on learning word order rules by adults with and without language-based learning disabilities. *Journal of Communication Disorders*, 39(2), 158–170. <https://doi.org/10.1016/j.jcomdis.2005.11.004>
- Haebig, E., Kaushanskaya, M., & Weismer, S. E. (2015). Lexical processing in school-age children with autism spectrum disorder and children with specific language impairment: The role of semantics. *Journal of Autism and Developmental Disorders*, 45(12), 4109–4123. doi: 10.1007/s10803-015-2534-2
- Haebig, E., Leonard, L. B., Deevy, P., Karpicke, J., Christ, S. L., Usler, E., ... & Weber, C. (2019). Retrieval-based word learning in young typically developing children and children with developmental language disorder II: A comparison of retrieval schedules. *Journal of Speech, Language, and Hearing Research*, 62(4), 944–964. 10.1044/2018_JSLHR-L-18-0071
- Haebig, E., Saffran, J. R., & Ellis Weismer, S. (2017). Statistical word learning in children with autism spectrum disorder and specific language impairment. *Journal of Child Psychology and Psychiatry*, 58(11), 1251–1263. <https://doi.org/10.1111/jcpp.12734>
- Hsu, H. J., & Bishop, D. V. (2010). Grammatical difficulties in children with specific language impairment: Is learning deficient? *Human Development*, 53(5), 264–277. <https://doi.org/10.1159/000321289>
- Hsu, H. J., Tomblin, J. B., & Christiansen, M. H. (2014). Impaired statistical learning of non-adjacent dependencies in adolescents with specific language impairment. *Frontiers in Psychology*, 5, 175. <https://doi.org/10.3389/fpsyg.2014.00175>
- Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35(1), 1–19. https://doi.org/10.1207/S15327906MBR3501_1
- Kachergis, G., Yu, C., & Shiffrin, R. M. (2013). Actively learning object names across ambiguous situations. *Topics in Cognitive Science*, 5(1), 200–213. <https://doi.org/10.1111/tops.12008>
- Kan, P. F., & Windsor, J. (2010). Word learning in children with primary language impairment: A meta-analysis. *Journal of Speech, Language, and Hearing Research*, 53, 739–756. [https://doi.org/10.1044/1092-4388\(2009\)08-0248](https://doi.org/10.1044/1092-4388(2009)08-0248)
- Kapa, L. L., Plante, E., & Doubleday, K. (2017). Applying an integrative framework of executive function to preschoolers with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 60(8), 2170–2184. 10.1044/2017_JSLHR-L-16-0027
- Kapantzoglou, M., Restrepo, M. A., & Thompson, M. S. (2012). Dynamic assessment of word learning skills: Identifying language impairment in bilingual children. *Language, Speech, and Hearing Services in Schools*, 43(1), 81–96. [https://doi.org/10.1044/0161-1461\(2011\)10-0095](https://doi.org/10.1044/0161-1461(2011)10-0095)
- Kaufman, A. S. (2004). *Kaufman brief intelligence test—second edition (KBIT-2)*. Circle Pines, MN: American Guidance Service.
- Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, 16(4), 705–710. doi: 10.3758/PBR.16.4.705
- Lahey, M., & Edwards, J. (1996). Why do children with specific language impairment name pictures more slowly than their peers? *Journal of Speech, Language, and Hearing Research*, 39(5), 1081–1098. <https://doi.org/10.1044/jshr.3905.1081>
- Lahey, M., & Edwards, J. (1999). Naming errors of children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 42(1), 195–205. <https://doi.org/10.1044/jslhr.4201.195>
- Leonard, L. B., Deevy, P., Karpicke, J. D., Christ, S. L., & Kueser, J. B. (2020). After initial retrieval practice, more retrieval produces better retention than more study in the word learning of children with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 63(8), 2763–2776. https://doi.org/10.1044/2020_JSLHR-20-00105
- Leonard, L. B., Karpicke, J., Deevy, P., Weber, C., Christ, S., Haebig, E., ... Krok, W. (2019). Retrieval-based word learning in young typically developing children and children with developmental language disorder

- I: The benefits of repeated retrieval. *Journal of Speech, Language, and Hearing Research*, 62(4), 932–943. https://doi.org/10.1044/2018_JSLHR-L-18-0070
- Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., & Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, 198, 104191. <https://doi.org/10.1016/j.cognition.2020.104191>
- Lüke, C., & Ritterfeld, U. (2014). The influence of iconic and arbitrary gestures on novel word learning in children with and without SLI. *Gesture*, 14(2), 204–225. <https://doi.org/10.1075/gest.14.2.04luk>
- Lum, J. A., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex*, 51, 1–10. <https://doi.org/10.1016/j.cortex.2013.10.011>
- MacDonald, K., Yurovsky, D., & Frank, M. C. (2017). Social cues modulate the representations underlying cross-situational learning. *Cognitive Psychology*, 94, 67–84.
