Received: 19 November 2021 | Revised: 1 May 2022 | Accepted: 12 May 2022

DOI: 10.1111/desc.13291

RESEARCH ARTICLE

Developmental Science WILEY

Not all is forgotten: Children's associative matrices for features of a word learning episode

Melina L. Knabe | Haley A. Vlach

Department of Educational Psychology, University of Wisconsin-Madison, Madison, Wisconsin, USA

Correspondence
Melina L. Knabe, Department of Educational Psychology, University of Wisconsin-Madison, 1025 W. Johnson St. Madison, WI 53706, USA.
Email: mknabe@wisc.edu

All raw data, code, analyses, and stimuli are available at https://osf.io/n3pxb/

Abstract

Word learning studies traditionally examine the narrow link between words and objects, indifferent to the rich contextual information surrounding objects. This research examined whether children attend to this contextual information and construct an associative matrix of the words, objects, people, and environmental context during word learning. In Experiment 1, preschool-aged children (age: 3;2–5;11 years) were presented with novel words and objects in an animated storybook. Results revealed that children constructed associations beyond words and objects. Specifically, children attended to and had the strongest associations for features of the environmental context but failed to learn word-object associations. Experiment 2 demonstrated that children (age: 3;0–5;8 years) leveraged strong associations for the person and environmental context to support word-object mapping. This work demonstrates that children are especially sensitive to the word learning context and use associative matrices to support word mapping. Indeed, this research suggests associative matrices of the environment may be foundational for children's vocabulary development.

KEYWORDS
context dependency, fast mapping, language development, word learning

1 | INTRODUCTION

The process of learning a new word involves many challenging steps. Children must segment a speech stream, encode the novel word form, identify the intended referent of a word, and store the word-referent pairing for subsequent retrieval. Despite the difficulty of these steps, preschool-aged children readily map words to referents, a behavior termed *fast mapping* (Axelsson et al., 2012; Carey, 2010; Markman, 1989; Mather & Plunkett, 2012; Spiegel & Halberda, 2011; Vlach & Sandhofer, 2012). In laboratory studies, researchers have used a consistent paradigm to determine how children fast map words to objects (see Figure 1 for a schematic of fast mapping tasks; e.g., Carey & Bartlett, 1978; Horst & Samuelson, 2008; Vlach & Sandhofer, 2012). Children are presented with novel labels (e.g., "*wug*", "*fep*") and are simultaneously shown novel objects during a learning phase. Children's ability to associate the labels with the respective objects is tested immediately after the learning phase using a recognition memory test. For instance, the experimenter shows children a series of novel and/or familiar objects and asks: "Which one is the *wug*?".

This body of work has revealed that children utilize a variety of linguistic (Mather & Plunkett, 2009; Woodward et al., 1994) and extralinguistic cues (Brooks & Meltzoff, 2008; Diesendruck & Markson, 2001; Grassman & Tomasello, 2010; Mather & Plunkett, 2012; Smith, 2000) during word mapping. Many of the cues children use to map

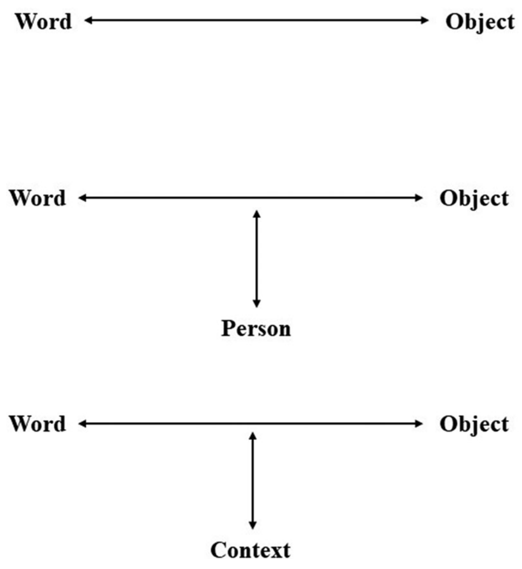

FIGURE 1 Traditional fast mapping studies assess how children map words to objects in a single encounter, as well as the use of social (e.g., person) and environmental (e.g., context) cues in word mapping

> **RESEARCH HIGHLIGHTS**
>
> - We examined children's attention to and memory for the associations between words, objects, people, and broader environmental context encountered during a word learning episode.
> - Experiment 1 revealed that children had the strongest associations for features of the word learning context (i.e., person and scene context).
> - Experiment 2 revealed that children could leverage stronger associations for the person and scene context to map words and objects.
> - Children construct contextually-grounded associative matrices to support word mapping and thus researchers should shift to focusing on contextual information when developing word learning theories.

words to objects are social in nature (Akhtar & Tomasello, 1996; Briganti & Cohen, 2011; Houston-Price et al., 2006; Koenig et al., 2004; Tomasello & Akhtar, 1995; Yu & Ballard, 2007). For instance, pointing and eye-gaze are two potent extralinguistic cues used to determine the referent of a label. Early on in development, infants and young children recognize that eye gaze and pointing are intentional acts that refer to shared events in the environment (Senju & Csibra, 2008). Furthermore, caregivers use children's attention to objects in the environment as opportunities for word learning (e.g., Chen et al., 2021; Tamis-LeMonda et al., 2014; Yu & Smith, 2017; Yu et al., 2019). The sociopragmatic cues used by children and caregivers can then be exploited to guide learning of novel word-object mappings, effectively reducing ambiguity during labeling (Baldwin, 1991; Woodward, 2003).

Children also learn words across several environments (e.g., home, school, play areas) and use these broad environmental cues to map words to objects (Clerkin et al., 2016; Roy et al., 2015; Samuelson et al., 2011; Tamis-LeMonda et al., 2019; Vlach & Sandhofer, 2011). Research has shown that even when the environmental context is irrelevant to the word learning event, children use these cues to guide word mapping. For example, Vlach and Sandhofer (2011) manipulated whether the physical background surrounding target objects (i.e., the pattern/color of a cloth placed underneath the object) matched between learning and test. Their results demonstrated that 2-year-olds' word mapping performance was lower at test when the context cues changed, whereas 3-year-olds' performance was supported when learning was distributed across a variety of contexts. These studies suggest that children encode the environmental context during word learning and use these cues to determine word-object associations.

Research on fast mapping has identified the social and environmental cues that impact children's mapping of words to objects in

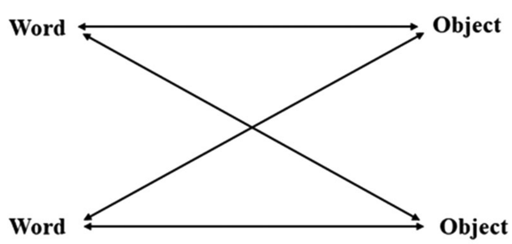

FIGURE 2 Cross-situational word learning studies assess how children use the co-occurrence of words and objects across multiple encounters to learn word-object associations

a *single* encounter. Other approaches to language development view word learning as the gradual accumulation of word meanings across *several* encounters (see Figure 2 for a schematic of CSWL paradigms; e.g., Smith & Yu, 2008; Suanda et al., 2014; Yurovsky et al., 2014). These cross-situational word learning (CSWL) paradigms present learners with several words and objects on-screen ("This is a *wug*. This is a *dax*."). At first, it is ambiguous which word corresponds to which object. Across the learning phase, however, words and objects co-occur in a reliable manner. If the learner attends to the presented co-occurrence statistics, they will successfully learn the presented word-object mappings.

Fast mapping and CSWL paradigms assess different timescales of word learning but share an important commonality: Both tasks are conducted in artificial word learning environments in which there is little exposure to information beyond words and objects. That is, researchers have primarily examined the narrow link between words and objects even though various cues promote learning of word-object associations. Indeed, what leads to better experimental control may diverge in important ways from the environments presented to the early word learner (see Smith et al., 2014, for a review). In real-world scenarios, children are presented with a rich set of cues during

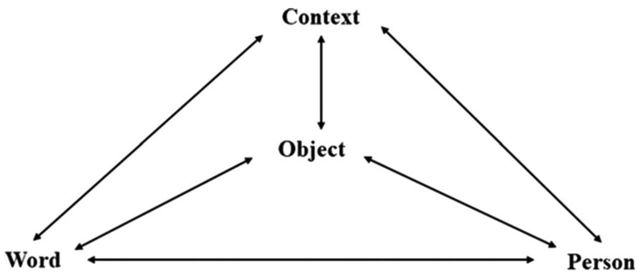

FIGURE 3 Schematic of the associative matrix of a word learning episode investigated in Experiment 1

word mapping. These cues provide the opportunity for children to generate many associations, such as between words, objects, people, and the broader environmental context. For example, when a child hears "Can you get the fork?" for the first time, they may encode the fork, their caregiver speaking, a nearby spoon, and the overall background context (i.e., kitchen), forming a rich matrix of associations between the object and its environment. Thus far, we know very little about the nature of these broader associative matrices during word learning, and how the interconnectedness of associations contributes to word mapping (for a review, see Yu & Smith, 2012).

We addressed this gap in the current research by characterizing children's associative matrices for word learning events and examining how children may use a broader system of associations to map words to objects. We grounded our research in the context of children's learning from screens, in which children organically learn words via e-storybooks, real-time conversations with relatives, etc. (Akhtar, 2005; Axelsson & Horst, 2014; Horst & Houston-Price, 2015; Horst et al., 2011; Houston-Price et al., 2014; Khu et al., 2014; Williams & Horst, 2014). We hypothesized that children acquire a system—or matrix—of associations between words, objects, people, and the broader environmental context (a simple version is presented in Figure 3). If children build a system of associations, we predicted that the children leverage this system to retrieve word-object mappings (Smith & Yu, 2008; Smith et al., 2014). One example of leveraging is *semantic bootstrapping*, in which children use prior associations between word position in a sentence and word meaning to learn a novel word (e.g., see Höhle, 2009 for a review; Kachergis et al., 2017). Children may also use other associations, such as the association between objects and their locations in scenes, to learn word-object mappings.

