Referential cues focus attention and constrain the input to cross-situational word learning mechanisms

Kyle MacDonald, Daniel Yurovsky, & Michael C. Frank

Stanford University

Abstract

Tracking word-object co-occurrence statistics can reduce referential uncertainty and support word learning. But human learners are constrained by limits on attention and memory, and therefore must store a subset of the information available in a single exposure — how do they select what information to store? Drawing on social-pragmatic theories of language acquisition, we hypothesize that the presence of a referential cue, like gaze, guides how learners allocate their attention and modulates the underlying representation stored in memory. In three large-scale experiments with adults, we test how the presence of referential cues affects cross-situational word learning. Referential cues shift learners away from multiple hypothesis tracking towards storing only a single hypothesis (Experiments 1 and 2). In addition, learners are sensitive to the reliability of a cue and when it is less reliable, they are less likely to use it and more likely to store multiple hypotheses (Experiment 3). Together, the data suggest that learners make a rational tradeoff: In conditions of greater uncertainty, they store a broader range of information.

1 Introduction

Learning the meaning of a new word requires inferring structure from noisy data. We focus here on the problem of mapping concrete nouns to objects, as opposed to other substantive inferential problems such as word segmentation and generalization. To make such a mapping, learners must resolve the core problem of referential uncertainty (Quine, 1960): that a speaker's intended meaning is largely unconstrained and a new utterance could refer to many possible objects in the visual scene, to parts of of those objects, or even to something that is not present. How do learners infer word meanings from data with this kind of uncertainty?

Statistical learning theories offer a solution to this learning problem by aggregating cross-situational statistics across labeling events to identify underlying word meanings (Siskind, 1996; Yu & Smith, 2007). Recent experimental work shows that both adults and young infants can use word-object co-occurrence statistics to learn words from individually ambiguous naming events (L. Smith & Yu, 2008; Vouloumanos, 2008). For example, L. Smith & 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.[1] Some theories hold that we accumulate graded, statistical evidence about multiple referents for each word (McMurray, Horst, & Samuelson, 2012), while others argue that we track only a single candidate referent (Trueswell, Medina, Hafri, & Gleitman, 2013). Recent experimental and modeling work by Yurovsky & Frank (under review) suggests

[1] For a detailed discussion of these debates, see Smith, Suanda, & Yu (2014).

an integrative explanation: learners allocate a fixed amount of their attention to one hypothesis, and the rest gets distributed evenly among the remaining alternatives. As the set of alternatives grows, the amount allocated to each object approaches zero.

Another ongoing debate in the literature is how to best characterize the input to cross-situational learning mechanisms. One way researchers have quantified the ambiguity in the input is to ask adults to guess the meaning of an intended referent from clips of caregiver-child interactions (Human Simulation Paradigm: HSP). Using the HSP, Medina, Snedeker, Trueswell, & Gleitman (2011) found that adults did not show evidence of aggregating multiple word–referent correspondences across trials, concluding that real world learning contexts are too noisy to support tracking of multiple correspondences over time. In contrast, Yurovsky, Smith, & Yu (2013) found a bimodal distribution, with half of the naming episodes being unambiguous to adults. In addition, Cartmill et al. (2013) showed that the proportion of unambiguous naming episodes varies across parents, with some parents' rarely providing highly informative contexts and others' doing so relatively often.

Thus, representations in cross-situational word learning can appear distributional or discrete, and the input to statistical learning can vary along a continuum of ambiguity. These results raise an interesting question: could learners be sensitive to the ambiguity of the input and use this information to guide how many word-object links they store in memory? An answer to this question requires taking cross-situational learning models from Marr (1982)'s "computational" level down to the "algorithmic" level, where the goal is to specify the nature of the input/output representations and the algorithm that performs the transformation between the two (see Yurovsky & Frank (under review)).

Our goal in the work reported here is to investigate the interaction between the ambiguity of the learning context and learners' underlying representations during cross-situational word learning. We hope to understand how aspects of the labeling event constrain the input to statistical learning mechanisms. To accomplish this, we vary the ambiguity of the learning context by manipulating the presence of a valid referential

cue – a speaker's gaze. This manipulation is inspired by social-pragmatic theories of language acquisition that emphasize the importance of using referential cues to infer word meanings (Bloom, 2002; Clark, 2009). Moreover, experimental work shows that even very young children are sophisticated intention-readers, preferring to map novel words to objects that are the target of a speaker's gaze and not their own (Baldwin, 1993). Together, the evidence suggests that referential cues could alter the underlying representations by focusing attention and memory to a subset of the statistics in the input.

