Social cues modulate the representations underlying cross-situational learning

Kyle MacDonald[1,*], Daniel Yurovsky[1], Michael C. Frank[1]

Department of Psychology, Stanford University, United States

Abstract

Because children hear language in environments that contain many things to talk about, learning the meaning of even the simplest word requires making inferences under uncertainty. A cross-situational statistical learner can aggregate across naming events to form stable word-referent mappings, but this approach neglects an important source of information that can reduce referential uncertainty: social cues from speakers (e.g., eye gaze). In four large-scale experiments with adults, we tested the effects of varying referential uncertainty in cross-situational word learning using social cues. Social cues shifted learners away from tracking multiple hypotheses and towards storing only a single hypothesis (Experiments 1 and 2). In addition, learners were sensitive to graded changes in the strength of a social cue, and when it became less reliable, they were more likely to store multiple hypotheses (Experiment 3). Finally, learners stored fewer word-referent mappings in the presence of a social cue even when visual inspection time was equivalent to naming events without a social cue present (Experiment 4). Taken together, our data suggest that the representations underlying cross-situational word learning are quite flexible: In conditions of greater uncertainty, learners store a broader range of information.

Keywords: statistical learning, social cues, word learning, language acquisition

*Corresponding author

Email address: kyle.macdonald@stanford.edu (Kyle MacDonald)

1. Introduction

Learning the meaning of a new word should be hard. Consider that even concrete nouns are often used in complex contexts with multiple possible referents, which in turn have many conceptually natural properties that a speaker could talk about. This ambiguity creates the potential for an (in principle) unlimited amount of referential uncertainty in the learning task.[1] Remarkably, word learning proceeds despite this uncertainty, with estimates of adult vocabularies ranging between 50,000 to 100,000 distinct words (P. Bloom, 2002). How do learners infer and retain such a large variety of word meanings from data with this kind of ambiguity?

Statistical learning theories offer a solution to this problem by aggregating cross-situational statistics across labeling events to identify underlying word meanings (Siskind, 1996; Yu & Smith, 2007). Recent experimental work has shown that both adults and young infants can use word-object co-occurrence statistics to learn words from individually ambiguous naming events (Smith & Yu, 2008; Vouloumanos, 2008). For example, Smith and Yu (2008) taught 12-month-olds three novel words simply by repeating consistent novel word-object pairings across 10 ambiguous exposure trials. Moreover, computational models suggest that cross-situational learning can scale up to learn adult-sized lexicons, even under conditions of considerable referential uncertainty (K. Smith, Smith, & Blythe, 2011).

Although all cross-situational learning models agree that the input is the co-occurrence between words and objects and the output is stable word-object mappings, they disagree about how closely learners approximate the input distribution (for review, see Smith, Suanda, & Yu 2014). One approach has been to model learning as a process of updating connection strengths between multiple word-object links (McMurray, Horst, & Samuelson, 2012), while other approaches have argued that learners store only a single word-object hypothesis (Trueswell, Medina, Hafri, & Gleitman, 2013). In recent experimental and modeling work Yurovsky and Frank (2015) suggest an integrative explanation: learners allocate a fixed amount of attention to a single hypothesis, and distribute the rest evenly among the remaining alternatives. As the set of alternatives grows, the amount of attention allocated to each object approaches zero.

In addition to the debate about representation, researchers have disagreed about how to char-

[1] This problem is a simplified version of Quine's *indeterminacy of reference* (Quine, 1960): That there are many possible meanings for a word ("Gavigai") that include the referent ("Rabbit") in their extension, e.g., "white," "rabbit," "dinner." Quine's broader philosophical point was that different meanings ("rabbit" and "undetached rabbit parts") could actually be extensionally identical and thus impossible to tease apart.

acterize the ambiguity of the input to cross-situational learning mechanisms. One way to quantify the uncertainty in a naming event is to show adults clips of caregiver-child interactions and measure their accuracy at guessing the meaning of an intended referent (Human Simulation Paradigm: HSP [Gillette, Gleitman, Gleitman, and Lederer, 1999]). Using the HSP, Medina, Snedeker, Trueswell, and Gleitman (2011) found that approximately 90% of learning episodes were ambiguous ($< 33\%$ accuracy) and only 7% were relatively unambiguous ($> 50\%$ accuracy). In contrast, Yurovsky, Smith, and Yu (2013) found a higher proportion of clear naming events, with approximately 30% being unambiguous ($> 90\%$ accuracy). Consistent with this finding, Cartmill, Armstrong, Gleitman, Goldin-Meadow, Medina, and Trueswell (2013) showed that the proportion of unambiguous naming episodes varies across parent-child dyads, with some parents rarely providing highly informative contexts and others' doing so relatively more often.[2]

Thus, representations in cross-situational word learning can appear distributional or discrete, and the input to statistical learning mechanisms can vary along a continuum from low to high ambiguity. These results raise an interesting question: could learners be sensitive to the ambiguity of the input and use this information to alter the representations they store in memory? In the current line of work, we investigated how the presence of referential cues in the social context might alter the ambiguity of the input to statistical word learning mechanisms.

Social-pragmatic theories of language acquisition emphasize the importance of social cues for word learning (P. Bloom, 2002; Clark, 2009; Hollich et al., 2000). Experimental work has shown that even children as young as 16 months prefer to map novel words to objects that are the target of a speaker's gaze and not their own (Baldwin, 1993). In an analysis of naturalistic parent-child labeling events, Yu and Smith (2012) found that young learners tended to retain labels that were accompanied by clear referential cues, which served to make a single object dominant in the visual field. And correlational studies have demonstrated strong links between early intention-reading skills (e.g., gaze following) and later vocabulary growth (Brooks & Meltzoff, 2005, 2008; Carpenter, Nagell, Tomasello, Butterworth, & Moore, 1998). Moreover, studies outside the domain of language acquisition have shown that the presence of social cues: (a) produce better spatial learning of

[2]The differences in the estimates of referential uncertainty in these studies could be driven by the different sampling procedures used to select naming events for the HSP. Yurovsky, Smith, and Yu (2013) sampled utterances for which the parent labeled a co-present object, whereas Medina, Snedeker, Trueswell, et al. (2011) randomly sampled any utterances containing concrete nouns. Regardless of these differences, the key point here is that variability in referential uncertainty across naming events exists and thus could alter the representations underlying cross-situational learning.

audiovisual events (Wu, Gopnik, Richardson, & Kirkham, 2011), (b) boost recognition of a cued object (Cleveland, Schug, & Striano, 2007), and (c) lead to preferential encoding of an object's featural information (Yoon, Johnson, & Csibra, 2008). Together, the evidence suggests that social cues could alter the representations stored during cross-situational word learning by modulating how people allocate attention to the relevant statistics in the input.

The goal of our current investigation is to ask whether the presence of a valid social cue – a speaker's gaze – can change the representations underlying cross-situational word learning. We used a modified version of Yurovsky and Frank (2015)'s paradigm to provide a direct measure of memory for alternative word-object links during cross-situational learning. In Experiment 1, we manipulated the presence of a referential cue at different levels of attention and memory demands. At all levels of difficulty, learners tracked a strong single hypothesis but were less likely to track multiple word-object links when a social cue was present. In Experiment 2, we replicated the findings from Experiment 1 using a more ecologically valid social cue. In Experiment 3, we moved to a parametric manipulation of referential uncertainty by varying the reliability of the speaker's gaze. Learners were sensitive to graded changes in reliability and retained more word-object links as uncertainty in the input increased. Finally, in Experiment 4, we equated the length of the initial naming events with and without the referential cue. Learners stored less information in the presence of gaze even when they had visually inspected the objects for the same amount of time. In sum, our data suggest that cross-situational word learners are quite flexible, storing representations with different levels of fidelity depending on the amount of ambiguity present during learning.