1.1 | The current research

To characterize the nature of children's associative matrices for word learning events, the current research examined children's visual attention to and recognition memory for features of a word learning episode. We assessed visual attention to gauge online processing of the word learning scene; understanding how children visually inspect their environment in-the-moment provides a window into how associations between words and objects are formed (Ellis et al., 2015; Weighall et al., 2017). Previous research has used eye tracking technology for capturing children's attention to cues in the word learning environment (e.g., Gliga et al., 2009; Yow et al., 2017), and thus we used a Tobii X3-120 eye-tracker to measure children's visual attention. We also assessed children's recognition memory for features of the word learning events as a measure of the associations children form and retrieve during word learning events. Children were tested on their memory for associations between words, objects, people, and the broader environmental context.

We extended the basic model of studying word-object associations (i.e., the fast mapping paradigm) by incorporating additional factors children encounter in everyday word learning: multiple objects, people, and an environmental context. In Experiment 1, children were taught six novel-object pairings across six animated storybook scenes. In line with research outlined above, we hypothesized that children would encode the social and broader environmental context in which words are presented, and thus remember more than just word-object mappings. In Experiment 2, we investigated whether children could use the broader person and environmental context to support the mapping of word-object associations. We hypothesized that children could leverage the environmental context for word-object mapping. Taken together, these experiments afforded the opportunity to characterize children's associative matrices more naturalistically, and assess how cues beyond words and objects can be exploited to learn novel words.

2 | EXPERIMENT 1

2.1 | Method

The experimental protocol was approved by the Education and Social/Behavioral Institutional Review Board. Signed consent was obtained from a parent or legal guardian of each participant and verbal assent was provided by each participant before the study began.

2.1.1 | Participants

The participants were 46 typically developing preschool-aged children (M_{age} = 55.57, *SD* = 9.16, range: 38–71 months, 21 girls) and were recruited from local day care centers and preschools. An additional four children were excluded due to missing eye-tracking data or gaze samples lower than 30% (i.e., the percentage of usable samples from the total number of recorded eye tracking samples). An additional child was excluded for missing important demographic information. The age range (i.e., 3- to 6-year-olds) was chosen because it is a period of development in which children are rapidly learning new words. A power analysis, using Cohen's s *d* and a conservative estimate of effect size ($d = 0.5$), was conducted to determine a sample size that would provide at least 80% power. Effect sizes were gathered from previous studies assessing children's learning of word-object associations ($ds < 1.0$;

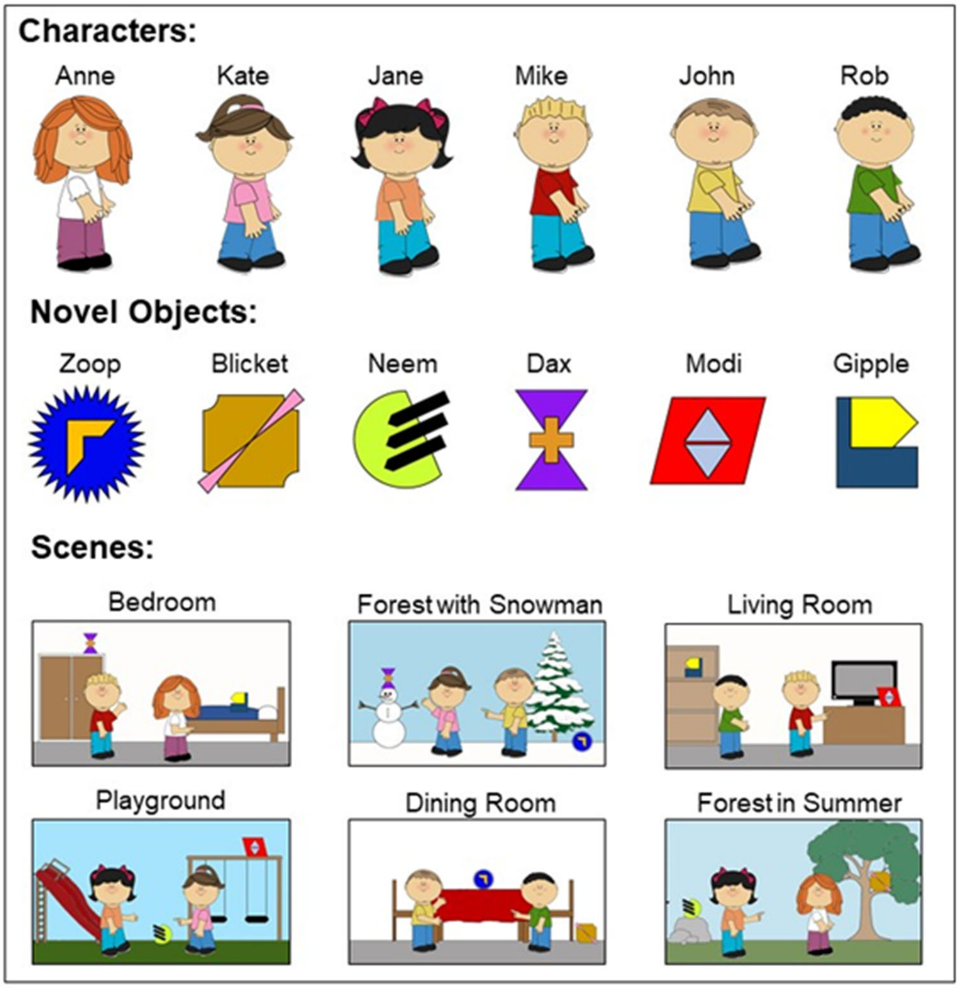

FIGURE 4 Stimuli used during the experiment. Stimuli included six characters (three males, three females), six novel objects, and six animated storybook scenes. Characters and novel objects were randomly assigned names and labels

e.g., Horst & Samuelson, 2008; Yow et al., 2017). A power analysis for a one-sample *t*-test, with $\alpha = 0.05$ and $d = 0.5$, yielded a sample size of at least 27 participants to achieve 80% power. Because prior research has found that attention to the social environment is affected by language learning history (e.g., Colunga et al., 2012; Yow et al., 2017), we recruited monolingual ($n = 26$) and second language exposed ($n = 20$) children. Participants were collapsed into one group for subsequent analyses because no differences between language groups were observed. Children received a storybook as a thank you for their participation.

2.1.2 | Apparatus and stimuli

The experimental stimuli (Figure 4) were presented on a Dell Precision 7510 laptop. Visual stimuli consisted of six animated storybook scenes, which were shown twice in the same order (Figure 5). Scene settings included a bedroom, a forest with a snowman, a living room, a playground, a dining room, and a forest in the summer. These scenes were selected because they are all familiar spaces for children in the local area. To ensure scenes were similarly interesting, each contained two salient background objects (e.g., tree and snowman, tree and rock, television and bookshelf etc.). Two novel objects were displayed in each scene, which included one target and one distractor object. Novel objects were tested for their nameability and familiarity by a sample of 34 adults. Each scene also consisted of two cartoon characters selected from a set of three female and three male characters. Every character appeared twice across the six scenes, labeling the target object once and serving as an additional speaker once. Target objects were randomly assigned a set of six novel labels (i.e., *gipple*, *dax*, *modi*, *neem*, *zoop*, and *blicket*) that followed American English phonotactics. A female, native American English speaker narrated the procedure.

Eye movement data was collected during the learning phase using a portable Tobii X-120 eye-tracker mounted on a 34.5 × 19.3 cm Dell Precision 7510 laptop monitor, sampling at 120 Hz. Cameras embedded in the eye tracker recorded the reflection of an infrared light source on the cornea relative to the pupil for each eye. The average accuracy of the eye tracking system is 0.4° with an average precision of 0.24°. Gaze recovery time following eye blinks is immediate, whereas a 100-ms recovery time is needed following lost tracking. Children were seated

