PAPER

Flexible fast-mapping: Deaf children dynamically allocate visual attention to learn novel words in American Sign Language

Amy M. Lieberman[1] | Allison Fitch[2] | Arielle Borovsky[3]

[1] Wheelock College of Education and Human Development, Boston University, 2 Silber Way, Boston, Massachusetts 02215, USA

[2] Department of Psychology, Rochester Institute of Technology, Rochester, New York, USA

[3] Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana, USA

Correspondence
Amy M. Lieberman, Boston University, Boston, MA, USA.
Email: alieber@bu.edu

Abstract

Word learning in young children requires coordinated attention between language input and the referent object. Current accounts of word learning are based on spoken language, where the association between language and objects occurs through simultaneous and multimodal perception. In contrast, deaf children acquiring American Sign Language (ASL) perceive both linguistic and non-linguistic information through the visual mode. In order to coordinate attention to language input and its referents, deaf children must allocate visual attention optimally between objects and signs. We conducted two eye-tracking experiments to investigate how young deaf children allocate attention and process referential cues in order to fast-map novel signs to novel objects. Participants were deaf children learning ASL between the ages of 17 and 71 months. In Experiment 1, participants ($n = 30$) were presented with a novel object and a novel sign, along with a referential cue that occurred either before or after the sign label. In Experiment 2, a new group of participants ($n = 32$) were presented with two novel objects and a novel sign, so that the referential cue was critical for identifying the target object. Across both experiments, participants showed evidence for fast-mapping the signs regardless of the timing of the referential cue. Individual differences in children's allocation of attention during exposure were correlated with their ability to fast-map the novel signs at test. This study provides first evidence for fast-mapping in sign language, and contributes to theoretical accounts of how word learning develops when all input occurs in the visual modality.

KEYWORDS
American Sign Language, deaf children, fast-mapping, novel word learning, referential cues, visual attention

1 | INTRODUCTION

Children learn words by attending to and connecting the input they perceive onto the objects and events around them. While this task may sound simple, young word learners are faced with the daunting challenge of linking an ever-changing array of linguistic and referential cues that may not align in space and time. The synchronous, multisensory alignment of the child's attention to linguistic information in the audi-

tory channel along with social referential cues in the visual channel (e.g., gaze and pointing), can make referent mapping more transparent and support word learning (Baldwin, 1991; Gogate et al., 2000; Suanda et al., 2016, 2019). Yet, for deaf children learning American Sign Language (ASL), the task of mapping input to its referents is necessarily asynchronous, as linguistic input, referential cues, and objects are all perceived sequentially via the visual modality. Thus, children's ability to allocate attention in a way that enables them to connect

language and its referents sequentially, rather than simultaneously, is a typical and necessary aspect of word learning in ASL. Deaf children learning ASL with early and rich exposure show vocabulary growth at rates that match hearing peers (Caselli et al., 2020), thus raising the question of how children learning ASL succeed at connecting asynchronous linguistic and visual information to learn new words. We address this question by focusing on how timing of referential cues in the input impacts fast-mapping of novel ASL signs, and investigating whether individual differences in children's ability to allocate attention between multiple visual inputs predicts their ability to learn new signs.

1.1 | Allocation of attention to language and referential cues

Infants learn to interpret the social referential cues of their adult interlocutors as meaningful connections between object labels and their referents (Baldwin, 1993; Booth et al., 2008). Although infants will follow gaze from a very early age, it is not until about 12 months that infants reliably make the connection between an adult's gaze and the referential intent of the object of their gaze (Woodward, 2003); and not until 18 months that children use gaze reliably as an entry point to joint attention (Baldwin, 1991; Moore & Corkum, 1998). Gaze continues to be a powerful social cue for children across the first several years of development (Yurovsky & Frank, 2017), and children rely on such cues to resolve referential ambiguity.

Infants learning ASL have enhanced gaze control (Bosworth & Stone, 2021), and deaf infants learning ASL have enhanced abilities to follow the gaze of their interlocutors relative to hearing infants (Brooks et al., 2020). By the age of two, deaf children acquiring ASL shift gaze meaningfully and frequently when interacting with adult interlocutors to connect language and visual information (Lieberman et al., 2014). This enhanced gaze cueing persists into childhood (Pavani et al., 2019). However, children learning ASL must learn when and how to best allocate their own visual attention to perceive both linguistic and non-linguistic input, particularly when following adult cues requires a shift in attention from one locus to another.

Most current theories of joint attention describe children's ability to follow cues as an important milestone in development, but cannot as easily account for children's developing ability to evaluate multiple sources of input and make in-the-moment decisions about whether to maintain or shift gaze. Observing children who are learning ASL can shine a light on this feature of development. Optimal attention allocation would enable children to perceive linguistic input when it is produced, and to attend objects and events during pauses in linguistic input. We do not currently know how this attention allocation develops. Additionally, we do not know whether or how individual differences in children's ability to allocate attention predicts word learning in ASL, in much the same way that individual differences in gaze following predict later vocabulary and other linguistic outcomes in spoken language (Markus et al., 2000). These gaps are addressed in the current study.

RESEARCH HIGHLIGHTS

- Investigated how young deaf children learning American Sign Language learn novel words by flexibly allocating visual attention between sign labels and objects.
- Deaf children between 17 and 71 months old participated in two eye-tracking experiments in which they were exposed to novel signs and variably-timed referential cues.
- Children showed evidence for fast-mapping novel signs to objects, regardless of the timing of referential cues. Individual differences in attention allocation predicted word learning.
- Findings advance understanding of word learning in the visual modality and contribute to theoretical accounts of the coordination of cues that effectively support word learning.

1.2 | Timing of referential cues relative to object labels

The timing of referential cues is critical to infants' ability to connect the cue, the linguistic input, and the referent object. In child-directed speech, parents often align their own speech with the toddler's gaze (Gogate et al., 2000), and this multisensory synchrony between linguistic labeling and visual attention can support word learning (Yu & Smith, 2012). Perreira and colleagues (Pereira et al., 2014) found that toddlers were more likely to learn the names of objects when the object being labeled was more centrally located and looming large in the child's view. Further, parent gesture towards an object just prior to object labeling, and continuing after, appears to increase referential transparency (Gogate & Hollich, 2010; Trueswell et al., 2016). Labeling an object *while* the infant is directing attention to it, manipulating it, touching it, or fixating can provide optimal moments for word learning (Yu & Smith, 2012).

While deaf children are adept at responding to referential cues, the timing of such cues that would most support word learning in ASL is unknown. In sign language, simultaneous perception of the label and the object is theoretically possible, but only when the parent has carefully positioned the object between themselves and the child. In addition, the "single, looming, object" described by Pereira et al. (2014) as optimal for labeling is not compatible with ASL-based interactions—a child cannot simultaneously have a single looming object in their visual field and perceive the ASL label for that object. Similarly, the shift in attention towards an object just prior to labeling, which has been found to be an important predictor of referential transparency in spoken language (Trueswell et al., 2016) would be problematic in an ASL interaction if the shift in attention led the child to look away from the interlocutor. On the other hand, if children only fixate on the interlocutor, they will miss the opportunity to map input to its referents. Instead, the child must perceive the object label and referent in sequence.

1.3 | Word learning in ASL

Deaf children acquiring ASL from birth acquire words at a rate and manner that is largely parallel to spoken language development (Anderson & Reilly, 2002; Caselli et al., 2020). Yet only two previous studies have examined novel word learning in ASL. Lederberg and colleagues (Lederberg et al., 2000) studied word learning as a function of children's use of sign and spoken language. They found that learning through direct reference with overt social pragmatic cues preceded learning through indirect mapping that draws on mutual exclusivity constraints. In a follow up study with deaf and hard-of-hearing preschoolers (Lederberg & Spencer, 2009), they found a similar developmental progression in which children who showed evidence for learning via indirect mapping had already mastered mapping via direct reference. In the direct reference conditions, the experimenter gazed at, pointed at, and manipulated the object while labeling it, and then handed it to the child to explore. The child's gaze or attention to the objects or the experimenter was not explicitly coded, though the researchers "used eye gaze, pointing, and manipulation to make reference clear while labeling the objects, and care was taken to ensure the child attended to the object and to the linguistic stimulus either simultaneously or in quick succession" (Lederberg & Spencer, 2009, p. 51).