- McGregor, K. (2017). The dynamics of word learning in children with developmental language disorder: A prospective cohort, ID# researchregistry3425. Research Registry.
- McGregor, K. K. (2021a). Dynamics of word learning. Retrieved from osf.io/r56s7
- McGregor, K. K. (2021b). Children's vocabulary project: Cross-situational learning. Retrieved from osf.io/u78pr
- McGregor, K., Arbisi-Kelm, T., & Eden, N. (2017). The encoding of word forms into memory may be challenging for college students with developmental language impairment. *International Journal of Speech-Language Pathology*, 19(1), 43–57. <https://doi.org/10.3109/17549507.2016.1159337>
- McGregor, K. K., Arbisi-Kelm, T., Eden, N., & Oleson, J. (2020). The word learning profile of adults with developmental language disorder. *Autism & Developmental Language Impairments*, 5, 1–19. <https://doi.org/10.1177/2396941519899311>
- McGregor, K. K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2020). The fast-mapping abilities of adults with developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 63(9), 3117–3129. https://doi.org/10.1044/2020_JSLHR-19-00418
- McGregor, K. K., Gordon, K., Eden, N., Arbisi-Kelm, T., & Oleson, J. (2017). Encoding deficits impede word learning and memory in adults with developmental language disorders. *Journal of Speech, Language, and Hearing Research*, 60(10), 2891–2905. https://doi.org/10.1044/2017_JSLHR-L-17-0031
- McGregor, K. K., Licandro, U., Arenas, R., Eden, N., Stiles, D., Bean, A., & Walker, E. (2013). Why words are hard for adults with developmental language impairments. *Journal of Speech, Language, and Hearing Research*, 56(6), 1845–1856. [https://doi.org/10.1044/1092-4388\(2013\)12-0233](https://doi.org/10.1044/1092-4388(2013)12-0233)
- McGregor, K. K., Newman, R. M., Reilly, R. M., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 45, 998–1014. [https://doi.org/10.1044/1092-4388\(2002\)081](https://doi.org/10.1044/1092-4388(2002)081)
- McGregor, K. K., Oleson, J., Bahnsen, A., & Duff, D. (2013). Children with developmental language impairment have vocabulary deficits characterized by limited breadth and depth. *International Journal of Language & Communication Disorders*, 48(3), 307–319. <https://doi.org/10.1111/1460-6984.12008>
- McGregor, K. K., Rost, G. C., Guo, L. Y., & Sheng, L. (2010). What compound words mean to children with specific language impairment. *Applied Psycholinguistics*, 31(03), 463–487. doi: 10.1017/S014271641000007X
- 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*, 108(22), 9014–9019. <https://doi.org/10.1073/pnas.1105040108>
- Montgomery, J. W., Evans, J. L., & Gillam, R. B. (2009). Relation of auditory attention and complex sentence comprehension in children with specific language impairment: A preliminary study. *Applied Psycholinguistics*, 30(1), 123. doi: 10.1017/S0142716408090061
- Mulak, K. E., Vlach, H. A., & Escudero, P. (2019). Cross-situational learning of phonologically overlapping words across degrees of ambiguity. *Cognitive Science*, 43(5), e12731. <https://doi.org/10.1111/cogs.12731>
- Nagy, W. E., & Anderson, R. C. (1982). The number of words in printed school English. Center for the Study of Reading Technical Report (Report no. 253). U.S. Department of Education.