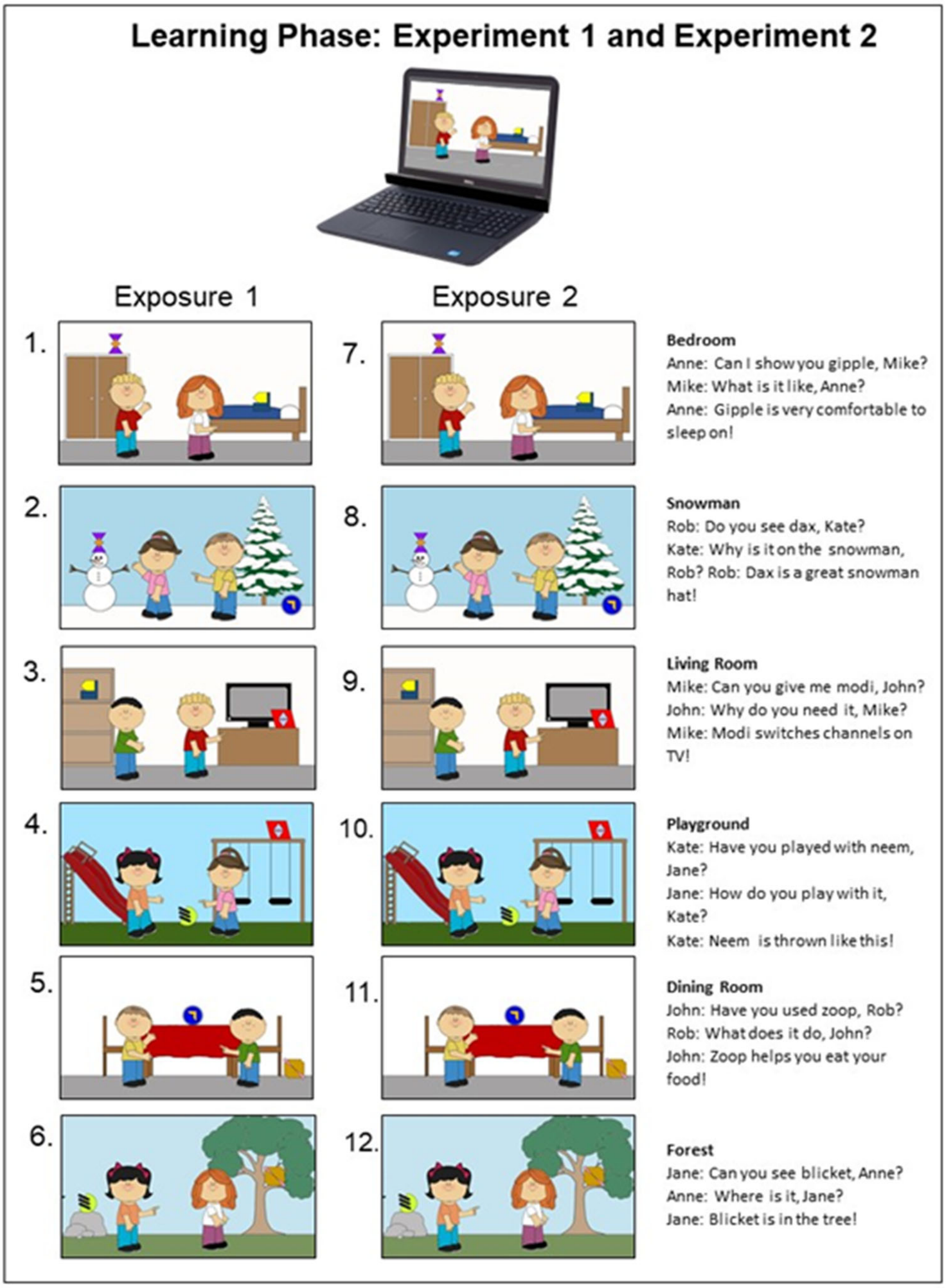

FIGURE 5 Stimuli and script used during the learning phase of Experiments 1 and 2. Each scene was shown twice, resulting in target objects being labeled four times each

at a viewing distance of 60 cm. With this viewing distance, the 34.5 × 19.3 cm stimuli on-screen subtended 32.08 × 18.27 ° visual angle. The eye-tracker was calibrated for each child using a 5-point calibration. The recording was started after each of the 5 points was calibrated for each eye.

2.1.3 | Procedure

There were two phases of the experiment: a learning phase and a testing phase. Before beginning the experiment, children were told that they would be watching videos of characters teaching them new things. They were also asked to stay as still as possible throughout the experiment.

2.1.4 | Learning phase

Following eye tracker calibration, the experiment began with a learning phase (Figure 5). In the learning phase, children viewed six animated videos twice, resulting in 12 learning trials. A black fixation cross (159 × 167 pixels) was displayed on a white screen for 1 s after each scene to orient the child's attention to the middle of the screen. The order of the videos (i.e., bedroom, forest with a snowman, living room, playground, dining room, and forest in the summer) stayed the same across both exposures. Furthermore, the pairings between the novel objects, characters, and scenes were consistent across participants. Each scene began with a target and distractor object in view. Immediately after the scene began, the target speaker entered from either the left or right side of the screen for 3 s. Following the entrance of the first character, the second character entered from the opposite side of the screen for 3 s. The side (left/right) from which each character entered was counterbalanced across scenes to ensure characters spent an equivalent amount of time on-screen. A dialogue began once both characters had entered (12 s after the start of the trial). The dialogue consisted of three sentences in which the target object was named twice. For example, in the first scene, the target speaker stated: "Can I show you *gipple*, Mike?". The additional speaker responded: "What is it like, Anne?". Finally, the target speaker claimed: "*Gipple* is very comfortable to sleep on" (see Appendix A for dialogue script). The characters moved their arm every time they spoke to indicate that they were the speaker, and the target speaker pointed at the target object during the last sentence to highlight the intended referent. After a 0.25 s delay following the last sentence, the target object spun for 2 s to highlight the intended referent once more. Visual attention was measured throughout the entire duration of each scene (23 s).

2.1.5 | Testing phase

Immediately following the learning phase, the experimenter presented children with six forced-choice recognition tests for six categories of associations: word-object, person-object, scene-object, scene-person, scene-word, person-word associations. Each test consisted of six trials (Figure 6). Children were assigned to one of two randomly generated test orders (Order 1, Order 2). For all trials, the image was displayed on screen until children selected one of the options. Thus, chance performance was at 33.33%. Each of the six novel objects was tested once and each trial was separated by a fixation cross displayed for 1 s. The experimenter did not provide feedback during testing. If no answer was selected, the trial was left blank and scored as incorrect.

2.1.6 | Word-object associations test trials

Children viewed three target objects along the bottom of the screen and heard the narrator request one of the objects (e.g., "Which one is *gipple*?"). The three objects consisted of the target object, the distractor object, and a third object that did not appear with the target object.

2.1.7 | Person-object associations test trials

Children saw the target object appear on the top of the screen while the narrator asked, "Can you point to who played with this?". Three characters appeared on screen after the narration, which included the target speaker, the additional speaker, and a third speaker that did not appear when the target object was named.

2.1.8 | Scene-object associations test trials

Children saw the target object appear on the top of the screen while the narrator asked, "Can you point to where you heard the kids talk about this?". Three scenes appeared at the bottom of the screen, including the scene where the object served as a target, the scene where the object served as a distractor, and a third scene in which the object did not appear.

2.1.9 | Scene-person associations test trials

Children saw the target speaker appear on the top of the screen while the narrator asked, "Where did this person teach you about a new toy?". Once again, three scenes appeared at the bottom of the screen, including the scene where the target speaker labeled the object, the scene where the character was the additional speaker, and a scene in which the character did not appear.

2.1.10 | Scene-word associations test trials

Children viewed three scenes along the bottom of the screen and heard the narrator ask where the characters taught them about a given novel object (i.e., "Can you point to where the kids taught you about *gipple*?"). The three options included the scene in which the intended referent was labeled, the scene in which the intended referent appeared but was not labeled, and a scene in which the referent did not appear.

2.1.11 | Person-word associations test trials

Children viewed three characters along the bottom of the screen and heard the narrator ask who taught them about a given novel object (i.e., "Can you point to who taught you about *gipple*?"). The three options included the target speaker, the additional speaker, and a character that did not appear with the intended referent.

2.1.12 | Data analysis

To characterize children's visual attention, we analyzed children's eye movements during learning using an Area of Interest (AOI) approach on Tobii's Pro Lab (Version 1.118). AOIs were defined manually around five areas in the learning phase: the target object (180 × 190 pixels), the distractor object (180 × 190 pixels), the target speaker (369 × 660 pixels), the additional speaker (369 × 660 pixels), and the entire scene (1919 × 1074 pixels). The scene AOI was drawn around the entire

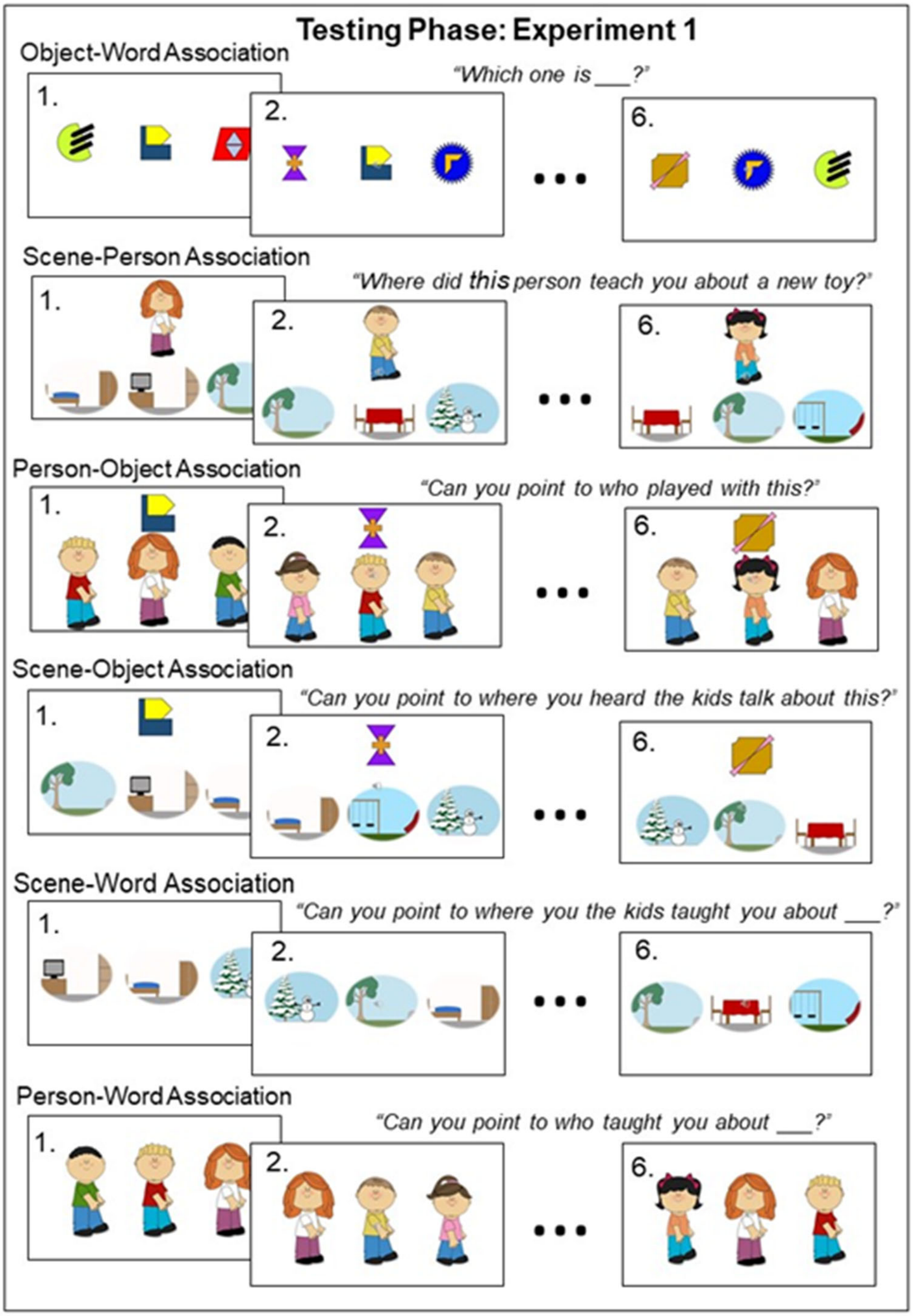

FIGURE 6 Example stimuli and script used during the testing phase of Experiment 1. The testing phase consisted of six forced-choice recognition tests (i.e., word-object, person-object, scene-object, sceneperson, scene-word, person-word), in which associations for all six scenes were tested

frame and the target object, distractor object, target speaker, and additional speaker AOIs were subtracted from the scene AOI. These AOIs were chosen because they represented important features of the word learning episode and associative matrix: the objects, speakers, and environmental context. AOI sizes for each feature were kept constant across all six storybook scenes and all looking measures were weighted based on AOI size (see calculations below). Looking to the AOIs was measured for the entire duration of the scene, which was consistent across all six scenes. No AOIs were drawn for the testing phase because our research aim focused on attentional mechanisms during learning.