2. Experiment 1

We set out to test the effect of a referential cue on the representations underlying cross-situational word learning. We used a version of Yurovsky and Frank (2015)'s paradigm where we manipulated the ambiguity of the learning context by including a gaze cue from a schematic, female interlocutor. Participants saw a series of ambiguous exposure trials where they heard one novel word that was either paired with a gaze cue or not and selected the object they thought went with each word. In subsequent test trials, participants heard the novel word again, this time paired with a new set of novel objects. One of the objects in this set was either the participant's initial guess (Same test trials) or one of the objects was *not* their initial guess (Switch test trials). Performance on Switch trials provided a direct measure of whether referential cues influenced the number of alternative word-object links that learners stored in memory. If learners performed worse on Switch trials after an exposure trial with gaze, this would suggest that they stored fewer

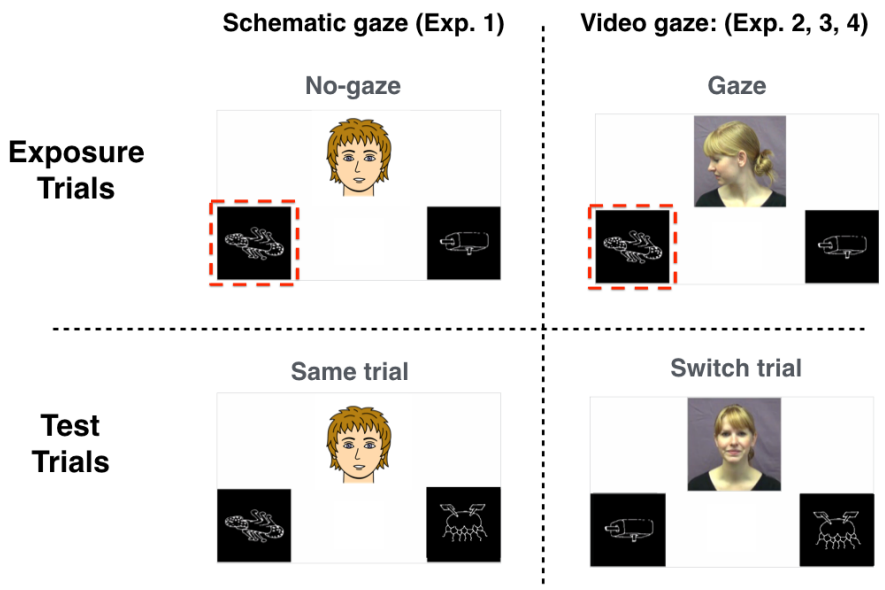

Figure 1. Screenshots of exposure and test trials from Experiment 1 (schematic gaze cue) and Experiments 2, 3 & 4 (video gaze cue). Participants saw exposure trials with or without a gaze cue depending on condition assignment. All participants saw both types of test trials: Same and Switch. On Same trials, the object that participants chose during exposure appeared with a new novel object. On Switch trials the object that participants did not choose appeared with a new novel object.

additional objects from the initial learning context.

2.1. Method

2.1.1. Participants

We posted a set of Human Intelligence Tasks (HITs) to Amazon Mechanical Turk. Only participants with US IP addresses and a task approval rate above 95% were allowed to participate, and each HIT paid 30 cents. 50-100 HITs were posted for each of the 32 between-subjects conditions. Data were excluded if participants completed the task more than once or if participants did not respond correctly on familiar object trials (131 HITs). The final sample consisted of 1438 participants.

2.1.2. Stimuli

Figure 1 shows screenshots taken from Experiment 1. Visual stimuli were black and white pictures of familiar and novel objects taken from Kanwisher, Woods, Iacoboni, and Mazziotta

(1997). Auditory stimuli were recordings of familiar and novel words by an AT&T Natural Voices ™(voice: Crystal) speech synthesizer. Novel words were 1-3 syllable pseudowords that obeyed all rules of English phonotactics. A schematic drawing of a human speaker was chosen for ease of manipulating the direction of gaze, the referential cue of interest in this study. All experiments can be viewed and downloaded at the project page: `https://kemacdonald.github.io/soc_xsit/`.

2.1.3. Design and Procedure

Participants saw a total of 16 trials: eight exposure trials and eight test trials. On each trial, they heard one novel word, saw a set of novel objects, and were asked to guess which object went with the word. Before seeing exposure and test trials, participants completed four practice trials with familiar words and objects. These trials familiarized participants to the task and allowed us to exclude participants who were unlikely to perform the task as directed, either because of inattention or because their computer audio was turned off.

After the practice trials, participants were told that they would now hear novel words and see novel objects and that their task was to select the referent that "goes with each word." Over the course of the experiment, participants heard eight novel words two times, with one exposure trial and one test trial for each word. Four of the test trials were *Same* trials in which the object that participants selected on the exposure trial was shown with a set of new novel objects. The other four test trials were *Switch* trials in which one of the objects was chosen at random from the set of objects that the participant did not select on exposure.

Participants were randomly assigned to one of the 32 between-subjects conditions (4 Referents X 4 Intervals X 2 Gaze conditions). Participants either saw 2, 4, 6, or 8 referents on the screen and test trials occurred at different intervals after exposure trials: either 0, 1, 3, or 7 trials from the initial exposure to a word. For example, in the 0-interval condition, the test trial for that word would occur immediately following the exposure trial, but in the 3-interval condition, participants would see three additional exposure trials for other novel words before seeing the test trial for the initial word. The interval conditions modulated the time delay and the number of intervening trials between learning and test, and the number of referents conditions modulated the attention demands present during learning.

Participants were assigned to either the Gaze or No-Gaze condition. In the Gaze condition, gaze was directed towards one of the objects on exposure trials; in the No-Gaze condition, gaze was always directed straight ahead (see Figure 1 for examples). At test, gaze was always directed straight ahead. To show participants that their response had been recorded, a red box appeared around the selected object for one second. This box always appeared around the selected object,

even if participants' selections were incorrect.

2.2. Results and Discussion

2.2.1. Analysis plan

The structure of our analysis plan is parallel across all four experiments. First, we examined accuracy and response time on exposure trials to provide evidence that learners were (a) sensitive to our experimental manipulation and (b) altered their allocation of attention in response to the presence of a social cue. Accuracy on exposure trials was defined as selecting the referent that was the target of gaze in the Gaze condition. (Note that there was no "correct" behavior for exposure trials in the No-Gaze condition.) Next, we examined accuracy on test trials to test whether learners' memory for alternative word-object links changed depending on the ambiguity of the learning context. Accuracy on test trials (both Same and Switch) was defined as selecting the referent that was present during the exposure trial for that word.

The key behavioral prediction of our hypothesis is that the presence of gaze would result in reduced memory for multiple word-object links, operationalized as a decrease in accuracy on Switch test trials after seeing exposure trials with a gaze cue. To quantify participants' behavior, we used mixed-effects regression models with the maximal random effects structure justified by our experimental design: by-subject intercepts and slopes for each trial type (Barr, 2013). We limited all models to include only two-way interactions because the critical test of our hypothesis was the interaction between gaze condition and trial type, and we did not have theoretical predictions for any possible three-way or four-way interactions. All models were fit using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2013), and all of our data and our processing/analysis code can be viewed in the version control repository for this paper at `https://github.com/kemacdonald/soc_xsit`.

2.2.2. Exposure trials

To ensure that our referential cue manipulation was effective, we compared participants' accuracies on exposure trials in the Gaze condition to a model of random behavior defined as a Binomial distribution with a probability of success $\frac{1}{NumReferents}$. Correct performance was defined as selecting the object that was the target of the speaker's gaze. Following Yurovsky and Frank (2015), we fit logistic regressions for each gaze, referent, and interval combination specified as `Gaze Target` $\sim$ `1 + offset(logit(1/Referents))`. The offset encoded the chance probability of success given the number of referents, and the coefficient for the intercept term shows on a log-odds scale how much more likely participants were to select the gaze target than would be