In the current set of studies, we use a modified version of Yurovsky & Frank (under review)'s paradigm to provide a direct test of the hypothesis that the presence of a social cue to intent, a speaker's gaze, will alter learners' allocation of attention and reduce the number of word-object links that are stored in memory. In Experiment 1, we manipulate 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 learners were less likely to track multiple word-object links when referential cues were present. In Experiment 2, we replicate the findings from Experiment 1 with a more ecologically valid stimulus set. In Experiment 3, we show that reducing the reliability of the referential cue increases learners multiple hypothesis tracking, providing evidence that learners were sensitive to a graded manipulation of the quality of the learning context. In sum, the data suggest that learners adaptively allocate attention and store representations with different levels of fidelity depending on the amount of referential uncertainty present during learning.

2 Experiment 1

We set out to test the effects of referential cues on cross-situational learning at different levels of attention and memory demands. Participants saw a series of ambiguous exposure trials that consisted of a set of novel objects (either 2, 4, 6, or 8) and an

image of a schematic, female interlocutor. On each trial they heard one novel word that was either paired with an gaze cue or not, and were asked to make guesses about which object went with each word. In subsequent test trials, participants heard the novel word again after different numbers of intervening trials (0, 1, 3, and 7), this time paired with a new set of novel objects. Importantly, test trials were contingent upon participants' selection during exposure such that one of the objects in the set was either the participant's initial guess (Same trials) or one of the objects that was *not* the initial guess (Switch trials). While both single and multiple referent trackers could succeed on Same trials, only participants who encoded multiple word-object links during their first encounter could succeed on Switch trials. This provides a direct measure of whether learners track multiple alternatives and if these representations are influenced by the presence of referential cues.

2.1 Method

2.1.1 Participants

This experiment was posted to Amazon Mechanical Turk as a set of Human Intelligence Tasks (HITs) to be completed only by participants with US IP addresses and an approval rate above 95%. Each HIT paid 30 cents. Approximately 50-130 HITs were posted for each of the 32 conditions (4 referents X 4 intervals X 2 gaze conditions) for total of approximately 2400 paid HITs. If a participant completed the experiment more than once, he or she was paid each time but only data from the first HITs completion was included in the final data set. In addition, data was excluded from the final sample if participants did not give correct answers for familiar trials (5 HITs excluded).

2.1.2 Stimuli

Figure 1 shows stimuli used in Experiment 1. These stimuli consisted of black and white pictures of familiar and novel objects drawn from the set of 140 first used in Kanwisher,

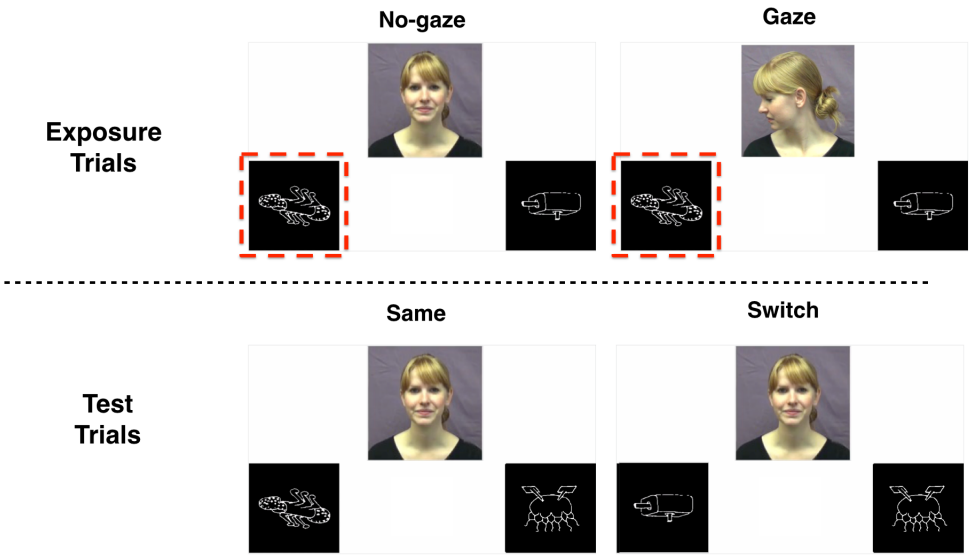

Figure 1: Examples of exposure and test trials. Participants saw exposure trials with or without a gaze cue depending on condition assignment. All participants saw both types of test trials: same and switch. On same trials the object that participants chose during exposure appeared with a new novel object. On switch trials the object that participants did not choose appeared with a new novel object.