Children's looking during learning was quantified as the average proportion of looking to each AOI (e.g., Gangopadhyay & Kaushanskaya, 2020; Pomper & Saffran, 2019; Yow et al., 2017) using the following steps: First, as larger AOIs result in more looking, all outcome measures *v* (e.g., total fixation duration, average fixation count) were weighted based on AOI size. Below, *a*, *s*, and *i* index the AOI,

TABLE 1 Weighted averages for eye tracking measures by AOI–Experiment 1

Variable	Target object	Distractor object	Target speaker	Additional speaker	Scene
Proportion of total fixation duration	0.22 (0.01) [b,d,e]	0.03 (0.00) [a,c,d,e]	0.21 (0.01) [b,d,e]	0.16 (0.01) [a,b,c,e]	0.30 (0.01) [a,b,c,d]
Average fixation count	15.17 (1.09) [b,e]	2.96 (0.29) [a,c,d,e]	19.27 (1.02) [c,d,e]	14.33 (0.70) [b,c,e]	29.24 (1.19) [a,b,c,d]

Note: Weighted proportion of total fixation duration and average fixation count (± SE) to the five AOIs. To adjust for multiple comparison, the *p*-value was set to 0.005.
[a]Significantly different from target object. [b]Significantly different from distractor object. [c]Significantly different from target speaker. [d]Significantly different from additional speaker. [e]Significantly different from scene.

scene, and individual participant, respectively. The weighted outcome $\tilde{v}$ is

$$\tilde{v}_{a,s,i} = \frac{v_{a,s,i}}{1 + p_{a,s}}, \quad (1)$$

where the weights p are the proportions each AOI comprised in the oversall scene: $p_{a,s} = \frac{area_{a,s}}{area_s}$. Next, the proportions of looking to each AOI for each individual and scene were calculated. Finally, because we were interested in the average proportion of looking to each AOI for each individual participant, the mean proportion across all scenes was computed:

$$y_{a,i} = \frac{1}{6}\sum_{s=1}^{6} \frac{\tilde{v}_{a,s,i}}{\sum_{a=1}^{5} \tilde{v}_{a,s,i}} \quad (2)$$

Thus, the outcome measures y capture the average proportion of looking to an AOI for each participant, controlling for the size of the AOI.

Children's memory performance on the forced-choice recognition tests was calculated as the proportion of correct trials out of six trials for each of the six tests. Because each testing trial included three response options, chance level was 0.33. One-sample *t*-tests were conducted to assess whether mean performance on each of the six tasks was significantly greater than chance, $ps > 0.05$. Furthermore, paired samples *t*-tests were conducted to compare the strength of associations. Finally, Pearson's correlations were conducted to assess the relation between children's visual attention and their memory for associations.

2.2 | Results and discussion

2.2.1 | Visual attention

A central aim of this study was to assess children's attention to features of a word learning episode (Figure 7). First, we analyzed children's eye movements during learning. We assessed looking using total fixation duration and average fixation count. Fixation duration captures the total duration of all individual fixations to an AOI and is a common eye-tracking metric used in psychological science (Gliga et al., 2009; Yow et al., 2017).

Results revealed that children spent a high proportion of the learning trials fixating on key features of the word learning episode: the target object, speakers, and overall scene (Table 1). Specifically, Bonferroni-corrected ($\alpha = 0.005$) paired samples *t*-tests revealed that children spent a significantly greater proportion of time attending to the overall scene in comparison to the target speaker, $p < 0.001$, $d = 0.84$, additional speaker, $p < 0.001$, $d = 2.71$, target object, $p < 0.001$, $d = 0.73$, and distractor object, $p < 0.001$, $d = 7.53$. Notably, children spent a significantly lower proportion of time attending to the distractor object in comparison to the scene, target speaker, $p < 0.001$, $d = 4.91$, additional speaker, $p < 0.001$, $d = 4.66$, and target object, $p < 0.001$, $d = 3.70$. The same pattern of results was observed when examining the average number of fixations across children: the highest number of fixations were to the scene and the lowest number of fixations were to the distractor object (Table 1). In sum, looking measures revealed that children attended to important features of the word learning environment and spent the most time attending to the overall scene and person context.

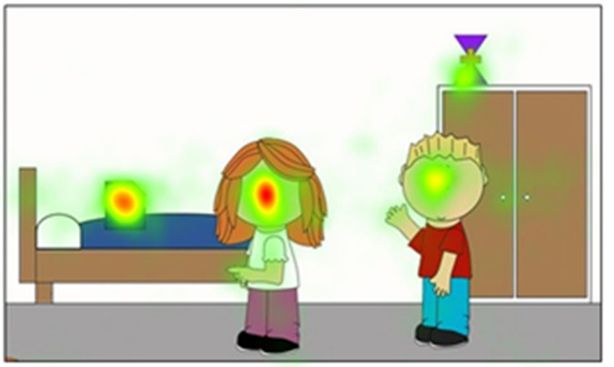

FIGURE 7 Example stimuli and script used during the testing phase of Experiment 2. The testing phase only tested word-object associations. Half of the objects were presented with their corresponding scene context; half of the objects were presented with their corresponding person context. Order of object presentation and type of context reinstatement was counterbalanced across participants

2.2.2 | Memory performance

Much like prior studies testing memory for word-object associations, we assessed children's forced-choice recognition performance at an immediate test. We hypothesized that children would have memory for more than just word-object mappings, and potentially have

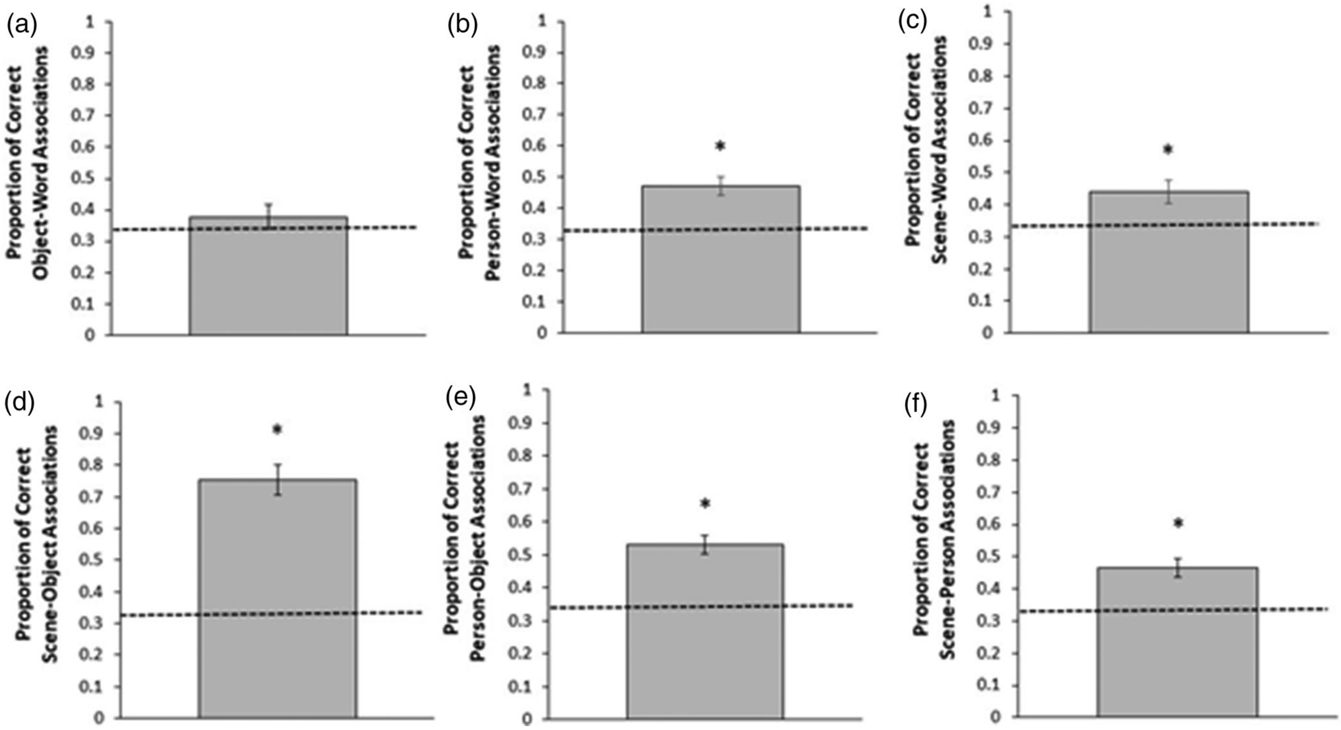

FIGURE 8 Proportion correct on the word-object (A), person-word (B), scene-word (C), scene-object (D), person-object), and scene-person (F) associations at the post-test, * indicates performance was significantly above chance, ps < 0.05

TABLE 2 One-sample T-tests comparing associations to chance level–Experiment 1

Variable	M	SD	df	T	p	d
Scene-object	0.75	0.32	45	9.10	< 0.001*	1.34
Word-object	0.38	0.26	45	1.23	0.227	0.18
Scene-person	0.42	0.23	45	2.84	0.007*	0.42
Scene-word	0.44	0.24	45	3.01	0.004*	0.44
Person-word	0.47	0.20	45	4.79	< 0.001*	0.71
Person-object	0.51	0.22	45	5.41	< 0.001*	0.79

Note: Results for one-sample *t*-tests comparing all associations to chance level (0.33).
**ps* < 0.05.

stronger memory for the social and scene context in which words were presented (Baldwin, 1991; Vlach & Sandhofer, 2011; Woodward, 2003).

