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 can use language to talk about many things with no gurantee that they refer to objects in the co-occuring visual context. This creates a scenario where the intended meaning of a word is largely unconstrained. And yet children, who are just learning how to walk, are quite capable mapping words to their corresponding conceptual representations. How do they process and learn words despite ambiguity in the input?

Research on early lexical development has pursued several solutions. First, lab-based studies and computational models show that learners can overcome referential uncertainty within a specific labeling event by tracking the elements of a context that remain consistent across multiple exposures to a new word (Roy & Pentland, 2002; Yu & Smith, 2007). Social-pragmatic accounts highlight research showing that children's social partners can reduce the complexity of the learning task by using gesture and eye gaze to coordinate language interactions with children (Bloom, 2002; Estigarribia & Clark, 2007). And, from a young age, children can use a social information to infer a speaker's intended meaning (Baldwin, 1993).

Thus, 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 could integrate the two information sources to facilitate acquisition. For example, computational modeling by Yu & Ballard (2007) found better word-object mapping performance if their model used social cues (eye gaze) to increase the strength of specific word-object associations stored from a given labeling event. Moreover, Frank, Goodman, & Tenenbaum (2009) showed that a model including inferences about a speaker's intended meaning was able to capture a variety of key behavioral findings in early language development.

The statistical and social accounts of word learning reviewed above reflect a somewhat passive construal of the learner. Children, however, take actions that shape their learning by choosing where to look, point, or whether to ask a question. Research on "active learning" shows that giving people control can accelerate learning because it allows them to integrate prior knowledge and their current uncertainty to seek more useful information (e.g., asking a question about something that is particularly confusing) (Gureckis & Markant, 2012). Moreover, recent empirical and modeling work has explored how active control could change word learning (Hidaka, Torii, & Kachergis, 2017; Partridge, McGovern, Yung, & Kidd, 2015). For example, Kachergis, Yu, & Shiffrin (2013) found that adults who selected the set of novel objects to be labeled learned more than adults who passively experienced the word-object pairings generated by an experimenter.

Here, we pursue the question of whether children integrate their prior knowledge of words when choosing to seek social information during real-time language processing. We focus on eye movements as a case study because visual fixations are important for the task of linking language to the co-occuring context. And recent work has shown that infants' ability to sustain visual attention on objects is a strong predictor of their novel word learning (Smith & Yu, 2013) and that social partners can facilitate this form of sustained attention (Yu & Smith, 2016).

We characterize eye movements as information seeking decisions that aim to minimize uncertainty about the world (Hayhoe & Ballard, 2005). Under this account, learners should integrate statistical and social information by considering the usefulness (i.e., information gained) of an eye movement for their current task goal. Overall, our goal is to use this framework to understand how statistical word learning operates over fundamentally social input. Most prior work has used linguistic stimuli from a disembodied voice, removing a rich set of multimodal cues (e.g., gaze, facial expressions, mouth movements) that occur during face-to-face communication. By including a social fixation target, we can ask how the opportunity to seek information from a commuinicative partner might change the input to 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 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.

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

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 falied 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. 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

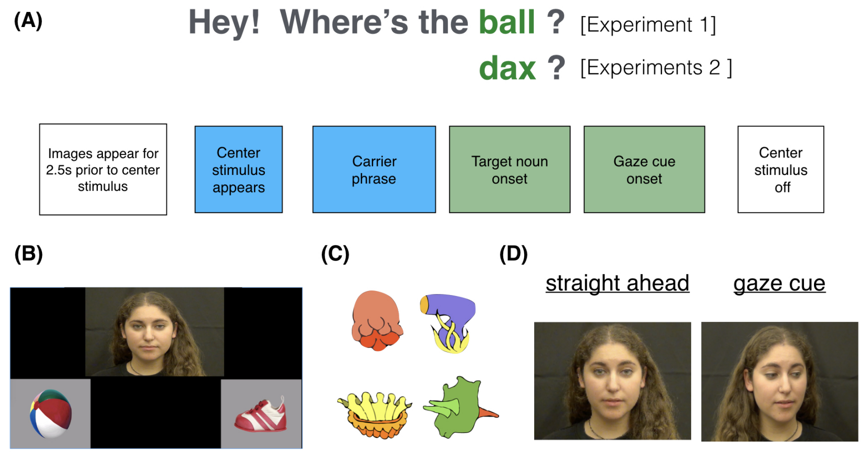

Figure 1: Stimuli for Experiments 1 and 2. Panel A shows the structure of the linguistic stimuli for a single trial. Panel B shows the layout of the fixation locations for all tasks: the center stimulus, the target, and the distracter. Panel C shows a sample of the images used as novel objects in Experiment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