1.4 | Current study

We sought to shed light on the conditions under which temporal alignment of attention supports word learning by carrying out two experiments in deaf children acquiring ASL. We measured fast-mapping skills (Carey & Bartlett, 1978; Halberda, 2003) by presenting children with novel objects and novel ASL labels. During exposure, we varied whether the referential cue occurred before or after the object label. We reasoned that children might learn novel words better when the sign label preceded the point, as this would enable them to perceive both the word and the point prior to shifting gaze from the signer to the object. We then assessed children's fast-mapping abilities. Here, we hypothesized that children would successfully fast-map a novel sign to a novel object, and that there would be a positive correlation between children's age and performance both in terms of allocating attention at exposure and recognizing the target at test. Finally, we asked whether deaf children's allocation of attention between the novel sign and object during the exposure phase would predict their performance on the fast-mapping task. We predicted that children who allocated attention most efficiently would show more robust mapping of the target signs. Experiments 1 and 2 varied in the degree of referential uncertainty during exposure, by presenting an array of one (Experiment 1) or two (Experiment 2) novel objects. We hypothesized that novel word learning in two-object arrays would not only be more challenging, but also might be more sensitive to the timing of referential cues with respect to the sign label, as the learner would be more reliant on the signer's referential cues to map the label onto the object.

2 | EXPERIMENT 1: FAST-MAPPING NOVEL ASL SIGNS TO NOVEL OBJECTS

2.1 | Methods

2.1.1 | Participants

We recruited 41 deaf children between the ages of 17 and 69 months (M_{age} = 43 months) who were learning ASL to participate in the study. Participants were recruited from the Northeast and Midwestern United States, through programs serving deaf children and through social media. Of the recruited participants, nine children did not complete any trials of the eye-tracking task. Two participants were excluded for excessive trackloss (described below). The final sample of participants (n = 30) was between 17 and 69 months (M_{age} = 44 months), with 22 males and 8 females. Reported race was White (n = 22), Asian (n = 2), and more than one race (n = 4), with two participants not reporting. Three participants reported Hispanic/Latinx ethnicity, 25 reported not Hispanic/Latinx, and 4 did not report ethnicity. Twenty-five children had deaf parents and five children had hearing parents. The children with deaf parents were exposed to ASL from birth; children with hearing parents were exposed to ASL between 0 and 36 months. All children attended an early intervention program that used ASL at the time of testing.

2.1.2 | Stimuli

Novel signs and objects
Stimuli consisted of six novel objects that were paired with six novel signs. The novel signs were developed by deaf signing researchers and obeyed the phonotactic constraints of ASL. Six novel objects were selected from a novel object database (Horst & Hout, 2016), and set against a white background square that measured 400 × 400 pixels (approximately 4 in. square). Objects and signs were paired arbitrarily, though we ensured that there was no iconic relationship between the novel signs and their labels. The six novel objects were then divided into three pairs. The assignment of object pairs to exposure condition was counterbalanced across versions of the experiment. In addition to the novel sign trials, children also saw trials with familiar sign-object pairs, and these were interleaved with the novel sign trials.

Video stimuli
Exposure videos. All videos featured a deaf, native ASL user signing the stimuli sentences. Exposure consisted of an attention-getting sign, a labeling utterance, a filler period, and then a second labeling utterance (Figure 1). After the first labeling utterance, the signer returned their hands to a resting position, then produced a filler sign (e.g., COOL or WOW), then returned their hands to a resting position again before producing the second labeling utterance. Each exposure video was edited to be exactly 10 s long by removing extraneous frames from the beginning and end of the video. We then identified a unique onset and offset point for each novel sign label. We defined sign onset as the first
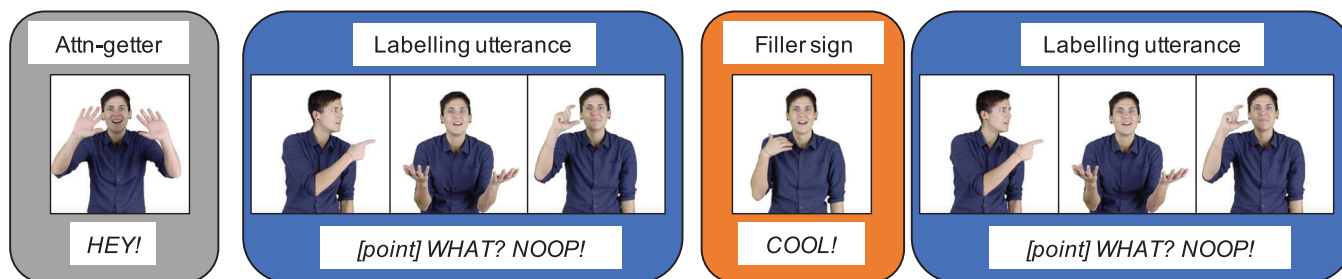
FIGURE 1 Structure of exposure trial sign videos. The signer first produced an attention-getting sign, followed by a labeling utterance, a filler period, and then a second identical labeling utterance. The sequence unfolded over 10 s

frame where the signer's hands moved away from the previous sign, and sign offset as the first frame where the signer's hands transitioned to the next sign. Similarly, we defined the onset of the filler period as the first frame where the signer's hands moved away from the final sign in the first labeling utterance, and the offset of the filler period as the first frame where the signer's hands transitioned to the first sign in the second labeling utterance. Each video was cropped to 400×400 pixels with the signer appearing against a white background.

Exposure trials consisted of three conditions. In the *point-word* condition, the signer simultaneously pointed and gazed at the novel object first, and then labeled it, for example: HEY! POINT WHAT? KOBA. COOL! POINT WHAT? KOBA ("Hey! What's that? That [point and gaze] is a KOBA! Cool! What's that? That [point and gaze] is a KOBA!"). In the *word-point* condition, the gaze and point cue occurred after the object label, as follows: HEY! KOBA WHAT? POINT. WOW! KOBA WHAT? POINT ("Hey! A KOBA, what is it? It's that [point and gaze]! Wow! A KOBA, what is it? It's that [point and gaze]!"). Finally, in the *word-only* condition, the signer did not produce a gaze and point cue, but instead used the word "see" while looking forward, as follows: HEY! SEE WHAT? KOBA. WOW! SEE WHAT? KOBA ("Hey! Do you see it? It's a KOBA. Wow! Do you see it? It's a KOBA."). Note that ASL does not include determiners, and the demonstrative pronoun "that" is realized as a point.

Test videos. The test video stimuli were identical across conditions. The video consisted of the signer producing the sentence WHERE KOBA? ("Where is the KOBA?"). Stimuli were edited so that the WHERE sign was 1000 ms and the object label was 1000 ms. The onset of the sign was identified as the first frame when the signer's hand transitioned away from the final movement of the WHERE sign. The videos were 400×400 pixels (4 in. square) with the signer appearing against a white background.

