Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

*J Mem Lang*. Author manuscript; available in PMC 2017 April 01.

Published in final edited form as:

*J Mem Lang*. 2016 April 1; 87: 105–127. doi:10.1016/j.jml.2015.09.005.

# Observational Word Learning: Beyond Propose-But-Verify and Associative Bean Counting

**Tanja Roembke** and

Dept. of Psychological and Brain Sciences University of Iowa

**Bob McMurray**

Dept. of Psychological and Brain Sciences Dept. of Communication Sciences and Disorders Dept. of Linguistics University of Iowa

## Abstract

Learning new words is difficult. In any naming situation, there are multiple possible interpretations of a novel word. Recent approaches suggest that learners may solve this problem by tracking co-occurrence statistics between words and referents across multiple naming situations (e.g. Yu & Smith, 2007), overcoming the ambiguity in any one situation. Yet, there remains debate around the underlying mechanisms. We conducted two experiments in which learners acquired eight word-object mappings using cross-situational statistics while eye-movements were tracked. These addressed four unresolved questions regarding the learning mechanism. First, eye-movements during learning showed evidence that listeners maintain multiple hypotheses for a given word and bring them all to bear in the moment of naming. Second, trial-by-trial analyses of accuracy suggested that listeners accumulate continuous statistics about word/object mappings, over and above prior hypotheses they have about a word. Third, consistent, probabilistic context can impede learning, as false associations between words and highly co-occurring referents are formed. Finally, a number of factors not previously considered in prior analysis impact observational word learning: knowledge of the foils, spatial consistency of the target object, and the number of trials between presentations of the same word. This evidence suggests that observational word learning may derive from a combination of gradual statistical or associative learning mechanisms and more rapid real-time processes such as competition, mutual exclusivity and even inference or hypothesis testing.

### Keywords

Observational learning; cross-situational learning; associative learning; word learning; statistical learning; eye movements

**Corresponding Author** Tanja Roembke, E11 SSH, Dept. of Psychological and Brain Sciences, University of Iowa, Iowa City, IA 52242, 319-335-0692 (voice), 319-335-0191 (fax), tanja-roembke@uiowa.edu.

## 1.1 Observational learning and referential ambiguity

Early in language acquisition, children are often assumed to learn the mapping between words and objects largely from observation (Gleitman, 1990) without reliable feedback. However, a fundamental problem for observational learning is referential ambiguity (Quine, 1960): in any naming event, there is a vast array of possible interpretations for a novel word. Consequently, learners may require strategies or biases to cope with this ambiguity (Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Markman, 1990). Recently, Yu and Smith (2007; see also Siskind, 1996) argued the problem of referential ambiguity may in part be an artificial consequence of restricting the analysis of word learning to one encounter with a word. Across multiple situations, there may be sufficient statistical information to support learning. For example, many words (e.g., objects) are more likely to co-occur with their referents than with other objects.

Yu and Smith (2007) tested this in adults (and later in infants, Smith & Yu, 2008): On each trial, participants saw a number of novel objects and heard novel names for each of them, creating considerable ambiguity. Across multiple trials, a word and its referent always co-occurred while its co-occurrence with other objects was lower. After a short training, participants showed above-change accuracy for selecting the words' referents, suggesting statistics were sufficient to support learning. This raises the possibility that learners have powerful mechanisms for inferring the words' meanings across multiple situations, even if any given situation is ambiguous.

## 1.2 How do people learn words in the cross-situational paradigm?

There has since been a large number of experiments examining how mostly adults learn words in observational paradigms (Medina, Snedeker, Trueswell, & Gleitman, 2011; Trueswell et al., 2013; Vouloumanos, 2008; Yurovsky, Yu, & Smith, 2013). This has led to a debate over the mechanism underlying such learning.

Originally, Yu and Smith (2007, 2012) described cross-situational learning as a process of tracking co-occurrence statistics between words and objects across many situations. This is a form of statistical or associative learning in which the word-object pairs with the highest co-occurrence are the correct mapping. However, more recent accounts suggest people could harness cross-situational information using propositional logic (Medina et al., 2011; Trueswell et al., 2013): The most prominent theory of this sort is "propose-but-verify", in which learners form a single explicit hypothesis after encountering a novel word, which is carried forward unless disconfirmed by later encounters.

Others have proposed hybrid accounts: For example, there are memory-based accounts in which such inferences are made over stored episodes of situations in long-term memory (Dautriche & Chemla, 2014). Bayesian accounts take a hypothesis-testing approach, but evaluate multiple probabilistic hypotheses simultaneously to find the most likely mapping given the data (Frank, Goodman, & Tenenbaum, 2009). Finally, McMurray, Horst and Samuelson (2012) propose that gradual associative learning may be buttressed with real-time decision making to account for both cross-situational learning and other developmental phenomena. These real-time processes may allow the system to engage in more inferential processes in the moment (e.g., mutual exclusivity), while long-term statistics are tracked via associations.

These theories are still developing with newer iterations of purely statistical accounts (Yu & Smith, 2012), propose-but-verify (Koehne, Trueswell, & Gleitman, 2014) and the dynamic associative account (McMurray, Zhao, Kucker, & Samuelson, 2013). While these theories may exhibit stark differences in their core commitments (e.g., whether learning is propositional or associative), they appear flexible in how these commitments get implemented. Consequently, it may be premature to experimentally disentangle them.

However, there are crucial open questions about the basic properties of observational learning, which may constrain how these theories are developed. Thus, we identified four such questions that have played (or may play) a crucial role in these debates and critically evaluated them across two experiments. These questions include the issues of 1) whether participants maintain multiple hypotheses for a given word[1]; 2) whether information is gradually accumulated; 3) the role of context, and 4) other factors that may shape learning.

