Journal of Experimental Child Psychology 241 (2024) 105859

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

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

Infants' attention during cross-situational word learning: Environmental variability promotes novelty preference

Kirsty J. Dunn [a,*], Rebecca L.A. Frost [b], Padraic Monaghan [a]

[a] Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK
[b] Department of Psychology, Edge Hill University, Ormskirk, Lancashire L39 4QP, UK

ARTICLE INFO

Article history:
Received 30 August 2023
Revised 21 November 2023

Keywords:
Word learning
Cross-situational statistics
Gesture
Joint attention
Prosody
Curiosity

ABSTRACT

Infants as young as 14 months can track cross-situational statistics between sets of words and objects to acquire word–referent mappings. However, in naturalistic word learning situations, words and objects occur with a host of additional information, sometimes noisy, present in the environment. In this study, we tested the effect of this environmental variability on infants' word learning. Fourteen-month-old infants (N = 32) were given a cross-situational word learning task with additional gestural, prosodic, and distributional cues that occurred reliably or variably. In the reliable cue condition, infants were able to process this additional environmental information to learn the words, attending to the target object during test trials. But when the presence of these cues was variable, infants paid greater attention to the gestural cue during training and subsequently switched preference to attend more to novel word–object mappings rather than familiar ones at test. Environmental variation may be key to enhancing infants' exploration of new information.

* Corresponding author.
E-mail address: k.dunn@lancaster.ac.uk (K.J. Dunn).

https://doi.org/10.1016/j.jecp.2023.105859
0022-0965/© 2024 The Authors. Published by Elsevier Inc.

Introduction

Infants are born into a ‘‘blooming, buzzing confusion” (James, 1890) and must learn to make sense of the vast array of information sources around them. In word learning, this is often characterized as a hugely difficult problem, where infants are required to determine which word in multiword utterances relates to which feature of their environment (Quine, 1960). Yet, from these multiple possibilities within the language and within the environment to which language refers, infants are able to accomplish the task of learning word–object pairings with impressive speed (Bloom, 2000; Fenson et al., 1994, Swingley & Aslin, 2000). How do infants navigate this complexity?

Infants have been shown to be able to track statistical regularities effectively within a modality, such as speech (Saffran et al., 1996), visual shape stimuli (Kirkham et al., 2002), or human action sequences (Baldwin et al., 2008; Monroy et al., 2017), and these skills begin even prior to birth (Reid et al., 2017). Infants are also able to respond to regularities across modalities, which has been proposed as a solution to the problem of word learning, by tracking cross-situational correspondences between co-occurring words and objects, thereby constraining the possible mappings between words and potential referents in the environment (Siskind, 1996).

Smith and Yu (2008) tested 12- and 14-month-old infants on their ability to track these cross-situational statistical correspondences between words and objects in a computer-based task. Over a series of trials, infants were presented with two words and two objects. On an individual trial, it was not possible to determine which word corresponded with which of the two objects; however, if infants were able to track these word–object co-occurrences across multiple trials (i.e., *cross-situational learning*), the word–object mappings could be acquired. After training, infants were tested by hearing a single word and seeing two objects on-screen. Both 12- and 14-month-olds looked longer to the target object than to the distractor object, indicating that infants' sensitivity to cross-situational statistics supports word learning (see also Vlach & DeBrock, 2017; Vlach & Johnson, 2013).

However, these studies of infant cross-situational learning (Smith & Yu, 2008; Vlach & DeBrock, 2019; Vlach & Johnson, 2013) tested children's ability to track cross-situational statistics in idealized circumstances, where only words and their referents occurred in a learning situation. But natural language learning environments are substantially noisier and more variable, with multiple referents and multiple words—some of which refer to observable features in the environment and some of which do not such as function words (Monaghan & Mattock, 2012; Yu & Ballard, 2007). Yu et al. (2021) showed that infants can navigate the environment to attend to a small set of potential referents during word naming incidences (see also Yurovsky et al., 2013). However, beyond the complexity of referents, there exist multiple other sources of variability in the communicative environment, such as prosody and gesture (e.g., Monaghan, 2017) and variations in object properties (e.g., Bazhydai & Westermann, 2020), that may affect word learning.

Fortunately, this complexity and variation in the environment provides many additional potential sources of information to support word learning. Within child-directed speech utterances, distributional information about how types of words commonly appear in relation to other types of words can provide critical cues to grammatical categories of words (Mintz, 2003; Monaghan et al., 2007). This directs focus to the target referent for a word by constraining the possible referents to which the word can map (Frost et al., 2019; Monaghan & Mattock, 2012). For instance, words following ‘‘the” tend to be nouns and will relate to objects in the environment, whereas words following ‘‘you” tend to be verbs, which will relate to actions, reducing the possible pairings among words and referents. Similarly, verbal deixis can also help to direct children's attention to which word is the focus of the communicative situation (e.g., ‘‘that toy”; Cameron-Faulkner et al., 2003). A further source of information, and variation, in child-directed speech is intonation, or prosody, which can again provide constraints for word–referent mappings by highlighting the communicative focus within the caregiver's speech. Messer (1981) reported that approximately 50% of child-directed utterances with a learning goal reached peak amplitude for the referring word. Furthermore, in a task involving caregivers teaching new words to children, pitch, variation, intensity, and duration of the new word tended to increase compared with other words used by the speaker (Fernald, 1991).

There are additional non-speech cues that provide a source of information about the focus of the communicative situation such as deictic gesture, for example pointing (Cheung et al., 2021; Iverson et al., 1999), and direction of eye gaze (Farroni et al., 2004). Unlike distributional or prosodic cues, deictic gestures and eye gaze cues constrain the target referent (rather than the target word) by directing visual attention to the referent that is being described. Infants follow gaze from birth (Farroni et al., 2004), and by 4 months of age they show improved object processing when attention to that object has been directed by another (Hoehl et al., 2008). The speaker's intention in word naming by observing speaker gaze is reliably used by infants (Baldwin & Moses, 2001; Bruner, 1983; Tomasello & Todd, 1983; Woodward, 2009), and caregiver gestures are frequently used to indicate the focus for communication by clarifying the intended referent (O'Neill et al., 2005). Indeed, in a meta-analysis, Çetinçelik et al. (2021) reported consistent evidence of the facilitatory effects of eye gaze and head turn following in vocabulary development, object–word mappings, and object processing from birth to 2 years of age.