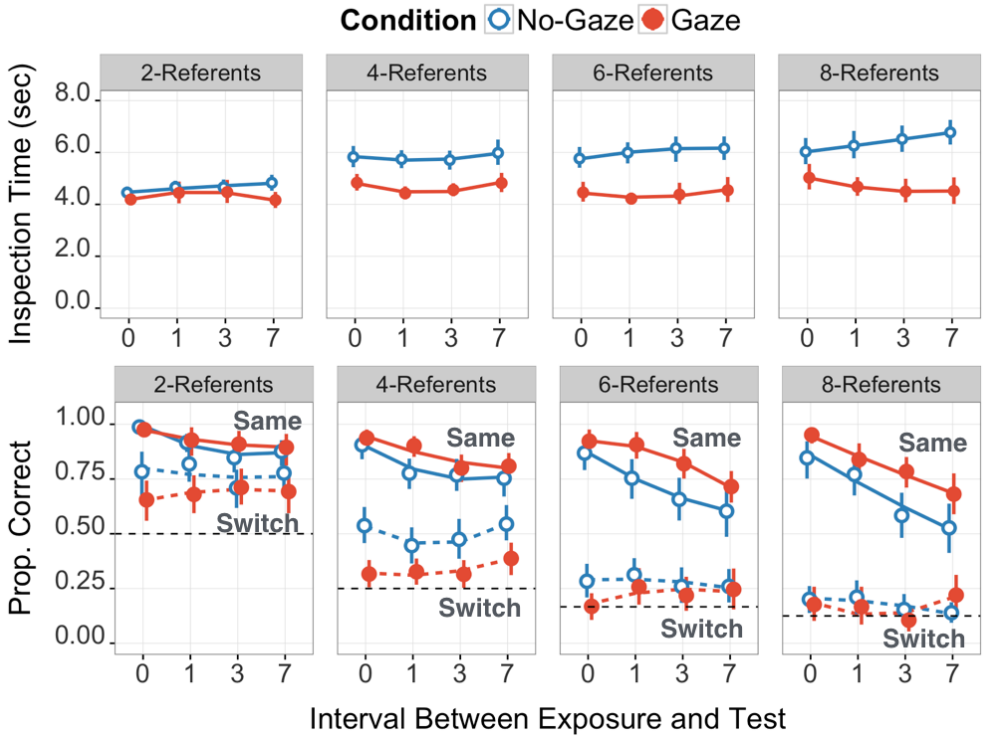

Figure 2. Experiment 1 results. The top row shows average inspection times on exposure trials for all experimental conditions as a function of the number of trials that occurred between exposure and test. Each panel represents a different number of referents, and line color represents the Gaze and No-Gaze conditions. The bottom row shows accuracy on test trials for all conditions as a function of the number of intervening trials. The horizontal dashed lines represent chance performance for each number of referents, and the type of line (solid vs. dashed) represents the different test trial types (Same vs. Switch). Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

expected if participants were selecting randomly. In all conditions, participants used gaze to select referents on exposure trials more often than expected by chance (smallest $\beta = 1.4$, z = 9.38, $p <$ .001). However, there was variability across conditions in the mean proportion of gaze cue (overall $M = 0.84$, range: 0.77–0.93).

We were also interested in differences in participants' response times across the experimental conditions. Since these trials were self-paced, participants could choose how much time to spend inspecting the referents on the screen, thus providing an index of participants' attention. To quantify the effects of gaze, interval, and number of referents, we fit a linear mixed-effects model that predicted participants' inspection times as follows: `Log(Inspection time)` $\sim$ `Gaze * Log(Interval) + Gaze * Log(Referents) + (1 | subject)`. We found a significant main

effect of the number of referents (β = 0.34, p < .001) with longer inspection times as the number of referents increased, a significant interaction between gaze condition and the number of referents (β = -0.27, p < .001) with longer inspection times in the No-Gaze condition, especially as the number of referents increased, and a significant interaction between gaze condition and interval (β = -0.08, $p = 0.004$) with slower inspection times in the No-Gaze condition, especially as the number of intervening trials increased (see the top row of Figure 2). Shorter inspection times on exposure trials with gaze provide evidence that the presence of a referential cue focused participants' attention on a single referent and away from alternative word-object links.

2.2.3. Test trials

Next, we explored participants' accuracy in identifying the referent for each word in all conditions for both kinds of test trials (see the bottom row of Figure 2). We first compared the distribution of correct responses made by each participant to the distribution expected if participants were selecting randomly defined as a Binomial distribution with a probability of success $\frac{1}{NumReferents}$. Correct performance was defined as selecting the object that was present on the exposure trial for that word. We fit the same logistic regressions as we did for exposure trials: `Correct ~ 1 + offset(logit(1/Referents))`. In 31 out of the 32 conditions for both Same and Switch trials, participants chose the correct object more often than would be expected by chance (smallest $\beta = 0.36$, $z = 2.44$, $p = 0.01$). On Switch trials in the 8-referent, 3-interval condition, participants' responses were not significantly different from chance ($\beta = 0.06$, z = 0.33, $p = 0.74$). Participants' success on Switch trials replicates the findings from Yurovsky and Frank (2015) and provides direct evidence that learners encoded more than a single hypothesis in ambiguous word learning situations even under high attentional and memory demands and in the presence of a referential cue.

To quantify the effects of gaze, interval, and number of referents on the probability of a correct response, we fit the following mixed-effects logistic regression model to a filtered dataset where we removed participants who did not reliably select the object that was the target of gaze on exposure trials:[3] `Correct ~ Trial Type * Gaze + Trial Type * Log(Interval) + Trial Type * Log(Referents) + offset(logit(`$^1/_{Referents}$`)) + (TrialType | subject)`.

[3]We did not predict that there would be a subset of participants who would not follow the gaze cue, thus this filtering criteria was developed posthoc. However, we think that the filter is theoretically motivated because we would only expect to see an effect of gaze if participants actually used the gaze cue. The filter removed 94 participants (6% of the sample). The key inferences from the data do not depend on this filtering criteria.

Predictor	Estimate	Std. Error	z value	p value	
Intercept	3.01	0.29	10.35	$< .001$	***
Switch Trial	-1.36	0.24	-5.63	$< .001$	***
Gaze Condition	0.12	0.26	0.47	0.64	
Log(Interval)	-0.45	0.11	-4.08	$< .001$	***
Log(Referents)	0.23	0.11	2.02	0.04	*
Switch Trial*Gaze Condition	-1.09	0.12	-9.07	$< .001$	***
Switch Trial*Log(Interval)	0.52	0.05	9.50	$< .001$	***
Switch Trial*Log(Referent)	-0.59	0.09	-6.49	$< .001$	***
Gaze Condition*Log(Interval)	0.06	0.06	1.00	0.32	
Gaze Condition*Log(Referent)	0.20	0.09	2.15	0.03	*
Log(Interval)*Log(Referent)	-0.04	0.04	-1.02	0.31	

Table 1. Predictor estimates with standard errors and significance information for a logistic mixed-effects model predicting word learning in Experiment 1.

We coded interval and number of referents as continuous predictors and transformed these variables to the log scale. We limited the model to include only two-way interactions because the critical test of our hypothesis is the interaction between gaze condition and trial type, and we did not have any theoretical predictions for possible three-way interactions. [4]

Table 1 shows the output of the logistic regression. We found significant main effects of the number of referents ($\beta = 0.23$, $p < .001$) and interval ($\beta = -0.45$, $p < .001$), such that as each of these factors increased, accuracy on test trials decreased. We also found a significant main effect of trial type ($\beta = -1.36$, $p < .001$), with worse performance on Switch trials. There were significant interactions between trial type and interval ($\beta = 0.52$, $p < .001$), trial type and referents ($\beta = -0.59$, $p < .001$), and gaze condition and referents ($\beta = 0.2$, $p < .05$). These interactions can be interpreted as meaning: (a) the interval between exposure and test affected Same trials more than Switch trials, (b) the number of referents affected Switch trials more than Same trials, and (c) participants performed slightly better at the higher number of referents in the Gaze condition. The interactions between gaze condition and referents and between referents and interval were not

[4] If we allowed for three-way interactions in the model, there was a marginally significant interaction between gaze condition, trial type, and interval ($\beta = 0.21$, $p = 0.058$). The two-way interaction between gaze condition and trial type remained significant in this more complex model ($\beta = -1.3$, $p = 0.006$).

significant. Importantly, we found the predicted interaction between trial type and gaze condition ($\beta = -1.09$, $p < .001$), with participants in the Gaze condition performing worse on Switch trials. This interaction provides direct evidence that the presence of a referential cue reduces participants' memory for alternative word-object links.

We were also interested in how the length of inspection times on exposure trials would affect participants' accuracy at test. So we fit an additional model where participants' inspection times were included as a predictor. We found a significant interaction between inspection time and gaze condition ($\beta = -0.17$, $p = 0.01$), such that longer inspection times provided a larger boost to accuracy in the No-Gaze condition. Importantly, the key test of our hypothesis, the interaction between gaze condition and trial type, remained significant in this alternative version of the model (β = -1.02, p = p $< .001$).

Taken together, the inspection time and accuracy analyses provide evidence that the presence of a referential cue modulated learners' attention during learning, and in turn made them less likely to track multiple word-object links. We saw some evidence for a boost to performance on Same trials in the Gaze condition at the higher number of referent and interval conditions, but reduced tracking of alternatives did not always result in better memory for learners' candidate hypothesis. This finding suggests that the limitations on Same trials may be different than those regulating the distribution of attention on Switch trials.

There was relatively large variation in performance across conditions in the group-level accuracy scores and in participants' tendency to *use* the referential cue on exposure trials. Moreover, we found a subset of participants who did not reliably use the gaze cue at all, potentially reducing the effect of gaze on cross-situational learning in this experiment. It is possible that the effect of gaze was reduced because the referential cue that we used – a static schematic drawing of a speaker – was relatively weak compared to the cues present in real-world learning environments. Thus we do not yet know how learners' memory for alternatives during cross-situational learning would change in the presence of a stronger and more ecologically valid referential cue. We designed Experiment 2 to address this question.