2.1.3 | Procedure

The protocol was approved by the Institutional Review Board at Boston University. Parents provided informed consent. Parents and children were seated in front of a 17-in. monitor and SR-Research Eyelink 1000+ eye-tracker that recorded eye movements at 500 Hz. A short, animated movie was played to att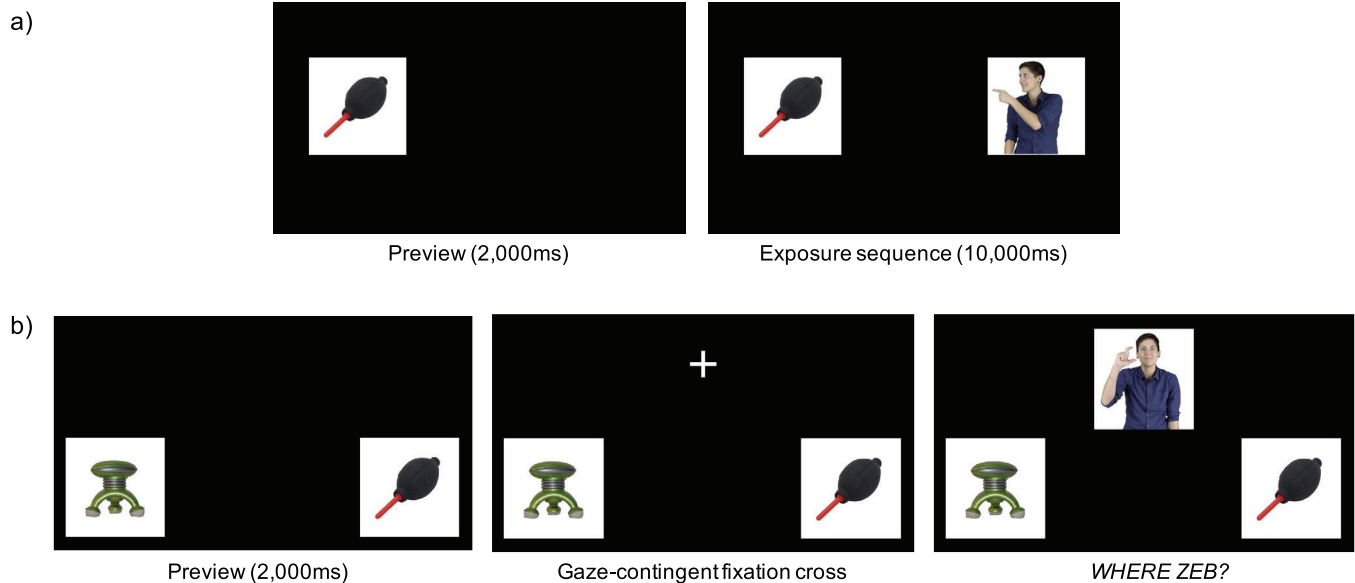ract the child's attention while the experimenter affixed a small sticker on the child's forehead and focused the camera. Next, a five-point calibration sequence was carried out. The experiment consisted of alternating blocks of eight exposure trials from a single condition followed by eight test trials. Both exposure and test blocks included four novel object trials used in the current experiment and four familiar object trials (not analyzed here). During exposure trial blocks, each object was labeled four times (twice in each of two trials). During test trial blocks, the same object pairs were presented four times, such that each object in the pair appeared twice as the target and twice as the distractor. This allowed us to control for the possibility that children could employ a strategy of attaching novel labels to objects based on amount of exposure, as each object was seen and labeled equally.

In the exposure trials, children first saw the picture of the novel object on one side of the screen for 2 s. Then, the exposure video appeared on the opposite side of the screen, while the novel object picture remained on screen (Figure 2a). The picture and video were 480 pixels (approximately 5 in.) apart on the monitor. Side placement of the picture and video was counterbalanced across trials. In the test trials, pairs of pictures consisting of a target and distractor first appeared on each side in the lower quadrant of the screen. Next, a fixation cross appeared centered in the top half of the screen. The cross was gaze-contingent, such that when children fixated on the cross, this triggered the onset of the video. The test video then played, and children had 2 s to fixate the pictures. The trial ended when the experimenter hit a key on the keyboard following the 2 s period after the offset of the test video (Figure 2b). Following each block of trials there was a break, during which children saw an engaging animated video. Between each trial, a small central image appeared on screen, and the experiment only advanced when the experimenter clicked on the image, allowing children to take breaks as needed. There were three blocks in total, one for each exposure condition. The order of blocks was counterbalanced.

2.2 | Results

2.2.1 | Approach to analysis

Analysis focused on gaze patterns during exposure and test trials, first independently and then in relation to one another. All analyses were done using R Verison 3.6.2 (R Core Team, 2019), with the lme4 (Bates

a)

Preview (2,000ms)　　　　　　　　　Exposure sequence (10,000ms)

b)

Preview (2,000ms)　　　　Gaze-contingent fixation cross　　　　*WHERE ZEB?*

FIGURE 2 Trial structure for (a) exposure trials and (b) test trials in Experiment 1. In exposure trials, participants first saw a single object on one side of the screen for 2000 ms. Then, the sign video appeared on the opposite side of the screen. In test trials, participants saw pairs of signs comprised of a target and distractor. After 2000 ms, a gaze-contingent fixation cross appeared. Once the child fixated the cross, the test video appeared, directed the child to the target object

et al., 2015), and EyetrackingR (Dink & Ferguson, 2015) packages. Following a trackloss analysis, we first examined gaze patterns during the exposure trials to investigate how children allocated attention between the novel sign and novel object, and whether attention to the novel sign varied as a function of condition. Next, we analyzed the test trials to determine whether children showed evidence of fast-mapping novel signs. Finally, we analyzed whether children's gaze patterns during exposure predicted their performance on the fast-mapping task. While we present data visualizations as fixation proportions, prior to analyzing fixations we first calculate log-gaze transformations, which address specific limitations of raw proportions: first, raw proportions are not linearly independent (i.e., greater looks to one AOI means fewer looks to other AOIs); second, as raw proportions are fixed between zero and one, they violate homogeneity of variance assumptions. This approach has been used in our prior eye-movement studies with both hearing (Borovsky, 2020; Borovsky et al., 2016) and deaf (Lieberman & Borovsky, 2020; Wienholz & Lieberman, 2019) children.

2.2.2 | Trackloss

We removed trials in which the child was not looking at any AOI for at least 20% of the time across the trial, which led to exclusion of 38 exposure trials and 53 test trials. We excluded any participants whose data did not include at least 25% of the experimental trials, which led to exclusion of two participants. All subsequent analyses were carried out on the remaining 322 exposure and 307 test trials from 30 participants.

2.2.3 | Allocation of attention during exposure

We first plotted fixations to the video and image during the full 10 s exposure video, to obtain a broad look at how children divided their gaze between the novel object and ASL input. We then homed in on two critical points during the trial: the sign label windows, during which children needed to be fixating the sign video to perceive the sign, and the filler period between labeling utterances, during which children could fixate the novel object without missing critical linguistic input.

Full trial
We plotted the time course across the full exposure trial window (Figure 3) by condition. Children largely fixated the sign video throughout the trial, with a sharp increase in looks to the novel object that coincided with the filler period between labeling utterances. Children spent an average of 60% of the time fixating the sign video and 12.8% of the time fixating the novel object across conditions.

Sign label
We analyzed fixations to the video during the novel sign label, that is, the window of time between sign onset and sign offset, which was uniquely identified for each trial. There were two instances of the novel sign during each trial, and these were included as separate data points because they occurred at different points in the trial. We fit a linear mixed-effects model of sign fixations as a log-transformed proportion of total fixations to the sign label [log(PropSign/1-PropSign)], with fixed effects for condition and age and random effects for participant and item (Table 1). There were no significant effects of condition, but there
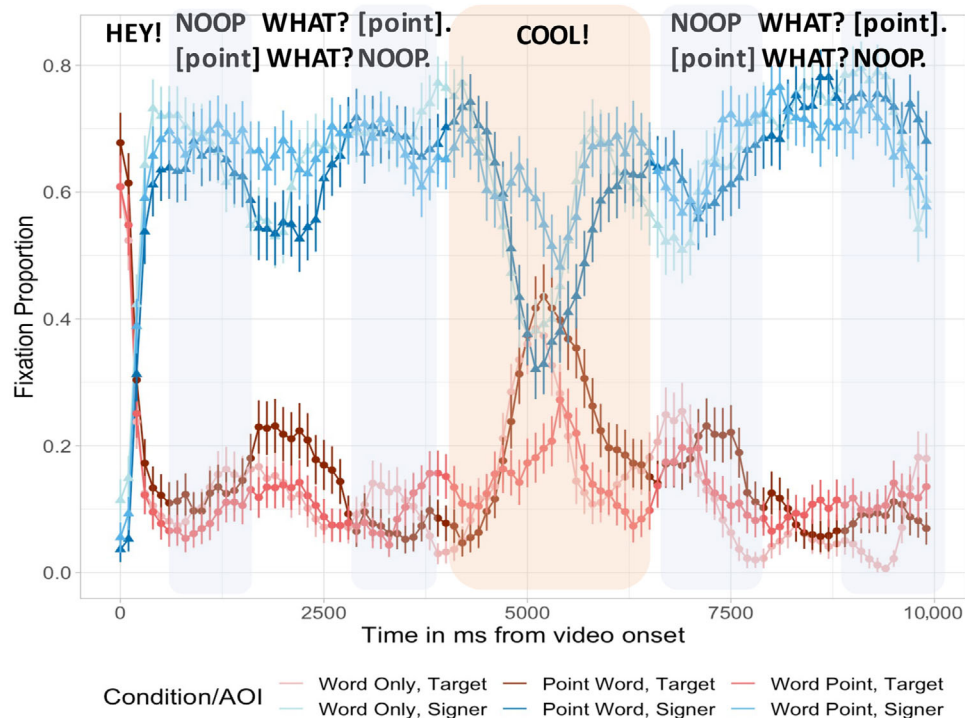
FIGURE 3 Experiment 1 timecourse of looking to video (signer) and target during exposure trials, by condition. Error bars indicate ±1 standard error. Sample exposure sentences are overlayed, with shaded areas showing the analyzed sign labels (blue shading) and filler period (orange shading)

TABLE 1 Experiment 1 model output: Fixations (LogGaze transformed) to the video during novel sign label as a function of condition and age during exposure

	Estimate	Standard error	df	t	p
Intercept	2.07	0.81	32.71	2.55	**0.02**
Point-word[a]	0.37	0.81	12.86	0.46	0.66
Word-point[a]	−0.59	0.81	12.75	−0.73	0.48
Age	0.11	0.04	27.10	2.90	**0.007**

[a]Compared to word-only condition.