Woods, Iacoboni, & Mazziotta (1997), a schematic drawing of a human interlocutor, and audio recordings of familiar and novel words. Familiar words consisted of the labels for the familiar objects as produced by AT&T Natural Voices ™(voice: Crystal). Novel words were 1-3 syllable pseudowords obeying the rules of English phonotactics produced using the same speech synthesizer. 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.

2.1.3 Design and Procedure

Participants were exposed to a series of 16 trials (8 exposure, and 8 test) in which they heard a speaker say one novel word, saw a set of novel objects, and were asked to guess which object went with the word. After a written explanation of the task, participants completed four practice trials that consisted of familiar words and objects. These trials also served to screen for participants who did not have their audio enabled or who were not attending to the task.

After the practice trials, participants were informed that they would now hear novel words, and see novel objects, and that they should continue selecting the correct referent for each word. Participants heard eight novel words twice, one exposure 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 appeared again amongst a set of new objects. The other four were *Switch* trials in which one of the objects in the set was selected randomly from the objects that the participant did not select on the previous exposure trial. All other objects were completely novel on each trial.

Participants were randomly assigned to see either 2, 4, 6, or 8 referents on each trial and to have either 0, 1, 3, or 7 trials in between exposure and test. Participants were also 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. On test trials, gaze was never informative. To

indicate that participants' selections had been registered, a red dashed box appeared around the object they selected for one second after their click was received. This box appeared around the selected object whether or not it was the "correct" referent.

2.2 Results and Discussion

2.2.1 Exposure trials

To ensure that our referential cue manipulation was effective we compared participant's performance on Exposure trials in the gaze condition against the distribution expected if participants were selecting randomly (defined by a Binomial guessing model with four trials and a probability of success of $\frac{1}{NumReferents}$). In all conditions, participants' responses differed from those expected by chance, exact binomial p(two-tailed) $< .001$, suggesting that gaze effectively directed participants' attention to the target referent.

We also analyzed participants' response times on exposure trials, which were self-paced and thus a proxy for attention allocated to the referents on the screen. Panel A of Figure 2 shows participants' response times across all conditions. We fit a linear mixed effects model[2] as follows: $RT \sim Gaze \times Log(Interval) \times Log(Referents) + (Trial\,type \mid subject)$. We found a significant main effect of referents ($\beta = 806.95$, p $< .001$) with slower responses as the number of referents increased. We also found a significant two-way interaction between gaze condition and number of referents ($\beta = -517.43$, p $< .001$) such that responses were faster in the gaze condition, especially as the number of referents increased. This analysis provides evidence that the presence of a referential cue focused participants' attention away from alternative word-object links.

[2] All models were fit using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2013).