We began our analyses by examining performance on the six forced-choice recognition tests. Performance was calculated as the proportion of correct responses out of six trials (Figure 8). One sample *t*-tests revealed that mean performance was significantly above chance (33.33%) for all associations, except word-object associations (Table 2). Overall, children's memory for word-object associations was numerically lowest and their memory for scene-object associations was numerically highest. A series of Bonferroni-corrected *t*-tests ($\alpha = 0.005$) revealed that children's performance on word-object associations was significantly lower than performance on scene-object and person-object associations (Table 3), whereas performance on scene-object associations was significantly greater than all other associations (Table 4). Taken together, these findings suggest that children did not learn the novel words. Instead, their memory performance was greatest for associations between objects and scenes, followed by objects and speakers.

2.2.3 | Linking visual attention and memory

The next question was whether children's visual attention was related to their memory performance. That is, did children who looked more to the scene and object also demonstrate better memory for scene-object associations? In contrast to our prediction, total fixation duration, $r = 0.29$, $p = 0.06$, and average fixation count, $r = 0.17$, $p = 0.27$, to the scene and object were not related to performance on scene-object associations. Similarly, total fixation duration, $r = 0.27$, $p = 0.09$, and

TABLE 3 Paired-samples T-tests comparing word-object associations to all other associations–Experiment 1

Association	*M*	*SD*	*df*	*T*	*p*	*d*
Scene-object	0.75	0.32	45	6.77	< 0.001*	1.27
Scene-person	0.42	0.23	45	−0.95	0.174	0.16
Scene-word	0.44	0.24	45	−1.35	0.092	0.24
Person-word	0.47	0.20	45	−2.24	0.015	0.39
Person-object	0.51	0.22	45	−2.72	0.005*	0.54

Note: Results for paired-samples *t*-tests comparing word-object associations to all other associations. To adjust for multiple comparison, the *p*-value was set to 0.01.
*$ps < 0.01$.

TABLE 4 Paired-samples T-tests comparing scene-object associations to all other associations–Experiment 1

Association	*M*	*SD*	*df*	*T*	*p*	*d*
Word-object	0.38	0.26	45	6.77	< 0.001*	1.27
Scene-person	0.42	0.23	45	6.58	< 0.001*	1.18
Scene-word	0.44	0.24	45	5.57	< 0.001*	1.10
Person-word	0.47	0.20	45	5.35	< 0.001*	1.05
Person-object	0.51	0.22	45	6.36	< 0.001*	0.87

Note: Results for paired-samples *t*-tests comparing scene-object associations to all other associations. To adjust for multiple comparison, the *p*-value was set to 0.01.
*$ps < 0.01$.

average fixation count, r = -0.01, p = 0.97, to the target object were not related to performance on word-object associations (see Appendix B for additional analyses). These results did not align with previous research on fast mapping, which has found that children's visual attention generally parallels the strength of their associations (e.g., Ellis et al., 2015; Weighall et al., 2017; Yow et al., 2017).

Taken together, the results from Experiment 1 revealed that children formed associations between multiple features of events during word learning, forming an associative matrix. When aggregating visual attention to and memory for features of the word learning episode across children, we found that children spent the most time looking at the scene and had the strongest association for the relationship between the object and the scene. Individual children's looking behavior to the overall scene and object; however, was not related to their memory for scene-object associations. These results suggest that there may be individual differences in children's looking to word learning scenes. Whereas some children may rapidly link objects with the scene context (e.g., Oliva, 2005) and then attend to different information within the word learning episode, other children may struggle to link objects with the overall scene context, resulting in longer dwelling times. Moreover, coarse measures of looking—as used in the present study—do not provide insight into children's dynamic visual attention. Future work should therefore investigate the relation between scan patterns and children's associations between words, objects, speakers, and the scene context (see Helo et al., 2014 for scene viewing behavior in children).

Why did children have weaker associations between objects and words in comparison to objects and the environmental context (i.e., person or scene context) when prior research has demonstrated robust fast mapping in children? Unlike prior studies, the word learning events in the current study were more visually complex. It is plausible that children spent time encoding other aspects of the scene and thus more weakly encoded the target object and/or word. In fact, although children fixated on the target object, the proportion of fixations to the speakers and overall scene either matched or exceeded looking to the target object. This suggests that, when learning environments are more complex, children may have difficulty mapping words and objects.

Another way in which this study diverged from real world contexts is that children did not receive contextual information at test. Thus, children were not cued to use other associations to access word-object associations. Can providing contextual cues that are strongly associated (e.g., person or scene context) aid word-object mapping? That is, can children use their associative matrices for word learning events to support the mapping of word-object associations? Experiment 2 sought to answer this question by presenting word-object trials with one of two contextual cues: the person context or the scene context. We predicted that mapping performance for word-object trials would be higher when children are presented with the person or scene context at test.

3 | EXPERIMENT 2

In this experiment, we examined whether stronger associations, such as associations grounded in social and environmental cues, could be used to map words and objects. Children viewed the same six storybook scenes from Experiment 1. At test, their word-object associations

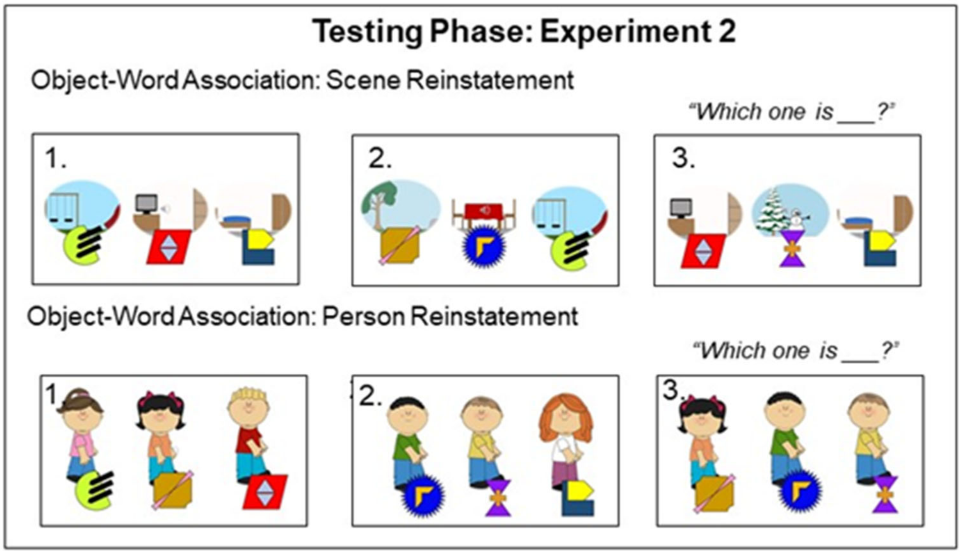

FIGURE 9 Example visualization of participants' total fixations to the target object, distractor object, target speaker, and additional speaker AOIs while watching an animated storybook scene

were assessed with one of two cues: the person who taught the novel object or the environment in which the object was presented. We hypothesized that children would use these contextual cues to map words and objects, resulting in higher performance on word-object associations in Experiment 2 than Experiment 1.

3.1 | Method

The general methods for Experiment 2 resembled Experiment 1, but no eye tracking paradigm was used. After all, an assessment of visual attention was not central to this study's research question. The Testing Phase differed in structure and number of trials, which will be described in further detail below.

3.1.1 | Participants

A total of 53 typically developing preschool-aged children (M_{age} = 50.17, *SD* = 7.59, range: 36–68 months, 30 girls) were recruited from local day care centers and preschools. This age range was chosen to match the age range for Experiment 1. Children did not participate in Experiment 1. All parents consented to their child's participation, and all children gave oral assent. Children received a storybook as a thank you for their participation.

A power analysis, using Cohen's s *d* and a conservative estimate of effect size ($d = 0.50$), was again conducted to determine a sample size that will provide at least 80% power. Effect sizes were gathered from previous studies assessing children's learning of word-object associations (*ds* < 1.0; e.g., Yow et al., 2017). A power analysis for a one-sample *t*-test, with $\alpha = 0.05$ and $d = 0.50$, yielded a sample size of at least 27 participants to achieve 80% power.

3.1.2 | Apparatus and stimuli

The experimental apparatus and stimuli were similar to Experiment 1, except that eye movements were not measured.

3.1.3 | Procedure

The Learning Phase resembled Experiment 1. However, the Testing Phase differed. The Testing Phase consisted of six word-object association trials (Figure 9). For three of the six trials, the objects were presented with their corresponding person context (i.e., the person who taught them the object label). For the other three trials, the objects were presented with their corresponding scene context (i.e., the scene where the object was labeled). Children were simultaneously presented with the word for one of the objects (i.e., "Which one is *wug*?") and were asked to select one of the objects. The image was displayed on screen until the child selected one of the objects.

Each of the six novel objects was tested once and each trial was separated by a fixation cross displayed for 1 s. The order in which the trials were presented, and whether the scene or person context was shown, was counterbalanced across participants. The experimenter did not provide feedback throughout testing but did encourage the child to select an answer. The child's answers were recorded on a response sheet.