TABLE 2 Experiment 1 model output: Fixations (LogGaze transformed) to target object during filler period as a function of condition and age

	Estimate	Standard error	df	t	p
Intercept	−2.86	0.42	14.67	−9.19	**<0.001**
Point-word[a]	0.97	0.59	15.01	1.64	0.12
Word-point[a]	−0.46	0.59	14.81	−0.78	0.45
Age	0.03	0.01	27.73	2.19	**0.04**

[a]Compared to word-only condition.

were significant age effects. Older children fixated more on the sign than younger children.

Filler period
The other critical window was the filler period, that is, the window of time between the two labeling utterances during which the signer returned her hands to a resting position, then produced a filler sign (e.g., COOL!). Fixations to the novel object during this window would enable children to map the recently provided label onto the sign. We fit a linear mixed-effects model of log-transformed fixations to the novel object with fixed effects for condition and age and random effects for participant and item (Table 2). There were no significant effects of condition, but there were significant age effects. Older children fixated more on the target picture during the filler period than younger children.

2.2.4 | Allocation of attention during test

To determine whether children had successfully mapped the novel signs onto the novel objects during test trials, we plotted the time-course of looks to the target and distractor pictures starting from the onset of the target word and continuing through the length of the trial (Figure 4). Children generally fixated the sign video until approximately 600 ms following sign onset, and then shifted gaze to the target picture. Following convention from previous studies of ASL lexical recognition (Lieberman & Borovsky, 2020; MacDonald et al., 2018), we analyzed looks from 600 to 2500 ms following sign onset, which we call the sign recognition window. However, given that recognition of novel signs is a more difficult task than recognition of familiar signs and thus often occurs later in the timecourse (Bion et al., 2013; Booth & Waxman, 2009; Borovsky et al., 2016;
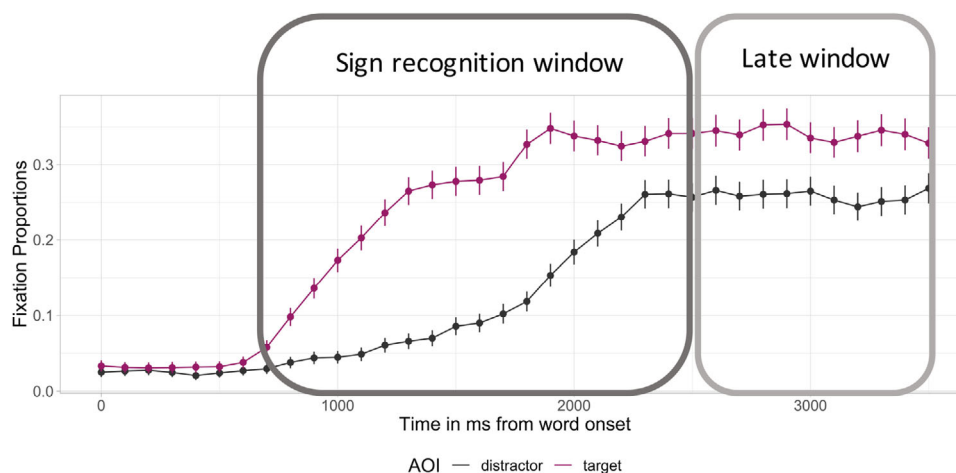
FIGURE 4 Experiment 1 timecourse of looks to target and distractor images during test trials starting at target sign onset. Boxes indicate the sign recognition window (600–2500 ms after onset) and the late window (2500–3500 ms after onset)

Houston-Price et al., 2010; Mather & Plunkett, 2010), we analyzed a second, late window, from 2500 to 3500 ms following sign offset. For each window, we fit a linear mixed-effects model of fixations as a log-transformed proportion of looking to the target relative to distractor [log10(ProportionTarget/ProportionDistractor)], with fixed effects for trial condition and child age and random effects for participant and item. In the sign recognition window, children looked to the target more than the distractor as indicated by a log gaze greater than zero ($M = 0.41$, $SD = 0.5$), suggesting they had successfully mapped the novel signs [$t(29) = 4.47$, $p < 0.001$, 95% CI (0.25, Infinite)]. In the late window, log gaze to the target was again greater than zero ($M = 0.23$, $SD = 0.74$), although the effect was smaller relative to the early window [$t(29) = 1.71$, $p = 0.049$, 95% CI (0.002, Infinite)]. There were no significant effects of exposure condition nor child age on target fixations.

2.2.5 | Relationship between gaze patterns at exposure and test

Finally, we conducted an analysis to determine whether children who allocated attention between the sign label and the object most efficiently would show greater evidence of word learning. We predicted that children who were more adept at allocating attention during exposure would show greater fixations to the target during test. For each child, we created two exposure trial predictors: (1) log-adjusted proportion of total looking time to the object label during the sign label window, for all instances of sign labeling across trials and conditions, and (2) log-adjusted proportion of total looking time to the target picture during the filler period, across trials and conditions. In the test trials, we averaged the log-adjusted proportion of total looking time to the target picture across trials and conditions in the sign recognition window. As fixations to specific locations are calculated as a proportion of total fixations to any AOI, this analysis controls for children's overall likelihood of fixating the screen. We conducted

TABLE 3 Experiment 1 model output: Proportion of fixations (LogGaze transformed) to the video during sign label windows and to the novel object during the filler period as predictors of target looking (600–2500 ms window) at test

	Estimate	SE	t	p
Intercept	−3.2	0.99	−3.21	**0.003**
Fixations to video during sign label	0.25	0.09	2.87	**0.007**
Fixations to target object during filler	0.24	0.24	0.99	0.33

a linear regression using each of the two critical exposure windows to predict proportion looking time to the target during test (Table 3). Proportion of looking to the video during the sign label, but not to the object during the filler period, was a significant predictor of looking at test (full model $R^2(27) = 0.23$, $p = 0.01$). Children who allocated gaze such that they were fixating the signer when the novel label was produced were better at mapping the novel sign to the novel object at test.

2.2.6 | Summary of Experiment 1

Our results illustrate how deaf children learning ASL divide attention between ASL input and visual stimuli. Children were drawn to the dynamic ASL input, and waited for a natural pause in the input to shift gaze away from the signer and towards the object. Contrary to our initial hypothesis, the timing of the referential gaze and point cue relative to the object label did not have a significant impact on children's gaze patterns. Children instead waited past the cue until the filler period, and then shifted gaze to the object. Our analysis of target fixations during test trials revealed that children successfully mapped the novel object onto the novel label. Critically, the only significant predictor of target looking at test was children's fixations to the signer

when the novel sign was produced during exposure; neither age nor exposure condition predicted fixations at test. This pattern suggests that children are rapidly assessing the structure of the input sentences, which are consistent in overall prosodic cadence, and then strategically allocating attention during natural prosodic breaks. Children who do this most effectively—that is, those who allocate attention such that they perceive the novel sign label—are better able to learn the novel words.

Experiment 1 allowed us to determine how children allocate visual attention in a highly simplified visual scene. However, the fact that there was only a single object presented with the novel label raises the possibility that children were simply ignoring the referential cue. That is, the object-sign pairing was such that children did not need to process the direction of the gaze and point, as there were no competing objects on the screen. If children are indeed ignoring the referential cues, then we would expect that when there is more than one potential referent during the exposure phase that children would not map the novel sign onto the novel object. In contrast, if children are processing the referential cues, then they would still show evidence of fast-mapping even when more than one potential referent is presented during exposure. To address this possibility, in Experiment 2, we investigated the effects of referential cue timing when the referential cue is necessary in order to identify the target object.