- Norbury, C. F. (2005). Barking up the wrong tree? Lexical ambiguity resolution in children with language impairments and autistic spectrum disorders. *Journal of Experimental Child Psychology*, 90(2), 142–171. <https://doi.org/10.1016/j.jecp.2004.11.003>
- Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., ... & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, 57(11), 1247–1257. <https://doi.org/10.1111/jcpp.12573>
- Noterdaeme, M., Amorosa, H., Mildenerger, K., Sitter, S., & Minow, F. (2001). Evaluation of attention problems in children with autism and children with a specific language disorder. *European Child & Adolescent Psychiatry*, 10(1), 58–66. <https://doi.org/10.1007/s007870170048>
- Oleson, J. J., Brown, G. D., & McCreery, R. (2019). The evolution of statistical methods in speech, language, and hearing sciences. *Journal of Speech, Language, and Hearing Research*, 62(3), 498–506. https://doi.org/10.1044/2018_JSLHR-H-ASTM-18-0378
- Plante, E., Bahl, M., Vance, R., & Gerken, L. (2010). Children with specific language impairment show rapid, implicit learning of stress assignment rules. *Journal of Communication Disorders*, 43(5), 397–406. <https://doi.org/10.1016/j.jcomdis.2010.04.012>
- Plante, E., Gomez, R., & Gerken, L. (2002). Sensitivity to word order cues by normal and language/learning disabled adults. *Journal of Communication Disorders*, 35(5), 453–462. [https://doi.org/10.1016/S0021-9924\(02\)00094-1](https://doi.org/10.1016/S0021-9924(02)00094-1)
- Poepsel, T. J., & Weiss, D. J. (2016). The influence of bilingualism on statistical word learning. *Cognition*, 152, 9–19. <https://doi.org/10.1016/j.cognition.2016.03.001>
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Redmond, S. M. (2005). Differentiating SLI from ADHD using children's sentence recall and production of past tense morphology. *Clinical Linguistics & Phonetics*, 19(2), 109–127. <https://doi.org/10.1080/02699200410001669870>
- Richardson, J., Harris, L., Plante, E., & Gerken, L. (2006). Subcategory learning in normal and language learning-disabled adults: How much information do they need? *Journal of Speech, Language, and Hearing Research*, 49(6), 1257–1266. [https://doi.org/10.1044/1092-4388\(2006\)090](https://doi.org/10.1044/1092-4388(2006)090)
- Richtsmeier, P. T., Gerken, L., Goffman, L., & Hogan, T. (2009). Statistical frequency in perception affects children's lexical production. *Cognition*, 111(3), 372–377. <https://doi.org/10.1016/j.cognition.2009.02.009>
- Romberg, A. & Yu, C. (2014). Interactions between statistical aggregation and hypothesis testing mechanisms during word learning. In Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 36, No. 36).
- Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of Memory and Language*, 87, 105–127. <https://doi.org/10.1016/j.jml.2015.09.005>
- Rudolph, J. M. (2017). Case history risk factors for specific language impairment: A systematic review and meta-analysis. *American Journal of Speech-Language Pathology*, 26(3), 991–1010.