How children isolate these multiple sources of useful information and integrate them to support word learning is a matter of debate. One view is that cues are used additively. In their computational model of word–referent mapping, Yu and Ballard (2007) found that combining joint attention, prosody, and cross-situational co-occurrences resulted in better learning than using any one cue alone. In a series of studies investigating 15-month-olds' word learning, Hollich et al. (2000) found that gaze direction alone did not lead to word learning in 12-month-olds until additional salience cues, such as higher frequency and longer duration of labelling, were also included, suggesting additive effects. However, infants were able to use both perceptual salience (movement of stimuli) and gestural cues (face turn toward stimuli) independently to learn words. Furthermore, when these types of cues were added complementarily (i.e., consistent word–referent cues occurring at the same time), learning was not superior to learning in the presence of one cue type alone. Thus, there is mixed evidence for the additive benefit of cues for word learning in infants at this age. An alternative view is that covariance among cues may be critical for effective learning and that children will be sensitive to this co-occurrence. In this respect, Bahrick et al.'s (2004) intersensory redundancy hypothesis proposed that information from multiple cues indicating the same structure will increase salience of each cue by reducing the likelihood that each cue is randomly distributed.

Both the additive and intersensory redundancy views of multiple cues predict that greater reliability of cues will improve learning. An alternative perspective is that reliability of individual cues may in fact impair learning. In machine learning systems, the presence of a reliable information source can result in a brittle system that is not robust to future variation in the environment. This is because the system becomes reliant on only a subset of features of the environment, and the system becomes less sensitive to the wide variety of information sources that may also be useful for learning. "Dropout" computational systems avoid this issue by stochastically removing parts of the system responding to different regions of the environment during training so that the wider environment continues to contribute to learning, resulting in greater robustness of the system (Ren et al., 2021; Srivastava et al., 2014).

Monaghan (2017) noted that variation in infants' word learning environment has this same effect—but the dropout is in the environment rather than in the individual. Thus, distributional, prosodic, and gestural cues and statistical correspondences are individually useful in the environment, but none of them turns out to be entirely reliable. For instance, a pointing gesture does not occur with a new word label in 85% of occurrences (Iverson et al., 1999). Hence, natural communicative environments likely prevent overreliance on a single source of information by the learner. Monaghan (2017) constructed a computational model of cross-situational word learning, derived from the multimodal integration model of speech processing (Smith et al., 2017), but with input from multiple environmental information sources. The model showed that variability in multiple information sources resulted in slightly slower, but much more robust, learning than when all information sources occurred with 100% reliability. Hence, noise and variability in the environment may even be conducive to infants' word learning if the multiple information sources can be accommodated by, rather than overwhelm, the learner (Hendrickson & Perfors, 2019).

The current study

In this study, we tested the way in which multiple information sources in infants' learning environment direct the focus of the communicative situation, and we examined how variability in those information sources affects word learning from cross-situational statistics. First, we determined whether word learning by infants is robust to learning situations that contain multiple information sources beyond those that have been previously tested where typically only isolated words and their referents are present (e.g., Smith & Yu, 2008; Vlach & DeBrock, 2019). Specifically, we determined whether infants are still sensitive to cross-situational statistical correspondences between words and their referents when the language comprises multiple-word utterances containing distributional and prosodic cues. In parallel, we tested whether infants can learn correspondences between words and objects when gaze cues to the intended referent are also present. We further tested whether word learning by infants is robust to variation in those sources of information when individual cues indicate communicative focus but their presence is not 100% reliable during training. Theories that multiple cues facilitate learning via either additivity (Hollich et al., 2000; Yu & Ballard, 2007) or intersensory redundancy (Bahrick et al., 2004) would predict that variation in cues would result in poorer learning, whereas the degeneracy model (Monaghan, 2017) predicts that variability will improve learning compared with when cues are perfectly reliable due to effects of dropout on enhanced sensitivity to individual cues.

We tested 14-month-old infants with a cross-situational word learning paradigm based on Smith and Yu (2008) but adapted to include some of the variation existing in infants' word learning environments. For each learning trial, infants saw two objects and heard a sentence comprising a label for each of the objects in view, but for some trials sentences also contained words that served as complementary lexical distributional cues that helped to highlight the target word. These words did not refer to an object themselves but had a role similar to deictic pronouns in English (e.g., in English "that" precedes the noun being used to refer to a target object). The speech also varied in prosody, where the word that was the focus of the learning situation for that trial could occur with similar amplitude and pitch to the other words in the sentence or could occur with higher amplitude and greater variation in pitch. Finally, we also included an additional visual gestural cue, where a video of a face appeared between the two object referents and in some trials turned and moved its gaze to the target referent, whereas in other trials the face remained with its gaze facing directly forward. The extent to which there was variation in the presence of these cues (distributional, prosodic, and gestural) was varied across two learning conditions with 100% and 67% reliability of cues during training trials.

Method

Participants

Based on sample sizes typically reported in previous literature for cross-situational word learning tasks (e.g., Smith & Yu, 2008; Vlach & Johnson, 2013), 32 typically developing, monolingual (native English) 14-month-old infants (*M* = 403.06 days, *SD* = 17.56; 14 female) with language comprehension scores within the typical range (*M* = 110.47, *SD* = 64.92) were assigned to one of two conditions; a reliable cue condition or a variable cue condition. Ethical approval regarding the recruitment, methodology, and data handling was sought and gained from the [Lancaster] University ethics committee. All infants were from the local population, recruited from the [blinded] area through phone calls from the [Lancaster] database compiled from caregivers who expressed an interest in taking part in developmental psychological research. Data from an additional 21 infants could not be included due to poor calibration (*n* = 4), fussiness (*n* = 16), or failure to complete the experiment (*n* = 1). Crying or prolonged inattention to events defined fussy behavior.

Materials

Six bisyllabic pseudowords (*bimdah, chelad, gorshell, kerwol, makkot,* and *raken*) were taken from Monaghan and Mattock's (2012) study to be used as referring words. Two monosyllabic pseudowords (*tha* and *noo*) were also used to operate as distributional (verbal deixis) cues to the word that was the communicative focus in that learning trial. Six novel objects (shown in Fig. 1) were drawn from the Novel Object and Unusual Name (NOUN) database (Horst & Hout, 2016) as the referents for the words. The words were recorded by a female native English speaker. Two versions of the bisyllabic words were recorded: one spoken in monotone and the other spoken with emphasis (for the prosodic cues), with greater amplitude and pitch variation than the monotone version. In online Supplementary Material 1, we report acoustic parameters of the auditory stimuli.

A forward-facing head was presented in the center of the screen between the two objects. For the gesture cue, the head turned toward the object that was the communicative focus for the learning trial, with movement commencing with the onset of the speech in a trial. As in Houston-Price et al. (2006), the face initially looked forward because infants are more likely to follow the gaze of another who has first initiated eye contact (Hains & Muir, 1996). When the gesture cue was not present, the face remained looking forward.

Two unique stimuli presentation lists per condition were used to ensure that particular pairings between words and objects or use of particular words for distributional cues did not affect behavior.

Caregivers were asked to complete the UK Communicative Development Inventory (UK-CDI; Alcock et al., 2017) as a measure of infant language comprehension and production ability at the time of testing.