3 | EXPERIMENT 2: FAST-MAPPING NOVEL ASL SIGNS WITH REFERENTIAL AMBIGUITY

We used an identical paradigm to Experiment 1, with one exception: during exposure to the novel sign, children saw two novel objects on the screen. Thus in order to map the novel sign label onto the correct object, children had to attend to the referential cues. This modification allows us to determine what factors children consider when deciding when and where to shift gaze. We consider two possibilities: If children strategically allocate their attention based solely on moments of semantically light linguistic content and prosodic pauses, then they might not make use of the referential cue to guide their looking behavior. Such a strategy would potentially negatively impact mapping of the novel object-label pairing. In contrast, if children are strategically using the best cues that are available in the moment, as we predict, then they will make use of the signer's point to reduce referential uncertainty in a two-referent condition, and will successfully map the novel sign onto the novel object.

3.1 | Methods

3.1.1 | Participants

We recruited 35 deaf children between the ages of 18 and 71 months (M_{age} = 43.5 months) to participate in the study. Of these, two participants did not complete the eye-tracking task, and one participant was excluded for excessive trackloss. The age range of the remaining 32

children was 18–71 months (M_{age} = 44 months). There were 21 males and 11 females. Reported race was White (n = 27), or more than one race (n = 1), with four participants not reporting. Eight participants reported Hispanic/Latinx ethnicity, 22 reported not Hispanic/Latinx, and 2 did not report ethnicity. Twenty-eight children had deaf parents and four children had hearing parents. Children with deaf parents were exposed to ASL from birth, and children with hearing parents were exposed to ASL between 0 and 36 months. All children attended an early intervention program that used ASL. Seven children had previously participated in Experiment 1, with at least 6 months between testing sessions.

3.1.2 | Stimuli

The novel object signs and videos were a subset (four out of six) of those used in Experiment 1. In Experiment 2, there were two conditions: word-point and point-word.

3.1.3 | Procedure

The procedure and trial structure were identical to Experiment 1, except for the following: during exposure trials, two novel objects (a target and a distractor) were displayed on either side of the screen, and the sign video was displayed in the center. In order to accommodate two pictures, we used a 24-in. monitor. During exposure, the images and video, each of which were 400 × 400 pixels (4 in. square), were displayed on the same horizontal plane, spaced approximately 7 in. apart (Figure 5). The object pairs were consistent between exposure and test; that is, the target and distractor objects that appeared during exposure were the same pairs that appeared during test. There were two blocks each of exposure and test trials. The block order was counterbalanced such that half the children saw the point-word condition first, and the other half saw the word-point condition first.

3.2 | Results

3.2.1 | Approach to analysis

Analysis followed the same structure as Experiment 1.

3.2.2 | Trackloss

We removed any individual trials in which the child was not looking at any AOI at least 20% of the time across the trial, which resulted in exclusion of 18 exposure trials and 27 test trials. Next, we excluded any participants whose data did not include at least 25% of the experimental trials, which led to exclusion of one participant. All subsequent analyses were carried out on the remaining 238 exposure and 229 test trials from 32 participants.

FIGURE 5 Schematic of exposure trials in Experiment 2. During the preview period, the target and distractor were presented on opposite sides of the screen for 2000 ms. This was followed by the exposure video, which was presented in the center of the screen

Preview (2,000ms)   Exposure sequence (10,000ms)
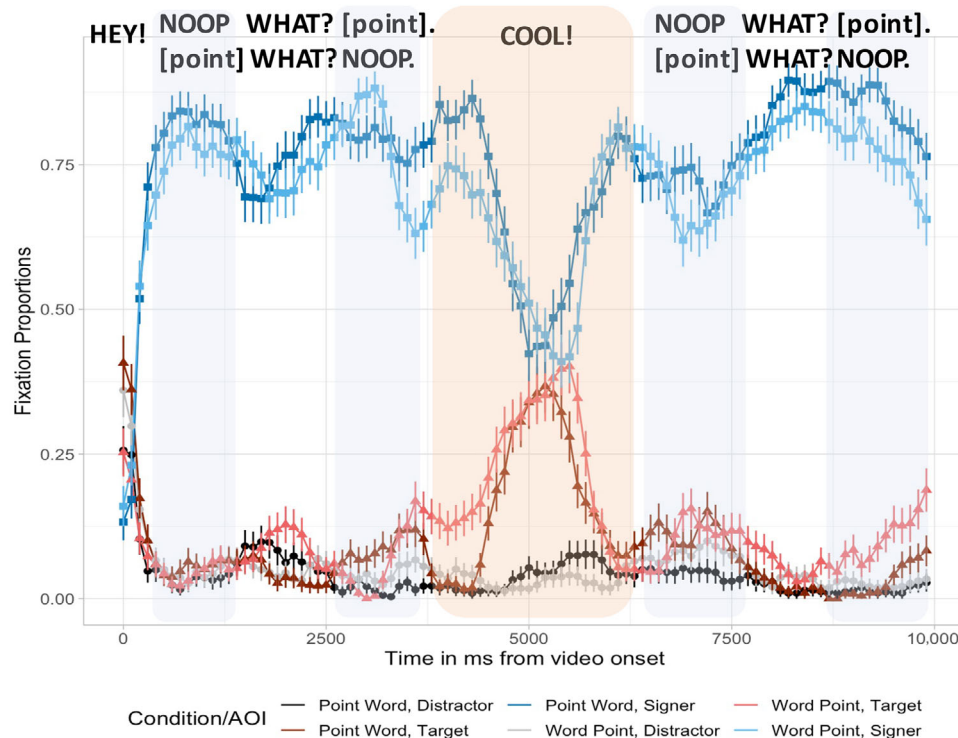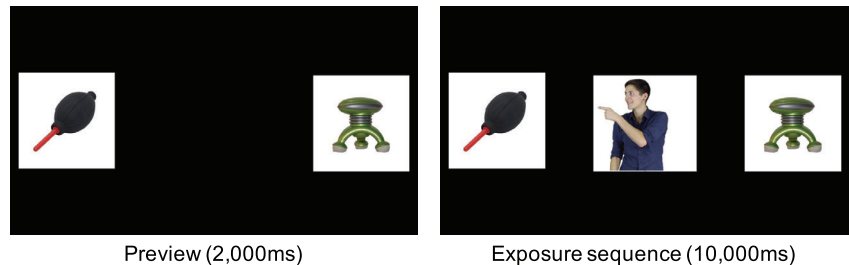
FIGURE 6 Experiment 2 timecourse of looking to sign video, target, and distractor during exposure trials, by condition. Error bars indicate ±1 standard error. Sample exposure sentences are overlayed, with shaded areas showing the analyzed sign labels (blue shading) and filler period (orange shading)

### 3.2.3 | Allocation of attention during exposure phase

Full trial

As in Experiment 1, children largely fixated on the sign video throughout the trial, with a shift to the novel object that coincided with the filler period (Figure 6). Over the full trial window, children spent an average of 72% of the time fixating the video, 10% of the time fixating the novel object, and 4% of the time fixating the distractor object [target vs. distractor: $t(31) = 5.87$, $p < 0.001$, 95% CI = (0.04, 0.08)]. Children were clearly perceiving and processing the referential cues despite the fact that they did not immediately shift their gaze to follow the cues.

Sign label

To analyze fixations to the novel signs, we fit a linear mixed-effects model of sign fixations as a log-transformed proportion of total fixations to the video, with fixed effects for condition and age and random effects for participant and item (Table 4). There were no significant effects of condition nor age.

Filler period

To analyze fixations to the target object during the filler period between labeling utterances, we fit a linear mixed-effects model of log-transformed proportion of total fixations to the target picture, with fixed effects for condition and age and random effects for participant and item (Table 5). There were no significant effects of condition, but
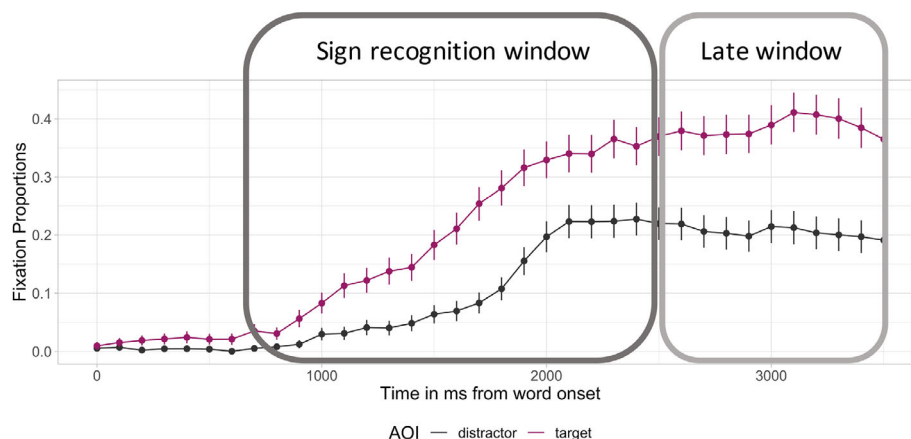
FIGURE 7 Experiment 2 timecourse of looks to target and distractor images during test trials starting at target sign onset. Boxes indicate the sign recognition window (600–2500 ms after onset) and the late window (2500–3500 ms after onset)

TABLE 4 Experiment 2 model output: Fixations (LogGaze transformed) to the video during novel sign label as a function of condition and age during exposure

	Estimate	Standard error	df	t	p
Intercept	3.46	0.53	12.67	6.49	**<0.001**
Word-point[a]	−1.08	0.59	5.39	−1.83	0.12
Age	0.04	0.03	27.04	1.64	0.11

[a]Compared to point-word condition.