### 1.2.1 Do learners maintain multiple hypotheses about the meaning of a word?

The first question is how many hypotheses learners maintain for a given word. For example, in a dinner table event, when *fork* is heard for the first time, do learners form a single hypothesis for *fork* (positing that it refers to *either* the fork or the spoon), or do they note that this word co-occurred with both objects (but not with a car or boat)? In an associative account, learners track the co-occurrence of multiple objects with a word (e.g. Yu & Smith, 2007), relying on the accumulation of data to resolve any ambiguity. At the dinner table, for example, the learner will eventually encounter the word *fork* without a spoon, pushing its statistical co-occurrence with fork above that with spoon. Consequently, learners must maintain multiple hypotheses with different degrees of strength. In contrast, early versions of propose-but-verify suggested learners posit a *single* hypothesis about a word, which can be updated on future encounters. However, more recent propositional accounts also admit multiple hypotheses: For example, learners may recall previously considered hypotheses in the face of memory failure or disconfirming evidence (e.g. Koehne, Trueswell, & Gleitman, 2014).

As an empirical issue, whether learners track one or many hypotheses remains unresolved. This is largely because most studies address this issue indirectly using trial-by-trial autocorrelation analyses. Such analyses infer what a learner may have learned about a word from previous trials' accuracy, and measure how it predicts performance on subsequent encounters (Trueswell et al., 2013): In propositional accounts, if learners previously selected the correct object, they must have arrived at the right hypothesis and should continue to select the correct object on present trials. However, if they were incorrect on a previous trial, they likely had the wrong hypothesis, and should now be at chance. In contrast, in statistical accounts, even on an incorrect trial, they accumulate more "data" and could show a benefit on subsequent trials. Autocorrelation analyses conducted by Trueswell et al. (2013)

[1]We here use the term *hypothesis* here to refer to any knowledge structure mapping a word to potential referents, including both abstract knowledge and associative links.

supported a single-hypothesis account, and even an analysis of participants' eye-movements (a potentially more sensitive measure) showed little evidence for any learning after an incorrect trial.

Dautriche and Chemla (2014) pointed out that prior incorrect trials may function differently depending on the information on the current trial: If the prior incorrect selection is present, people may continue to select it and be *below* chance. Trueswell et al.'s (2013) choice of four foils made it likely that prior selections were repeated, leading them to potentially underestimate what people were learning from incorrect trials. Dautriche and Chemla (2014) decreased the number of foils, and found that people were now above chance in selecting the correct referent even after an incorrect previous encounter.

This offers tentative evidence that learners track multiple hypotheses for a given word. However, these experiments have several shortcomings: First, as Dautriche and Chemla (2014) point out, indirectly inferring what people might know on a previous trial from their overt response(s) oversimplifies the complex mapping between prior and present trials. Moreover, these analyses assume that prior accuracy is a robust (and uniform) index of knowledge. However, early in training a correct response may be due to chance, and response accuracy is therefore confounded by a trial's position in the learning curve. Thus, the trial-by-trial analyses of Trueswell et al. (2013) and Dautriche and Chemla (2014) may not be sufficient to evaluate the claim of multiple hypotheses.

Moreover, this approach fails to address a second important question. If listeners are retaining multiple hypotheses, what do they do with them? Prior studies focus on whether multiple hypotheses are retained, but they do not address whether these hypotheses are also simultaneously *activated* in the moment when a novel word is processed.

What is needed is a more direct measure that addresses what listeners bring to bear in the moment: On a single trial, are multiple objects under *active* consideration as referents for a word? The present study achieves this by examining eye-movements to potential referents relative to the participant's response. If the participant clicks on the correct referent but simultaneously fixates a second referent (more than some baseline), this offers strong evidence that multiple hypotheses are not only tracked, but influence behavior simultaneously.

### 1.2.2 What do learners carry forward from prior encounters with a word?

A related issue is whether listeners gradually accumulate information across trials. For example, after a third and fourth encounter with *fork*, do listeners have stronger associations or more confidence than after the first (even if all encounters favored the correct interpretation)? If so, learners may even accumulate evidence from ambiguous encounters (e.g., the *fork*/ *spoon* example).

Evidence for this comes from Yurovsky, Fricker, Yu and Smith (2014), in which adults first learned a small set of words. Then, words that were still at chance received additional training in a second phase along with new words. This initial exposure, even though it seemed to yield no measurable learning for the original words, improved learning in the second phase for *all* words. This suggests that learners must have acquired some partial knowledge about the original words during the first phase despite at-chance performance.

This contrast with Trueswell et al.'s (2013) autocorrelation analyses showing that if learners were incorrect on a prior encounter with a word they were at chance on subsequent trials (though see Dautriche and Chemla, 2014). However, this style of analysis focuses largely on a single type of information that could be retained - the prior responses to a word (and by inference, hypotheses). .

The converse of this question - whether listeners gradually accumulate statistics about words and objects - has not been properly examined for two reasons. First, the training paradigms themselves were very short (e.g. five repetitions of each word in Trueswell et al., 2013). Statistical and associative accounts are most likely to be accurate when the contribution of any given trial is small (to avoid over-committing to an erroneous prediction). Consequently under these theories, any single trial's contribution could be small, and a few repetitions may be sufficient to see these effects.

Second, many of these studies did not include analyses that tested for gradual learning. There is a statistical confound with the most versions of the autocorrelation analyses: Trials in which participants were incorrect most likely came from early portions of training , while correct trials were most likely later in training. Thus, a comparison of accuracy as a function of last-encounter performance would be heavily confounded with position in the learning curve. Again, this makes it very difficult to see any effects of gradual learning.

The solution is to simultaneously examine the effect of last-encounter performance and the effect of the number of prior exposures to the word (the contribution of gradual learning) (c.f. Wasserman, Brooks, & McMurray, 2015). This latter factor serves as a covariate to account for where the participant is in the learning curve, and simultaneously offers a statistical test of the gradual learning. The only example of such an analysis we are aware of is Experiment 3 of Trueswell et al. (2013) which found no effect of the number of encounters with a word (though there was an interaction with last-encounter performance). But, as noted earlier, five trials may not be sufficient to observe gradual learning effects. By coupling a positive statistical test of gradual learning with a longer training period, we may find clearer evidence for the influence of the gradual accumulation of statistics on single trial accuracy.