Procedure

Infants were seated on their caregiver's lap 50 to 70 cm from the screen (21.5 inches, 1920 × 1080 pixels). A Tobii x50 eye tracker positioned below the screen recorded infants' gaze patterns during the experiment, which began upon successful completion of a 5-point calibration.

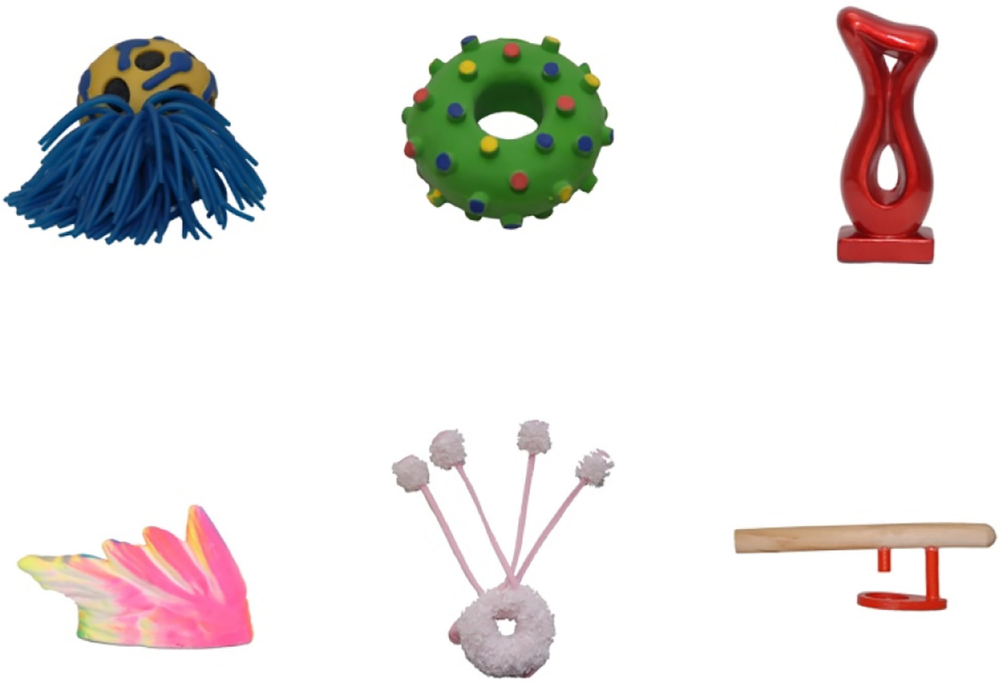

Fig. 1. The six novel objects selected from Horst and Hout's (2016) NOUN (Novel Object and Unusual Name) database.

Training trials

For each training trial, participants viewed two objects appearing on either side of a face that was centered on the screen. Participants heard a sentence including two words that individually reliably occurred with one of the two objects in view. The order of words and the position of objects were counterbalanced.

When the distributional focus cue was present, one of the monosyllabic words (e.g., noo) succeeded the word that was the focus of the communicative situation, and the other monosyllabic word (e.g., tha) succeeded the non-focused word. When the prosodic cue was present, the word that was the focus of the communicative situation was the emphasized version and the non-focus word (and monosyllabic deictic words if present) were monotone versions. When the gesture cue for focus of the communicative target was present, the face moved to look toward the object that was the target of the focus.

For the reliable cue condition, in all trials the distributional cue, prosodic cue, and gesture cue were present. For the variable cue condition, each cue was present for 67% of the trials. For the variable cue condition, cues were selected individually, such that there were trials in the variable cue condition where there were no cues (11% of trials), one cue (22% of trials), two cues (44% of trials), and three cues (22% of trials) present. An example of a trial containing all cues is given in Fig. 2 for the sentence "Chelad noo GORSHELL tha." The two words and two objects for each trial were balanced, such that each object and each word occurred with every other word and object at least once across the training trials. Hence, the conditional probability for target word–object pairs [note that p(object|word) = p(word|object)] was.50 and for non-target pairs was.20.

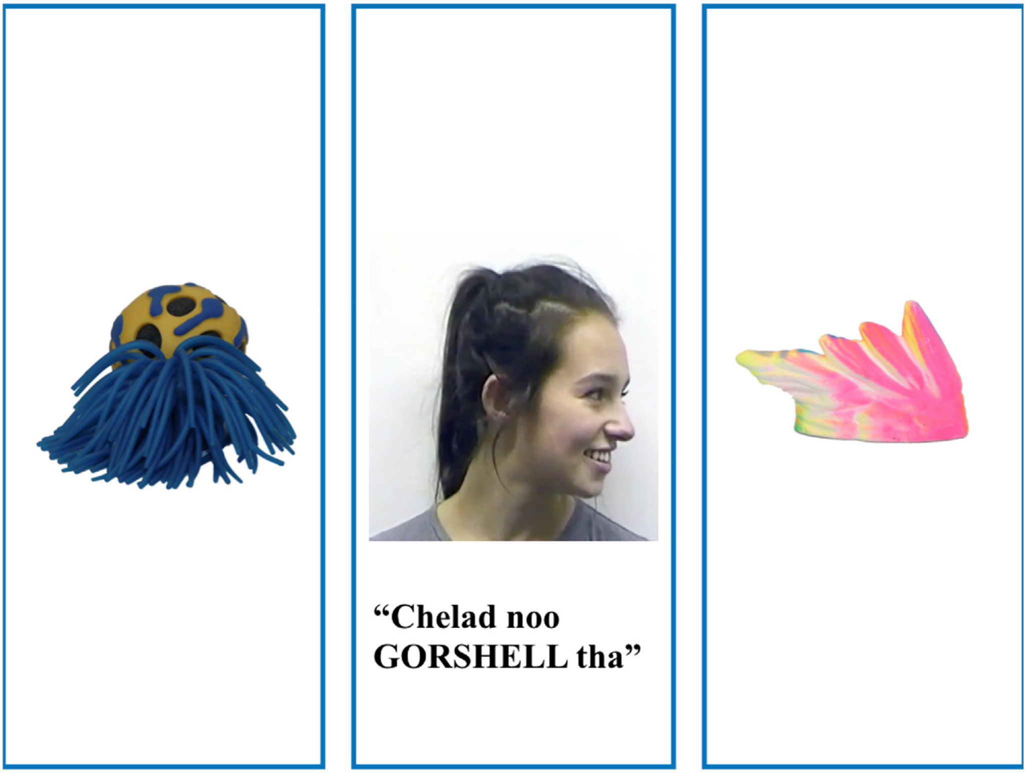

Fig. 2. Example of a training trial containing distributional, prosodic, and gesture cues. The use of "tha" (distributional deictic focus word), exaggerated prosody (indicated by capital letters), and head turn (gesture) can be used to identify the communicative focus of the learning trial in terms of the target word and object. Areas of Interest are outlined in blue rectangles.