3.2 | Results and discussion

This study assessed whether displaying the person or scene would help children map words and objects. Performance was calculated as the proportion of correct trials out of three person and three scene context

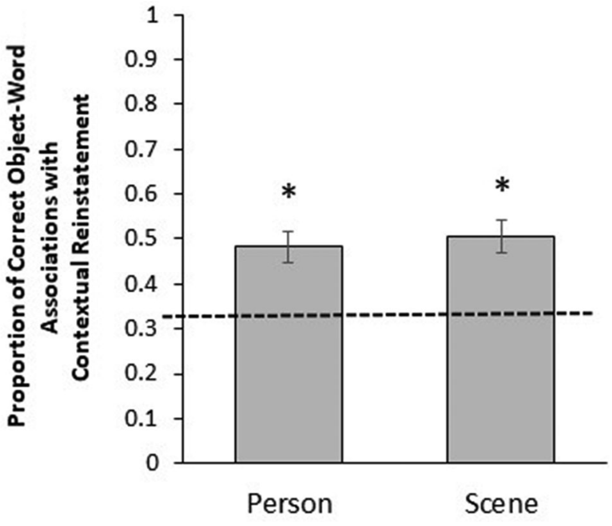

FIGURE 10 Proportion correct on word-object associations with person and scene reinstatement, ps < 0.001

trials (Figure 10). First, we assessed whether children performed above chance for each trial type. One sample *t*-tests confirmed that children performed significantly above chance (0.33) for the person ($M = 0.48$, $SD = 0.31$), $t(52) = 3.62$, $p = 0.001$, $d = 0.48$, and scene context trials ($M = 0.51$, $SD = 0.30$), $t(52) = 4.40$, $p < 0.001$, $d = 0.60$.

Next, we were interested in whether performance on the word-object associations with contextual cues in Experiment 2 was significantly better than the word-object associations without contextual cues from Experiment 1. For this purpose, we calculated the proportion of correct trials out of the six total trials and conducted a one-way ANCOVA with Experiment subtask as the between-subjects factor and age as a covariate. We controlled for age to account for possible age differences between the children recruited in Experiment 1 and Experiment 2. This analysis revealed that providing contextual cues led to significantly better performance in Experiment 2 ($M = 0.50$, $SD = 0.25$) than Experiment 1 ($M = 0.38$, $SD = 0.26$), $F(1, 96) = 6.57$, $p = .012$, $d = 0.06$. These results suggest that children used their memory for scene-object and person-object associations to map word-object associations. That is, children encoded an associative matrix during word learning and utilized stronger associations (i.e., person-object, scene-object) to map previously weaker associations (i.e., word-object). The implications of these findings will be discussed in greater detail in the General Discussion.

There is an alternative explanation for these findings. One difference between the two experiments was that the recognition memory test in Experiment 2 was shorter than in Experiment 1. That is, children completed six test trials (i.e., only the word-object association trials) in Experiment 2 as opposed to the 36 test trials in Experiment 1. This difference in time and trials may have led to more interference during Experiment 1, and thus weaker performance in Experiment 1. Indeed, children at this age have been shown to experience verbal interference and rapidly forget verbal material across even short delays (Brainerd et al., 2012; Filippi et al., 2020; Hale et al., 1997; Horst & Samuelson, 2008; Vlach & DeBrock, 2019).

To investigate this possibility, we compared the two test orders children encountered in Experiment 1. In one test order, the word-object association trials appeared first. In the other test order, word-object associations appeared fifth out of six. We predicted that performance on the word-object associations would not differ significantly between the two test orders in Experiment 1, despite the difference in testing phase duration. We conducted an independent-samples *t*-test with test order as the between-subjects factor (Test Order 1; Test Order 2) to test our prediction. Results supported our prediction: Children's memory performance did not differ significantly when they saw word-object associations first ($M = 0.36$, $SD = 0.25$) versus last ($M = 0.39$, $SD = 0.28$), $t(44) = -0.39$, $p = 0.35$, $d = 0.11$, in Experiment 1. Moreover, when we compared performance on the word-object associations from Test Order 1 in Experiment 1 ($M = 0.36$, $SD = 0.25$) to the performance on word-object associations in Experiment 2 ($M = 0.50$, $SD = 0.25$), we also found significantly better memory performance for word-object associations in Experiment 2, $F(1, 78) = 5.30$, $p = 0.024$, $d = 0.064$. Together, these results provide evidence against the possibility that testing phase duration solely contributed to the difference in performance on word-object associations in Experiment 1 and 2.

4 | GENERAL DISCUSSION

The current research extended traditional word learning paradigms by investigating children's associations between words, objects, people, and the broader environmental context encountered during word learning. Experiment 1 revealed that children attended to several features of the word learning environment and had the strongest association for features of the overall scene. Children's association between words and objects—an association typically measured in word learning studies—was comparatively weak; indeed, children failed to learn the words. Experiment 2 demonstrated that the scene and person context could be used to support the mapping of words and objects. Taken together, this work suggests that (1) children privilege associations with the broader environmental context, and (2) the broader environmental context can be leveraged to map word-object associations. We suggest that researchers should further examine the contribution of the environmental context to children's early word learning and account for the role of children's memory for the context when developing word learning theories. This will lead to a more robust understanding of how regularities in the environment interact with the learner to facilitate word learning.

Using an eye-tracking paradigm, the key finding from Experiment 1 was that children attended to all aspects of the word learning episode but spent the greatest proportion of time fixating on and visiting aspects of the overall scene. This was observed even when controlling for the size of the AOIs. These results suggest that children do not ignore contextual information during word learning episodes. In addition to the overall scene, children demonstrated a high proportion of looking to the two speakers in the scene. These results corroborate a robust literature on the salience of social cues in children's early word learning (Yu & Ballard, 2007; Yurovsky & Frank, 2017). Taken together, eye gaze measures from Experiment 1 revealed that children attended to the central features of a word learning episode and constructed an associative matrix of these features.

A notable characteristic of the matrix was that the associations were not equivalent in strength, even if children attended to all features of the word learning episode. That is, certain associations were stronger—or weaker—than other associations. Children demonstrated significantly stronger memory for scene-object associations than for all other associations (i.e., word-object, scene-word, person-word, scene-person, object-person), whereas their memory performance for word-object associations was lowest in magnitude. In fact, their memory for word-object associations was significantly lower than their memory for the association between objects and scenes, as well as objects and speakers. These results from Experiment 1 suggest that the scene and person context may play an important role during word learning, and word learning suffers in more complex learning environments

Why might the learner privilege the association between objects and the broader scene and person context? One explanation is that children spend the initial months of their life exploring the visual world before they utter their first word. These explorations provide a rich data set for the developing system, which comprises the caretakers and speakers in the environment, the typical settings and routines in which word learning occurs, the objects that co-occur, and semantic information about objects that are commonly labeled (Bahrick, 1988; Bushnell, 2001; Jayaraman et al., 2015; Perry & Saffran, 2017; Roy et al., 2015; Slater et al., 1999; Smith et al., 2018; Wojcik & Saffran, 2013; 2015; Zettersten et al., 2018). That is, children accrue knowledge about the world prior to comprehending and producing words, and this world knowledge may in turn aid word learning.

Two important aspects of this world knowledge are the association between objects and the contexts in which they appear, as well as the speakers that label the objects. Might children use their robust associations between the object, speaker, and scene in service of word learning? Experiment 2 sought to answer this question. Results revealed that word-object associations were strengthened when the corresponding person or scene context was displayed. That is, performance on word-object associations was enhanced when the corresponding context was displayed in comparison to when no context was displayed. We also ruled out the possibility that a longer testing phase duration may have led to poorer memory for word-object associations in Experiment 1 in comparison to Experiment 2. Instead, these results suggest that children used stronger scene-object and person-object associations to map weaker word-object associations.

There are several possible ways that children leveraged their knowledge of scene-object and person-object associations. For instance, children could have used the scene and person context as a cue to retrieve fragile word-object associations. Alternatively, children might have used their knowledge of the scene and person context to make inferences about word-object associations. That is, rather than mapping the words to the objects, children mapped the word to the scene or person presented in the trial. The data in the present study do not distinguish between these possibilities. Therefore, future studies should identify the specific process by which children leverage their knowledge of the context in complex word learning environments.

One reason why the scene and person context may support mapping of words to objects is that objects reliably co-occur with specific social and environmental contexts. That is, children frequently encounter objects in consistent environments (e.g., a fork in the kitchen) and with consistent people (e.g., caretakers). In fact, studies have found that word learning occurs within and is aided by predictable routines (Benitez & Saffran, 2018; Benitez & Smith, 2012; Hirsh-Pasek et al., 2015; Roy et al., 2015). During these predictable routines and interactions, children are gathering data about the visual properties, functions, and locations of objects, as well as the speakers labeling the objects. The associations children encode between objects and their corresponding scenes and speakers may then be used to support word-object mappings.

Constructing a system of associations between these features of a word learning episode has a notable advantage for the word learner. If the word learner only encoded a single association between a label ("fork") and an object, the word learning problem would become arduous and slow. Instead, if the word learner constructs associations between a label, object, social context (e.g., speaker), and scene context (e.g., kitchen), partial knowledge of these associations could be used to accelerate learning of word-object mappings (Smith & Yu, 2008; Yu, 2008; Yurovsky et al., 2014). In fact, evidence suggests that seeing an object in its typical environment facilitates visual processing and activates corresponding semantic and phonological information, thereby strengthening subsequent memory for word-object mappings (e.g., Bar, 2004; Biederman et al., 1982; Chun & Jiang, 1998; Gronau et al., 2008; Hollingworth, 2009; Huettig & McQueen, 2007; Huettig et al., 2011; Meints et al., 2004). Thus, encoding multimodal information in a matrix of associations may increase the efficiency of the language learning system, contributing to the rapid vocabulary gains observed in early childhood.

Future studies should investigate whether and how children integrate associative matrices. For instance, researchers should examine how children integrate word-object associative matrices (e.g., Figure 2) into more contextually-rich matrices (e.g., Figure 3). We predict that children integrate a multitude of associations between words, objects, people, and the scene context, and use these associative matrices to guide word mapping. However, there could be consequences of integrating associative matrices; integration could lead to competing and spurious associations. For instance, children may map the word "fork" to competing objects in the kitchen, such as a knife, because these objects share several contextual features.