3. Experiment 2

In Experiment 2, we set out to replicate the findings from Experiment 1 using a more ecologically valid stimulus set. We replaced the static, schematic drawing with a video of a female actress. While these stimuli were still far from actual learning contexts, they included a real person who provided both a gaze cue and a head turn towards the target object. To reduce the across-conditions

variability that we found in Experiment 1, we introduced a within-subjects design where each participant saw both Gaze and No-Gaze exposure trials in a blocked design. We selected a subset of the conditions from Experiment 1 and tested only the 4-referent display with 0 and 3 intervening trials as between-subjects manipulations. Our goals were to replicate the reduction in learners' tracking of alternative word-object links in the presence of a referential cue and to test whether increasing the ecological validity of the cue would result in a boost to the strength of learners' recall of their candidate hypothesis.

3.1. Method

3.1.1. Participants

Participant recruitment and inclusion/exclusion criteria were identical to those of Experiment 1. 100 HITs were posted for each condition (1 Referent X 2 Intervals X 2 Gaze conditions) for total of 400 paid HITs (33 HITs excluded).

3.1.2. Stimuli

Audio and picture stimuli were identical to Experiment 1. The referential cue in the Gaze condition was a video (see Figure 1). On each exposure trial, the actress looked out at the participant with a neutral expression, smiled, and then turned to look at one of the four images on the screen. She maintained her gaze for 3 seconds before returning to the center. On test trials, she looked straight ahead for the duration of the trial.

3.2. Design and Procedure

Procedures were identical to those of Experiment 1. The major design change was a within-subjects manipulation of the gaze cue where each participant saw exposure trials with and without gaze. The experiment consisted of 32 trials split into 2 blocks of 16 trials. Each block consisted of 8 exposure trials and 8 test trials (4 Same trials and 4 Switch trials) and contained only Gaze or No-gaze exposure trials. The order of block was counterbalanced across participants.

3.3. Results and Discussion

We followed the same analysis plan as in Experiment 1. We first analyzed inspection times and accuracy on exposure trials and then analyzed accuracy on test trials.

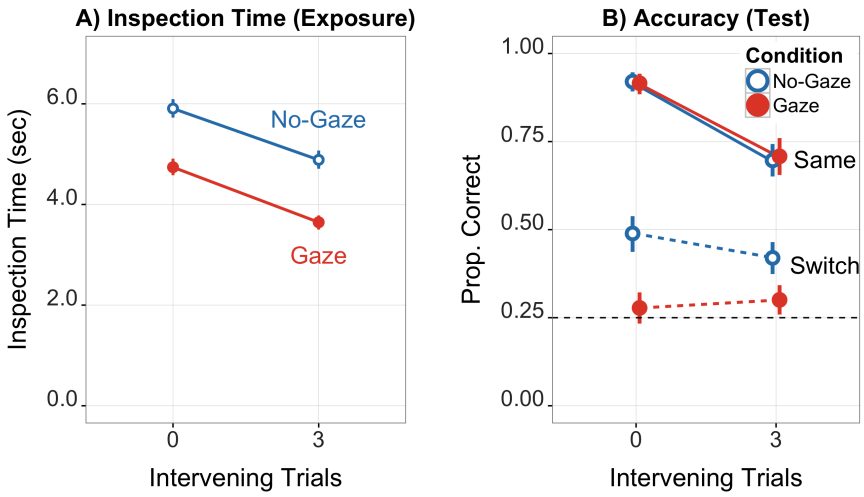

Figure 3. Experiment 2 results. Panel A shows inspection times on exposure trials with and without gaze. Panel B shows accuracy on Same and Switch test trials. All plotting conventions are the same as in Figure 2. Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

3.3.1. Exposure trials

Similar to Experiment 1, participants' responses on exposure trials differed from those expected by chance (smallest $\beta = 3.39$, z = 31.99, $p < .001$), suggesting that gaze was effective in directing participants' attention. Participants in Experiment 2 were more consistent in their use of gaze with the video stimuli compared to the schematic stimuli used in Experiment 1 ($M_{Exp1} = 0.8$, $M_{Exp2} = 0.91$\$), suggesting that using a real person increased participants' willingness to follow the gaze cue.

We replicated the findings from Experiment 1. Inspection times were shorter in the Gaze ($\beta =$ -1.1, $p < .001$) and the 3-interval condition ($\beta =$ -0.48, $p < .001$). The interaction between gaze and interval was not significant, meaning that gaze had the same effect on participants' inspection times at both intervals (see Panel A of Figure 3).

3.3.2. Test trials

Across all conditions for both trial types, participants selected the correct referent at rates greater than chance (smallest $\beta = 0.58$, z = 9.32, $p < .001$). We replicated the critical finding from Experiment 1: after seeing exposure trials with gaze, participants performed worse on Switch

Predictor	Estimate	Std. Error	z value	p value	
Intercept	2.94	0.18	16.00	$< .001$	***
Switch Trial	-2.99	0.19	-16.11	$< .001$	***
Gaze Condition	-0.10	0.16	-0.63	0.53	
Log(Interval)	-0.93	0.10	-9.23	$< .001$	***
Switch Trial*Gaze Condition	-0.71	0.16	-4.49	$< .001$	***
Switch Trial*Log(Interval)	0.79	0.10	8.03	$< .001$	***
Gaze Condition*Log(Interval)	0.15	0.08	2.05	0.04	*

Table 2. Predictor estimates with standard errors and significance information for a logistic mixed-effects model predicting word learning in Experiment 2.

trials, meaning they stored fewer word-object links ($\beta = -0.71$, $p < .001$).[5] Participants were also less accurate as the interval between exposure and test increased (β = -0.93, $p < .001$) and on the Switch trials overall ($\beta = -2.99$, $p < .001$).

In addition, there was a significant interaction between trial type and interval ($\beta = 0.79$, $p < .001$), with worse performance on Switch trials in the 3-interval condition. The interaction between gaze condition and interval was also significant ($\beta = 0.15$, $p = 0.041$), such that participants in the gaze condition were less affected by the increase in interval. Similar to Experiment 1, we did not see evidence of a boost to performance on Same trials in the gaze condition.

Next, we added inspection times on exposure trials to the model. Similar to Experiment 1, the key interaction between gaze and trial type remained significant in this version of the model (β = -0.54, $p < .001$). However, we found an interaction between inspection time and trial type (β = 0.21, $p = 0.05$), with longer inspection times providing a larger boost to performance on Switch trials. This result differs slightly from what we found in Experiment 1 where longer inspection times led to better accuracy in the No-Gaze condition. It seems plausible that more time spent visually exploring the objects during learning would lead to better performance on Switch trials, which depend on encoding multiple alternatives. Thus, the interaction between inspection time and gaze condition found in Experiment 1 might have been driven by the fact that longer inspection times were more likely to occur in the absence of a gaze cue.

The results of Experiment 2 provide converging evidence for our primary hypothesis that the

[5] As in Experiment 1, we fit this model to a filtered dataset removing participants who did not reliably use the gaze cue.

presence of a referential cue reliably focuses learners' attention away from alternative word-object links and shifts them towards single hypothesis tracking. Moving to the video stimulus led to higher rates of selecting the target of gaze on exposure trials, but did not result in a boost to performance on Same trials. This finding suggests that the level of attention and memory demand present in the learning context might modulate the effect of gaze on the fidelity of learners' single hypothesis.

Thus far we have shown that people store different amounts of information in response to a categorical manipulation of referential uncertainty. In both Experiments 1 and 2, the learning context was either entirely ambiguous (No-Gaze) or entirely unambiguous (Gaze). But not all real-world learning contexts fall at the extremes of this continuum. Could learners be sensitive to more subtle changes in the quality of the input? In our next experiment, we tested a prediction of our account: whether learners would store more word-object links in response to graded changes in referential uncertainty during learning.

4. Experiment 3

In Experiment 3, we explored whether learners would allocate attention and memory flexibly in response to *graded* changes in the referential uncertainty that was present during learning. To test this hypothesis, we moved beyond a categorical manipulation of the presence/absence of gaze, and we parametrically varied the reliability of the referential cue. We manipulated cue reliability by adding a block of familiarization trials where we varied the proportion of Same and Switch trials. If participants saw more Switch trials, this provided direct evidence that the speaker's gaze was a less reliable cue to reference because the gaze target on exposure trials would not appear at test. This design was inspired by a growing body of experimental work showing that even young children are sensitive to the prior reliability of speakers and will use this information to decide whom to learn novel words from (e.g., Koenig, Clement, & Harris, 2004).