A total of 17 attention-getter trials (a colorful rattle leaning from left to right with a series of musical tones) were interspersed throughout in order to orient and maintain infants' attention to the screen. These preceded the first 4 trials and were then interspersed every 4 trials as per Smith and Yu (2008). There were 30 training trials altogether, such that each word–object target focus pair occurred five times and each trial ended after 5 s.

Test trials

After training, infants were given 12 test trials where no cues were available to guide them to the word–referent mappings. For each test trial, infants heard one word repeated four times, with the target object and another distractor object appearing on the screen. Each of the six objects occurred twice as the target and twice as the distractor. Each trial ended after 7 s.

Open practices

All data, materials, and RStudio analysis code have been made publicly available via the Open Science Framework (https://osf.io/6v7y2/?view_only=15ca68e0c2a04dc29f3ed07f3c01604a).

Results

Areas of interest (AOIs; outlined in blue rectangles in Figs. 2 and 3) were generated around the target focus object, the non-focused object, and the face (for the training trials), and infants' looking time

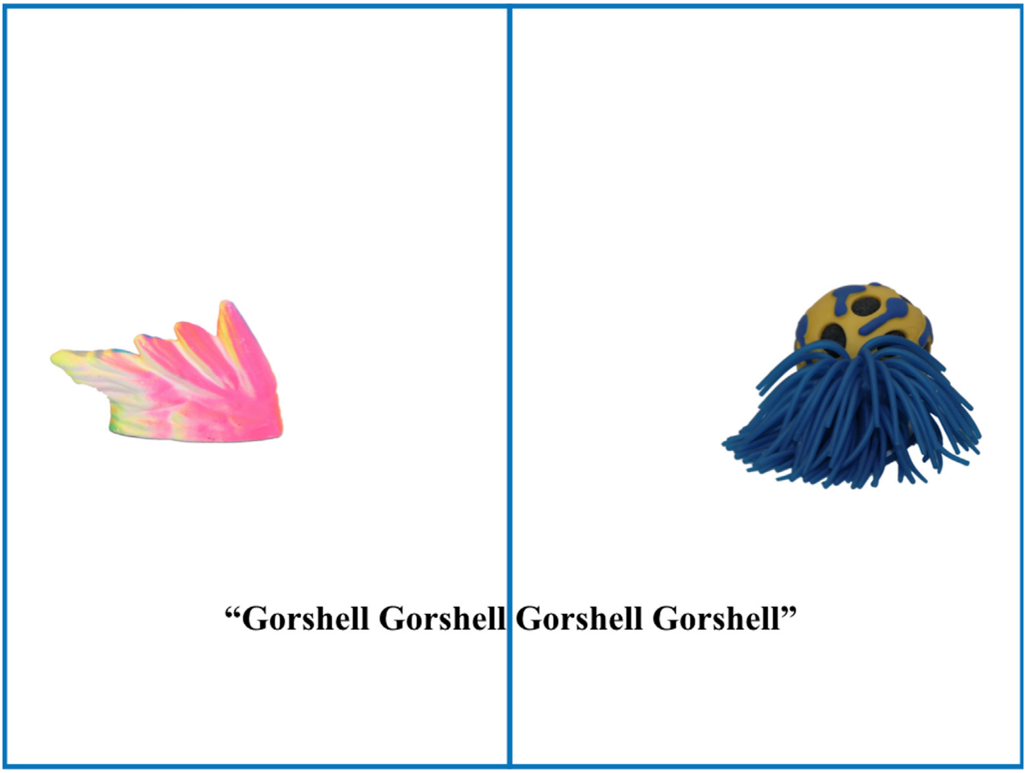

Fig. 3. Example of a test trial containing target object, distractor object, and four repetitions of the target word. Areas of interest are outlined in blue rectangles.

to each of these areas was the dependent variable in the analyses. We analyzed the data using linear mixed-effects models, with participant and target focus object as random effects (including the non-focused object as well resulted in a singular fit, so this was omitted from the random-effects structure). We tested fixed effects using an incremental approach (Baayen et al., 2008) by adding each fixed effect and interaction to determine whether it significantly improved model fit using likelihood ratio test comparisons.

Fixed effects were trial number (to determine whether infants' looking time changed over the course of the study), the cue condition (reliable or variable), the AOI (target focus object, non-focus object, or face [for the training trials]), and the interaction between cue condition and the AOI.

We first analyzed training trials to determine how infants' looking behavior changed over the course of the study (i.e., with exposure to the cues), and then we analyzed testing trials to determine how infants responded to the cross-situational statistics and whether cue variability influenced learning, enabling a comparison with previous looking time studies of infants' cross-situational learning without additional cues (e.g., Smith & Yu, 2008).

Training trials

For the training trials, 144 of 1501 observations (9.59 %) were identified as outliers across the dataset using Tukey's (1977) method, defined as observations outside 1.5 times the interquartile range.

The starting model included only random effects, and then we added trial number (1–30) as a fixed effect to determine whether overall looking time changed during training. There was no significant effect of trial number on looking time, $\chi^2(1) = 2.34$, $p = .127$.

We next added cue condition (reliable or variable) as a fixed effect to determine whether overall looking time differed across conditions. There was no significant effect of condition, $\chi^2(1) = 1.64$, $p = .200$. Furthermore, there was no significant interaction between trial number and condition, $\chi^2(1) = 0.45$, $p = .505$. Thus, there was no observable difference in looking time changes over training between the conditions.

Adding AOI (target, distractor, or face) resulted in a significant improvement in fit, $\chi^2(2) = 155.22$, $p < .001$. Infants spent more time looking at the face (M = 0.48 s, SD = 0.25) than at the target focus object (M = 0.33 s, SD = 0.20) and non-focus object (M = 0.31 s, SD = 0.20), both ps <.001. Critically, adding the interaction between reliable or variable cue condition and AOI to a model containing main effects of condition and AOI resulted in a significant improvement in fit, $\chi^2(2) = 71.89$, $p < .001$ (see Table 1 for estimates in the final model). This indicated that the cue condition affected infants' looking during cross-situational learning (Fig. 4).

Table 1
Summary of the linear mixed-effects model on looking times during training trials.

Fixed effects	Estimated coefficient	*SE*	Confidence intervals		*t*	*p*
			2.5 %	97.5 %		
(Intercept)	.25	.02	.21	.29	13.20	<.001*
Area of the screen						
Non-focus object vs. face	.28	.02	.24	.32	13.06	<.001*
Non-focus object vs. target (Focus) Object	.03	.02	−.02	.08	1.20	.229
Cue condition	.10	.02	.05	.14	4.19	<.001*
Area of the Screen * Cue Condition						
Face * Cue Condition	−.22	.02	−.27	−.16	−7.40	<.001*
Target (Focus) Object * Cue Condition	−.01	.03	−.07	.05	−0.29	.774

Random effects	Variance	*SD*
Participant (intercept)	.00	.02
Target word (intercept)	.00	.01

Note. N = 1357 observations and 32 participants. R syntax for the final model: lmer(OBS – (1|PPT) + (1|TargetWord) + Object*Condition_Overall, data = multiplecueskids_training_ready).

