

# Children integrate social and statistical information during language processing

Anonymous CogSci submission

## Abstract

How do children learn words from input that has many possible word-object mappings? Statistical learning allows children to aggregate consistent word-object co-occurrences to reduce uncertainty over time, while social-pragmatic cues can constrain ambiguity within a labeling event. Here, we present two eye-tracking studies asking how children integrate statistical and social information during real-time language processing. When processing familiar words, children and adults did not delay their gaze shifts to seek a post-nominal social cue to reference (eye gaze). When processing novel words, however, children and adults fixated longer on a speaker who produced a gaze cue, which, in turn, led to an increase in looking to a named object and less looking to the other objects in the scene. These results suggest that learners integrate their knowledge of object labels when deciding how to allocate visual attention to social partners, which in turn can shape the input to word learning mechanisms.

**Keywords:** eye movements; language processing; information-seeking; word learning; gaze following

## Introduction

People use language to talk about many things, with no guarantee that they refer to objects in the co-occurring visual context. The flexibility of language creates a challenge where learners often encounter new words whose intended meanings are mostly unconstrained. But children are quite capable word learners. How does word learning unfold despite ambiguity in the input?

Research on lexical development has pursued several solutions. Lab-based studies and computational models show that learners can overcome referential uncertainty within a labeling event by tracking the elements of a context that remain consistent across multiple exposures to a new word (cross-situational learning) (Roy & Pentland, 2002; Yu & Smith, 2007). Social-pragmatic accounts highlight research showing that children's social partners reduce the complexity of the learning task by using gesture and eye gaze to coordinate language interactions with children (Bloom, 2002; Estigarribia & Clark, 2007). Moreover, even 16-month-olds can use cues from other people (e.g., the direction of their gaze) to infer the meaning of a new word (Baldwin, 1993).

Both social and statistical information can reduce uncertainty about reference during language processing. These processes, however, do not operate in isolation, and a sophisticated learner might integrate information from both sources to learn words. Computational models show better learning

by integrating social information with cross-situational learning mechanisms (Frank, Goodman, & Tenenbaum, 2009; Yu & Ballard, 2007).

The statistical and social accounts of word learning reviewed above reflect a somewhat passive construal of the learner. Children, however, take actions to shape learning (e.g., choosing where to look or point). Empirical work shows that "active" control can speed learning because it allows people to integrate prior knowledge and uncertainty to seek more useful information (Gureckis & Markant, 2012). For example, Kachergis, Yu, & Shiffrin (2013) found that adults who selected objects to be labeled learned more word-object associations compared to adults who passively experienced labels.

Here, we pursue an active, integrative account and ask how children's social information seeking changes as they experience multiple exposures to a word-object link. Our second goal is to test whether children and adults would show a similar pattern of adaptation to gather a social cue to reference. It could be that children rely more on information from social partners because they have to understand and learn language despite having partial knowledge of word-object links and without a fully-developed language model. Moreover, adults have stronger statistical word learning skill, which could diminish the usefulness of seeking a social cue to reference when they detect a consistent association between a word and object.

In our experiments, we analyze fixations to a speaker's face who either does or does not provide a valid cue to reference: eye gaze. We chose this behavior as a case study of social information seeking because gaze following is thought to be relevant for the ecological task of linking language to the world. Moreover, recent work has found a reliable link between sustained visual attention on objects and word learning (Smith & Yu, 2013). In Experiment 1, we explore social information seeking in the context of processing concrete, familiar words, where we assume prior exposure to statistical information about word-object links outside the lab. Experiment 2 extends our approach to the case of novel word learning, allowing us to control the number of exposure to a new word and ask whether social information seeking becomes more useful in contexts with more uncertainty.

We characterize eye movements as a series of information seeking decisions that aim to minimize uncertainty (Hayhoe

& Ballard, 2005). Using this account, children should integrate statistical and social information by considering the utility of an eye movement for their current goal. We hypothesized that when a communicative partner produces a social cue to reference, they increase the value of looking to the speaker for the task of disambiguating reference. Our key behavioral prediction is that listeners will delay generating an eye movement away from a speaker until they have gathered information about the direction of their gaze. This delay will lead to an increase in fixations to the named object, which, in turn, could change the time course of statistical word learning.