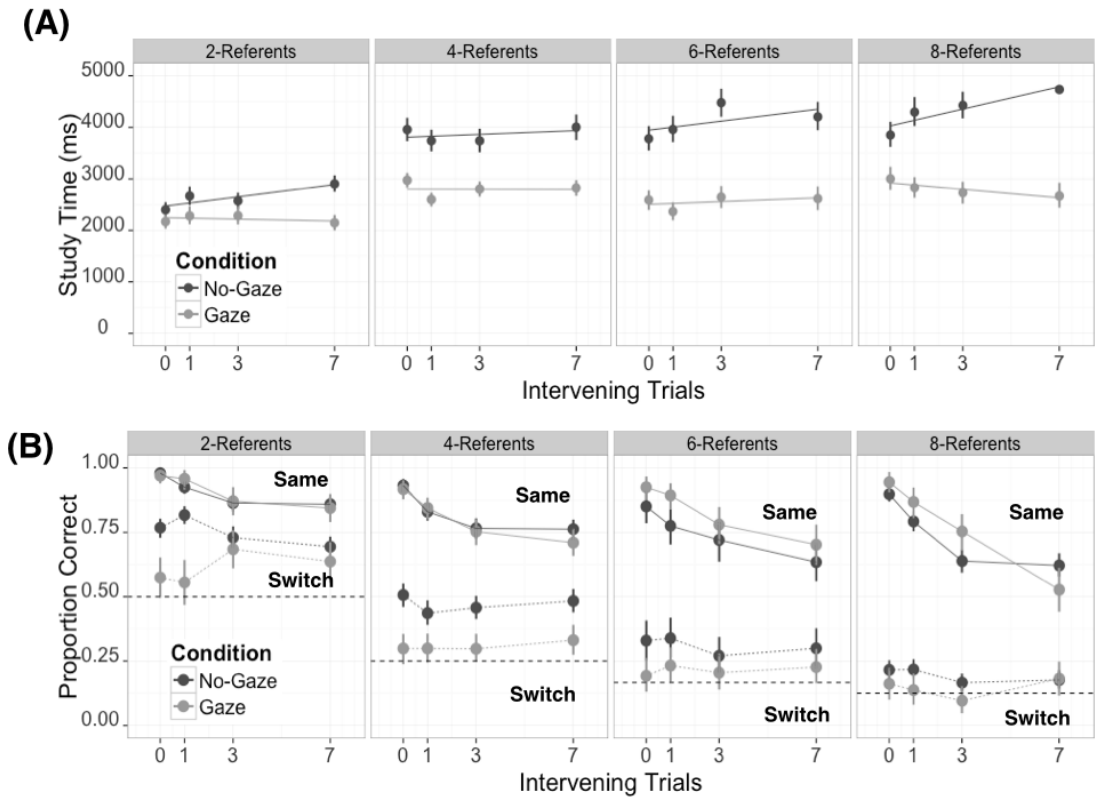

Figure 2: Experiment 1 results. Panel A shows study time on exposure trials across all experimental conditions: gaze and no-gaze, number of referents (2, 4, 6, and 8), and number of intervening trials (0, 1, 3, and 7). Panel B shows accuracy on test trials for both trial types (Same and Switch) across all conditions. Each datapoint represents approximately 35-130 participants. Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

2.2.2 Test trials

To analyze performance on Test trials, we compared the distribution of correct responses made by each participant to the distribution expected if participants were selecting randomly. Figure 2 shows participants' accuracies in identifying the referent of each word in all conditions for both kinds of trials (Same and Switch) and in each referential cue condition (Gaze and No-gaze). We replicate the finding from Yurovsky & Frank (under review): at all Referent and Interval levels, both for Same and for Switch trials, participants' responses differed from those expected by chance (smallest χ^2 2(4) = 12.07, p < .01). Participants' success on Switch trials provides direct evidence that learners encoded more than a single hypothesis in ambiguous word learning situations, even under high attentional and memory demands.

To quantify the effect of each factor on the likelihood of a correct response, we fit a mixed-effects logistic regression model to the full dataset as follows: $Accuracy \sim Trialtype \times Gaze \times Log(Interval) \times Log(Referents) + (Trial\,type \mid subject)$. We found significant main effects of number of referents ($\beta = -0.66$, p < .001) and interval ($\beta = -0.49$, p < .001), such that as each of these factors increased, accuracy on test trials decreased. We also found significant main effects of trial type ($\beta = -1.3$, p < .001), with worse performance on switch trials.

Next we examined the interactions between each factor. There were significant two-way interactions between trial type and interval ($\beta = 0.32$, p < .001) and trial type and number of referents ($\beta = -0.69$, p < .001) such that the interval between exposure and test affected same trials more than switch trials, and the number of referents affected switch trials more than same trials. The two-way interaction between trial type and gaze condition was not significant in this model, but trended in the correct direction ($\beta = -0.52$, p = 0.13).

The interaction between trial type and gaze condition is the critical test of our hypothesis because it shows that the presence of a referential cue reduced learners

multiple hypothesis tracking, resulting in lower accuracy scores on switch trials. But we would only expect to see this effect if participants were actually following the gaze cue on exposure trials. Thus, we fit a mixed-effects logistic regression to a filtered dataset, removing those participants who were not reliably selecting the referent that was the target of gaze on exposure trials. This filter removed 90 participants who selected the gaze target at below chance levels on exposure trials. The analysis of the filtered dataset showed a reliable interaction between trial type and gaze condition ($\beta = -0.8$, $p < .05$), with worse performance on switch trials in the gaze condition.