* $p < .05$.

Further analysis of the interaction revealed that there was significantly more looking toward the target focus object in the 100 % cue condition (*M* = 0.37 s, *SD* = 0.21) than in the 67 % variable cue condition (*M* = 0.28 s, *SD* = 0.17), *z* = 7.24, *p* <.001. There was no significant difference in looking to the non-focus object across conditions, *z* = 0.77, *p* =.956. There was significantly more looking toward the face in the 67 % variable cue condition (*M* = 0.53 s, *SD* = 0.25) than in the 100 % cue condition (*M* = 0.41 s, *SD* = 0.24), *z* = 3.05, *p* =.021. In the 67 % variable cue condition, there was more looking to the face than to the target, *z* = 5.67, or the distractor, *z* = 5.43, both *ps* <.001, but there was also more looking to the target focus object than to the non-focus object, *z* = 18.96, *p* <.001. In the 100 % cue condition, there was more looking to the face than to the non-focus object, z = 9.27, *p* <.001, but not to the target focus object, *z* = 2.17, *p* =.200, and there were more looks to the target focus object than to the non-focus object, *z* = 6.51, *p* <.001, indicating that for both cue conditions infants had learned to look more toward the focused object. Indeed, repeating the analysis but omitting looks to the face indicated no significant interaction between looking to the target focus object or the non-focus object and cue condition, $\chi^2(2)$ = 0.02, *p* =.883.

To investigate whether individual cues were having different effects on learning in the 67 % variable cue condition, we investigated cue presence or absence for each of the distributional, prosodic, and gesture cue conditions. There were no significant improvements in model fit when including the presence of any cues [distributional cue: $\chi^2(1)$ = 2.00, *p* =.157; prosodic cue: $\chi^2(1)$ = 1.03, *p* =.309; gesture cue: $\chi^2(1)$ = 1.39, *p* =.238], and the overall effect of number of cues present (from zero to three) also did not improve model fit, $\chi^2(1)$ = 0.78, *p* =.376. Interactions between the presence of individual cues and looking to different areas on the screen were also not significant [distributional cue: $\chi^2(2)$ = 1.61, *p* =.447; prosodic cue: $\chi^2(2)$ = 5.98, *p* =.050; gesture cue: $\chi^2(1)$ = 0.82, *p* =.665, although model fits were singular for these models], indicating that behavior was likely to be consistent across trials that varied the presence or absence of individual cues.

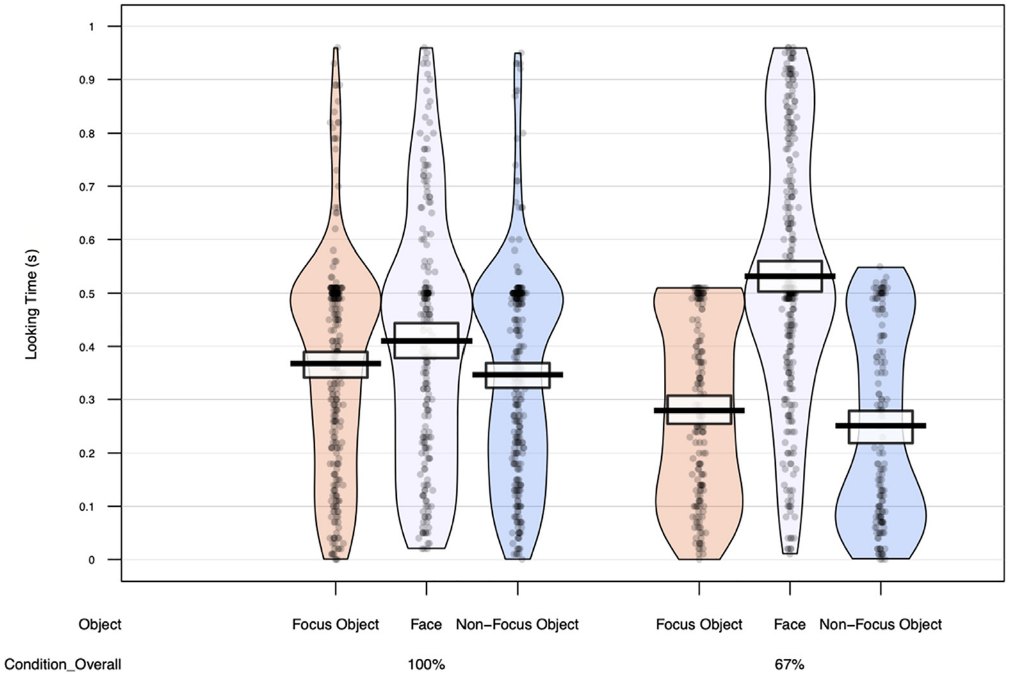

Fig. 4. Pirate plot depicting mean looking time to target focus object, non-focus object, and face on training trials. Black lines represent the means, shaded boxes represent standard errors, and individual data points are plotted by gray dots.

Finally, we checked whether the position of the word that was the communicative focus (i.e., whether it occurred as the first or second word in the sentence) affected behavior across both cue conditions. There was no significant improvement in fit, $\chi^2(1) = 0.04$, p =.837. We also checked whether the different versions of the word -object mappings resulted in different behavior, but this also did not significantly improve model fit, $\chi^2(1) = 0.24$, p =.621, indicating that the precise pairings used in the study did not affect behavior.

In Supplementary Material 2, we also tested the interaction between reliable or variable cue condition and area of the screen for trials where the face was animated (thereby omitting the variable trials when the face remained immobile). The results were similar.

Test trials

For the test trials, 1 of 705 observations (0.14%) was identified as an outlier using the same criteria as used for the training trials, and it was omitted from further analyses.

The analysis was conducted in a similar way to the training trials, constructing a model with only random effects (participant and target object as random effects) and then testing the improvement in model fit by adding in each fixed effect. We first tested the effect of trial order to see whether looking times changed across testing and found no significant improvement in model fit, $\chi^2(1) = 3.06$, p =.080.

We next added reliable or variable condition as a fixed effect to identify whether there was a difference in overall looking across conditions, but there was no significant improvement in model fit, $\chi^2(1) = 0.89$, p =.344. Adding area of the screen (target or distractor object) also resulted in no significant improvement in model fit, $\chi^2(1) = 0.40$, p =.526. Critically, the interaction between cue condition and area of the screen resulted in a significant improvement in model fit, $\chi^2(1) = 4.71$, p =.030 (see Table 2).