## Experiment 1

In Experiment 1, we measured the time course of children and adults’ decisions about visual fixation as they processed sentences with familiar words (“Where’s the ball?”). We manipulated whether the speaker produced a post-nominal gaze cue to the named object. The visual world consisted of three fixation targets (a center video of a person speaking, a target picture, and a distracter picture; see Figure 1). The primary question of interest is whether listeners would delay shifting their gaze away from the speaker’s face when she was likely to generate a gaze cue. We predicted that fixating longer on the speaker would allow listeners to gather more language-relevant visual information to facilitate comprehension. In contrast, if listeners show parallel gaze dynamics across the gaze and no-gaze conditions, this pattern suggests that hearing the familiar word was the primary factor driving shifts in visual attention.

### Analytic approach

To quantify evidence for our predictions, we present two analyses. First, we analyze the time course of listeners’ looking to each area of interest (AOI). Proportion looking reflects the mean proportion of trials on which participants fixated on the speaker, the target image, or the distracter image at every 33-ms interval of the stimulus sentence. We tested condition differences in the proportion looking to the language source – signer or speaker – using a nonparametric cluster-based permutation analysis, which accounts for the issue of taking multiple comparisons across many time bins in the timecourse (Maris & Oostenveld, 2007). A higher proportion of looking to the language source in the gaze condition would indicate listeners’ prioritization of seeking visual information from the speaker.

Next, we analyzed the RT and Accuracy of participants’ initial gaze shifts away from the speaker to objects in the scene. RT corresponds to the latency of shifting gaze away from the central stimulus to either object measured from the onset of the target noun. All reaction time distributions were trimmed to between zero and two seconds, and RTs were modeled in log space. Accuracy corresponds to whether participants’ first gaze shift landed on the target or the distracter object. If listeners generate slower but more accurate gaze

Table 1: Age distributions of children in Experiments 1 and 2. All ages are reported in months.

Experiment	n	Mean	Min	Max
Exp. 1 (familiar words)	38	55.50	35.60	71.04
Exp. 2 (novel words)	54	52.60	36.26	70.94

shifts, this provides evidence that gathering more visual information from the speaker led to more robust language processing in the gaze context.

We used the `brms` (Bürkner, 2017) package to fit Bayesian mixed-effects regression models. The mixed-effects approach allowed us to model the nested structure of our data – multiple trials for each participant and item, and the within-participants manipulation. We used Bayesian estimation to quantify uncertainty in our point estimates, which we communicate using a 95% Highest Density Interval (HDI), providing a range of credible values given the data and model.

## Methods

**Participants** Participants were native, monolingual English-learning children ( $n = 38$ ; 19 F) and adults ( $n = 33$ ; 23 F). All participants had no reported history of developmental or language delay and normal vision. 12 participants (9 children, 3 adults) were run but not included in the analysis because either the eye tracker failed to calibrate (8 children, 2 adults) or the participant did not complete the task (1 children, 1 adults).

**Materials** *Linguistic stimuli.* The video/audio stimuli were recorded in a sound-proof room and featured two female speakers who used natural child-directed speech and said one of two phrases: “Hey! Can you find the (target word)” or “Look! Where’s the (target word). The target words were: ball, bunny, boat, bottle, cookie, juice, chicken, and shoe. The target words varied in length (shortest = 411.68 ms, longest = 779.62 ms) with an average length of 586.71 ms.

*Gaze manipulation.* To create the stimuli in the gaze condition, the speaker waited until she finished producing the target sentence and then turned her head to gaze at the bottom right corner of the camera frame. After looking at the named object, she then returned her gaze to the center of the frame. We chose to allow the length of the gaze cue to vary to keep the stimuli naturalistic. The average length of gaze was 2.12 seconds with a range from 1.78 to 3.07 seconds.

*Visual stimuli.* The image set consisted of colorful digitized pictures of objects presented in fixed pairs with no phonological overlap between the target and the distracter image (cookie-bottle, boat-juice, bunny-chicken, shoe-ball). The side of the target picture was counterbalanced across trials.