Our goal here was to evaluate whether the amount of exposure predicts accuracy, over and above prior-trial performance (while simultaneously accounting for the statistical confound in the influence of prior trial accuracy). Thus, we replicated Trueswell et al.'s (2013) analyses adding the number of trials as a factor.

### 1.2.3 How does context influence observational word learning?

An additional factor of recent interest is context. Many words appear in consistent contexts (the kitchen, a farm, etc.), and while context does not tell the listener what object a word refers to, it can have complex effects on learning: For example, if the learner knows that *fork* is a kitchen word, he or she can rule out referents that do not fall into that category (e.g., a dog), even if they do not precisely know which of the remaining ones is a *fork*. Under a statistical account, however, context may also impede learning by raising the likelihood of spurious correlations. The fact that forks and spoons frequently appear together, for instance, means that the word *fork* may be linked to both objects.

Dautriche and Chemla (2014) manipulated contextual consistency by presenting a set of word-referent pairs as a consistent context in the first block of trials. For example, trials 1-4 may have included the objects *dog*, *cat*, *rabbit*, *cow*, with each one the target on a trial. This established a set (all four are animals) that then served as context. This contextual manipulation improved learning on subsequent blocks when words were presented "out of context" (with other competitors). Later experiments made the grouping of the objects into contexts completely arbitrary, with similar results. This suggests context serves as a memory cue for ruling out competing hypotheses in the moment. Moreover, this appears to challenge associative or statistical account.

However, Dautriche and Chemla's (2014) context manipulation creates potential benefits for context without the statistical costs. Context only occurred on one block of trials with the same four competitors. Subsequent presentations of a word had random foils, meaning that context was very salient and did not create spurious correlations that could have hurt learning. It is unclear how behavior would be affected if contextual consistency was manipulated across the whole experiment in a more natural, probabilistic manner. Thus, it is possible that context serves as both a source of information that can be used in real-time to eliminate competitors, even as it simultaneously exerts a cost on learning. Our third goal therefore was to investigate the potential associative *costs* that come with context. This was done by comparing learning if highly co-occurring competing objects were present throughout learning or not.

### 1.2.4 Do other factors from prior encounters with a word play a role in performance?

Finally, the autocorrelation analysis used until now focused on factors that directly propositional inference and/ or associative learning: whether the learner knew the word on the previous trial and the number of encounters. However, there may be considerably more information on a given encounter that learners could use and could reveal aspects of the mechanism. For example, in the kitchen example above, the spatial arrangement of the fork and spoon could match the current encounter or differ. Would such information matter for learning?

This was examined in an animal learning study by Wasserman et al. (2015). They taught pigeons to map 16 different categories onto 16 unique responses (in a problem analogous to word learning using an operant learning paradigm. They conducted detailed auto-correlation analyses showed both the same effect of last-trial accuracy as Trueswell et al. (2013), and unambiguous evidence for gradual learning. However, they also included a number of other factors not considered by prior work in their autocorrelation analyses: This included the spatial location of the target response (and foils) on the last encounter, how many trials lapsed between repetitions of a category, and learner's knowledge of the foils. All of these played a role in pigeon learning.

Given the similarities they report between pigeons' learning and human word learning (as well as the obvious differences), our fourth goal here is to investigate these issues, by examining a range of other factors in our autocorrelation analysis. Such findings may unify cross-situational learning with other forms of learning and memory. For example, well established effects of spacing of training trials (e.g. Ebbinghaus, 1992; Pavlik Jr & Anderson, 2005; Smith, Smith, & Blythe, 2011), predict that the number of trials between successive presentations of the target impact learning. Similarly, recent studies indicate that children's word representations are initially bound in space (e.g. Samuelson, Smith, Perry, & Spencer, 2011), suggesting the consistency of the spatial arrangement may matter. Identifying such factors at work in cross-situational learning may thus show how this paradigm maps onto broader ideas in learning and development.

## 1.3 The present study

The present study addresses these four questions across two experiments. Experiment 1 primarily asks whether people maintain multiple hypotheses and activate them in a given moment during learning (Question 1). Learners were exposed to a set of word-object mappings in which each word had two additional objects that serve as highly co-occurring foils (though they were the named target on other trials). We used a variant of the visual world paradigm (VWP) (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) during cross-situational learning to measure simultaneous activation of multiple hypotheses. Our logic was that if participants click on one object but simultaneously fixate another object *on the same trial*, this offers strong evidence that both hypotheses were not only retained (as suggested by Dautriche and Chemla, 2014), but also simultaneously *activated* in the moment as potential referents. Such an analysis was not possible in prior eye-tracking designs (e.g., Trueswell et al., 2013) which did not manipulate co-occurrence, nor condition eye-movement analyses on the response. It is important to note that generally in the VWP, looks to competitors are driven by auditory ambiguity (e.g., words that overlap phonologically) or semantic similarity. However, neither of these factors is present here, and we are filtering trials to only analyze those in which the correct object was chosen. Thus, we expected to see very small differences in looking between the high-co-occurrence competitors and the control foils.

We addressed the second question (concerning gradual learning) by replicating the Trueswell et al. (2013) analyses, and then looking for an effect of gradual learning above and beyond that. This analysis was carried out both for Experiment 1 and 2.

Comparing Experiment 1 and 2 addresses the third question, the issue of whether context can impede learning. In Experiment 2, foil co-occurrence was completely random. Here, the high-co-occurrence competitors of Experiment 1 could have either served as a contextual cue and speeded learning (as in Dautriche and Chemla, 2014) or created statistical uncertainty and slowed it. However, unlike Dautriche and Chemla (2014), this co-occurrence was repeated throughout the experiment to potentially magnify any interfering effects. By comparing the Experiments, we address the possibility that consistent context may also interfere with slower statistical learning process.