Taken together, the response time and accuracy analyses provide evidence that learners use of a referential cue modulated their attention during learning, thus making them less likely to track multiple word-object links. Interestingly, we did not see strong evidence that reduced tracking of alternatives resulted in improved performance on same trials. This finding suggests that the limitations on same trials may be different than those regulating the distribution of attention on switch trials, since the presence of a referential cue selectively reduced learners tracking of alternatives but apparently did not lead learners to form a stronger memory of their single candidate hypothesis.

It is important to point out that participants' tendency to *use* the referential cue modulated the effect of gaze on attention and memory in this experiment. It is interesting that a subset of the particpants did not reliably use gaze to make their selections on exposure trials. Perhaps, the effect of gaze was reduced in our experiment 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. Hence, we do not yet know whether learners' cross-situational learning change in the presence of a stronger and more ecologically valid referential cue. Our next experiment tests this hypothesis.

3 Experiment 2

In Experiment 2, we attempt to replicate the findings from Experiment 1 using a more ecologically valid stimulus set. To move closer to a real word referential cue, we replaced the static, schematic gaze cue with a live actress and introduced a within-subjects design where each participant saw both gaze and no-gaze exposure trials. We selected a subset of conditions from Experiment 1, testing only the 4-referent display with 0 and 3 intervening trials as between-subjects manipulations. Our goals were to replicate the reduction in learners' multiple hypothesis tracking in the presence of referential cues, and to test whether increasing the ecological validity of the cue would result in a boost to the strength of learners' recall of their single candidate hypothesis.

3.1 Method

3.1.1 Participants

Participant recruitment and inclusionary/exclusionary criteria were identical to those of Experiment 1 (excluded 36 HITs). 100 HITs were posted for each condition (1 referent X 2 intervals X 2 gaze conditions) for total of 400 paid HITs.

3.1.2 Stimuli

Audio and picture stimuli were identical to Experiment 1. The referential cue in the gaze condition was a film of a live actress (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. That is, participants saw exposure trials with and without gaze. The experiment consisted of 32 trials broken down into 2 blocks of 16 trials. Each block consisted of 8 exposure trials and 8 test trials (4 same test trials and 4 switch test 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. First, we analyze performance on exposure trials to ensure that participants were using the referential cue and to test if the cue changed response times. Then we analyze performance on test trials to measure the effect of the presence of referential cues on the number of word-object links learners stored in memory.

3.3.1 Exposure trials

Similar to Experiment 1, participants' responses on exposure trials differed from those expected by chance, exact binomial p(two-tailed) $< .001$, suggesting that gaze was effective in directing attention to the target referent. Participants in Experiment 2 were numerically more consistent in their use of gaze with the live action stimuli compared to the schematic stimuli used in Experiment 1 (M1 = .76, M2 = .81). Panel A of figure 3 shows participants' response times. We replicate the effect of effect gaze, with faster response times in gaze condition. We fit a linear mixed effects model to response times with the same specification as Experiment 1, finding main effects for gaze condition ($\beta = -1112.83$, p $< .001$) and interval ($\beta = -498.96$, p $< .001$) with faster responses in the gaze condition and in the the longer interval condition. The two-way interaction

between gaze condition and interval was not significant, showing that gaze had the same effect on participants' response times at both intervals.