**Procedure** Participants viewed the task on a screen while their gaze was tracked using an SMI RED corneal-reflection eye-tracker mounted on an LCD monitor, sampling at 30 Hz.

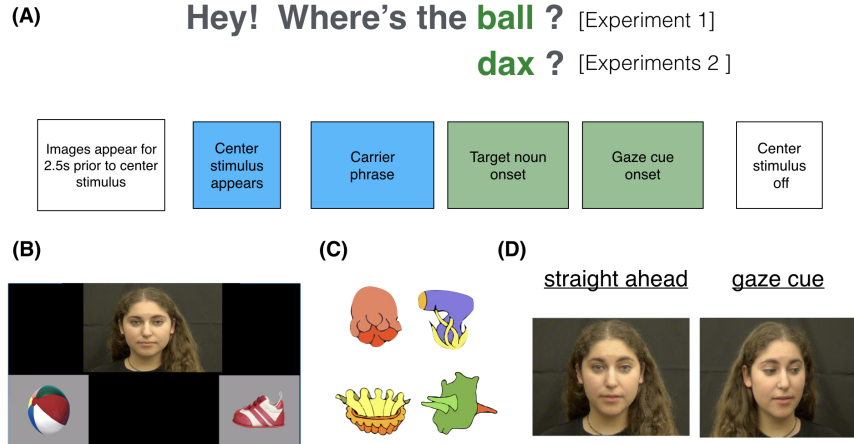


Figure 1: Stimuli for Experiments 1 and 2. Panel A shows the structure of the linguistic stimuli for a single trial. Panel B shows the layout of the fixation locations for all tasks: the center stimulus, the target, and the distracter. Panel C shows a sample of the images used as novel objects in Experiment 2. Panel D shows an example of the social gaze manipulation.

The eye-tracker was first calibrated for each participant using a 6-point calibration. On each trial, participants saw two images of familiar objects on the screen for two seconds before the center stimulus appeared. Next, they processed the target sentence – which consisted of a carrier phrase, a target noun, and a question – followed by two seconds without language to allow for a response. Both children and adults saw 32 trials (16 gaze trials; 16 no-gaze trials) with several filler trials interspersed to maintain interest. The gaze manipulation was presented in a blocked design with the order of block counterbalanced across participants.

## Results and Discussion

**Timecourse looking** We first analyzed how the presence of gaze influenced listeners' distribution of attention across the three fixation locations while processing familiar words. At target-noun onset, listeners tended to look more at the speaker than the objects. As the target noun unfolded, the mean proportion looking to the center decreased as participants shifted their gaze to the images. Proportion looking to the target increased sooner and reached a higher asymptote compared to proportion looking to the distracter for both gaze conditions with adults spending more time looking at the target compared to children. After looking to the named referent, listeners tended to shift their gaze back to the speaker's face.

We did not see evidence that the presence of a post-nominal gaze cue changed how children or adults allocated attention early in the target word. Children in the gaze condition, however, tended to shift their focus back to the speaker earlier after shifting gaze to the named object and spent more time fixating on the speaker's face throughout the rest of the trial ( $p < .001$ ). Next, we ask how these different processing contexts changed the timing and accuracy of children's initial decisions to shift away from the center stimulus.

**First shift RT and Accuracy** To quantify differences across the groups, we fit a Bayesian linear mixed-effects regression predicting first shift RT as a function of gaze condition and age group:  $\text{Log}(\text{RT}) \sim \text{gaze condition} + \text{age group} + (\text{gaze condition} + \text{item} \mid \text{subject})$ . Both children and adults generated similar RTs in the gaze (children  $M_t = 563.159$  ms, adults  $M_t = 652.405$  ms) and no-gaze (children  $M_t = 575.762$  ms, adults  $M_t = 608.314$  ms) conditions, with the null value of zero condition differences falling within the 95% credible interval ( $\beta = -0.36$ , 95% HDI  $[-0.89, 0.06]$ ). Next, we fit the same model to estimate first shift accuracy. Adults generated more accurate gaze shifts ( $M = 0.9$ ) compared to children ( $M = 0.64$ ) with the null value falling outside the 95% HDI ( $\beta_{\text{age}} = -1.76$ , 95% HDI  $[-2.19, -1.34]$ ). Similar to the RT analysis, we did not find strong evidence of a difference in performance across the gaze conditions ( $\beta = 0.10$ , 95% HDI  $[-0.18, 0.41]$ ).