4.1. Method

4.1.1. Participants

Participant recruitment and inclusion/exclusion criteria were identical to those of Experiment 1 and 2 (27 HITs excluded). 100 HITs were posted for each reliability level (0%, 25%, 50%, 75%, and 100%) for total of 500 paid HITs.

4.1.2. Design and Procedure

Procedures were identical to those of Experiments 1 and 2. We modified the design of our cross-situational learning paradigm to include a block of 16 familiarization trials (8 exposure trials and 8 test trials) at the beginning of the experiment. These trials served to establish the reliability of the speaker's gaze. To establish reliability, we varied the proportion of Same/Switch trials that occurred during the familiarization block. Recall that on Switch trials the gaze target did not show up at test, which provided evidence that the speaker's gaze was not a reliable cue to reference. Reliability was a between-subjects manipulation such that participants either saw 8, 6, 4, 2, or 0 Switch trials during familiarization, which created the 0%, 25%, 50%, 75%, and 100% reliability conditions. After the familiarization block, participants completed another block of 16 trials (8 exposure trials and 8 test trials). Since we were no longer testing the effect of the presence or absence of a referential cue, all exposure trials throughout the experiment included a gaze cue. Finally, at the end of the task, we asked participants to assess the reliability of the speaker on a continuous scale from "completely unreliable" to "completely reliable."

4.2. Results and Discussion

4.2.1. Exposure trials

Participants reliably chose the referent that was the target of gaze at rates greater than chance (smallest $\beta = 2.62$, z = 31.99, $p < .001$). We fit a mixed effects logistic regression model predicting the probability of selecting the gaze target as follows: `Correct-Exposure` $\sim$ `Reliability Condition * Subjective Reliability + (1 | subject)`. We found an effect of reliability condition ($\beta = 3.28$, $p = 0.03$) such that when the gaze cue was more reliable, participants were more likely to use it ($M_{0\%} = 0.83$, $M_{25\%} = 0.82$, $M_{50\%} = 0.87$, $M_{75\%} = 0.9$, $M_{100\%} = 0.94$). We also found an effect of subjective reliability ($\beta = 7.26$, $p < .001$) such that when participants thought the gaze cue was reliable, they were more likely to use it. The interaction between reliability condition and subjective reliability assessments was marginally significant (β = -4.58, p= 0.092). This analysis provides evidence that participants were sensitive to the reliability manipulation both in how often they used the gaze cue and in how they rated the reliability of the speaker at the end of the task.

4.2.2. Test trials

Next, we tested whether the reliability manipulation altered the strength of participants' memory for alternative word-object links. Across all conditions, participants selected the correct referent at rates greater than chance (smallest $\beta = 0.42$, z = 3.69, $p < .001$). Our primary prediction was an

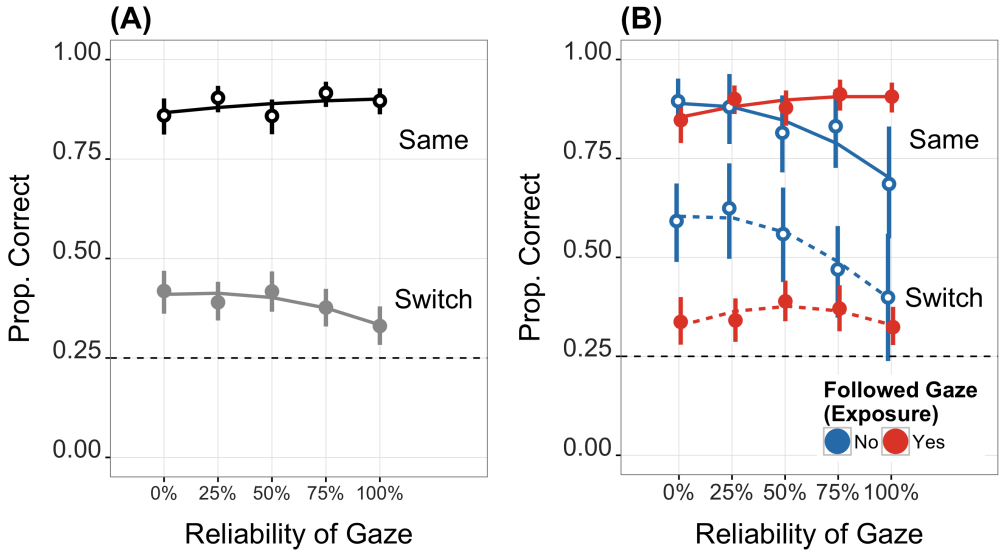

Figure 4. Primary analyses of test trial performance in Experiment 3. Panel A shows performance as a function of reliability condition. Panel B shows performance as a function of reliability condition and whether participants chose to follow gaze on exposure trials. The horizontal dashed lines represent chance performance, and error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

interaction between reliability and test trial type, with higher levels of reliability leading to worse performance on Switch trials (i.e., less memory allocated to alternative word-object links). To explore this prediction, we performed four complementary analyses: our primary analysis, which tested the effect of the reliability manipulation, and three secondary analyses, which explored the effects of participants' (a) use of the gaze cue, (b) subjective reliability assessments, and (c) inspection time on exposure trials.

Reliability condition analysis. To test the effect of reliability, we fit a model predicting accuracy at test using reliability condition and test trial type as predictors. We found a significant main effect of trial type ($\beta = -3.95$, $p < .001$), with lower accuracy on Switch trials. We also found the key interaction between reliability condition and trial type (β = -0.76, $p = 0.044$), such that when gaze was more reliable, participants performed worse on Switch trials (see Panel A of Figure 4). This interaction suggests that people stored more word-object links as the learning context becomes more ambiguous. However, the interaction between reliability and trial type was not particularly strong, and – similar to Experiment 1 – there was variability in performance across conditions (see the 50% reliable condition in Panel A of Figure 4). So to provide additional support for our hypothesis, we conducted three follow-up analyses.

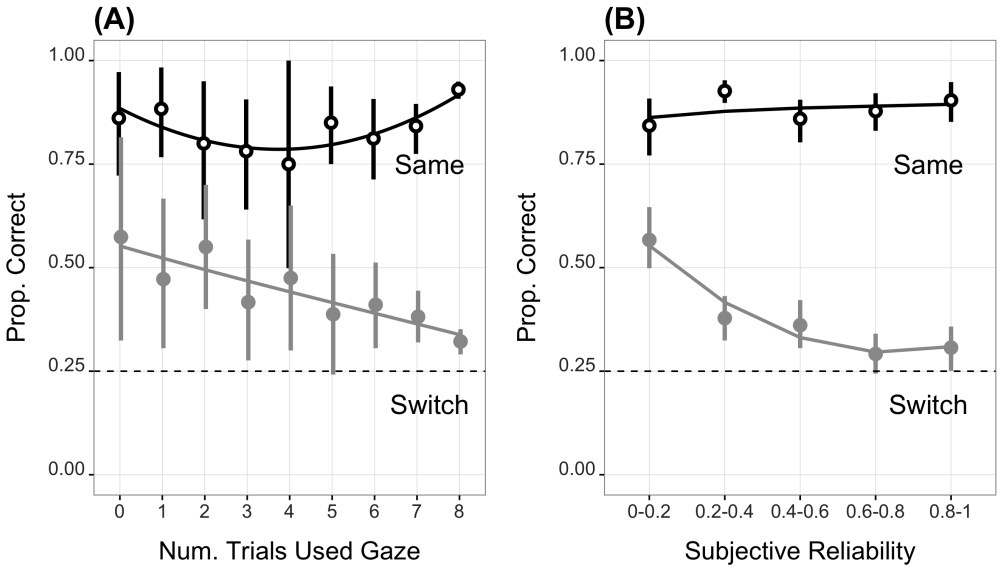

Figure 5. Secondary analyses of test trial performance in Experiment 3. Panel A shows accuracy as a function of the number of exposure trials on which participants chose to use the gaze cue. Panel B shows accuracy as a function of participants' subjective reliability judgments. The horizontal dashed lines represent chance performance, and error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