TABLE 5 Experiment 2 model output: Fixations (LogGaze transformed) to target object during filler period as a function of condition and age

	Estimate	Standard error	df	t	p
Intercept	−3.8	0.3	230	−12.57	**<0.001**
Word-point[a]	0.54	0.42	230	1.29	0.2
Age	0.04	0.01	230	2.99	**0.003**

[a]Relative to point-word condition.

there were significant age effects. Older children fixated more on the novel object during the filler period than younger children.

3.2.4 | Allocation of attention during test

We plotted the timecourse of looks to the video, target, and distractor pictures starting from the onset of the target word and continuing for 3500 ms (Figure 7). The data visualization revealed that looks to the target and distractor did not diverge until almost 1000 ms after sign onset, and that looks to the target picture continued to increase throughout the trial, with greatest looks occurring between 2000 and 3500 ms following sign onset. We conducted a linear mixed-effects regression with log-transformed ratios of target to distractor looking, with fixed effects for exposure condition and child age, and random

effects for participant and item for both the sign recognition (600–2500 ms) and late (2500–3500 ms) windows. In both the sign recognition and late windows, children looked to the target more than the distractor [sign recognition: $t(31) = 4.2$, $p < 0.001$, 95% CI (0.29, Inf), $M = 0.49$, $SD = 0.66$; late: $t(31){:}3.36$, $p = 0.001$, 95% CI (0.24, Inf), $M = 0.48$, $SD = 0.81$], suggesting they had successfully mapped the novel signs. There were no significant effects of exposure condition nor age in either window.

3.2.5 | Relationship between gaze patterns at exposure and test

To determine whether children's looking patterns during exposure predicted performance at test, we used two predictors from the exposure trials: fixations to the sign video during the sign label windows, and fixations to the target picture during the filler period. We conducted a linear regression using the log-transformed proportion looking to sign during the sign window, and log-transformed proportion looking to target picture during the filler period, to predict log-transformed proportion looking time to the target picture during test trials (Table 6). Looking to the target picture during the filler period was particularly important as children had now been provided the referential information necessary to identify which of the two pictures was the target, which they did not have during the preview period. As target looking at test was significant in both the sign recognition window and the late window, we repeated the model with both outcome measures (i.e., target looking in sign recognition and late windows). In the sign recognition window, looking to the object during the filler period, but not looking to the video during the sign label, predicted proportion looking time to the target picture at test (full model $R^2(29) = 0.16$, $p = 0.033$). In the late window, the results were even more pronounced: both exposure trial looking variables were robust predictors of target recognition at test (full model $R^2(29) = 0.38$, $p < 0.001$). Children who allocated attention in a way that enabled them to perceive the target label and follow the referential cues to map the label onto the correct

TABLE 6 Experiment 2 model output: Proportion of fixations (LogGaze transformed) to the video during sign label windows and to the novel object during the filler period as predictors of target looking at test in (a) sign recognition (600–2500 ms) and (b) late (2500–3500 ms) windows

(a)				
	Estimate	SE	t	p
Intercept	−1.85	0.95	−1.95	0.06
Fixations to video during sign label	0.06	0.12	0.52	0.61
Fixations to target object during filler	0.56	0.22	2.56	**0.02**
(b)				
	Estimate	SE	t	p
Intercept	−0.09	1.13	−0.08	0.94
Fixations to video during sign label	0.38	0.15	2.58	**0.02**
Fixations to target object during filler	.82	0.26	3.15	**0.004**

object were more successful at novel word learning as evidenced by target recognition during test.

3.2.6 | Summary of Experiment 2

Children successfully mapped a novel ASL sign onto a novel object, as illustrated by increased looks to the target versus distractor pictures in the test phase. As with Experiment 1, the timing of the referential cues that accompanied the novel sign labels (point-word vs. word-point) did not have an effect on target recognition during either exposure or test. Although the timing of these gaze shifts did not differ by condition, the children were clearly attending to and interpreting the referential cue, as indicated by increased fixations to the target versus the unnamed distractor during the exposure trial. Once again, children's ability to successfully allocate and alternate attention between the sign video and target picture robustly predicted their ability to learn new signs. Specifically, in the exposure phase, children's attention to the signer during the object labels and children's attention to the target picture during the filler period each predicted overall attention to the target picture at test.

4 | DISCUSSION

There is a growing consensus that early word learning is supported by aligning synchronous cues in the visual and auditory domains (Gogate et al., 2000; Tomasello & Todd, 1983; Yu & Smith, 2012, 2013). However, it is clear that children can show typical rates of vocabulary growth even when these cues appear sequentially in the visual domain only, as is the case in deaf children learning ASL (Caselli et al., 2020; Lillo-Martin, 1999). What consequences do de-coupling synchronous social referential and linguistic cues into sequential, but temporally

adjacent events have on early language learning? We sought to gain insight into this question in two studies, where we measured how deaf children learn words by allocating visual attention between linguistic and referential cues that varied in sequential ordering. We presented children with a novel object and paired it with a novel ASL sign, to determine whether and how children would map the label to the object. Children successfully mapped the novel sign onto the novel object in both experiments, providing a first demonstration that fast-mapping can be used to map ASL signs onto their referents. Age was a significant predictor of fixation patterns during exposure but not at test.

We varied the timing of the referential cues relative to the object label during exposure. Contrary to our expectations, timing of the referential cue did not impact gaze patterns during exposure nor target fixations at test. Our results revealed a different story: regardless of when the cue appeared, children maintained fixation to the signer throughout the labeling utterance and only shifted gaze to the target object during a prosodic pause that contained minimal input. Crucially, children varied in how effectively they allocated attention during exposure to perceive the target sign, and—in Experiment 2—the target object, and these individual differences in allocation predicted performance at test. Thus, while children as a whole showed evidence for fast-mapping novel words, individual children's effective allocation of attention when exposed to new words predicted their success in mapping those words onto their referents.

4.1 | Allocation of attention to signers and objects

Children learning new words in ASL must alternate attention strategically to perceive the sign label, referential cues, and the referent object. Although linguistic input is dynamic, structured, and highly salient, children must either shift away from ongoing input or wait for pauses in the input to connect the linguistic content to the surrounding visual scene. Recent work has shown that young ASL-learning children rapidly appraise the incoming language and shift gaze away from linguistic input and towards a target precisely when they have sufficient information to identify the target (MacDonald et al., 2020). Our work adds to this growing compendium that children learning ASL are exquisitely attuned to the temporal dynamics of the communicative context, and that they can leverage this attentional flexibility to support language acquisition. Children's ability to allocate attention improved with age; we speculate that vocabulary would be an equally if not more powerful predictor of this ability.

Children used a unique strategy to decide when to shift gaze: children encoded the referential cue when it was presented, but maintained gaze on the unfolding linguistic input until there was a prosodic break in the input. Only at this point did children shift gaze from the signer to the referent object. In the case of Experiment 1, a possible explanation for the observed pattern of results is that children ignored the referential cue completely. However, this is unlikely, given the salience of these referential cues for both hearing (Baldwin, 1993; Deák et al., 2014; Houston-Price et al., 2006; Moore, 2008) and deaf

children (Brooks et al., 2020). In addition, our findings from the exposure phase of Experiment 2, in which children fixated the target more than the distractor, indicate that children did in fact encode the referential cue. Thus, unlike in spoken language, where children use visual referential cues to make an immediate gaze shift to the target (Bertenthal et al., 2014; Brooks & Meltzoff, 2005), children with exposure to ASL waited for optimal moments to shift gaze such that they did not miss any meaningful linguistic input, potentially demonstrating inhibition of gaze shift.