The time course and first shift analyses suggest that hearing a familiar noun was sufficient for both adults and children to shift visual attention away from the speaker and seek a named referent. Neither age group showed evidence of delaying eye movements to gather a social cue to reference. Children did allocate more attention to the speaker after processing the familiar noun. While we did not predict these results, listeners' behavior seems reasonable if eye movements during familiar language processing are highly-practiced visual routines such that seeking a post-nominal gaze cue becomes less-relevant to disambiguating reference.

The results of Experiment 1 suggest that listeners do not always seek social information when it is available; instead, they might take their uncertainty into account and use social information when ambiguity is higher. This interpretation raises an interesting question: Would listeners adapt to gather social information when they do not already know the

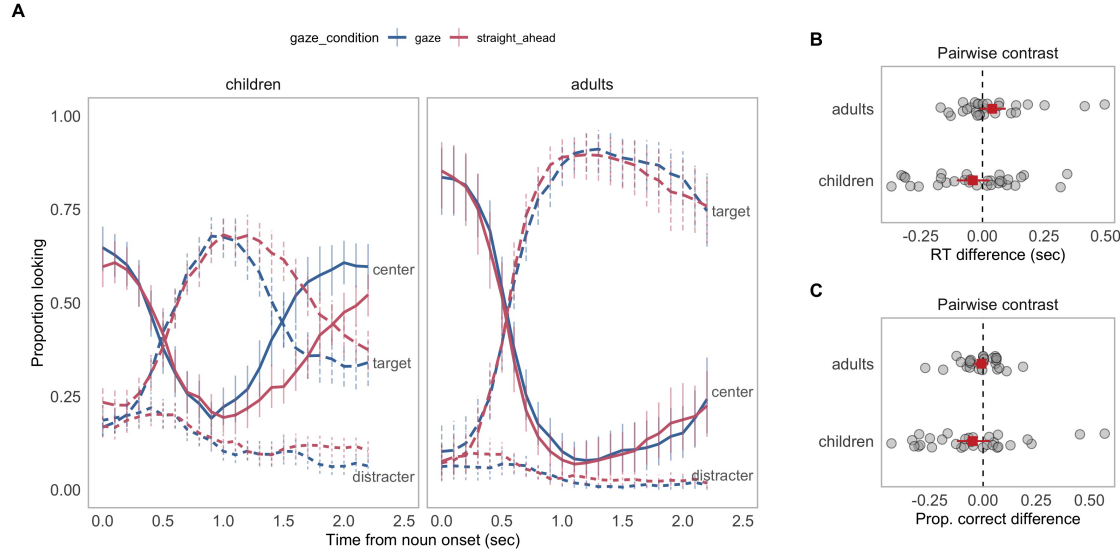


Figure 2: Timecourse looking, first shift Reaction Time (RT), and Accuracy results for children and adults in Experiment 1. Panel A shows the overall looking to the center, target, and distracter stimulus for each gaze condition and age group. Panel B shows the distribution of pairwise contrasts between each participant's RT in the gaze and no-gaze conditions. The square point represents the group means. The vertical dashed line represents the null model of zero condition difference. Error bars represent the 95% HDI. Panel C shows the same information but for first shift accuracy.

meaning of a word? That is, when surrounded by unfamiliar objects, the value of fixating on a social partner may increase since this action could provide access to critical disambiguating information such as eye gaze or pointing – an idea emphasized by social-pragmatic theories of language acquisition (Bloom, 2002). Experiment 2 explores this case to ask whether learners adapt the timing of gaze shifts to seek information from social partners when encountering a novel word.

## Experiment 2

In Experiment 2, we ask three questions: (1) do listeners adapt their gaze to seek social information in the context of processing novel words? (2) Does social information seeking change as a function of gaining more exposures to a word-object link? And (3) does following a gaze cue change the link between attention during labeling and word learning? To answer these questions, we compared the timing and accuracy of eye movements during a real-time cross-situational word learning task where participants processed sentences containing a novel word (“Where’s the dax?”) while looking at a simplified visual world.