Finally, the fourth question — what information is carried forward across trials — is addressed by conducting an autocorrelation analysis on the combined data from Experiments 1 and 2. We consider a range of additional exploratory factors highlighted by Wasserman et al. (2015).like the distance between presentations of a word, spatial layout and foil knowledge.

In both experiments, the learning paradigm was similar to other cross-situational experiments with three objects presented on each trial, one of which was named. As in Trueswell et al. (2013), participants made an overt response on each trial (and they received no feedback). As described, this was crucial for both the fixation and trial-by-trial analyses. Unlike prior studies supporting propositional or memory-based accounts of learning, words were mapped to novel objects, rather than photographs of known objects to disencourage a paired-associative-process (linking novel names to existing names).

One of the most important differences between our experiments and previous studies is that we included significantly more trials (60 encounters with each word). This was motivated by one of our questions: the contribution of the gradual accumulation of information. In unsupervised statistical learning paradigms, the changes from trial to trial are likely to be very small. Hence, we were concerned that with only a few repetitions, the effect of number of exposures would be too difficult to observe. Moreover, a longer experiment was also necessary for our eye movement design: As we wanted to look at correct trials for evidence of activation of alternative hypotheses, learners needed to show very high levels of accuracy for a substantial proportion of the experiment. This is also why we only included eight words to be learned. Finally, the higher number of trials also gave us more power to carry out the extended autocorrelation analyses we used to examine our fourth question.

We acknowledge that this highly repetitive and simplified learning paradigm may be unrepresentative of everyday word learning, which likely features more ambiguity, more items and more variability. However, these decisions were based on theoretical and methodological considerations. In that regard, most cross-situational word learning studies (or word learning studies in general) make similarly reductionist assumptions, though perhaps in different ways (use of a small number of trials, mapping words to already known objects, e.g. Trueswell et al., 2013; Yu & Smith, 2007). Given these sorts of simplifications, we do not claim that these types of experiments capture the phenomenon of word learning as a whole; rather, they isolate and distill critical learning mechanisms that were not possible to investigate in previous experiments and that are likely involved in various forms of observational word learning.

## 2. Experiment 1

### 2.1 Method

**2.1.1 Participants—**Thirty-two native English speakers took part in this experiment. Participants were students at the University of Iowa. Thirty received course credit as compensation, two received gift cards worth $15. Participants underwent informed consent in accord with an IRB approved protocol.

Author Manuscript
Author Manuscript
Author Manuscript



**2.1.2 Design and materials—**Participants learned eight word-referent pairs over approximately 40 minutes. Referents were novel objects, presented on a black background (Figure 1 for examples). Words were two-syllable, CVCV pseudo words, which were phonologically legal words in English. There was no phonological overlap among any words at onset (Table 1). The specific mapping between each word and its referent was randomized for each subject at the beginning of the experiment.

During training trials, each word was strongly correlated with a single target referent (co-occurring on 100% of trials). To build "spurious" associations between the word and incorrect competitor referents, high and low co-occurrence competitors were also included. The high co-occurrence (HC) competitor was 60% likely to be seen with the target word; the low co-occurrence (LC) competitor was 40% likely to co-occur with the target word (see Table 2). HC and LC competitors were neither phonologically nor visually related to the target word/object pairs. All of the other five objects were randomly selected from trial-to-trial with a co-occurrence rate of approximately 20%. All words and objects were equally likely to appear throughout the experiment. The random objects (ROs) for each trial were chosen without replacement to avoid spuriously increasing the co-occurrence of an RO with a word.

Each trial included three objects and a single target word. To manipulate the co-occurrence of targets and competitors, we controlled frequency of four trial-types, defined by which competitors were on the screen (Table 3) relative to the target word. In HCLC trials, the target referent, the high co-occurrence competitor and the low co-occurrence competitor were present. In HC trials, the target referent, the high co-occurrence competitor and a randomly chosen object were included. In LC trials, the target referent, the low co-occurrence competitor and a randomly chosen object were present. In RO trials, the target referent was accompanied by two randomly chosen objects. The number of trials of each type was manipulated to obtain the desired co-occurrence frequencies (Table 3).

Participants responded at the end of each trial by clicking on the object that they thought mapped onto the word they heard. There was not a separate testing phase at the end of training. The experiment was separated into four blocks of 120 trials (resulting in an overall number of 480 trials). Trial-types were randomized within one block; the interval between learning instances for one word was thus completely random within block. Each block consisted of 16 HCLC trials, 56 HC trials, 32 LC trials, and 16 RO trials.

**2.1.3 Procedure—**Participants were told that their task was to discover which object goes with what word, and that on each trial, they were to indicate their best guess by clicking on that picture. Participants were also told that while we expected them to guess at the beginning, their response should become more informed over time.

At the start of each trial, three pictures were presented on a 19" monitor operating at 1280 × 1024 resolution. Simultaneously, a small blue circle appeared at the center of the screen. Participants were given 1050 msec to inspect the objects. Afterwards, the circle turned red, cueing the participant to click on it to cue the auditory stimulus. When the participant clicked on it, the red circle disappeared and the target word was played via headphones.

Participants then clicked on the picture corresponding to the word, and the trial ended. Trials were not time-limited and participants were told to take their time and perform accurately.

Pilot results suggested that participants made fewer eye-movements over the course of the experiment as they learned to identify the objects, and as they learned the exact positions the objects could be in. To minimize this reduction in fixations, objects were presented in a triangle configuration, and the orientation of this triangle was randomly selected on each trial (either two objects on top or one object on top). The location of each object was randomized across the three possible locations of a given triangle on each trial.