Further analysis of the interaction revealed that there was no significant difference in looking toward the target across conditions, $z = 1.86$, p =.222. However, there was significantly longer looking toward the distractor item in the variable cue condition ($M = 2.80$, $SD = 1.82$) than in the reliable cue condition ($M = 2.34$, $SD = 1.51$), $z = 8.91$, p <.001. Infants in the 100 % reliable cue condition looked marginally longer toward the target item ($M = 2.70$, $SD = 1.62$) than toward the distractor item ($M = 2.34$, $SD = 1.51$), z = 2.35, p =.077, which partially replicated Smith and Yu's [2008] finding for longer looking times to the target than to the distractor objects. However, infants in the 67 % variable cue condition showed the reverse effect, looking significantly longer toward the distractor item ($M = 2.80$, $SD = 1.82$) than toward the target item ($M = 2.61$, $SD = 1.70$), $z = 10.01$, p <.001, demonstrating that the variability in cues reversed the preference away from looking at the co-occurring words and objects (which were preferred during the training trials) (Fig. 5).

As anticipated, we found no significant improvement in fit when adding whether the target object was on the left or right of the screen, $\chi^2(1) = 2.44$, p =.118, or when the different versions of the stimuli lists were compared, $\chi^2(3) = 1.53$, p =.676. In Supplementary Material 3, we report an additional

Table 2
Summary of the linear mixed-effects model on looking times during training trials.

Fixed effects	Estimated coefficient	*SE*	Confidence intervals		*t*	*p*
			2.5 %	97.5 %		
(Intercept)	2.77	0.17	2.44	3.10	16.52	<.001*
Cue condition	−0.44	0.22	−0.88	−0.01	−1.94	.053
Area of screen	−0.19	0.17	−0.53	0.15	−1.10	.197
Cue Condition * Area of Screen	0.53	0.24	0.05	1.01	2.17	.030*
Random effects			Variance			*SD*
Participant (intercept)			.17			.41
Target object (intercept)			.02			.13

Note. N = 704 observations and 32 participants. R syntax for the final model: lmer(OBS ~ (1|PPT) + (1|TargetWord) + Condition_Overall*Object, data = multiplecueskids_overall).

* p <.05.

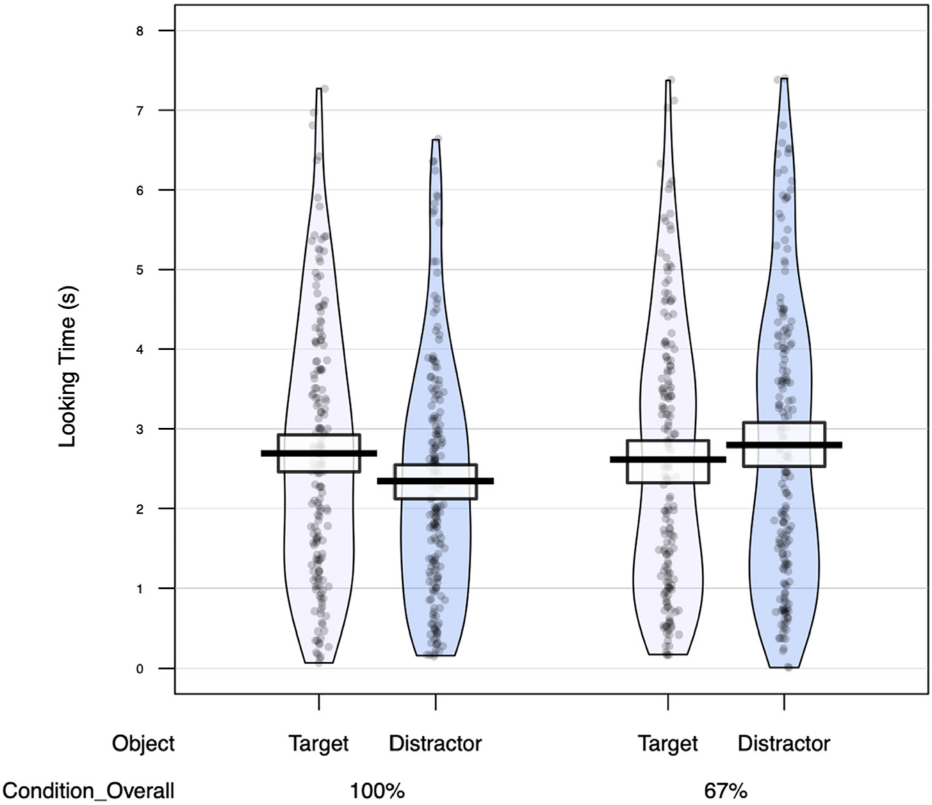

Fig. 5. Pirate plot depicting mean looking time to target and distractor items per condition on test trials. Black lines represent the means, shaded boxes represent standard errors, and individual data points (for each participant for each trial) are plotted in gray dots.

exploratory analysis where we tested the effect of infants' language development on their performance. Adding language comprehension group (high and low groups based on a median split) resulted in a significant improvement in model fit, $\chi^2(1) = 9.40$, p =.002 (see Table 2). Adding language comprehension group as an interaction did not significantly improve model fit, $\chi^2(1) = 3.29$, p =.35. Overall, infants in the high comprehension group looked longer to the screen (M = 2.85, SD = 1.74) than those in the low language comprehension group (M = 2.35, SD = 1.55).

Discussion

Children's language learning takes place in a rich and variable environment in the presence of multiple sources of information—each occurring with a changeable degree of reliability. Yet, from a young age, children can skillfully navigate the multiple possibilities for relating words to potential referents in order to learn new word meanings (Bhat et al., 2022). Children have been shown to use cross-situational statistics to learn mappings between words and referents by tracking their co-occurrences over multiple situations under circumstances where only words and their referents occur (Smith & Yu, 2008; Vlach & DeBrock, 2017; Vlach & Johnson, 2013). However, cross-situational instances in natural learning environments are replete with other variable sources of information that may support children in determining which word and which aspect of the environment are the focus of the communicative situation (Monaghan, 2017). In this study, we tested how multiple additional environmental cues, indicating the focus of the word in speech and the focus of the object in the visual environment in a communicative situation, may affect infants' learning of word–referent mappings.

Distributional, prosodic and gestural cues all were presented on 100% of cross-situational training trials for half of the infants (reliable condition). The remaining half received a spread of cues that were each present on 67% of training trials (variable condition). Whether these cues were variable or not had an influence on infants' looking behavior during exposure to uncertain word–referent mappings. Infants in the variable cue condition looked significantly longer to the gestural cue—the face—that indicated the focus for that trial than those in the reliable cue condition. Thus, if a head-turn gesture occurs all the time, infants attend to this cue less than if that cue occurs with less reliability; variability increases attention to the cue. This result, in this context of language learning, is consistent with studies of visual attention showing that variable features of stimuli are attended to more than features that occur reliably in adult studies (Garner, 1978) and showing longer looking times to stimuli that vary from a standard in infants (Fagan, 1990).