Furthermore, the present study investigated children's mapping of words to objects when a single speaker labeled the object in a single scene context. Prior research has found that variability in the speaker and scene context can aid children's word mapping and retention (Crespo & Kaushanskaya, 2021; Goldenberg & Sandhofer, 2013; Goldenberg et al., 2022). A fruitful extension of the current study would therefore be to present objects in several contexts and with several speakers. How does this variability in contextual cues impact the mapping of word-object associations? These investigations would reveal the level of variability in the person and scene context that is desirable for novel word learning.

In conclusion, the present study demonstrates that children acquire a system of associations between words, objects, people, and the broader environmental context during word learning. This work highlights that word learning is a contextually grounded process, and the contextually grounded nature of word learning can be used to leverage novel word mapping. Indeed, while Quine (1960) posited that the rich information presented by the world should deter children's word mapping, the accumulation of this vast information may be what makes language learning possible.

ACKNOWLEDGMENTS

Many thanks to the participating families, as well as the undergraduate research assistants of the UW-Madison Learning, Cognition, & Development Lab for their help with this project. This work was funded by the James S. McDonnell Foundation and the Wisconsin Alumni Research Foundation.

CONFLICT OF INTEREST

Finally, the authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://osf.io/n3pxb/.

ORCID

Melina L. Knabe https://orcid.org/0000-0003-1686-8878
Haley A. Vlach https://orcid.org/0000-0001-6905-2106

REFERENCES

Akhtar, N. (2005). The robustness of learning through overhearing. *Developmental Science*, *8*, 199–209. https://doi.org/10.1111/j.1467-7687.2005.00406.x

Akhtar, N., & Tomasello, M. (1996). Two-year-olds learn words for absent objects and actions. *British Journal of Developmental Psychology*, *14*(1), 79–93. https://doi.org/10.1111/j.2044-835X.1996.tb00695.x

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

Axelsson, E. L., & Horst, J. S. (2014). Contextual repetition facilitates word learning via fast mapping. *Acta Psychologica*, *152*, 95–99. https://doi.org/10.1016/j.actpsy.2014.08.002

Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development*, *59*, 197–209. https://doi.org/10.2307/1130402

Baldwin, D. A. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, *62*(5), 875–890. https://doi.org/10.1111/j.1467-8624.1991.tb01577.x

Bar, M. (2004). Visual objects in context. *Nature reviews. Neuroscience*, *5*(8), 617–629. https://doi.org/10.1038/nrn1476

Benitez, V. L., & Saffran, J. R. (2018). Predictable events enhance word learning in toddlers. *Current Biology*, *28*(17), 2787–2793. https://doi.org/10.1016/j.cub.2018.06.017

Benitez, V. L., & Smith, L. B. (2012). Predictable locations aid early object name learning. *Cognition*, *125*(3), 339–352. https://doi.org/10.1016/j.cognition.2012.08.006

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*(2), 143–177. https://doi.org/10.1016/0010-0285(82)90007-X

Brainerd, C. J., Aydin, C., & Reyna, V. F. (2012). Development of dual-retrieval processes in recall: Learning, forgetting, and reminiscence. *Journal of Memory and Language*, *66*(4), 763–788. https://doi.org/10.1016/j.jml.2011.12.002

Briganti, A. M., & Cohen, L. B. (2011). Examining the role of social cues in early word learning. *Infant Behavior & Development*, *34*, 211–214. https://doi.org/10.1016/j.infbeh.2010.12.012

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: Longitudinal, growth curve modeling study. *Journal of Child Language*, *35*, 207–220. https://doi.org/10.1017/s030500090700829x

Bushnell, I. W. R. (2001). Mother's face recognition in newborn infants: Learning and memory. *Infant and Child Development: An International Journal of Research and Practice*, *10*(1-2), 67–74. https://doi.org/10.1002/icd.248

Carey, S. (2010). Beyond fast mapping. *Language Learning and Development*, *6*(3), 184–205. https://doi.org/10.1080/15475441.2010.484379

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, *15*, 17–29.

Chen, C. H., Houston, D. M., & Yu, C. (2021). Parent–child joint behaviors in novel object play create high-quality data for word learning. *Child Development*, *92*(5), 1889–1905. https://doi.org/10.1111/cdev.13620

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*(1), 28–71. https://doi.org/10.1006/cogp.1998.0681

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2016). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society: Biological Sciences*, *372*, 1–10. https://doi.org/10.1098/rstb.2016.0055

Colunga, E., Brojde, C., & Ahmed, S. (2012). Bilingual and monolingual children attend to different cues when learning new words. *Frontiers in Psychology*, *3*, 1–11. https://doi.org/10.3389/fpsyg.2012.00155

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*, 1–14. https://doi.org/10.1080/15475441.2021.1906680

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, *37*, 630–641. https://doi.org/10.1037/0012-1649.37.5.630

Ellis, E. M., Borovsky, A., Elman, J. L., & Evans, J. L. (2015). Novel word learning: An eye- tracking study. Are 18-month-old late talkers really different from their typical peers? *Journal of Communication Disorders*, *58*, 143–157. 10.1016/j.jcomdis.2015.06.011

Filippi, R., Ceccolini, A., Periche-Tomas, E., Papageorgiou, A., & Bright, P. (2020). Developmental trajectories of control of verbal and non-verbal interference in speech comprehension in monolingual and multilingual children. *Cognition*, *200*, e104252. https://doi.org/10.1016/j.cognition.2020.104252

Gangopadhyay, I., & Kaushanskaya, M. (2020). The role of speaker eye gaze and mutual exclusivity in novel word learning by monolingual and bilingual children. *Journal of Experimental Child Psychology*, *197*, 104878. https://doi.org/10.1016/j.jecp.2020.104878

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy*, *14*(5), 550–562. https://doi.org/10.1080/15250000903144199

Goldenberg, E. R., & Sandhofer, C. M. (2013). Same, varied, or both? Contextual support aids young children in generalizing category labels. *Journal* *of Experimental Child Psychology*, *115*(1), 150–162. https://doi.org/10.1016/j.jecp.2012.11.011

Goldenberg, E. R., Repetti, R. L., & Sandhofer, C. M. (2022). Contextual variation in language input to children: A naturalistic approach. *Developmental Psychology*, *58*, 1051–1065. https://doi.org/10.1037/dev0001345

Grassman, S., & Tomasello, M. (2010). Young children follow pointing over words in interpreting acts of reference. *Developmental Science*, *13*, 252–263. https://doi.org/10.1111/j.1467-7687.2009.00871.x

Gronau, N., Neta, M., & Bar, M. (2008). Integrated contextual representation for objects' identities and their locations. *Journal of Cognitive Neuroscience*, *20*(3), 371–388. https://doi.org/10.1162/jocn.2008.20027

Hale, S., Bronik, M. D., & Fry, A. F. (1997). Verbal and spatial working memory in school-age children: Developmental differences in susceptibility to interference. *Developmental Psychology*, *33*(2), 364–371. https://doi.org/10.1037//0012-1649.33.2.364

Helo, A., Pannasch, S., Sirri, L., & Rämä, P. (2014). The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research*, *103*, 83–91.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*(7), 1071–1083. 10.1177/0956797615581493

Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition*, *17*(1-2), 273–291. https://doi.org/10.1080/13506280802193367

Horst, J. S., & Houston-Price, C. (2015). An open book: What and how young children learn from picture and storybooks. *Frontiers in Psychology*, *6*, 1–4. https://doi.org/10.3389/fpsyg.2015.01719

Horst, J. S., Parsons, K. L., & Bryan, N. M. (2011). Get the story straight: Contextual repetition promotes word learning from storybooks. *Frontiers in Psychology*, *2*, 1–11. https://doi.org/10.3389/fpsyg.2011.00017

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention in 24-month-old infants. *Infancy*, *13*, 128–157. https://doi.org/10.1080/15250000701795598

Houston-Price, C., Howe, J. A., & Lintern, N. J. (2014). Once upon a time, there was a fabulous funambulist…: What children learn about the "high-level" vocabulary they encounter while listening to stories. *Frontiers in Psychology*, *5*, 1–14. 10.3389/fpsyg.2014.00075

Houston-Price, C., Plunkett, K., & Duffy, H. (2006). The use of social and salience cues in early word learning. *Journal of Experimental Child Psychology*, *95*, 27–55. https://doi.org/10.1016/j.jecp.2006.03.006

Höhle, B. (2009). Bootstrapping mechanisms in first language acquisition. *Linguistics*, *47*(2), 359–382. https://doi.org/10.1515/LING.2009.013

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language*, *54*, 460–482. https://doi.org/10.1016/j.jml.2007.02.001

Huettig, F., Olivers, C. N., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*(2), 138–150. https://doi.org/10.1016/j.actpsy.2010.07.013

Jayaraman, S., Fausey, C. M., & Smith, L. B. (2015). The faces in infant-perspective scenes change over the first year of life. *Plos One*, *10*(5), e0123780. 10.1371/journal.pone.0123780

Kachergis, G., Yu, C., & Shiffrin, R. M. (2017). A bootstrapping model of frequency and context effects in word learning. *Cognitive Science*, *41*(3), 590–622. https://doi.org/10.1111/cogs.12353

Khu, M., Graham, S. A., & Ganea, P. A. (2014). Learning from picture books: Infant's use of naming information. *Frontiers in Psychology*, *5*, 1–11. https://doi.org/10.3389/fpsyg.2014.00144

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*(10), 694–698. https://doi.org/10.1111/j.0956-7976.2004.00742.x

Markman, E. M. (1989). *Categorization and naming in children: Problems of induction*. Cambridge, MA: Bradford/MIT Press.

Mather, E., & Plunkett, K. (2009). Learning words over time: The role of stimulus repetition in mutual exclusivity. *Infancy*, *14*, 60–76. https://doi.org/10.1080/15250000802569702

Mather, E., & Plunkett, K. (2012). The role of novelty in early word learning. *Cognitive Science*, *36*, 1157–1177. https://doi.org/10.1111/j.1551-6709.2012.01239.x

Meints, K., Plunkett, K., Harris, P. L., & Dimmock, D. (2004). The cow on the high street: Effects of background context on early naming. *Cognitive Development*, *19*(3), 275–290. https://doi.org/10.1016/j.cogdev.2004.03.004

Oliva, A. (2005). Gist of the scene. In *Neurobiology of Attention* (pp. 251–256). Academic press.