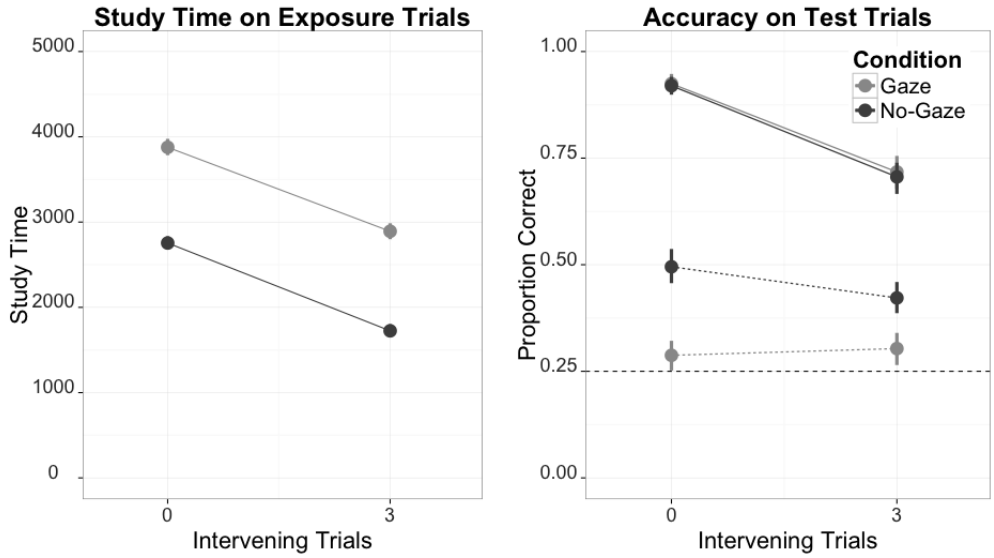

Figure 3: Experiment 2 results. Panel A shows study times for exposure trials. Panel B shows accuracy on test trials. Each datapoint represents 182 participants. Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

3.3.2 Test trials

Panel B of figure 3 shows performance on test trials in Experiment 2. We replicate the critical finding from Experiment 1: after seeing gaze exposure exposure trials participants stored fewer word-object links and performed worse on switch trials. We fit a mixed-effects logistic regression model and found significant main effects of interval ($\beta = -0.59$, $p < .001$) and trial type ($\beta = -2.74$, $p < .001$). Participants were less accurate as the interval between exposure and test increased and on the switch trials overall.

In addition, the model showed a significant two-way interaction between trial type and interval ($\beta = 0.51$, $p < .001$), with worse performance on switch trials at the

higher interval, and a marginal two-way interaction betwen gaze condition and interval ($\beta = 0.09$, p = 0.07) such that the number of intervening trials had a smaller effect on participants' performance in the gaze condition. Importantly, the model showed a reliable interaction between gaze condition and trial type ($\beta = -0.73$, $p < .001$) with switch trials being more difficult after gaze exposure trials. [3] Similar to Experiment 1, we did not find evidence of a boost to performance on same trials in the gaze condition.

Taken together, the data from Experiment 1 and 2 suggest that the presence of a referential cue reliably focuses learners' attention away from alternative word-object links and shifts them towards single hypothesis tracking strategy. Changing to a live action stimulus set led to slightly higher rates of selecting the target of gaze on exposure trials, but did not result in a boost to performance on Same trials, providing additional evidence that the fidelity of participants' single hypothesis was unaffected in our paradigm by the presence of a referential cue.

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 contexts were either entirely ambiguous or entirely certain. However, 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 learning contexts? In our next experiment, we test whether learners store different amounts of information in response to a *graded* manipulation of referential uncertainty. This provides an important test for our account that learners can move along a continuum of cross-situational word learning strategies, from single to multiple hypothesis tracking.

[3] We fit models to both the unfiltered and filtered datasets and found no difference between the two analyses, suggesting that increasing the ecological validity of the referential cue and switching to a within-subjects design reduced noise in our measurements.

4 Experiment 3

In Experiment 3, we explore the effects of a parametric manipulation of referential cues on cross-situational word learning. To accomplish this, we varied the reliability of the gaze cue. This design was inspired by experimental work showing that learners are sensitive to the prior reliability of speakers and use this information when deciding whom to ask for new information (Koenig, Clément, & Harris, 2004). By parametrically manipulating reliability, we hoped to test a clear prediction of our account: that learners will allocate attention and memory rationally in response to graded changes in the referential uncertainty present during learning.

4.1 Method

4.1.1 Participants

Participant recruitment, and inclusionary/exclusionary criteria were identical to those of Experiment 1 and 2 (excluded 4 HITs). 50 HITs were posted for each reliability level (0%, 25%, 50%, 75%, and 100%) for total of 250 paid HITs.

4.1.2 Design and Procedure

Procedures were identical to those of Experiment 1 and 2. We modified our cross-situational learning paradigm to include a block of 16 familiarization trials (8 exposure and 8 test), which established the reliability of the referential cue. To establish reliability, we varied the proportion of same/switch trials that occurred during this familiarization block. Switch trials provide evidence that gaze is not a reliable predictor of the object that will appear at test. Participants either saw 0, 2, 4, 6, or 8 switch trials. After the familiarization block, participants completed a block of 16 trials (8 exposure and 8 test). Importantly, since we were no longer testing the effect of presence or absence

of referential cues, all exposure trials in Experiment 3 included gaze, but this cue was more or less reliable depending on which familiarization block participants saw.