The fact that children successfully learned words even when the linguistic and referential cues were decoupled and presented sequentially requires an expanded understanding of the necessary conditions for language learning. There is robust evidence from spoken language that children learn words best when labels and their referents line up in time and space (Chen et al., 2020; Yu & Smith, 2013, 2017). The current results provide initial evidence that children may learn equally well when linguistic and referential cues occur in direct sequence. Children who are learning language under typical circumstances, as the majority of deaf children in the current study were, can adapt to variations in sensory experience: instead of sustained visual attention to objects, children in the current study alternated visual attention, and did so in a way that enabled them to encode and process all of the meaningful information in the input.

4.2 | Individual differences

Children's gaze patterns during exposure trials were correlated with their ability to fast-map the target item at test. In Experiment 1, children's fixation to the novel sign label within the labeling utterance predicted target looking at test. In Experiment 2, fixation to the novel sign label as well as fixation to the target picture during the filler period each predicted unique variance to target looking at test. Our findings reveal that it is not just attending to the input, but rather attending to the right type of stimulus (language vs. object) at the right moments that helps children map the referents effectively.

While previous work on deaf children's development of visual attention has typically included small participant samples (Chasin & Harris, 2008; Harris & Mohay, 1997; Lieberman et al., 2014), the larger sample here allowed us to explore individual differences in attention allocation. Our results provide initial evidence that word learning is enhanced when children strategically alternate attention between language and objects. These findings are a first step in understanding how children vary in their ability to allocate attention optimally. Our study highlights several avenues for future inquiry as to when and how language learning and attentional allocation interact in development. Children may exhibit flexible development of attention based on the specific demands of their input modality. One likely source of attention development is parental scaffolding during early interactions. In particular, deaf parents are known to structure early interactions in a way that supports children's perception of linguistic input. At the earliest ages, such scaffolding takes the form of modified signs and overt attention-getters to direct the child's attention (Lartz & Lestina, 1995;

Pizer & Meier, 2008; Spencer & Harris, 2005). However, by 18 months, parents are less likely to rely on these directive behaviors and instead wait for children's spontaneous gaze shifts to provide input (Swisher, 2000). With this scaffolding, children become adept at shifting gaze between linguistic and visual input such that they make rapid gaze shifts and respond to subtle linguistic and prosodic cues to decide when to do so (Lieberman et al., 2014; Pavani et al., 2019). We speculate that experience participating in interactions with skilled adult interlocutors, along with the child's developing skills in evaluating and responding to cues in the input, are two likely contributors to children's ability to allocate attention optimally.

4.3 | Word learning in ASL across development

While fast-mapping is a known and robust phenomenon among children learning spoken language (Bion et al., 2013; Carey & Bartlett, 1978; Halberda, 2003; Horst & Samuelson, 2008), the current study is the first to our knowledge to demonstrate that children can rapidly map and retain word meanings, irrespective of the sensory modality of linguistic input. Across both experiments, there were increased fixations to the target picture relative to the distractor picture at test, both in a sign recognition window that has been established in previous studies of sign recognition in deaf children (Lieberman & Borovsky, 2020; MacDonald et al., 2018), as well as in a later window. The late window provided more robust target fixations in Experiment 2, which likely reflects the more difficult nature of the task relative to Experiment 1. In Experiment 1, children only had to encode that an object and label were paired, whereas in Experiment 2, children saw the same object pairs during exposure and test trials, and thus had to retain not only an association between the sign and object, but the mapping of the specific sign-object pair through the use of referential cues. The pattern of fixations in Experiment 2 is also consistent with prior studies that report a more extended time course of novel (as opposed to familiar) word recognition in spoken language (Bion et al., 2013; Booth & Waxman, 2009; Borovsky et al., 2016; Houston-Price et al., 2006; Mather & Plunkett, 2010). Our findings shed light on mapping over the course of the experiment; future work should focus on retention of these associations that support robust word learning over time.

Given the broad age range of the children in the present study, it is likely that children can leverage additional word-learning strategies as they get older and gain more experience linking objects to their referents across situations. Specifically, though even very young children are capable of encoding and retrieving long term memories, these skills improve significantly across childhood (Lukowski & Bauer, 2013). These cognitive mechanisms may interact with children's performance on the task, and likely account for some of the age-related differences observed in the current study. Similarly, the child's own experiences with ASL input might influence their abilities. Deaf children, particularly those with hearing parents, have highly variable early language learning environments, as parents may be learning ASL alongside their children. Emerging evidence suggests that deaf children with hearing parents can acquire ASL at age-appropriate rates (Caselli et al., 2021),

yet it remains unknown how the processes and mechanisms underlying word learning are affected by variable exposure to non-native input. We speculate that many of the mechanisms that support word learning in spoken language would be equally supportive of word learning in ASL.

5 | CONCLUSION

Word learning is a complex phenomenon, and yet children seemingly accomplish this task with ease. Sign language input, unlike spoken language, competes with the visual input through which children encounter objects in the world. Instead of perceiving multi-modal information simultaneously, children acquiring ASL must interpret and connect sequential information to map labels onto objects. Therefore, the study of word learning processes in deaf children acquiring ASL can enrich our understanding of how all learners adapt fundamental mechanisms of word learning. The findings presented here demonstrate that children can evaluate and connect the unfolding input to the surrounding visual world by strategically and actively allocating visual attention across time and space. Simultaneity of linguistic and attentional cues is therefore not a requirement for successful word learning. Our findings suggest that recent accounts of language acquisition that prioritize temporal synchrony of linguistic and referential cues may need refinement. We find that learning may be similarly supported by sequential and temporally adjacent cues and the learner's own ability to optimize attentional flexibility given the sensory landscape of the learning environment. That is, word learning can proceed successfully even in a case where language input and the objects to which it refers are perceived sequentially, attesting to the remarkable adaptability of children's ability to divide attention in order to connect word to world.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT
The data that support the findings of this study are available upon reasonable request from the corresponding author (alieber@bu.edu).

ORCID
Amy M. Lieberman https://orcid.org/0000-0001-8217-9216

REFERENCES

Anderson, D., & Reilly, J. (2002). The MacArthur Communicative Development Inventory: Normative data for American Sign Language. *Journal of Deaf Studies and Deaf Education*, 7(2), 83–106. https://doi.org/10.1093/deafed/7.2.83

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

Baldwin, D. A. (1993). Early referential understanding: Infants' ability to recognize referential acts for what they are. *Developmental Psychology*, 29(5), 832–843. https://doi.org/10.1037/0012-1649.29.5.832

Bates, D., Maechler, M., & Bolker, B. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bertenthal, B. I., Boyer, T. W., & Harding, S. (2014). When do infants begin to follow a point? *Developmental Psychology*, 50(8), 2036–2048. https://doi.org/10.1037/a0037152

Bion, R. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word–object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, 126(1), 39–53. https://doi.org/10.1016/j.cognition.2012.08.008

Booth, A. E., McGregor, K. K., & Rohlfing, K. J. (2008). Socio-pragmatics and attention: Contributions to gesturally guided word learning in toddlers. *Language Learning and Development*, 4(3), 179–202. https://doi.org/10.1080/15475440802143091

Booth, A. E., & Waxman, S. R. (2009). A horse of a different color: Specifying with precision infants' mappings of novel nouns and adjectives. *Child Development*, 80(1), 15–22. https://doi.org/10.1111/j.1467-8624.2008.01242.x

Borovsky, A. (2020). When slowing down processing helps learning: Lexico-semantic structure supports retention, but interferes with disambiguation of novel object-label mappings. *Developmental Science*, 23(6), e12963. https://doi.org/10.1111/desc.12963

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Lexical leverage: Category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science*, 19(6), 918–932. https://doi.org/10.1111/desc.12343

Bosworth, R. G., & Stone, A. (2021). Rapid development of perceptual gaze control in hearing native signing Infants and children. *Developmental Science*, 24(4), e13086. https://doi.org/10.1111/desc.13086

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, 8(6), 535–543. https://doi.org/10.1111/j.1467-7687.2005.00445.x

Brooks, R., Singleton, J. L., & Meltzoff, A. N. (2020). Enhanced gaze-following behavior in deaf infants of deaf parents. *Developmental Science*, 23(2), e12900. https://doi.org/10.1111/desc.12900