**2.1.4 Eye-tracking Recording and Analysis—**Eye-movements were recorded throughout the experiment using an SR Research Eyelink II head-mounted eye-tracker operating at 250 Hz. Both corneal reflection and pupil were used to obtain point of gaze whenever possible, though for some participants only good pupil readings could be obtained. At the start of the experiment, participants were calibrated with the standard 9-point display. The Eyelink II compensates for head-movements using infra-red light emitters on the edge of the computer screen and a camera on the head to track head position. This yields a real-time record of gaze in screen coordinates. The resulting fixation record was automatically parsed into saccades and fixations using the default “psychophysical” parameter set. We then combined adjacent saccades and fixations into a single “look” which started at the onset of the saccade and ended at the offset of the fixation as in prior studies (McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; McMurray, Samelson, Lee, & Tomblin, 2010). To account for noise and/ or head-drift in the eye-track record, the boundaries of the ports containing the objects were extended by 100 pixels when computing the point of gaze. No overlap between the objects resulted from this.

### 2.2 Results: Overview

For Experiment 1, we first looked at overall accuracy and how performance differed across different trial-types to partially address Question 1 (multiple hypotheses). We then turn to the eye movements using both the standard statistical paradigm as well as a ratio measure to avoid common statistical issues when analyzing eye movements in the VWP. Finally, we replicate Trueswell et al.’s (2013) autocorrelation analyzes in our final part of the results’ section (Question 2: gradual learning).

Three participants were excluded from all analyses, as their learning plateaued at 45% correct (chance = 33.33%), with no improvement in performance over time. This left 29 participants for analysis[2].

Data were analyzed separately within each trial-type for the eye movement analyses. The analyses we report here focus on the two most important trial-types, those involving the HC competitor (the HC and HCLC trials). This is because the HC trials are experimentally the most important condition, as there is the largest difference in co-occurrence between the

[2]Their accuracy was more than 2SD below the mean in block 4. Furthermore, the data of these participants showed high levels of looks to the target before the auditory file was played, indicating that participants’ selection of the target was due to chance (and not driven by what word they heard). The overall small number of correct trials as well as these early looks reinforced us to not include them in the analysis. Additionally, one participant decided to stop the experiment during block 3; these data were retained.

spuriously associated object and the random object. Moreover, as such trials made up almost half of the experiment (a natural consequence of the co-occurrence manipulation) they had the greatest power to reveal any effects. We also ran an extensive set of analyses for the LC and RO trials which are presented in supporting online material and generally showed no differences (see supplement).

**2.2.1 Accuracy—**Figure 2 shows the accuracy of the 29 participants' overt responses (across the four trial-types). Participants' accuracy was well above chance in each block of the experiment, and 79% of the participants performed above 90% correct by block 3. Overall, accuracy between the different trial-types did not differ substantially (HCLC = 83.35%; HC = 84.67%; LC = 86.66%; RO = 85.51%). Participants' reaction time (RT) decreased over blocks (overall average RT = 1480.74 msec).

Figure 3 shows the distribution of responses (both correct responses and the distribution of errors) for the HC and HCLC trials. In the HC trials (Panel A), even though participants were highly accurate (the large light gray region), when they make an error, they were more likely to select the high co-occurrence competitor (dark grey) than a randomly chosen object (black), suggesting sensitivity to the co-occurrence manipulation. A similar pattern was observed in the HCLC trials (Panel B). In summary, participants learned the word-object mappings successfully and there are hints in their pattern of overt responses (and errors) that participants were sensitive to the co-occurrence manipulation.

We next turn to the primary analysis addressing Question 1 (maintenance of multiple hypotheses): We examined the fixation record to determine if there is evidence that participants are activating two hypothesized referents for a given word on the same trial. Next, we report a focused autocorrelation analysis designed to replicate the analysis of Trueswell et al. (2013) to determine whether prior choice and number of encounters influence trial-by-trial behavior (Question 2).

### 2.3 Analysis of Fixations

Figure 4A shows the typical time course of fixations on the HC trials: after a brief period of initial uncertainty, participants' looks rapidly converge on the target. More importantly, when we consider looks to the competitors, participants are more likely to look at the HC competitor than the RO. However, this figure combines trials with different responses, potentially conflating different kind of eye-movements. Some of the increased looking to the HC competitor is likely due to trials in which participants clicked on that competitor (and hence would have fixated it heavily). Thus, we restricted our analysis of the eye movements to only trials in which participants clicked on the correct object. This is extremely conservative for two reasons. First, eye-movements reflect motor planning as well as activation dynamics (Salverda, Brown, & Tanenhaus, 2011), and we have restricted ourselves to trials in which this motor plan reflects the correct referent; this should significantly inflate target looking. Second, and perhaps more importantly, this eliminates the trials with the most robust evidence that the HC competitor was under consideration; as a result the absolute magnitude of the effect will be quite small. However, if participants are still fixating that object more than chance even as they click the target, it offers the strongest evidence that multiple hypotheses were not only maintained but that both referents were under consideration at least partially *on the same trial*.

Figure 4B and C show the time course of fixations after excluding these trials. They display just the looks to the competitor objects as a function of the two trial-types containing HC competitors (for correct trials only). In both the HC and HCLC trials there are somewhat more looks to the HC than the random object (or the LC object). In both cases, this difference appears largely late in the time course of processing.

**2.3.1 Statistical Analysis—**To examine this statistically, we conducted two mixed effects models examining the HC and HCLC trials separately. For each, we computed the average proportion of fixations to the HC competitor and to either the random object or the LC competitor in the time window between 250 and 2000ms (Figure 5). These were submitted to a linear mixed effects model using the LME4 (version 1.1-5) (Bates & Maechler, 2009), lmerTest (version 2.0-6) packages in R (R Development Core Team, 2011). The independent variables included block (1-4, centered), and object-type (HC vs. RO or the LC, contrast coded as ±0.5). Significance for the coefficients was established using the reported *t*-statistics with degrees of freedom computed by the Satterwaithe approximation. Random effects included both participant and stimulus. We computed effect sizes (*d*) by doubling the *t*-value and dividing it by the square root of the estimated degrees of freedom.