We predicted that the presence of gaze would increase the value of looking to a speaker, leading to a higher proportion of fixations to the social target and slower first shift reaction times to the objects, especially earlier in learning. We operationalize this prediction as a main effect of gaze condition on proportion looking to the speaker, first shift RT, and a trial number by gaze condition interaction such that the decrease in RT will be greater on exposure trials in the gaze condition.

We also predicted that the presence of gaze would lead to faster learning of the novel word-object links, which we operationalized as more accurate first shifts, faster RTs, and a higher proportion looking to the target object on test trials in the gaze condition.

## Methods

**Participants** Participants were native, monolingual English-learning children ( $n = 54$ ; 30 F) and adults ( $n = 30$ ; 20 F). All participants had no reported history of developmental or language delay and normal vision. 6 adults were run but not included in the analysis because they were not native speakers of English. 7 children participants were run but not included in the analysis because the participant did not complete more than half of the trials in the task.

**Materials** *Linguistic stimuli.* The video/audio stimuli were recorded in a sound-proof room and featured two female speakers who used natural child-directed speech and said one of two phrases: “Hey! Can you find the (novel word)” or “Look! Where’s the (novel word). The target words were four pseudo-words: bosa, modi, toma, and pifo. The novel words varied in length (shortest = 472.00 ms, longest = 736.00 ms) with an average length of 606.31 ms.

*Gaze manipulation.* The gaze manipulation was identical to Experiment 1. The average length of gaze was 2.06 seconds with a range from 1.74 to 2.67 seconds.

*Visual stimuli.* The image set consisted of 28 colorful digitized pictures of objects that were selected such that they would be interesting to and that children would be unlikely to have already a label associated with the objects. The side

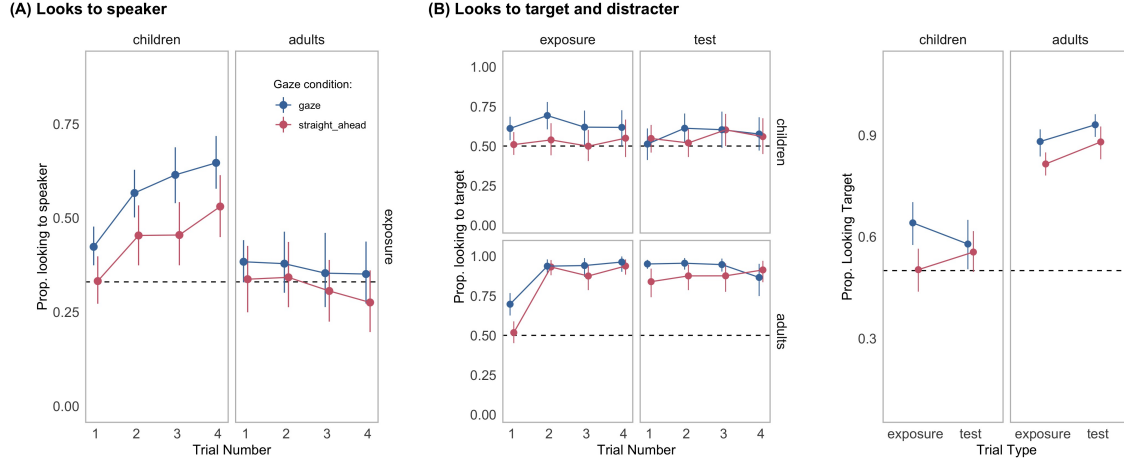


Figure 3: Panel A shows participants’ tendency to look at the speaker on exposure and test trials as a function of the trial number within a learning block. The horizontal, dashed line represents the tendency to distribute attention equally across the three AOIs. Color indicates gaze condition and error bars represent 95% credible intervals. Panel B shows the same information but for target and distracter looking across the learning block (left) and aggregated over all trials (right).

of the target picture was counterbalanced across trials.