After exposure to these cross-situational statistics, we tested whether infants responded to the trained word–referent mappings. There was a significant effect of whether cues were reliable or variable on infants' behavior in this test. Infants in the reliable condition looked marginally longer to target items than to distractor objects. This effect did not interact with infant language ability, with longer looking overall in those with higher parental-reported comprehension scores. This was a result consistent with the significant effect of longer looking to target objects over distractor objects in Smith and Yu (2008). In their study, without additional cues to communicative focus, infants demonstrated that they had computed the cross-situational statistics by responding with longer looks to mappings that occurred with higher frequency over multiple trials. In our study, when adding numerous additional sources of information into the learning situations—distributional cues in speech in multiword sentences, prosodic cues in terms of variation in pitch and amplitude of words in those sentences, and a gesture cue in terms of a moving face—resulted in similar behavior.

In contrast, as reflected in the interaction between cue condition and the areas of the screen that the infants looked at, the infants who had been exposed to variable cues behaved very differently. When the additional sources of information were variable during the exposure trials, during testing (when these cues were absent) infants demonstrated a preference for looking to the object that did not co-occur frequently with the word that they were hearing. Hence, variable cues increased infants' interest in pairings between words and objects that only occasionally occurred together.

This increased attention to stimuli that are novel has been described in terms of "curiosity" (Berlyne, 1954; Chen et al., 2022; Loewenstein, 1994) and has been suggested to be advantageous for learning. For instance, Twomey et al., 2018 found that children who were exposed to word–object pairings against a background that varied in color retained those pairings better than when the background remained the same color. In our study, it was the variability in the cues that were indicating the focus of the scene that resulted in enhanced learning, showing that curiosity is enhanves through variation in communicative cues.

How do these results relate to models of multiple cue integration in language learning? The intersensory redundancy hypothesis (Bahrick et al., 2007) predicted that performance would be better when cues are reliable, with enhanced attention to individual cues when they reliably co-occur. In the current study, this could have been indicated by greater differences in looks to the target over the distractor object during testing. We did not find evidence for this difference. However, we did find that looking behavior was qualitatively different rather than quantitatively different during testing. There was evidence to indicate that variability enhances attention to a greater extent than reliable redundant cues by increasing children's looks to the gesture cue (when this was variable) during exposure. Similarly, proposals that multiple cues have an additive effect (Hollich et al., 2000; Yu & Ballard, 2007) predict that performance would be better for the reliable cue condition but cannot explain different looking behavior during testing.

The degeneracy model (Monaghan, 2017) predicted enhanced use of individual cues when they are variable, and this was observed in the looking behavior during exposure—more looks to the variable gesture cue when it indicated the communicative focus 67% of the time compared with 100% of the time. However, as noted above, there was no evidence of better learning in the variable cue condition as a consequence of the learner remaining sensitive to multiple cues from the linguistic and visual environment, and so each of these theories of multiple cue use need to take into account changes in explorative behavior as a consequence of cue variability, consistent with the curiosity literature, rather than focusing on quantitative changes in learning.

The results indicate that the variability in caregivers' production of cues that provide communicative focus may result in enhanced attention to the individual cues and results in infants attending more to novel stimuli than to familiar stimuli in their environment. However, different environmental constraints may affect the way in which cues are produced and used in order to support focus on word–referent mappings during language learning. For instance, Cheung et al. (2021) examined caregivers' use of gestures in a word learning situation when there was no referential uncertainty (just one object in view) compared with when there was referential uncertainty (two or six objects in view). They found that caregivers used gestures more reliably when there was uncertainty about the referent for the word, and the degeneracy model predicted that gestures are more useful under these same conditions. Although the use of gesture was still variable—gestures to the focus of the communicative situation occurred only a proportion of the time when the word was uttered—this indicates that the optimal variability of communicative cues is likely to be a complex function of environmental constraints as well as attentional constraints relating to enhanced attention to variable—but useful—information.

Conclusion

We found that infants are able to learn cross-situational correspondences between words and objects and are sensitive to multiple environmental cues to shape their learning. These cues are varied and noisy across communicative situations for infants, and such variability may boost infants' interest in, and search for, novel information in learning language.

CRediT authorship contribution statement

Kirsty J. Dunn: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Visualization, Writing – original draft, Writing – review & editing. **Rebecca L.A. Frost:** Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Padraic Monaghan:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Data availability

An anonymised link is provided to the raw, anonymised data.

Acknowledgment

This work was supported by the International Centre for Language and Communicative Development (LuCiD) at Lancaster University and the University of Liverpool, funded by the Economic and Social Research Council (UK) [ES/S007113/1].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jecp.2024.105865.

References

Alcock, K. J., Meints, K., & Rowland, C. F. (2017). *UK-CDI Words and Gestures—Preliminary norms and manual.* Retrieved from http://lucid.ac.uk/ukcdi.

Baayen, R. H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Bahrick, L. E., Lickliter, R., & Flom, R. (2004). Intersensory redundancy guides the development of selective attention, perception, and cognition in infancy. *Current Directions in Psychological Science, 13*(3), 99–102. https://doi.org/10.1111/j.0963-7214.2004.00283.x.

Baldwin, D., Andersson, A., Saffran, J., & Meyer, M. (2008). Segmenting dynamic action via statistical structure. *Cognition, 106*(3), 1382–1407.

Baldwin, D. A., & Moses, J. A. (2001). Links between social understanding and early word learning: Challenges to current accounts. *Social Development, 10*, 311–329. https://doi.org/10.1111/1467-9507.00168.

Bazhydai, M., & Westermann, G. (2020). From curiosity, to wonder, to creativity: A cognitive developmental psychology perspective. In A. Schinkel (Ed.), *Wonder, education, and human flourishing.* VU University Press.

Berlyne, D. E. (1954). A theory of human curiosity. *British Journal of Psychology, 45*, 180–191.

Bhat, A., Spencer, J. P., & Samuelson, L. K. (2022). Word-Object Learning via Visual Exploration in Space (WOLVES): A neural process model of cross-situational word learning. *Psychological Review, 129*(4), 640–695. https://doi.org/10.1037/rev0000313.

Bloom, L. (2000). Pushing the limits on theories of word learning. *Monographs of the Society for Research in Child Development, 65* (3), 124–135. https://doi.org/10.1111/1540-5834.00100.

Bruner, J. (1983). Play, thought, and language. *Peabody Journal of Education, 60*(3), 60–69. https://doi.org/10.1080/01619568309538407.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science, 27*, 843–873. https://doi.org/10.1016/j.cogsci.2003.06.001.

Çetinçelik, M., Rowland, C. F., & Snijders, T. M. (2021). Do the eyes have it? A systematic review on the role of eye gaze in infant language development. *Frontiers in Psychology, 11*, 589096. https://doi.org/10.3389/fpsyg.2020.589096.