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 ($p < .001$; nonparametric cluster-based permutation analysis). 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: $Log(RT) \sim gaze\ condition + age\ group + (gaze\_condition + item \mid subject)$. Both children and adults generated similar RTs in the gaze (children M_{rt} = 563.159 ms, adults M_{rt} = 652.405 ms) and no-gaze (children M_{rt} = 575.762 ms, adults M_{rt} = 608.314 ms) conditions, with the null value of zero condition differences falling within the 95% credible interval (β = -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 (β_{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 (β = 0.10, 95% HDI [-0.18, 0.41]).

Taken together, 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 their eye movements to gather a social cue to reference that could have provided additional disambiguating information. The presence of gaze, however, did change children's looking behavior such that they were more likely to allocate attention to the speaker after processing the familiar noun. While we did not predict these results, it is interesting that listeners did not delay their responses to seek social information when processing familiar words. This 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 and grounding reference. Moreover, if listeners developed an expectation that their goal was to seek out named objects quickly, then fixating on the speaker for longer becomes less goal-relevant.

In previous work, we found that both children and adults fixated longer on a speaker when processing familiar words in the presence of background noise (MacDonald, Marchman, Fernald, & Frank, 2018). We explained this result as listeners adapting to the informational demands of the environment such that they gathered additional visual information when it

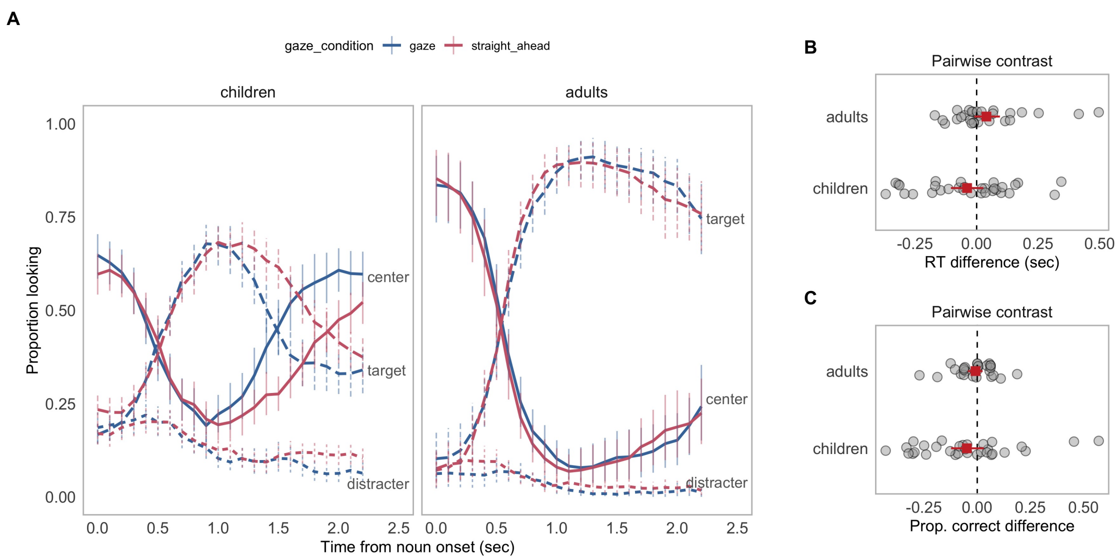

Figure 2: Timecourse looking, first shift Reaction Time (RT), and Accuracy results for children and adults in Experiment 1. Panel A shows the overall looking to the center, target, and distracter stimulus for each gaze condition and age group. Panel B shows the distribution of pairwise contrasts between 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.

was useful for language comprehension. The results of Experiment 1 help to constrain this information seeking explanation by showing that listeners do not always seek social information when it is available; instead, children may take their uncertainty into account and only adapt their information seeking when ambiguity is higher.