Before examining the fixed effects, we compared several models that differed on how these two random effects were implemented to determine the most appropriate one for our data. We compared a series of models with random intercepts for participant and stimulus or with random slopes of the fixed effects on participant. With only eight items, it was difficult to estimate random slopes on item, and this often resulted in very high correlations between random slopes, suggesting the model was overfitting the data; thus only models with intercepts on item were considered. The model containing random slopes of block and object-type on participant (but not their interaction), and random intercepts on stimulus, was the most conservative model (following Barr, Levy, Scheepers, & Tily, 2013) that converged; thus, it was used for analysis.

For the HC trials, we found a significant effect of block (B = −.020, SE = .0029, $t(26.8) = 6.80$, $p < .001$, $d = 2.63$). This was due to the fact that as the experiment progressed, participants were less likely to make eye movements at all as they got better at recognizing the objects and were better able to perform the task with peripheral vision (see Figure 5A). Crucially, there was a main effect of object-type (B = .0085, SE = .0039, $t(71) = 2.17$, $p = .033$, $d = 0.52$) with more fixations to the HC than the RO. The interaction between block and object-type was not significant ($p = .278$).

For the HCLC trials, the most conservative model included object-type and block as well as their interaction as random slopes on participant. We found a significant effect of block (B = −.013, SE = .0040, $t(28) = −3.19$, $p = .0035$, $d = 1.20$). No significant effect of object-type was observed ($p = .446$). However, the interaction of block and object-type was marginally significant (B = .011, SE = .005, $t(135.8) = 1.93$, $p = .056$, $d = 0.33$)[3], suggesting that participants looked more at the HC competitor (in comparison to the LC competitor) the longer the experiment lasted (c.f. Figure 5). Figure 5 suggests that the interaction is largely driven by an effect of object in block 4 of the experiment. Thus, to examine this interaction, we conducted post-hoc tests in which we split the data into block 1-3 and block 4 and examined the effect of object-type only. On blocks 1-3, there was no effect of object-type ($p$ = .917). On block 4, the effect of object-type was marginally significant (B = .021, SE = .011, $t(25.9) = 1.79$, $p = .086$, $d = 0.70$), though it is worth noting its large effect size.

**2.3.2 Fixation Odds-Ratios—**While the analytic approach used above is fairly standard for the VWP, it is not ideal for two reasons. First, the data come from an underlying binomial distribution, and particularly when the values are near zero (or one), such data may not meet the assumptions of linear models. The empirical logit transformation is often applied to transform the proportions of interest into log-odds ratios, which scale more linearly. However, this ignores a second factor which is commonly overlooked in the VWP. Looks to the HC and to the RO are not independent of each other: If the participant is looking at one, they cannot be looking at the other. Conventional analyses like the one just presented generally compare the proportion of looks to one object to the proportion of looks to the other. Since these are not independent of each other, this is problematic. In this case, the empirical logit transformation is insufficient to solve the problem: it simply allows us to compare two appropriately scaled variables that are still not independent.

The solution is to collapse the two looking measures into one (e.g., a difference score of some sort). In order to construct such a measure, but still respect the need for a more Gaussian distributed variable, we developed a new transformation based on the empirical logit. We replaced the odds ratio (*p/[1-p]*) with the odds ratio of looks to the HC over looks to the random object (or the LC competitor; whichever object-type that served as the baseline on that trial-type), as shown in (1).

$$DV = ln\left(\frac{p_{hc}}{p_{ro}}\right) \quad (1)$$

This can also be written in terms of the absolute number of looks as

$$DV = ln\left(\frac{p_{hc}}{p_{ro}}\right) \quad (2)$$

Here, $M_{hc}$ is the number of looks to the HC object, and $N_{hc}$ is the number of total looks. Since $N_{hc} = N_{ro}$, this term can be dropped. However, one problem is that of 0s on either the numerator or the denominator (a log of 0 is negative-infinity). In the empirical logit transformation, this is solved by adding .5 (half a success) to the numerator and denominator when computing the probabilities from the counts. However, doing this created highly skewed distributions, since eye-movements do not occur every four milliseconds. Thus, we added the equivalent of half of a fixation as the correction factor (C). Half a look was

[3]Using the same model as the HC trials (with no random slope of the interaction), the interaction was significant ($p$=.044).

calculated by taking the average fixation duration within each trial type (for correct trials only) and then dividing it by four (since the eye-tracker sampled at 4 msec) and two (to create half a look). This resulted for a correction of 45.33 for HC trials and 45.08 for HCLC trials. The denominator is the same with respect to the random object.

$$DV=ln\left(\frac{p_{hc}}{p_{ro}}\right) \quad (3)$$

As a whole, this is the log of the odds ratio of looks to the HC over looks to the RO, with a positive log-odds-ratio indicating more looks to the HC object. If this ratio was found to be significantly above 0, this would be evidence for participants looking more often at the HC competitor than baseline. This eliminates the non-independence issue since the two values are combined into one, and creates a more linear scaling.

Figure 6 shows the mean log-odds ratio as a function of block and trial-type, and confirms the results from the previous analysis. This measure was above zero for both trial-types on the later blocks, indicating more looks to the HC competitor than the baseline items (the RO or LC items). Data from the two trial-types were combined and examined with a linear mixed effects model with block (centered) and trial-type (HCLC vs. HC, +/−0.5) as fixed factors and subject and item intercepts. More complex (conservative) random effects structures showed a very high correlation (r=−1.00) between random slopes and intercepts, suggesting they may be overfitting the data. This model found that the intercept was significant (B=.046, SE= .022, $t(28.5) = 2.05$, $p = .049$, $d = 0.80$), indicating that this ratio was above 0 (equal looking for the HC competitor and the baseline comparison). This can be seen as further evidence that participants looked more at the HC competitor than the random object or the LC objects across the whole experiment. We also observed a significant effect of block (B=.042, SE= .019, $t(1700.6) = 2.16$, $p = .031$, $d = 0.11$). However, neither trial-type ($p = .265$) nor the trial-type by block interaction ($p = .563$) reached significance.