**Procedure** The procedure was identical to Experiment 1. Participants watched a series of ambiguous word learning events organized into pairs of one exposure and one test trial. On each trial, participants saw a set of two unfamiliar objects and heard one novel word. Each word occurred in a block of four exposure-test pairs for a total of eight trials for each novel word. On each trial within a learning block, one of the objects in the set had consistently appeared on the previous trials (target object), while the other object was a randomly generated novel object that had not been shown in the experiment (distracter object). Both children and adults saw 32 trials (16 gaze trials; 16 no-gaze trials) with several filler trials interspersed to maintain interest. The gaze manipulation was presented in a blocked design with the order of block counterbalanced across participants.

## Results and Discussion

**Proportion looking** *Learning effects.* Both children ( $M_{gaze} = 0.57$ ,  $M_{no-gaze} = 0.55$ ) and adults ( $M_{gaze} = 0.91$ ,  $M_{no-gaze} = 0.89$ ) showed evidence of learning the novel word-object links, with the null value of 0.5 falling below the lower bound of the lowest credible interval for children’s target looking in the No-gaze context (95% HDI [0.51, 0.60]). Our primary question of interest was how exposure to multiple co-occurrences of word-object pairs would change learners’ distribution of attention between the speaker and objects. Figure 3 shows proportion looking to the speaker and the target and distracter objects as a function of trial number within a word learning block. Both children and adults were more likely to fixate on the speaker when she provided a gaze cue ( $\beta_{gaze} = 0.09$ , 95% HDI [0.16, 0.01]). Moreover, there was a developmental difference such that children, but not adults, were

more likely to increase their fixations to the speaker over the course of the learning block ( $\beta_{age*tr.num} = -0.07$ , 95% HDI [-0.11, -0.04]).

Overall, looking to the target increased as learners were exposed to more word-object pairings ( $\beta_{tr.num} = 0.16$ , 95% HDI [0.09, 0.24]) and was higher when the novel word was accompanied by a gaze cue ( $\beta_{gaze} = 0.14$ , 95% HDI [0.21, 0.06]). Visual inspection of Figure 3 shows that on the first exposure trial, both adults and children used the gaze cue to disambiguate reference, fixating more on the target in the gaze condition. For children, higher target looking on exposure trials with gaze remained relatively constant across the learning block. In contrast, adults target looking reached ceiling for both gaze and no-gaze conditions by trial number two, indicating that they had successfully used the co-occurrence information across trials to map the novel word to its referent. We found an interaction between gaze condition and trial number such that looking to the target increased more quickly in the No-gaze condition ( $\beta_{gaze*tr.num} = 0.02$ , 95% HDI [0.00, 0.04]), which reflects (1) the higher intercept of target looking in the presence of gaze and (2) rapid learning of the word-object association via cross-situational information. Finally, visual inspection of the proportion looking plot suggests that adults tended to look more the target when learning from a gaze cue, only reaching similar levels of accuracy in the no-gaze condition at the end of the learning block. There was not strong evidence for an effect of the gaze manipulation on children’s looking behavior on Test trials.

*Relationship between looking on exposure and test.* For both children and adults, more time attending to the target object on exposure trials led to a higher proportion of looking to the target on test trials, especially for adults ( $\beta_{exposure*age} = 0.16$ , 95% HDI [0.05, 0.28]) and as the number of word-object exposures increased over the course a learning block

( $\beta_{\text{exposure} \times \text{tr.num}} = 0.07$ , 95% HDI [0.02, 0.12]). There was evidence that participants in the No-gaze condition showed less learning over the course of each word block ( $\beta_{\text{gaze} \times \text{tr.num}} = -0.02$ , 95% HDI [-0.04, 0.00]). This result dovetails with the findings from Experiment 2, providing evidence that the presence of social information did more than change attention on exposure trials but instead modulated the relationship between attention during learning and later memory for word-object links.

The proportion looking analyses suggest that the presence of gaze changed how children and adults allocated attention while processing novel words. In the context of unfamiliar objects, children tended to fixate more on a speaker's face when she provided a post-nominal social cue to reference, a difference in looking behavior that increased as they were exposed to more word-object co-occurrences. This result is different from the parallel looking behavior that we found in Experiment 1 where listeners processed highly familiar nouns. Moreover, in the presence of a speaker who provided a gaze cue, children and adults spent less time fixating on the distracter image, which modulates the word-object connections that learners could store from labeling event. These changes in gaze patterns, however, did not generalize to performance differences on Test trials for children. Finally, we found that the presence of a social cue increased the strength of the link between attention on exposure and fixations at test.