This interpretation raises an interesting question: Would children adapt gaze patterns to gather more social information when they do not already have labels for the objects? That is, when surrounded by unfamiliar objects, the value of fixating on a social partner should increase since this action could provide relevant disambiguating information via their gaze or pointing – an idea that has long been emphasized by social-pragmatic theories of language acquisition (Bloom, 2002). Experiments 2 and 3 explore this case and ask whether learners would adapt their gaze patterns to seek information from social partners in the context of processing novel words.

Experiment 2

Experiment 2 explores whether learners' real-time information seeking from social partners adapts as they accumulate knowledge of word-object links. We aimed to answer the following specific research questions: (1) Do young learners seek social information in the context of processing novel words?,(2) Does social information seeking change as a function of repeated exposures to a word-object link? And (3) does following a gaze cue change the relationship between visual attention during labeling and learning of novel word-object links?

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 with three fixation targets (video of a speaker and two images of unfamiliar objects).

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 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. The eye-tracker was first calibrated for each participant using a 6-point calibration. Then, participants watched a series of ambiguous word learning events organized into pairs of one exposure and one test trial. On each trial, participants saw of 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. Critically, 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 (β_{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 (β_{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 aze 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 ($\beta_{exposure*tr.num}$ = 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 ($\beta_{gaze*tr.num}$ = -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, as in Experiment 2, we found that the presence of a social cue increased the strength of the link between attention on exposure and

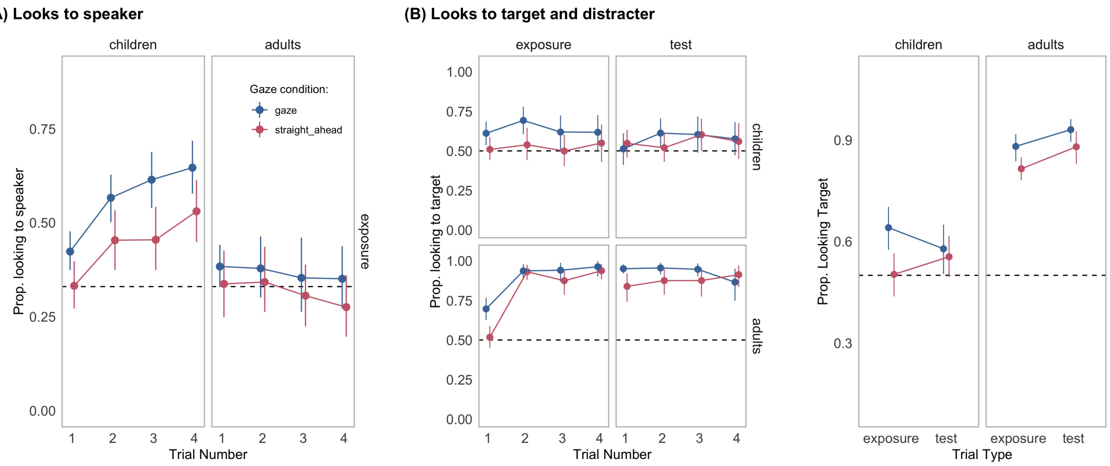

Figure 3: Panel A shows participants' tendency to look at the speaker on exposure and test trials as a function of the trial number within a learning block. The horizontal, dashed line represents the tendency to distribute attention equally across the three AOIs. Color indicates gaze condition and error 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).

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} = 1136.7668097 ms, No-gaze M_{rt} = 878.3664459 ms) and adults (Gaze M_{rt} = 878.6516854 ms, No-gaze M_{rt} = 726.9908386 ms) fixated longer on the speaker when she provided a gaze cue (β_{gaze} = -0.20, 95% HDI [-0.38, -0.01]). With no evidence of an interaction between gaze condition and age group ($\beta_{age*gaze}$ = 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 (β_{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 (β_{age} = 2.24, 95% HDI [1.50, 3.03]), that first shifts became more accurate as learners experienced repeated exposures to word-object pairings ($\beta_{tr.num}$ = 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 (β_{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 ($\beta_{gaze*tr.num}$ = -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