Perry, L. K., & Saffran, J. R. (2017). Is a pink cow still a cow? Individual differences in toddlers' vocabulary knowledge and lexical representations. *Cognitive Science*, *41*(4), 1090–1105. https://doi.org/10.1111/cogs.12370

Pomper, R., & Saffran, J. R. (2019). Familiar object salience affects novel word learning. *Child Development*, *90*(2), e246–e262. https://doi.org/10.1111/cdev.13053

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

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663–12668. https://doi.org/10.1073/pnas.1419773112

Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding word learning in space. *Plos One*, *6*, 1–13. https://doi.org/10.1371/journal.pone.0028095

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*, 668–671. https://doi.org/10.1016/j.cub.2008.03.059

Slater, A., Quinn, P. C., Brown, E., & Hayes, R. (1999). Intermodal perception at birth: Intersensory redundancy guides newborn infants' learning of arbitrary auditory— Visual pairings. *Developmental Science*, *2*(3), 333–338. https://doi.org/10.1111/1467-7687.00079

Smith, L. B. (2000). Learning how to learn words: An associative crane. In R. M. Golinkoff & K. Hirch-Pasek (Eds.), *Becoming a word learner: A debate on lexical acquisition*, pp. 51–80. Oxford: Oxford University Press.

Smith, L. B., Jayaraman, S., Clerkin, E., & Chen, Y. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*, 325–336. https://doi.org/10.1016/j.tics.2018.02.004

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word- referent learning. *Trends in Cognitive Sciences*, *18*, 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

Spiegel, C., & Halberda, J. (2011). Rapid fast-mapping abilities in 2-year-olds. *Journal of Experimental Child Psychology*, *109*, 132–140. https://doi.org/10.1016/j.jecp.2010.10.013

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. 10.1016/j.jecp.2014.06.003

Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science*, *23*(2), 121–126. https://doi.org/10.1177/0963721414522813

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child Development*, *90*(6), 2135–2152. https://doi.org/10.1111/cdev.13089

Tomasello, M., & Akhtar, N. (1995). Two-year-olds use pragmatic cues to differentiate reference to objects and actions. *Cognitive Development*, *10*, 201–224. https://doi.org/10.1016/0885-2014(95)90009-8

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., & Sandhofer, C. M. (2011). Developmental differences in children's context- dependent word learning. *Journal of Experimental Child Psychology*, *108*, 394–401. https://doi.org/10.1016/j.jecp.2010.09.011

Vlach, H., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in Psychology*, *3*, 1–8. 10.3389/fpsyg.2012.00046

Weighall, A. R., Henderson, L. M., Barr, D. J., Cairney, S. A., & Gaskell, M. G. (2017). Eye- tracking the time-course of novel word learning and lexical competition in adults and children. *Brain and Language*, *167*, 13–27. https://doi.org/10.1016/j.bandl.2016.07.010

Williams, S. E., & Horst, J. S. (2014). Goodnight book: The benefit of sleep consolidation on word learning via storybooks. *Frontiers in Psychology*, *5*, 1–12. 10.3389/fpsyg.2014.00184

Wojcik, E. H., & Saffran, J. R. (2013). The ontogeny of lexical networks: Toddlers encode the relationships among referents when learning novel words. *Psychological Science*, *24*(10), 1898–1905. https://doi.org/10.1177/0956797613478198

Wojcik, E. H., & Saffran, J. R. (2015). Toddlers encode similarities among novel words from meaningful sentences. *Cognition*, *138*, 10–20. https://doi.org/10.1016/j.cognition.2015.01.015

Woodward, A. L., Markman, E. M., & Fitzsimmons, C. M. (1994). Rapid word learning in 13- and 18-month-olds. *Developmental Psychology*, *30*, 553–566. https://doi.org/10.1037/0012-1649.30.4.553

Woodward, A. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science*, *6*, 297–311. https://doi.org/10.1111/1467-7687.00286

Yow, W. Q., Li, X., Lam, S., Gliga, T., Tan, K. H., Shek, L., Yap, F., Chong, Y. S., & Broekman, B. F. P. (2017). Effects of bilingualism on children's use of social cues in word learning. *Proceedings of the 41st Annual Boston University Conference on Language Development*, ed. Maria LaMendola & Jennifer Scott pp. 780–793. Somerville, MA: Cascadilla Press.

Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language, Learning, and Development*, *4*, 31–62. https://doi.org/10.1080/15475440701739353

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*, 2149–2165. https://doi.org/10.1016/j.neucom.2006.01.034

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yu, C., & Smith, L. B. (2017). Hand-eye coordination predicts joint attention. *Child Development*, *88*(6), 2060–2078. https://doi.org/10.1111/cdev.12730

Yu, C., Suanda, S. H., & Smith, L. B. (2019). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental Science*, *22*(1), e12735. https://doi.org/10.1111/desc.12735

Yurovsky, D., & Frank, M. C. (2017). Beyond naïve cue combination: Salience and social cues in early word learning. *Developmental Science*, *20*(2), e12349. https://doi.org/10.1111/desc.12349

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), 1–22. 10.3758/s13423-013-0443-y

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

How to cite this article: Knabe, M. L, & Vlach, H. A (2023). Not all is forgotten: Children's associative matrices for features of a word learning episode. *Developmental Science*, *26*, e13291. https://doi.org/10.1111/desc.13291

APPENDIX A

Scene script

Task introduction:

Are you ready to play a game? I will show you a few cartoons with different people, places, and things. The people will teach you something. Try to remember as much as you can from what you see!

Learning phase script:

1. **Bedroom**

   Anne: Can I show you gipple, Mike?
   Mike: What is it like, Anne?
   Anne: Gipple is very comfortable to sleep on!

2. **Snowman**

   Rob: Do you see dax, Kate?
   Kate: Why is it on the snowman, Rob?
   Rob: Dax is a great snowman hat!

3. **Living room**

   Mike: Can you give me modi, John?
   John: Why do you need it, Mike?
   Mike: Modi switches channels on TV!

4. **Playground**

   Kate: Have you played with neem, Jane?
   Jane: How do you play with it, Kate?
   Kate: Neem is thrown like this!

5. **Dining room**

   John: Have you used zoop, Rob?
   Rob: What does it do, John?
   John: Zoop helps you eat your food!

6. **Forest**

   Jane: Can you see blicket, Anne?
   Anne: Where is it, Jane?
   Jane: Blicket is in the tree!

Testing phase introduction:

I will now ask you a few questions about what you saw. Are you ready?

Testing phase script:

1. **Word-object association**

   "Which one is __?"

2. **Person-object association**

   "Can you point to who played with this?"

3. **Scene-object associations**

   "Can you point to where you heard the kids talk about this?"

4. **Scene-person association**

   "Where did this person teach you about a new toy?"

5. **Scene-word association**

   "Can you point to where the kids taught you about __?"

6. **Person-word association**

   "Can you point to who taught you about __?").

APPENDIX B

Linking children's visual attention and memory performance

We examined the relation between children's total fixation duration and average fixation count to AOIs (target object, target speaker, scene) and their memory performance for associations that contained the AOI (word-object, person-object, scene-object, scene-person, scene-word, person-word). For associations that contained two AOIs (e.g., person-object, scene-object, scene-person), we examined the association between the combined total fixation duration and average fixation count of the AOIs and children's memory performance for the associations.

Target object

Total fixation duration to the target object was not significantly related to children's memory performance on the object-word association, $r = 0.27$, $p = 0.09$. Total fixation duration to the target object and target speaker was not significantly related to children's memory performance on the person-object association, $r = 0.26$, $p = 0.09$. Similarly, total fixation duration to the target object and scene was not significantly related to children's memory performance on the scene-object association, $r = 0.29$, $p = 0.06$.

Average fixation count to the target object was not significantly related to children's memory performance on the object-word association, $r = 0.04$, $p = 0.80$. However, average fixation count to the target object and target speaker was significantly related to children's performance on the person-object association, $r = 0.30$, $p = 0.048$. Average fixation count to the target object and scene was not significantly related to children's memory performance on the scene-object association, $r = 0.17$, $p = 0.27$.

Speaker

Total fixation duration to the speaker was not significantly related to children's memory performance on the person-word association, $r = -0.31$, $p = 0.052$. Similarly, total fixation duration to the target speaker and scene was not significantly related to children's memory performance on the scene-person association, $r = -0.23$, $p = 0.13$. As shown above, total fixation duration to the target speaker and target object was not significantly related to children's memory performance on the person-object association, $p = 0.09$.

Average fixation count to the target object was not significantly related to children's memory performance on person-word association, $r = -0.27$, $p = 0.09$. Similarly, average fixation count to the target speaker and scene was not significantly related to children's memory performance on the scene-person association, $r = 0.26$, $p = 0.10$. As shown above, average fixation count to the target speaker and target object was significantly related to children's performance on the person-object association, $p = 0.048$.

Scene

Total fixation duration to the scene was not significantly related to children's memory performance on the scene-word association, $r = -0.25$, $p = 0.12$. As shown above, total fixation duration to the scene and target object was not significantly related to children's memory performance on the scene-object association, $p = 0.06$. Similarly, total fixation duration to the scene and target speaker was not significantly related to children's memory performance on the scene-person association, $p = 0.13$.

Average fixation count to the scene was not significantly related to children's memory performance on the scene-word association, $r = 0.23$, $p = 0.15$. As shown above, average fixation count to the scene and target object was not significantly related to children's memory performance on the scene-object association, $p = 0.27$. Similarly, average fixation count to the scene and target speaker was not significantly related to children's memory performance on scene-person association, $p = 0.10$.

Finally, we assessed whether greater looking to the scene negatively impacted performance on word-object associations. We found that total fixation duration, $r = -0.18$, $p = 0.25$, and average fixation count, $r = -0.14$, $p = 0.37$, were not significantly negatively correlated with performance on the word-object associations.