**First shift RT and Accuracy** We next asked how the presence of gaze influenced learners' decision to stop gathering visual information from the speaker and start fixating on the novel objects. To quantify the effect the gaze, we fit a Bayesian linear mixed-effects regression predicting first shift RT as a function of whether there was a gaze cue present on the trial and age group. Both children (Gaze  $M_{rt} = 1,136.77$  ms, No-gaze  $M_{rt} = 878.37$  ms) and adults (Gaze  $M_{rt} = 878.65$  ms, No-gaze  $M_{rt} = 726.99$  ms) fixated longer on the speaker when she provided a gaze cue ( $\beta_{\text{gaze}} = -0.20$ , 95% HDI [-0.38, -0.01]). With no evidence of an interaction between gaze condition and age group ( $\beta_{\text{age} \times \text{gaze}} = 0.27$ , 95% HDI [0.11, 0.44]). Moreover, both (Gaze  $M_{acc} = 0.64$ , No-gaze  $M_{acc} = 0.49$ ) and adults (Gaze  $M_{acc} = 0.89$ , No-gaze  $M_{acc} = 0.81$ ) generated more accurate first shifts in the gaze condition, indicating they were following the gaze cue on exposure trials ( $\beta_{\text{gaze}} = -0.57$ , 95% HDI [-1.13, 0.00]).

Finally, we asked whether the presence of gaze affected learning by predicting first shift accuracy on Test trials. We found that adults were more accurate than children ( $\beta_{\text{age}} = 2.24$ , 95% HDI [1.50, 3.03]), that first shifts became more accurate as learners experienced repeated exposures to word-object pairings ( $\beta_{\text{tr.num}} = 0.21$ , 95% HDI [-0.02, 0.44]). We did not see evidence for two of our predictions: (1) that children and adults would generate more accurate first shifts when learning from social gaze ( $\beta_{\text{gaze}} = -0.50$ , 95% HDI [-1.14, 0.14]) and (2) that learning from gaze would modulate the relationship between accuracy over the course of learning ( $\beta_{\text{gaze} \times \text{tr.num}} = -0.30$ , 95% HDI [-0.74, 0.12]), with the null

value falling within each credible interval.

Returning to our three behavioral predictions, we found evidence that both children and adults spent more time fixating on a speaker when she provided a useful social cue to reference. Moreover, adults decreased the amount of time fixating on the speaker as they gained more exposures to the word-object pairings, but children showed the opposite pattern, increasing their fixations to the speaker later in the task. This developmental difference suggests that looking to a social partner may have been more useful for children who were still trying to disambiguate the novel words; whereas adults showed evidence of successful disambiguation after the second exposure trial and could focus attention on the objects instead. Finally, we found mixed evidence that the presence of gaze modulated the relationship between visual attention during labeling and learning of the novel word-object mappings. Both children and adults generated a higher proportion of shifts landing on the target when there was post-nominal gaze cue available. But only adults spent more time fixating on the target object and generated more accurate first shifts for words learned with a gaze cue.

## General Discussion

Does social information seeking change as children gain more statistical information about the link between a word and object? Here, we pursued the idea that learners flexibly adapt their eye movements to gather social gaze when it was useful. We found that children and adults did not delay their gaze shifts to seek a post-nominal gaze cue while processing familiar words. When processing novel words, however, both children and adults fixated more on a speaker to seek a post-nominal gaze cue. This delay resulted in more attention allocated to the named object and less looking to the distracter object, an effect that increased throughout the task for children. Moreover, adults, but not children, showed evidence of stronger learning in the presence of social gaze while both age groups were capable of learning the word-object pairings from cross-situational statistics alone.