Gaze use analyses. We would only expect to see a strong interaction between reliability and trial type if learners chose to use the gaze cue during exposure trials. To test this hypothesis, we fit two additional models that included two different measures of participants' use of the gaze cue. First, we added accuracy on exposure trials as a predictor in our model. (Recall that correct performance on exposure trials was defined as using the gaze cue.) We found a significant interaction between accuracy on exposure trials and trial type ($\beta = -1.43$, $p < .001$) with worse performance on Switch test trials when participants used gaze on exposure trials (see Panel B of Figure 4). We also found an interaction between gaze use and reliability ($\beta = 0.97$, $p = 0.004$) such that when gaze was more reliable, participants were more likely to use it. The interaction between trial type and reliability became marginally significant in this model ($\beta = -0.62$, $p = 0.086$), suggesting that participants' use of the gaze cue was a stronger predictor of memory for alternative word-object links.[6]

We also hypothesized that the reliability manipulation might change how often individual participants chose to use the gaze cue throughout the task. To explore this possibility, we fit a model with the same specifications, but we included a predictor that we created by binning

[6]We are grateful to an anonymous reviewer for suggesting this analysis, but we would like to note that it is exploratory.

participants based on the number of exposure trials on which they chose to follow gaze (i.e., a gaze following score). We found a significant interaction between how often participants chose to follow gaze on exposure trials and trial type ($\beta = -0.32$, $p < .001$), such that participants who were more likely to use the gaze cue performed worse on Switch trials, but not Same trials (see Panel B of Figure 5).[7] Taken together, the two analyses of participants' use of the gaze cue provide converging evidence that when the speaker's gaze was reliable participants were more likely to use the cue, and when they followed gaze, they tended to store less information from the initial naming event.

Subjective reliability analysis. The strong interaction between use of the gaze cue and memory for alternative word-object links suggests that participants' subjective experience of reliability in the experiment mattered. Thus, we fit the same model but substituted subjective reliability for the frequency of gaze use as a predictor of test trial performance. We found a significant interaction between trial type and participants' subjective reliability assessments ($\beta = -1.63$, $p = 0.01$): when participants thought the speaker was more reliable, they performed worse on Switch trials, but not Same trials (see Panel B of Figure 5).

Inspection time analyses. Finally, we analyzed the effect of inspection times on exposure trials, fitting a model using inspection time, trial type, and reliability condition to predict accuracy at test. We found a main effect of inspection time ($\beta = 0.31$, $p = 0.001$), with longer inspection times leading to better performance for both Same and Switch trials. There was a marginally significant interaction between inspection time and reliability condition (β = -0.2, p = 0.067) with longer inspection times providing a larger boost to accuracy when the speaker was less reliable.

Next, we explored the factors that influenced inspection time on exposure trials by predicting inspection times using reliability condition and participants' use of the gaze cue as predictors. We found a main effect of participants' use of the gaze cue (-0.32, $p < .001$) with shorter inspection times when participants followed gaze. The main effect of reliability condition and the interaction between reliability and use of gaze were not significant. These analyses provide evidence that inspection times were similar across the different reliability conditions and that use of the gaze cue was the primary factor affecting how long participants explored the objects on exposure trials.

Together, these four analyses show that when the speaker's gaze was more reliable, participants were more likely to: (a) use the gaze cue, (b) rate the speaker as more reliable, and (c) store

[7] We found this interaction while performing exploratory data analysis on a previous version of this study with an independent sample (N = 250, $\beta = -0.29$, $p < .001$). The results reported here are from a follow-up study where testing this interaction was a planned analysis.

fewer word-object links, showing behavior more consistent with single hypothesis tracking. These findings support and extend the results of Experiments 1 and 2 in several important ways. First, similar to Experiment 2, participants' performance on Same trials was relatively unaffected by changes in performance on Switch trials. The selective effect of gaze on Switch trials provides converging evidence that the limitations on Same trials may be different than those regulating the distribution of attention on Switch trials. Second, learners' use of a referential cue was a stronger predictor of reduced memory for alternative word-object links compared to our reliability manipulation. Although we found a significant effect of reliability on participants' use of the gaze cue, participants' tendency to use the cue remained high. Consider that even in the 0% reliability condition the mean proportion of gaze following was still 0.82. It is reasonable that participants would continue to use the gaze cue in our experiment since it was the only cue available and participants did not have a strong reason to think that the speaker would be deceptive.

The critical contribution of Experiment 3 is to show that learners respond to a graded manipulation of referential uncertainty, with the amount of information stored from the initial exposure tracking with the reliability of the cue. This graded accuracy performance shows that learners stored alternative word-object links with different levels of fidelity depending on the amount of referential uncertainty present during learning.

Across Experiments 1-3, learners tended to store fewer word-object links in unambiguous learning contexts when a clear referential cue was present. However, in all three experiments, participants' responses on exposure trials controlled the length of the trial, meaning that when participants used the gaze cue, they also spent less time visually inspecting the objects. Thus, we do not know whether there is an independent effect of referential cues the representations underlying cross-situational learning, or if the effects found in Experiments 1-3 are entirely mediated by a reduction in inspection time. In Experiment 4, we addressed this possibility by removing participants' control over the length of exposure trials, which made the inspection times equivalent across the Gaze and No-Gaze conditions.

5. Experiment 4

In Experiment 4, we asked whether a reduction in visual inspection time in the gaze condition could completely explain the effect of social cues on learners' reduced memory for alternative word-object links. To answer this question, we modified our paradigm and made the length of exposure trials equivalent across the Gaze and No-Gaze conditions. In this version of the task, participants saw the objects for a fixed amount of time regardless of whether gaze was present.

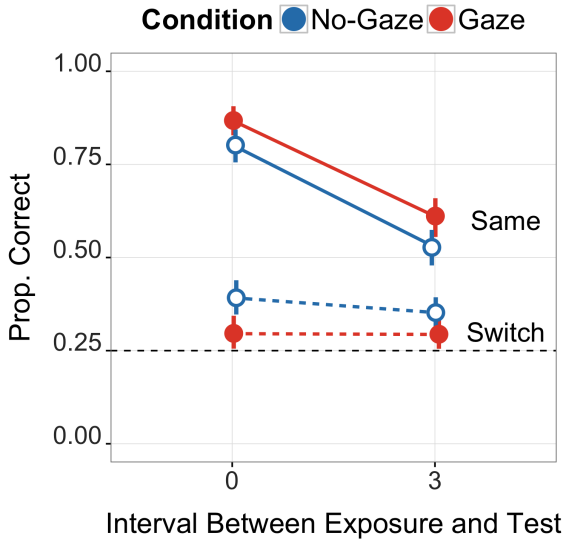

Figure 6. Experiment 4 results. Accuracy on test trials in Experiment 4 collapsed across the Long and Short inspection time conditions. The dashed line represents chance performance. Color and line type indicate whether there was gaze present on exposure trials. Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

We also included two different exposure trial lengths in order to test whether gaze would have a differential effect at shorter vs. longer inspection times. If the presence of gaze reduces learners' memory for multiple word-object links, then this provides evidence that referential cues affected the underlying representations over and above a reduction in inspection time.

5.1. Method

5.1.1. Participants

Participant recruitment and inclusion/exclusion criteria were identical to those of Experiments 1, 2, and 3. 100 HITs were posted for each condition (1 Referent X 2 Intervals X 2 Inspection Time conditions) for a total of 400 paid HITs (37 HITs excluded).

5.1.2. Stimuli

Audio, picture, and video stimuli were identical to Experiments 2 and 3. Since inspection times were fixed across conditions, we wanted to ensure that participants were aware of the time remaining on each exposure trial. So we included a circular countdown timer located above the center video. The timer remained on the screen during test trials but did not count down since participants could take as much time as they wanted to respond on test trials.

5.1.3. Design and Procedure

Procedures were identical to those of Experiment 1-3. The design was identical to that of Experiment 2 and consisted of 32 trials split into 2 blocks of 16 trials. Each block consisted of 8 exposure trials and 8 test trials (4 Same trials and 4 Switch trials) and contained only Gaze or No-Gaze exposure trials. The order of block was counterbalanced across participants.

The major design change was to make the length of exposure trials equivalent across the Gaze and No-Gaze conditions. We randomly assigned participants to one of two inspection time conditions: Short (6 seconds) or Long (9 seconds). These times were selected based on participants' self-paced inspection times in the Gaze and No-Gaze conditions in Experiment 2. After pilot testing, we added three seconds to each condition to ensure that participants had enough time to respond before the experiment advanced. If participants did not respond in the allotted time, an error message appeared informing participants that time had run out and encouraged them to respond within the time window on subsequent trials.

5.2. Results and Discussion

We did not see strong evidence of an effect of the different inspection times. Thus, all of the results reported here collapse across the short and long inspection time conditions. For all analyses, we removed the trials on which participants did not respond within the fixed inspection time on exposure trials (0.05% of trials).