**2.3.3 Visual Co-Occurrence—**The analyses thus far suggest that learners fixate the HC competitors more than other objects (even as they are clicking on the target). However, one final concern is that this could be driven by purely visual co-occurrence (ignoring which object was named) since visual targets appeared with their HC competitors more frequently than with other objects. If participants were sensitive to co-occurrence at a purely visual level, they may direct eye-movements to these objects independently of the auditory stimulus (and these in turn could mirror the effects of mapping the name onto each object). To rule this out, we investigated fixations to the objects *before* the stimulus was heard (between −1000 to 0 msec, during the pre-scanning period) of the HC trials. We used raw proportions of fixations as the dependent variable, and included all three objects (target, HC and RO) in the analysis as at this point the participant should have no information as to the eventual target. The fixed effects included block (centered) and object-type (two dummy codes). As in our prior analyses, we used random slopes of block and object for participant, and random intercepts for items. This analysis found no significant effect of object-type[4] ($F(2, 147.8) = 0.60$, $p = .551$), no significant effect of block ($F(1,28)=.18$, p=.67), and no interaction (F(2,2722)=.38, $p$ = .682) for HC trials. Thus, participants were not biased to look at one object-type or the other as a result of the co-occurrence manipulation.

We replicated these analyses by combining raw fixations from both the pre-scanning and the post-stimulus periods of the trial and adding *trial-period* as a factor (pre- or post- stimulus, centered), along with block and object-type. For maximum sensitivity to the hypothesized HC vs. RO difference, the target object was dropped from this analysis and object-type was coded as HC=+.5, RO=−.5. Again, models with random slopes of all three fixed effects and their interactions did not converge; as we were interested primarily in the trial period × object-type interaction, we used random slopes of both of these factors and their interaction on participant (but no slopes of block). There was a main effect of trial-period with more fixations post-stimulus than during the pre-scanning period (B=.038, SE=.0089, $t$(28)=4.24, $p$=.0002, d=1.6). While there was no main effect of object-type (B = .0029, SE= .0025, $t$(181)=1.18, $p$=.24, d=.18), we did find a significant interaction between object-type and trial period (B=.011, SE=.0054, $t$(74)=2.04, $p$=.045, d=.47). Given the prior analysis on just the pre-scanning period, this suggests the effect of object-type was only present during the post-stimulus period. This finding is further evidence for our assertion that participants' looks were driven by the auditory stimulus and not visual co-occurrence. (We did not investigate other trial types given that if such visual co-occurrence effect existed, the HC trials would have had most power to reveal it.)

**2.3.4 Summary of Fixation Analyses—**The foregoing analyses present strong evidence that even on the trials on which participants were selecting the correct referent, they are nonetheless considering the HC competitors more than chance. While this effect was numerically small (as expected), it was robust across two analyses and the effect size was moderate. We also showed that it could not be attributed to visual co-occurrence. Thus, learners are clearly maintaining two hypothesized referents for a given word. Our next set of analyses turns to Question 2, i.e. what information participants are carrying forward from trial to trial.

### 2.4 Autocorrelation analysis

We next examined the trial-by-trial accuracy data using an autocorrelation analysis similar to the approach of Trueswell et al. (2013). This analysis examined performance on the current trial as a function of what happened the last time the same target was seen (we conduct a more complete analysis to address Question 4 more thoroughly after Experiment 2).

We started with the simplest model, in order to replicate these prior approaches (Figure 7). This analysis started by including only the first five repetitions of each word in order to achieve the same number of repetitions used in Trueswell et al. (2013). Here, the outcome variable was binary, indicating whether the participant clicked on the correct target on the current trial. The only fixed effect was a dummy coded variable indicating if the participant

[4]Note that we included target as an object-type in this analysis and that we used all trials (correct and incorrect). We dummy coded object-type (with looks to the RO as the baseline).

had been correct the last time they heard that word (*last-target-correct*). The first trial with each target was excluded since there was no prior trial with the same target.

This model was constructed in a mixed models framework, using the binominal linking function. *P*-values were computed directly from the Wald *Z* statistics; there are no commonly accepted measures of effect size for logistic models. As before, we first explored the random effect structure that best fit the data using the full data set and a more complete model. Here, we found that random slopes of *last-target-correct* on participant and word, along with a random intercept of target object best fit the data. (Note that this model also included a fixed effect for *trial-type* to account for potential additional difficulties of some of the trial-types.) In addition, it should be highlighted that while the typical statistical tests on the intercept for a logistic model compare the coefficient to 0, this assumes a chance level of .5, whilst chance was here .333. Thus, to evaluate against.333, we subtracted −ln(2) from the original intercept (this will be reported as $B_{33}$ or the adjusted intercept), and then computed a Wald Z-statistic by dividing this modified intercept by the original SE estimated from the model.

This model replicated the effects of Trueswell et al. (2013). We found a significant effect of *last-target-correct* (B= 1.74, SE= .240, Z= 7.22, $p < .001$), with much greater performance after a correct trial than an incorrect. The intercept was not significantly different from 0 (B= −0.618, $B_{33}$=.075, SE= .210, Z= 0.355, $p = .723$). Since *last-target-correct* is dummy coded, the intercept tells us the degree to which performance is above chance when the last target was incorrect, indicating that performance was at chance after an incorrect trial (Figure 7A).