- Sciberras, E., Mueller, K. L., Efron, D., Bisset, M., Anderson, V., Schilpzand, E. J., ... Nicholson, J. M. (2014). Language problems in children with ADHD: A community-based study. *Pediatrics*, 133(5), 793–800. <https://doi.org/10.1542/peds.2013-3355>
- Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, 122(2), 163–180. doi: 10.1016/j.cognition.2011.10.010
- Sheng, L., & McGregor, K. K. (2010). Lexical-semantic organization in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 53(1), 146–159. doi: 10.1044/1092-4388(2009)08-0160)
- Smit, E. A., Milne, A. J., Dean, R. T., & Weidemann, G. (2019). Perception of affect in unfamiliar musical chords. *PLoS One*, 14(6), e0218570. <https://doi.org/10.1371/journal.pone.0218570>
- 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>

- Smith, L. B., & 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*(1), 25–49. <https://doi.org/10.1080/15475441.2012.707104>
- Smolak, E., McGregor, K. K., Arbisi-Kelm, T., & Eden, T. (2020). Sustained attention in developmental language disorder and its relation to working memory and language. *Journal of Speech, Language, and Hearing Research*, *63*(12), 4096–4108. [10.1044/2020_JSLHR-20-00265](https://doi.org/10.1044/2020_JSLHR-20-00265)
- Spaulding, T. J., Plante, E., & Vance, R. (2008). Sustained selective attention skills of preschool children with specific language impairment: Evidence for separate attentional capacities. *Journal of Speech, Language, and Hearing Research*, *51*(1), 16–34. [https://doi.org/10.1044/1092-4388\(2008\)002](https://doi.org/10.1044/1092-4388(2008)002)
- Stevens, J. S., Gleitman, L. R., Trueswell, J. C., & Yang, C. (2017). The pursuit of word meanings. *Cognitive Science*, *41*(S4), 638–676. <https://doi.org/10.1111/cogs.12416>
- Storkel, H. L., & Lee, S. Y. (2011). The independent effects of phonotactic probability and neighbourhood density on lexical acquisition by preschool children. *Language and Cognitive Processes*, *26*(2), 191–211. <https://doi.org/10.1080/01690961003787609>
- 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>
- Tonidandel, S., & LeBreton, J. M. (2015). RWA web: A free, comprehensive, web-based, and user friendly tool for relative weight analyses. *Journal of Business and Psychology*, *30*(2), 207–216. [doi 10.1007/s10869-014-9351-z](https://doi.org/10.1007/s10869-014-9351-z)
- Tonidandel, S., LeBreton, J. M., & Johnson, J. W. (2009). Determining the statistical significance of relative weights. *Psychological Methods*, *14*(4), 387–399. <https://doi.org/10.1037/a0017735>
- Torkildsen, J. V. K., Dailey, N. S., Aguilar, J. M., Gómez, R., & Plante, E. (2013). Exemplar variability facilitates rapid learning of an otherwise unlearnable grammar by individuals with language-based learning disability. *Journal of Speech, Language, and Hearing Research*, *56*(2), 618–629. [https://doi.org/10.1044/1092-4388\(2012\)11-0125](https://doi.org/10.1044/1092-4388(2012)11-0125)
- Toro, J. M., Sinnott, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition*, *97*(2), B25–B34. <https://doi.org/10.1016/j.cognition.2005.01.006>
- 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>
- Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, *41*(3), 399–433. [https://doi.org/10.1016/S0010-9452\(08\)70276-4](https://doi.org/10.1016/S0010-9452(08)70276-4)
- Van de Schoot, R., & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychologist*, *16*(2), 75–84.
- Vlach, H. A., & DeBrock, C. A. (2017). Remember dax? Relations between children’s cross-situational word learning, memory, and language abilities. *Journal of Memory and Language*, *93*, 217–230. <https://doi.org/10.1016/j.jml.2016.10.001>
- 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>
- Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants’ cross-situational statistical learning. *Cognition*, *127*(3), 375–382. <https://doi.org/10.1016/j.cognition.2013.02.015>
- Wechsler, D. (2011). *WASI-II: Wechsler abbreviated scale of intelligence*. PsychCorp.
- Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., ... & Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(11), S54–S64. [doi: https://doi.org/10.1212/WNL.0b013e3182872ded](https://doi.org/10.1212/WNL.0b013e3182872ded)
- 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/10.1037/a0016134>

- Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*(2), 245–272. <https://doi.org/10.1037/0033-295X.114.2.245>
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420. <https://doi.org/10.1111/j.1467-9280.2007.01915.x>
- 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>
- 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>
- Yurovsky, D., Fricker, D. C., Yu, C., & Smith, L. B. (2014). The role of partial knowledge in statistical word learning. *Psychonomic Bulletin & Review*, *21*, 1–22. <https://doi.org/10.3758/s13423-013-0443-y>