Do children integrate prior knowledge of words when deciding to seek social information? And how does social information seeking change as children gain more exposure to statistical information about the connection between a word and object? Here, we pursued the idea that learners flexibly adjust 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 target object and less looking to the distracter object during labeling, an effect that increased over the course

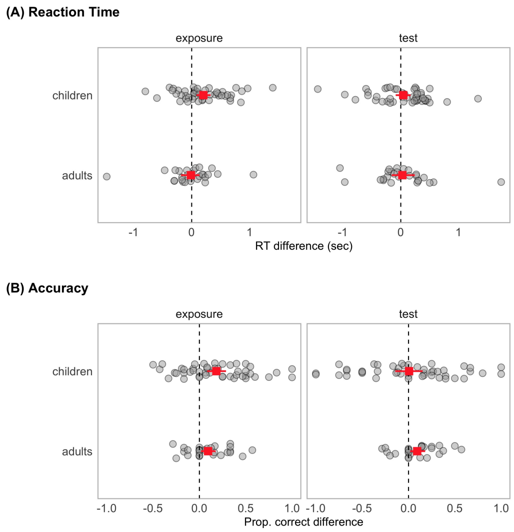

Figure 4: First shift Reaction Time (RT), and Accuracy results for children and adults in Experiment 2. Panel A shows the distribution of pairwise contrasts between RTs in the gaze and no-gaze conditions. The square point represents the mean value for each measure. The vertical dashed line represents the null model of zero condition difference. The width each point represents the 95% HDI. Panel B shows the same information but for participants' first shift accuracy.

of 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 without gaze.

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. Moreover, gaze focused attention on a single object and decreased looking to the other objects. This shift in looking behavior generalized to test trials where there was not a gaze cue present, illustrating how social gaze effects can accumulate over time and modulate the information that comes into contact with statistical learning mechanisms. Finally, seeking a social gaze cue increased the strength of the relationship between fixations during learning and strength of word learning on test trials, suggesting that learners extracted more information out of each fixation when it was directed by social gaze. This finding dovetails with other empirical work showing that the presence of social information changes how children process information (Wu, Gopnik, Richardson, & Kirkham, 2011; Yoon, Johnson, & Csibra, 2008).

This work has several limitations. First, we did not see strong evidence that the effects of seeking social gaze generalized to contexts without gaze for children in Experiment 2. Moreover, children did not show evidence of strong uptake of the novel word-object links overall. Future work could modify the paradigm to by increasing the strength of the social information. For example, Yurovsky, Wade, & Frank (2013) found that 3-5 year-olds show stronger word learning from an extended, as opposed to brief, social cue to reference. Following this work, we 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 a gaze cue. These extremes do not reflect the complexity of children's input from social interactions. 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). Moreover, our prior work suggests that adults are sensitive to the graded changes in the strength of a speaker's gaze cue, storing word-object links with greater fidelity when they expected the gaze cue to be reliable (MacDonald, Yurovsky, & Frank, 2017). 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.

These studies integrate social-pragmatic and statistical accounts of language acquisition with ideas from goal-based accounts of vision. 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 behavior led to increased visual attention to a single object and less attention distributed across potential spurious word-object links. Moreover, following gaze modulated the relationship between learners' real-time looking behavior during labeling and their learning of novel words. More generally, our approach sheds light on how children can integrate social and statistical information when using eye movements to gather information during language processing, 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.

Hidaka, S., Torii, T., & Kachergis, G. (2017). Quantifying the impact of active choice in word learning. In. Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2013). Actively learning object names across ambiguous situations. *Topics in Cognitive Science*, *5*(1), 200–213.

MacDonald, K., Marchman, V., Fernald, A., & Frank, M. C. (2018). Children seek visual information during signed and spoken language comprehension. *Preprint PsyArXiv*.

MacDonald, K., Yurovsky, D., & Frank, M. C. (2017). Social cues modulate the representations underlying cross-situational learning. *Cognitive Psychology*, *94*, 67–84.

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.

Partridge, E., McGovern, M. G., Yung, A., & Kidd, C. (2015). Young children's self-directed information gathering on touchscreens. In *Proceedings of the 37th annual conference of the cognitive science society*.

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.

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current Biology*, *26*(9), 1235–1240.

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).