How should we characterize the effects of gaze on information seeking and word learning in our task? Children selectively gathered social information when they were uncertain about the meaning of a new word, focusing attention on a single object. This pattern of behavior generalized to trials without a gaze cue, showing how the effect of gaze could accumulate over time. Finally, seeking social gaze increased the rate of word learning for adults. This finding dovetails with other work showing that the presence of social information changes information processing (Wu, Gopnik, Richardson, & Kirkham, 2011; Yoon, Johnson, & Csibra, 2008).

These studies have several important limitations. First, we did not find strong evidence that the effects of gaze generalized to contexts without gaze for children in Experiment 2. Moreover, children did not show strong learning of the novel word-object links overall. Prior work has shown that 3-5 year-olds learn words better from an extended, as op-



posed to brief, social cue to reference (Yurovsky, Wade, & Frank, 2013). Future work could increase the length of the gaze cue, which was relatively short in these studies (~2 sec). Second, we used a binary manipulation of the social context – a fully disambiguating gaze cue or entirely ambiguous label without gaze. These extremes do not reflect the complexity of children’s social interactions. For example, observational studies of child-caregiver play sessions show that social cues such as eye gaze or pointing are noisy (Frank, Tenenbaum, & Fernald, 2013) and that caregivers tend to provide a mixture of ambiguous and clear labeling events (Medina, Snedeker, Trueswell, & Gleitman, 2011). It would be interesting to know how children’s real-time information selection responds to graded changes in the utility of social information for reducing referential ambiguity.

This work integrates social-pragmatic and statistical accounts of language acquisition. We found that listeners’ decisions to seek social information varied depending on their uncertainty over word-object mappings. In the context of processing novel, but not familiar words, listeners adapted their gaze to seek a post-nominal social cue to reference. This change led to increased visual attention on a single object and less attention distributed across potential spurious word-object links. Moreover, following gaze modulated the relationship between attention during labeling and learning of word meaning. This approach sheds light on how children might integrate social and statistical information when deciding where to look, which, in turn, shapes the information that comes into contact with statistical learning processes.

Data/code available at <https://bit.ly/2FgIbsW>  
 E1 preregistration at <https://osf.io/2q4gw/>  
 E2 preregistration at <https://osf.io/nfz85/>

## Acknowledgements

We are grateful to the people who participated in this research. Thanks to Kayla Constandse, Tami Alade, and Hannah Slater for help with data collection. This work was supported by an NSF GRFP to KM and a Jacobs Foundation Fellowship to MCF.

## References

- Baldwin, D. A. (1993). Infants’ ability to consult the speaker for clues to word reference. *Journal of Child Language*, 20(02), 395–418.
- Bloom, P. (2002). *How children learn the meaning of words*. The MIT Press.
- Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), 1–28.
- Estigarribia, B., & Clark, E. V. (2007). Getting and maintaining attention in talk to young children. *Journal of Child Language*, 34(4), 799–814.
- Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers’ referential intentions to model early cross-situational word learning. *Psychological Science*, 20(5), 578–585.
- Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, 9(1), 1–24.
- Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning a cognitive and computational perspective. *Perspectives on Psychological Science*, 7(5), 464–481.
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4), 188–194.
- Kachergis, G., Yu, C., & Shiffrin, R. M. (2013). Actively learning object names across ambiguous situations. *Topics in Cognitive Science*, 5(1), 200–213.
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of eeg-and meg-data. *Journal of Neuroscience Methods*, 164(1), 177–190.
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences*, 108(22), 9014–9019.
- Roy, D. K., & Pentland, A. P. (2002). Learning words from sights and sounds: A computational model. *Cognitive Science*, 26(1), 113–146.
- Smith, L. B., & Yu, C. (2013). Visual attention is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development*, 9(1), 25–49.

- Wu, R., Gopnik, A., Richardson, D. C., & Kirkham, N. Z. (2011). Infants learn about objects from statistics and people. *Developmental Psychology*, 47(5), 1220.
- Yoon, J. M., Johnson, M. H., & Csibra, G. (2008). Communication-induced memory biases in preverbal infants. *Proceedings of the National Academy of Sciences*, 105(36), 13690–13695.
- Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, 70(13), 2149–2165.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, 18(5), 414–420.
- Yurovsky, D., Wade, A., & Frank, M. (2013). Online processing of speech and social information in early word learning. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 35).