4.2 Results and Discussion

4.2.1 Exposure trials

First, we analyzed exposure trials across both familiarization and test blocks. Similar to Experiments 1 and 2, as a group participants reliably chose the referent that was the target of gaze at rates greater than those that would be predicted by a guessing model p(two-tailed) $< .001$. Next we analyzed exposure trials in the post-familiarization block to see if participants were sensitive to our reliability manipulation. We fit a mixed effects linear regression model and found a significant effect of reliability level ($\beta = 1.39$, $p < .001$) such that as the reliability of the gaze cue during familiarization increased, participants were more likely to select the target of gaze on exposure trials in the post-familiarization block. Thus, learners were sensitive to the reliability of the gaze cue.

4.2.2 Test trials

Figure 4 shows participants accuracy on test trials within the test block. To quantify the effect of reliability on accuracy, we fit a mixed-effects logistic regression model and found a significant main effect of trial type ($\beta = -2.25$, $p < .001$), with participants responding less accurately on switch trials. In this analysis, we found a trend towards a significant interaction between reliability and trial type ($\beta = -0.72$, $p = 0.11$).

Similar to Experiment 1, we would only expect to see an interaction between reliability and trial type if learners used the gaze cue during exposure trials. Thus, we conducted a follow-up analysis where we modeled accuracy on test trials as a function of how often participants chose the target of gaze on exposure trials. We fit a mixed

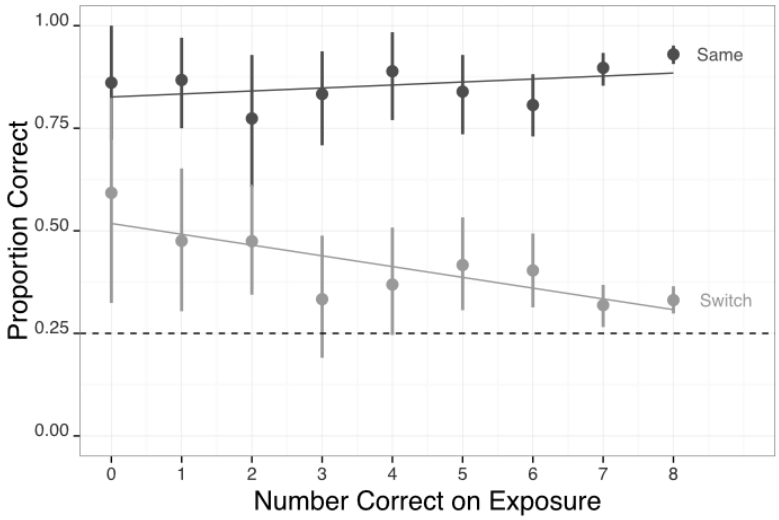

Figure 4: Accuracy on test trials in Experiment 3 for both trial types (Same and Switch) as a function of participants' use of gaze on exposure trials. Each datapoint represents approximately 50 participants. Error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

effects logistic regression model with the same specifications, but substituting accuracy on exposure trials for reliability condition as a predictor. With this analysis we found a robust two-way interaction between performance 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.[4] These analyses show that as a referential cue becomes more reliable, participants were more likely to use it, and that learners who used the referential cue were less likely to store multiple word-object links.

The findings from Experiment 3 support and extend the results of Experiments 1 and 2 in several ways. First, participants' performance on same trials was again relatively unaffected by changes in peformance on switch trials. This provides converging evidence that the limitations on same trials may be different than those regulating the distribution of attention on switch trials. Second, it was learners' *use* of a referential cue that predicted a reduction in memory for alternative word-object links. It is important to note that although we found a significant effect of reliability on participants' use of the gaze cue, participants still tendency to use the cue remained high, even in the 0% reliability condition ($M = 0.81$). It is reasonable that participants would continue to use the gaze cue in our experiment because it was the only cue available and participants did not have reason to think the speaker would be deceptive.