5.2.1. Exposure Trials

Participants' responses on exposure trials differed from those expected by chance (smallest β = 2.95, z = 38.08, $p < .001$), suggesting that gaze was again effective in directing participants' attention. Similar to Experiment 2, participants were quite likely to use the gaze cue when it was a video of an actress ($M_{0-interval} = 0.93$, $M_{3-interval} = 0.95$).

5.2.2. Test Trials

Figure 6 shows performance on test trials in Experiment 4. In the majority of conditions, participants selected the correct referent at rates greater than chance (smallest $\beta = 0.2$, z = 2.2, $p < .05$). However, participants' responses were only marginally different from chance on Switch trials after exposure trials with gaze in the 3-interval condition ($\beta = 0.17$, $p = 0.06$).

We replicate the key finding from Experiments 1-3: after seeing exposure trials with gaze, participants were less accurate on Switch trials ($\beta = 0.9$, $p < .001$). Since inspection times were fixed across the Gaze and No-Gaze conditions, this finding provides evidence that the presence

of a referential cue did more than just reduce the amount of time participants' spent inspecting the potential word-object links. In contrast to Experiments 2 and 3, visual inspection of Figure 6 suggested that the referential cue provided a boost to accuracy on Same trials. To assess the simple effect of gaze on trial type, we computed pairwise contrasts using the *lsmeans* package in R with a Bonferroni correction for multiple comparisons (Lenth, 2016). Accuracy was higher for Same trials in the Gaze condition ($\beta = 0.49$, $p < .001$), but lower for Switch trials ($\beta = -0.41$, $p < .001$). The boost in accuracy on Same trials differs from Experiments 2 and 3 and suggests that making inspection times equivalent across conditions allowed the social cue to affect the strength of learners' memory for their candidate hypothesis.

The results of Experiment 4 help to clarify the effect of gaze on memory in our task, providing evidence that the presence of a referential cue did more than just reduce participants' visual inspection time. Instead, gaze reduced memory for alternative word-object links even when people had the same opportunity to visually inspect and encode them. We also found evidence of a boost for learners' memory of their candidate hypothesis in the gaze condition, an effect that we saw at the higher number of referents and the longer intervals in Experiment 1, but that we did not see in Experiments 2 or 3. One explanation for this difference is that in Experiment 4, since participants' use of gaze was independent of the length of exposure trials, inspection times in the gaze condition were longer compared to those in Experiments 1-3. Thus, it could be that the combination of a gaze cue coupled with the opportunity to continue attending to the gaze target led to a boost in performance on Same trials relative to trials without gaze.

6. General Discussion

Tracking cross-situational word-object statistics allows word learning to proceed despite the presence of individually ambiguous naming events. But models of cross-situational learning disagree about how much information is actually stored in memory, and the input to statistical learning mechanisms can vary along a continuum of referential uncertainty from unambiguous naming instances to highly ambiguous situations. In the current line of work, we explore the hypothesis that these two factors are fundamentally linked to one another and to the social context in which word learning occurs. Specifically, we ask how cross-situational learning operates over social input that varies the amount of ambiguity in the learning context.

Our results suggest that the representations underlying cross-situational learning are quite flexible. In the absence of a referential cue to word meaning, learners tended to store more alternative word-object links. In contrast, when gaze was present learners stored less information, showing

behavior consistent with tracking a single hypothesis (Experiments 1 and 2). Learners were also sensitive to a parametric manipulation of the strength of the referential cue, showing a graded increase in the tendency to use the cue as reliability increased, which in turn resulted in a graded decrease in memory for alternative word-object links (Experiment 3). Finally, learners stored less information in the presence of gaze even when they spent the same amount of time visually inspecting the objects during learning (Experiment 4).

In Experiments 2 and 3 reduced memory for alternative hypotheses did not result in a boost to memory for learners' candidate hypothesis. This pattern of data suggests that the presence of a referential cue selectively affected one component of the underlying representation: the number of alternative word-object links, and not the strength learners' candidate hypothesis. However, in Experiments 1 and 4, we did see some evidence of stronger memory for learners' initial hypothesis in the presence of gaze: at the higher number of referents and interval conditions (Experiment 1), and when the length of exposure trials was equivalent across the Gaze and No-Gaze conditions (Experiment 4). We speculate that the relationship between the presence of a referential cue and the strength of learners' candidate hypothesis is modulated by how the cue interacts with attention. In Experiment 1, gaze may have provided a boost because, in the absence of gaze, attention would have been distributed across a larger number of alternatives. And, in Experiment 4, gaze may have led to better memory because it was coupled with the opportunity for sustained attention to the gaze target. More work is needed in order to understand precisely when the presence of gaze affects this particular component of the representations underlying cross-situational learning.

In Experiments 1-3, longer inspection times (i.e., more time spent encoding the word-object links during learning) led to better memory at test. We did, however, find slightly different interaction effects across our studies. In Experiment 1, longer inspection times led to higher accuracy in the No-Gaze condition for both Same and Switch trials. In Experiment 2, longer inspection times provided a larger boost to performance on Switch trials compared to Same trials, regardless of gaze condition. And in Experiment 3, we found some evidence that longer inspection times led to better memory when the gaze cue was less reliable. Despite these differences, we speculate that inspection time played a similar role across these studies: When a social cue was present, learners' attention was focused and inspection times tended to be shorter, which led to worse performance on Switch trials (i.e., reduced memory for alternative word-object links). Interestingly, in Experiment 4, we found an effect of social cues on memory for alternatives even when inspection times were equivalent, suggesting that gaze does more than just modulate visual attention during learning.

6.1. Relationship to previous work

Why might a decrease in memory for alternatives fail to increase the strength of learners' memory for their candidate hypothesis? One possibility is that participants did not shift their cognitive resources from the set of alternatives to their single hypothesis, but instead chose to use the gaze information to reduce inspection time, thus conserving their resources for future use. Griffiths, Lieder, and Goodman (2015) formalize this behavior by pushing the rationality of computational-level models down to the psychological process level. In their framework, cognitive systems are thought to be adaptive in that they optimize the use of their limited resources, taking the cost of computation (e.g., the opportunity cost of time or mental energy) into account. For example, Vul, Goodman, Griffiths, and Tenenbaum (2014) showed that as time pressure increased in a decision-making task, participants were more likely to show behavior consistent with a less cognitively challenging strategy of matching, rather than with the globally optimal strategy. In the current work, we found that learners showed evidence of altering how they allocated cognitive resources based on the amount of referential uncertainty present during learning, spending less time inspecting alternative word-object links and reducing the number of links stored in memory when uncertainty was low.

Our results fit well with recent experimental work that investigates how attention and memory can constrain infants' statistical word learning. For example, Smith and Yu (2013) used a modified cross-situational learning task to show that only infants who disengaged from a novel object to look at both potential referents were able to learn the correct word–object mappings. Moreover, Vlach and Johnson (2013) showed that 16-month-olds were only able to learn from adjacent cross-situational co-occurrence statistics, and unable to learn from co-occurrences that were separated in time. Both of these findings make the important point that only the information that comes into contact with the learning system can be used for cross-situational word learning, and this information is directly influenced by the attention and memory constraints of the learner. These results also add to a large literature showing the importance of social information for word learning (P. Bloom, 2002; Clark, 2009) and to recent work exploring the interaction between statistical learning mechanisms and other types of information (Frank, Goodman, & Tenenbaum, 2009; Koehne & Crocker, 2014; Yu & Ballard, 2007). Our findings suggest that referential cues affect statistical learning by modulating the amount of information that learners store in the underlying representations that support learning over time.

Is gaze a privileged cue, or could other, less-social cues (e.g., an arrow) also affect the representations underlying cross-situational learning? On the one hand, previous research has shown

that gaze cues lead to more reflexive attentional responses compared to arrows (Friesen, Ristic, & Kingstone, 2004), that gaze-triggered attention results in better learning compared to salience-triggered attention (Wu & Kirkham, 2010), and that even toddlers readily use gaze to infer novel word meanings (Baldwin, 1993). Thus, it could be that gaze is an especially effective cue for constraining word learning since it communicates a speaker's referential intent and is a particularly good way to guide attention. On the other hand, the generative process of the cue – whether it is more or less social in nature – might be less important; instead, the critical factor might be whether the cue effectively reduces uncertainty in the naming event. Under this account, gaze is placed amongst a set of many cues that could produce similar effects as those reported here. Future work could explore a wider range of cues to see if they modulate the representations underlying cross-situational learning in a similar way.