Chen, X., Twomey, K. E., & Westermann, G. (2022). Curiosity enhances incidental object encoding in 8-month-old infants. *Journal of Experimental Child Psychology, 223*, 105508. https://doi.org/10.1016/j.jecp.2022.105508.

Cheung, R. W., Hartley, C., & Monaghan, P. (2021). Caregivers use gesture contingently to support word learning. *Developmental Science, 24*(4), e13098. https://doi.org/10.1111/desc.13098.

Fagan, J. F. (1990). The paired-comparison paradigm and infant intelligence. In A. Diamond (Ed.), *Annals of the New York Academy of Sciences,* Vol. 68: *The development and neural bases of higher cognitive functions* (pp. 337–364). New York Academy of Sciences.

Farroni, T., Massaccesi, S., Pividori, D., & Johnson, M. H. (2004). Gaze following in newborns. *Infancy, 5*(1), 39–60. https://doi.org/10.1207/s15327078in0501_2.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5, Serial No. 242). https://doi.org/10.2307/1166093.

Fernald, A. (1991). Prosody in speech to children: Prelinguistic and linguistic functions. *Annals of Child Development, 8*, 43–80.

Frost, R. L., Monaghan, P., & Christiansen, M. H. (2019). Mark my words: High frequency marker words impact early stages of language learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(10), 1883–1898. https://doi.org/10.1037/xlm0000683.

Garner, W. R. (1978). Selective attention to attributes and to stimuli. *Journal of Experimental Psychology: General, 107*(3), 287–308. https://doi.org/10.1037/0096-3445.107.3.287.

Hains, S. M., & Muir, D. W. (1996). Infant sensitivity to adult eye direction. *Child Development, 67*(5), 1940–1951. https://doi.org/10.1111/j.1467-8624.1996.tb01836.x.

Hendrickson, A. T., & Perfors, A. (2019). Cross-situational learning in a Zipfian environment. *Cognition, 189*, 11–22. https://doi.org/10.1016/j.cognition.2019.03.005.

Hoehl, S., Reid, V., Mooney, J., & Striano, T. (2008). What are you looking at? Infants' neural processing of an adult's object-directed eye gaze. *Developmental Science, 11*(1), 10–16. https://doi.org/10.1111/j.1467-7687.2007.00643.x.

Hollich, G., Hirsh-Pasek, K., & Golinkoff, R. M. (2000). I. What does it take to learn a word? *Monographs of the Society for Research in Child Development, 65*(3), 1–16. https://doi.org/10.1111/1540-5834.00091.

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.

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.

Iverson, J. M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in mother–child interactions. *Cognitive Development, 14*, 57–75. https://doi.org/10.1016/S0885-2014(99)80018-5.

James, W. (1890). *The principles of psychology*. Henry Holt.

Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*(2), B35–B42. https://doi.org/10.1016/s0010-0277(02)00004-5.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*, 75–98. https://doi.org/10.1037/0033-2909.116.1.75.

Messer, D. J. (1981). The identification of names in maternal speech to infants. *Journal of Psycholinguistic Research, 10*(1), 69–77. https://doi.org/10.1007/BF01067362.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*(1), 91–117. https://doi.org/10.1016/s0010-0277(03)00140-9.

Monaghan, P. (2017). Canalization of language structure from environmental constraints: A computational model of word learning from multiple cues. *Topics in Cognitive Science, 9*(1), 21–34. https://doi.org/10.1111/tops.12239.

Monaghan, P., Christiansen, M. H., & Chater, N. (2007). The phonological distributional coherence hypothesis: Cross-linguistic evidence in language acquisition. *Cognitive Psychology, 55*, 259–305. https://doi.org/10.1016/j.cogpsych.2006.12.001.

Monaghan, P., & Mattock, K. (2012). Integrating constraints for learning word–referent mappings. *Cognition, 123*, 133–143. https://doi.org/10.1016/j.cognition.2011.12.010.

Monroy, C. D., Gerson, S. A., & Hunnius, S. (2017). Toddlers' action prediction: Statistical learning of continuous action sequences. *Journal of Experimental Child Psychology, 157*, 14–28. https://doi.org/10.1016/j.jecp.2016.12.004.

O'Neill, M., Bard, K. A., Linnell, M., & Fluck, M. (2005). Maternal gestures with 20-month-old infants in two contexts. *Developmental Science, 8*(4), 352–359. https://doi.org/10.1111/j.1467-7687.2005.00423.x.

Quine, W. V. O. (1960). *Word and object*. MIT Press.

Reid, V. M., Dunn, K., Young, R. J., Amu, J., Donovan, T., & Reissland, N. (2017). The human fetus preferentially engages with face-like visual stimuli. *Current Biology, 27*(12), 1825–1828. https://doi.org/10.1016/j.cub.2017.05.044.

Ren, P., Xiao, Y., Chang, X., Huang, P. Y., Li, Z., Gupta, B. B., Chen, X., & Wang, X. (2021). A survey of deep active learning. *ACM Computing Surveys (CSUR), 54*(9), 1–40. https://doi.org/10.1145/3472291.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*, 39–61. https://doi.org/10.1016/s0010-0277(96)00728-7.

Smith, A. C., Monaghan, P., & Huettig, F. (2017). The multimodal nature of spoken word processing in the visual world: Testing the predictions of alternative models of multimodal integration. *Journal of Memory and Language., 93*, 276–303. https://doi.org/10.1016/j.jml.2016.08.005.

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

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research, 15*(1), 1929–1958 https://www.cs.toronto.edu/~rsalakhu/papers/srivastava14a.pdf.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*(2), 147–166. https://doi.org/10.1016/S0010-0277(00)00081-0.

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

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Twomey, K. E., Ma, L., & Westermann, G. (2018). All the right noises: Background variability helps early word learning. *Cognitive Science, 42*, 413–438. https://doi.org/10.1111/cogs.12539.

Vlach, H. A., & DeBrock, C. A. (2017). Remember dax? Relations between children's cross-situational word learning, memory, and language abilities. *Journal of Memory and Language, 93*, 217–230. https://doi.org/10.1016/j.jml.2016.10.001.

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

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition, 127*(3), 375–382. https://doi.org/10.1016/j.cognition.2013.02.015.

Woodward, A. L. (2009). Infants' grasp of others' intentions. *Current Directions in Psychological Science, 18*(1), 53–57. https://doi.org/10.1111/j.1467-8721.2009.01605.x.

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., Zhang, Y., Slone, L. K., & Smith, L. B. (2021). The infant's view redefines the problem of referential uncertainty in early word learning. In *Proceedings of the National Academy of Sciences of the United States of America, 118*(52), Article e2107019118. https://doi.org/10.1073/pnas.2107019118.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science, 16*, 959–966. https://doi.org/10.1111/cogs.12035.