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

Caselli, N., Pyers, J., & Lieberman, A. M. (2021). Deaf children of hearing parents have age-level vocabulary growth when exposed to ASL by six-months. *The Journal of Pediatrics*, 232, 229–236. https://doi.org/10.1016/j.jpeds.2021.01.029

Caselli, N. K., Lieberman, A. M., & Pyers, J. E. (2020). The ASL-CDI 2.0: An updated, normed adaptation of the MacArthur Bates Communicative Development Inventory for American Sign Language. *Behavior Research Methods*, 52, 2071–2084. https://doi.org/10.3758/s13428-020-01376-6

Chasin, J., & Harris, M. (2008). The development of visual attention in deaf children in relation to mother's hearing status. *Polish Psychological Bulletin*, 39(1), 1–8. https://doi.org/10.2478/v10059-008-0001-z

Chen, C., Castellanos, I., Yu, C., & Houston, D. M. (2020). What leads to coordinated attention in parent–toddler interactions? Children's hearing status matters. *Developmental Science*, 23(3), e12919. https://doi.org/10.1111/desc.12919

Deák, G. O., Krasno, A. M., Triesch, J., Lewis, J., & Sepeta, L. (2014). Watch the hands: Infants can learn to follow gaze by seeing adults manipulate objects. *Developmental Science*, 17(2), 270–281. https://doi.org/10.1111/desc.12122

Dink, J. W., & Ferguson, B. (2015). *Eye-trackingR: An R library for eye-tracking data analysis.* http://www.eyetracking-r.com/

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, 71(4), 878–894. https://doi.org/10.1111/1467-8624.00197

Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review*, 117(2), 496–516. https://doi.org/10.1037/a0019049

Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87(1), B23–B34. https://doi.org/10.1016/S0010-0277(02)00186-5

Harris, M., & Mohay, H. (1997). Learning to look in the right place: A comparison of attentional behavior in deaf children with deaf and hearing mothers. *The Journal of Deaf Studies and Deaf Education*, 2(2), 95–103. https://doi.org/10.1093/oxfordjournals.deafed.a014316

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, 48(4), 1393–1409. https://doi.org/10.3758/s13428-015-0647-3

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), 128–157. https://doi.org/10.1080/15250000701795598

Houston-Price, C., Caloghiris, Z., & Raviglione, E. (2010). Language experience shapes the development of the mutual exclusivity bias. *Infancy*, 15(2), 125–150. https://doi.org/10.1111/j.1532-7078.2009.00009.x

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(1), 27–55. https://doi.org/10.1016/j.jecp.2006.03.006

Lartz, M. N., & Lestina, L. J. (1995). Strategies deaf mothers use when reading to their young deaf or hard of hearing children. *American Annals of the Deaf*, 140(4), 358–362. https://doi.org/10.1353/aad.2012.0358

Lederberg, A. R., Prezbindowski, A. K., & Spencer, P. E. (2000). Word-learning skills of deaf preschoolers: The development of novel mapping and rapid word-learning strategies. *Child Development*, 71(6), 1571–1585. https://doi.org/10.1111/1467-8624.00249

Lederberg, A. R., & Spencer, P. E. (2009). Word-learning abilities in deaf and hard-of-hearing preschoolers: effect of lexicon size and language modality. *The Journal of Deaf Studies and Deaf Education*, 14(1), 44–62. https://doi.org/10.1093/deafed/enn021

Lieberman, A. M., & Borovsky, A. (2020). Lexical recognition in deaf children learning American Sign Language: Activation of semantic and phonological features of signs. *Language Learning*, 70, 935–973. https://doi.org/10.1111/lang.12409

Lieberman, A. M., Hatrak, M., & Mayberry, R. I. (2014). Learning to look for language: Development of joint attention in young deaf children. *Language Learning and Development*, 10(1), 19–35. https://doi.org/10.1080/15475441.2012.760381

Lillo-Martin, D. (1999). Modality effects and modularity in language acquisition: The acquisition of American Sign Language. In W. Ritchie & T. Bhatia (Eds.), *Handbook of child language acquisition* (pp. 531–567). Academic Press.

Lukowski, A. F., & Bauer, P. J. (2013). Long-term memory in infancy and early childhood. In P. J. Bauer & R. Fivush (Eds.), *The Wiley handbook on the development of children's memory* (pp. 230–254). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118597705.ch11

MacDonald, K., LaMarr, T., Corina, D., Marchman, V. A., & Fernald, A. (2018). Real-time lexical comprehension in young children learning American Sign Language. *Developmental Science*, 21(6), e12672. https://doi.org/10.1111/desc.12672

MacDonald, K., Marchman, V. A., Fernald, A., & Frank, M. C. (2020). Children flexibly seek visual information to support signed and spoken language comprehension. *Journal of Experimental Psychology: General*, 149(6), 1078–1096. https://doi.org/10.1037/xge0000702

Markus, J., Mundy, P., Morales, M., Delgado, C. E. F., & Yale, M. (2000). Individual differences in infant skills as predictors of child-caregiver joint attention and language. *Social Development*, 9(3), 302–315. https://doi.org/10.1111/1467-9507.00127

Mather, E., & Plunkett, K. (2010). Novel labels support 10-month-olds' attention to novel objects. *Journal of Experimental Child Psychology*, 105(3), 232–242. https://doi.org/10.1016/j.jecp.2009.11.004

Moore, C. (2008). The development of gaze following. *Child Development Perspectives*, 2(2), 66–70. https://doi.org/10.1111/j.1750-8606.2008.00052.x

Moore, C., & Corkum, V. (1998). Infant gaze following based on eye direction. *British Journal of Developmental Psychology*, 16(4), 495–503. https://doi.org/10.1111/j.2044-835X.1998.tb00767.x

Pavani, F., Venturini, M., Baruffaldi, F., Caselli, M. C., & van Zoest, W. (2019). Environmental learning of social cues: Evidence from enhanced gaze cueing in deaf children. *Child Development*, 90(5), 1525–1534. https://doi.org/10.1111/cdev.13284

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, 21(1), 178–185. https://doi.org/10.3758/s13423-013-0466-4

Pizer, G., & Meier, R. P. (2008). Child-directed signing in ASL and children's development of joint attention. *Proceedings of the 9th International Conference on Theoretical Issues in Sign Language Research, Florianopolis, Brazil*. Petrópolis/Rio de Janeiro, Brazil: Editora Arara Azul, 459–474.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Spencer, P. E., & Harris, M. (2005). Patterns and effects of language input to deaf infants and toddlers from deaf and hearing mothers. In B. Schick, M. Marschark, & P. E. Spencer (Eds.), *Advances in the sign-language development of deaf children* (pp. 71–101). Oxford University Press. https://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195180947.001.0001/acprof-9780195180947

Suanda, S. H., Barnhart, M., Smith, L. B., & Yu, C. (2019). The signal in the noise: The visual ecology of parents' object naming. *Infancy*, 24(3), 455–476. https://doi.org/10.1111/infa.12278

Suanda, S. H., Smith, L. B., & Yu, C. (2016). The multisensory nature of verbal discourse in parent–toddler interactions. *Developmental Neuropsychology*, 41(5–8), 324–341. https://doi.org/10.1080/87565641.2016.1256403

Swisher, M. V. (2000). Learning to converse: How deaf mothers support the development of attention and conversational skills in their young deaf children. In P. Spencer, C. J. Erting, & M. Marschark (Eds.), *The deaf child in the family and at school: Essays in honor of Kathryn P. Meadow-Orlans* (pp. 21–39). Lawrence Erlbaum Associates Publishers.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, 4(12, Pt 3), 197–211. https://doi.org/10.1177/014272378300401202

Trueswell, J. C., Lin, Y., Armstrong, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, 148, 117–135. https://doi.org/10.1016/j.cognition.2015.11.002

Wienholz, A., & Lieberman, A. M. (2019). Semantic processing of adjectives and nouns in American Sign Language: Effects of reference ambiguity and word order across development. *Journal of Cultural Cognitive Science*, 3(2), 217–234. https://doi.org/10.1007/s41809-019-00024-6

Woodward, A. L. (2003). Infants' developing understanding of the link between looker and object. *Developmental Science*, 6(3), 297–311. https://doi.org/10.1111/1467-7687.00286

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. (2013). Joint Attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. *PLoS One*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

Yu, C., & Smith, L. B. (2017). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive Science*, *41*(S1), 5–31. https://doi.org/10.1111/cogs.12366

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