The critical contribution of Experiment 3 is the finding that learners respond to a *graded* manipulation of referential uncertainty, with the amount of information stored from the intital exposure tracking with both the reliability of the cue and participants' use of the cue. This provides support for a critical prediction of our account: that learners store a strong single candidate word meaning and a set of alternatives with different levels of fidelity depending the amount of referential uncertainty present during the initial exposure to a word.

[4]Initially, this analysis was post-hoc. So we ran a follow-up study, with all results reported here from a planned analysis of that follow-up experiment.

5 General Discussion

An ideal statistical learner with unlimited attention and memory could track all possible word-object co-occurrences, making cross-situational word learning a simple problem of getting enough data points. But human learners are constrained by limited cognitive resources, making it important to decide which statistics to store from an individual learning moment. Models of cross-situational learning disagree about how much information is actually stored in memory, with recent work suggesting that learners maintain both a strong candidate hypothesis and a set of weaker alternative word-object links (Yurovsky & Frank, under review).

Our results suggest that the representations underlying cross-situational learning are quite flexible. In the absence of a referential cue to word meaning, learners were more likely to store alternative word-object links. In contrast, when gaze was present, they 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 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). Interestingly, across all three experiments, reduced memory for alternative hypotheses did not result in a boost in 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 learners candidate hypothesis.

Why did we not see an increase in the strength of learners' candidate hypothesis? One possibility is that participants did not shift their cognitive resources from the set of alternatives to their single hypothesis, but instead rationally conserved their resources for future use. Griffiths, Lieder, & 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., opportunity cost of time or mental opportunity) into account. For example, Vul, Goodman, Griffiths, & 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. Here, we show evidence that learners adapt their allocation of cognitive resources to the level of referential uncertainty in the learning context, spending less time studying alternative word-object links and reducing the number of links stored in memory when uncertainty is low.

Our results also fit well with recent experimental work that investigates how attention and memory can constrain infants' statistical word learning. For example, Smith & 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 & 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 data that comes into contact with the learning system can be used for cross-situational word learning, and this data is directly influenced by the attention and memory constraints of the learner. Our findings suggest that referential cues could play an important role in focusing learners' limited attention to the relevant statistics in the input.

How should we characterize the effect of social information 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 mechansims (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 (M. C. Frank, Goodman, & Tenenbaum, 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 scenario, 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.

There are several limitations to the current study that are worth noting. First, the social context we used was relatively impoverished. Although we moved beyond a simple manipulation of the presence or absence of social information, we 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, M. C. Frank, Tenenbaum, & 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, because no single cue was a consistent predictor of reference in their corpus. In our data, we did see a more reliable effect of referential cues when we used a live 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 real-world learning environments that contain a rich combination of social cues.

Second, we do not yet know how these results would generalize to young word learners. Research with infants' shows rapid development of visual attention and memory in the first years of life (Colombo, 2001; Ross-sheehy, Oakes, & Luck, 2003). Moreover, experimental work shows that infants' attention is often stimulus driven and

sticky (Oakes, 2011), suggesting that very young word learners might not effectively explore the visual scene to extract the necessary statistics for effective cross-situational word learning. Our findings suggest that referential cues might play even more of an important role in overcoming young learners' limited and sticky attention, guiding them to the relevant statistics in the input.

And third, in the current experiments we tested a minimal cross-situational learning scenario. Our task contained only one exposure for each novel word-object pairing. In contrast, real world naming events are best characterized by discourse, where an object is likely to be named repeatedly in a short amount of time (M. C. Frank et al., 2013; Rohde & Frank, 2014). Thus, we need more evidence to understand how learners flexibly allocate cognitive resources in response to referential uncertainty at different timescales that more accurately reflect language learning environments.

Word learning proceeds despite the potential for high levels of referential uncertainty and 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 store more word-object links. Overall, these results bring together aspects of both social and statistical accounts of word learning, and increase our understanding of how statistical learning mechanisms operate over social input.

6 Acknowledgements

We are grateful 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 and an NIH NRSA Postdoctoral Fellowship to DY.

References

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

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.

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.

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.

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.

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.

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

Marr, D. (1982). Vision: A computational approach. Freeman & Co., San Francisco.

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. (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.

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

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (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.

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13), 2149–2165.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420.

Yurovsky, D., & Frank, M. (under review). An integrative account of constraints on cross-situational learning.

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.