How should we characterize the effect of gaze on attention and memory in our task? One possibility is that the referential cue acts as a filter, only allowing likely referents to contact statistical learning mechanisms (Yu & Ballard, 2007). This 'filtering account' separates the effect of social cues from the underlying computation that aggregates cross-situational information. Another possibility is that referential cues provide evidence about a speaker's communicative intent (Frank et al., 2009). In this model, the learner is reasoning about the speaker and word meanings simultaneously, which places inferences based on social information as part of the underlying computation. A third possibility is that participants thought of the referential cue as pedagogical. In this context, learners assume that the speaker will choose an action that is most likely to increase the learner's belief in the true state of the world (Shafto, Goodman, & Frank, 2012), making it unnecessary to allocate resources to alternative hypotheses. Experiments show that children spend less time exploring an object and are less likely to discover alternative object-functions if a single function is demonstrated in a pedagogical context (Bonawitz et al., 2011). However, because the results from the current study cannot distinguish between these explanations, these questions remain topics for future studies specifically designed to tease apart these possibilities.

6.2. Limitations

There are several limitations to the current study that are worth noting. First, the social context that we used was relatively impoverished. Although we moved beyond a simple manipulation of the presence or absence of social information in Experiment 3, we nevertheless isolated just a single cue to reference, gaze. But real-world learning contexts are much more complex, providing learners access to multiple cues such as gaze, pointing, and previous discourse. In fact, Frank, Tenenbaum, and Fernald (2013) analyzed a corpus of parent-child interactions and concluded that learners

would do better to aggregate noisy social information from multiple cues, rather than monitor a single cue since no single cue was a consistent predictor of reference. In our data, we did see a more reliable effect of referential cues when we used video of an actress, which included both gaze and head turn as opposed to the static, schematic stimuli, which only included gaze. It is still an open and interesting question as to how our results would generalize to learning environments that contain a rich combination of social cues.

Second, we do not yet know how variations in referential uncertainty during learning would affect the representations of young word learners, the age at which cross-situational word learning might be particularly important. Recent research using a similar paradigm as our own did not find evidence that 2- or 3-year-olds stored multiple word-object links; instead, children only retained a single candidate hypothesis (Woodard, Gleitman, & Trueswell, 2016). However, performance limitations on children's developing attention and memory systems (Colombo, 2001; Ross-sheehy, Oakes, & Luck, 2003) could make success on these explicit response tasks more difficult. Moreover, our work suggests that different levels of referential uncertainty in naturalistic learning contexts (see Medina, Snedeker, Trueswell, & Gleitman, 2011; Yurovsky & Frank, 2015) might evoke different strategies for information storage, with learners retaining more information as ambiguity in the input increases. Thus, we think that it will be important to test a variety of outcome measures and learning contexts to see if younger learners show evidence of storing multiple word meanings during learning.

In addition, previous work with infants has shown that their attention is often stimulus-driven and sticky (Oakes, 2011), suggesting that very young word learners might not effectively explore the visual scene in order to extract the necessary statistics for storing multiple alternatives. It could be that referential cues play an even more important role for young learners by filtering the input to cross-situational word learning mechanisms and guiding children to the relevant statistics in the input. In fact, recent work has shown that the precise timing of features such as increased parent attention and gesturing towards a named object and away from non-target objects were strong predictors of referential clarity in a naming event (Trueswell et al., 2016). It could be that the statistics available in these particularly unambiguous naming events are the most useful for cross-situational learning.

Finally, the current experiments used a restricted cross-situational word learning scenario, which differs from real-world language learning contexts in several important ways. One, we only tested a single exposure for each novel word-object pairing; whereas, real-world naming events are best characterized by discourse where an object is likely to be named repeatedly in a short amount

of time (Frank, Tenenbaum, & Fernald, 2013; Rohde & Frank, 2014). Two, the restricted visual world of 2-8 objects on a screen combined with the forced-choice response format may have biased people to assume that all words in the task must have referred to one of the objects. But, in actual language use, people can refer to things that are not physically co-present (e.g., Gleitman, 1990), creating a scenario where learners would not benefit from storing additional word-object links in the absence of clear referential cues. Finally, we presented novel words in isolation, removing any sentential cues to word meaning (e.g., verb-argument relations). In fact, previous work with adults has shown that cross-situational learning mechanisms only operate in contexts where sentence-level constraints do not completely disambiguate meaning (Koehne & Crocker, 2014). Thus, we need more evidence to understand how the representations underlying cross-situational learning change in response to referential uncertainty at different timescales and in richer language contexts that more accurately reflect real-world learning environments.

6.3. Conclusions

Word learning proceeds despite the potential for high levels of referential uncertainty and despite learners' limited cognitive resources. Our work shows that cross-situational learners flexibly respond to the amount of ambiguity in the input, and as referential uncertainty increases, learners tended to store more word-object links. Overall, these results bring together aspects of social and statistical accounts of word learning to increase our understanding of how statistical learning mechanisms operate over fundamentally social input.

7. Acknowledgements

We are grateful to Rose Schneider for helping record stimuli and to the members of the Language and Cognition Lab for their feedback on this project. This work was supported by a National Science Foundation Graduate Research Fellowship to KM, an NIH NRSA Postdoctoral Fellowship to DY, and a John Merck Scholars Fellowship to M.C.F.

8. References

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*(02), 395–418.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). Lme4: Linear mixed-effects models using eigen and s4. *R Package Version*, *1*(4).

Bloom, P. (2002). *How children learn the meaning of words*. The MIT Press.

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition*, *120*(3), 322–330.

Brooks, R., & Meltzoff, A. N. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of Child Language*, *35*(01), 207–220.

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, i–174.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283.

Clark, E. V. (2009). *First language acquisition*. Cambridge University Press.

Cleveland, A., Schug, M., & Striano, T. (2007). Joint attention and object learning in 5-and 7-month-old infants. *Infant and Child Development*, *16*(3), 295–306.

Colombo, J. (2001). The development of visual attention in infancy. *Annual Review of Psychology*, *52*(1), 337–367.

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.

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and

arrow cues. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(2), 319.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135–176.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229.

Hollich, G. J., Hirsh-Pasek, K., Golinkoff, R. M., Brand, R. J., Brown, E., Chung, H. L., ... Bloom, L. (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*, i–135.

Kanwisher, N., Woods, R. P., Iacoboni, M., & Mazziotta, J. C. (1997). A locus in human extrastriate cortex for visual shape analysis. *Journal of Cognitive Neuroscience*, *9*(1), 133–142.

Koehne, J., & Crocker, M. W. (2014). The interplay of cross-situational word learning and sentence-level constraints. *Cognitive Science*.

Koenig, M. A., Clement, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*(10), 694–698.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33. `http://doi.org/10.18637/jss.v069.i01`

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831.

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.

Oakes, L. M. (2011). *Infant perception and cognition: Recent advances, emerging theories, and future directions*. Oxford University Press, USA.

Quine, W. V. (1960). 0. word and object. *111e MIT Press*.

Rohde, H., & Frank, M. C. (2014). Markers of topical discourse in child-directed speech. *Cognitive Science*, *38*(8), 1634–1661.

Ross-sheehy, S., Oakes, L. M., & Luck, S. J. (2003). The development of visual short-term memory capacity in infants. *Child Development*, *74*(6), 1807–1822.

Shafto, P., Goodman, N. D., & Frank, M. C. (2012). Learning from others the consequences of psychological reasoning for human learning. *Perspectives on Psychological Science*, *7*(4), 341–

351.

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*(1), 39–91.

Smith, K., Smith, A. D., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*(3), 480–498.

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.

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.

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word–referent learning. *Trends in Cognitive Sciences*, *18*(5), 251–258.

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

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156.

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382.

Vouloumanos, A. (2008). Fine-grained sensitivity to statistical information in adult word learning. *Cognition*, *107*(2), 729–742.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science*, *38*(4), 599–637.

Woodard, K., Gleitman, L. R., & Trueswell, J. C. (2016). Two-and three-year-olds track a single meaning during word learning: Evidence for propose-but-verify. *Language Learning and Development*, *12*(3), 252–261.

Wu, R., & Kirkham, N. Z. (2010). No two cues are alike: Depth of learning during infancy is dependent on what orients attention. *Journal of Experimental Child Psychology*, *107*(2), 118–136.

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.

Yoshida, K., Rhemtulla, M., & Vouloumanos, A. (2012). Exclusion constraints facilitate statistical

word learning. *Cognitive Science*, *36*(5), 933–947.

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. (2012). Embodied attention and word learning by toddlers. *Cognition*.

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*.

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*(6), 959–966.