However, Figure 7B (which shows repetitions 56-60 at the end of training) suggests this finding may not hold if we allow more time for gradual learning to unfold. Here, when we run the same analysis *on the entire set of trials*, we again see a highly significant effect of *last-target-correct* (B= 3.02, SE= .205, Z= 14.72, $p < .001$), but now the intercept was significantly different from 0 (B=0.10, $B_{33}$=.79, SE= .141, Z= 5.605, $p < .001$), indicating that overall performance was above chance even if people responded incorrectly. Thus, by the end of training, while we replicated an effect of *last-target-correct*, we also find that performance was above chance even when learners were incorrect on the prior encounter, contradicting Trueswell et al. (2013). Thus, the difference in findings between our and Trueswell et al.'s (2013) results is a difference in the number of trials considered. This confirms our intuition that the gradual learning effect may simply be too small to be observable with a small number of trials.

More importantly, as we described, this statistical model (as well as many of those used previously), may underestimate effects of gradual learning as it confounds position in the learning curve with *last-target-correct.* Incorrect prior trials are more likely to come from the early portion of the learning curve, and correct trials from later points. More importantly, it cannot assess whether any gradual learning effect can be seen over and above last-trial performance. Thus, we extended the prior analysis to ask if there was a further effect of accumulated experience over and above the prior trial behavior. To do this, we added the centered log of the number of exposures to that word up to the current trial (*log-target-*

*count*) and its interaction with *last-target-correct* to the model (both were also added as random slopes on participant and word). For this model, we centered *last-target-correct* to facilitate the interpretation of the included interactions.

When we used this new model on only the first five repetitions of each word, it did *not* provide a better fit to the data ($p$ = .937). However, using our whole data set, the new model including *log-target-count* offered a significantly better fit than the prior model ($\chi^2(16)$ = 735.65, $p$ < .001), providing robust evidence for an influence of a gradual element to learning, over and above the effect of last-encounter responding. In this model, there was a main effect of *log-target-count* (B = 1.18, SE = 0.15, Z = 8.13, $p$ < .001), indicating that performance was better with more experience, independently of prior trial behavior. There also continued to be a significant main effect of *last-target-correct* (B = 2.40, SE = 0.26, Z = 9.25, $p$ < .001) with again better performance if the participant was correct on the prior encounter. This reveals that even when accounting for position in the learning curve, there was unique variance associated with the participants' behavior on prior trials with that target. Finally, the interaction between *last-target-correct* and *log-target-count* was significant (B = 0.45, SE = 0.22, Z = 2.05, $p$ = .04)[5].

There are two equally accurate descriptions for this interaction: First, the effect of amount of exposure (*log-target-count*) may depend on whether one was correct on the previous encounter (*last-target-correct*), and in the extreme, there is *only* an effect of exposure for trials on which the learner was correct on the prior encounter with that word (e.g., Trueswell et al., 2013). That is, people are building more confidence with each correct response, but not accumulating anything from an incorrect one. Alternatively, the effect of last-encounter performance (*last-target-correct*) may depend on the amount of exposure. The first interpretation (particularly the extreme version) accords with a propositional account, whilst the latter would concur with associative learning.

To investigate this statistically, we conducted a post-hoc test, in which we asked whether the effect of exposure persisted even for trials that followed an incorrect response; if so, this would indicate that the effect of *log-target-count* cannot completely depend on previous performance. For this purpose, we split the data by *last-target-correct*, and only analyzed trials that followed an incorrect response on the previous encounter with the same target word. We used the same model as before, but dropped *last-target-correct* as a main effect and slope on participant and target word. In this model, there was a main effect of *log-target-count* (B = 0.66, SE = 0.17, Z = 4.0, $p$ < .001), indicating that an effect of exposure remained even for trials that had not followed a correct response.

The interaction thus indicates that the effect of prior trials *increased* over the course of training, but that the effect of exposure was present in both types of trials. This suggests that the *last-target-correct* effect may have been a product of learning, not the cause of it. That is, if the effect of *last-target-correct* grows with more exposure, it appears more the result of the accumulation of evidence. In contrast, if the *last-target-correct* effect derives from a sudden inference that drives learning, one would have expected much bigger effects early

[5]There were also significant effects of *trial-type, p* < .001, but these are not relevant to the hypotheses being tested.

(when there was more learning to do) than later. Thus, this analysis suggests that the number of repetitions of a word (gradual learning) accounts for variability in participants' accuracy in addition to the contributions of last-encounter performance (proposing / verifying): participants must bring forward more information from prior encounters with a word.

### 2.5 Discussion

Experiment 1 revealed two primary findings. First, the eye-tracking results demonstrated that participants were simultaneously considering both the correct (target) referent and a competitor (e.g., the HC competitor) during the same naming event. Specifically, on trials in which participants ultimately selected the correct target, they still fixated the HC competitor more than other objects. This effect was numerically quite small, which was expected given the lack of ambiguity in the displays and the fact that we filtered out the trials on which listeners were considering this competitor so strongly they actually clicked it. Nonetheless, it was significant across several analyses and had a moderate effect size, suggesting that at least sometimes and on some trials, both possible referents were being considered. For HCLC trials, this was only seen at the end of the experiment (reflected in the marginally significant interaction between object-type and block). This offers some tentative evidence that these associations may grow over training. No such differences were observed in the LC trials (see supplement): This may indicate that these differences in co-occurrence (LC competitor versus RO) were too subtle to result in observable differences in association strength. However, for the HC competitors, there was clear evidence for the simultaneous consideration of multiple hypotheses.

It should also be noted that the difference in looks to HC over RO/ LC competitors was driven by eye movements generated late in the trial. This suggests that participants may have had difficulty fully suppressing the HC competitor. In addition, this indicates that a consistent context throughout learning (i.e. the presence of the HC competitor) may impede performace; however, this will become clearer when comparing Experiment 1 and 2 (see below).

Our autocorrelation analyses addressed our second question. They replicated previous work showing an influence of prior responding on current trial accuracy (Dautriche & Chemla, 2014; Trueswell et al., 2013). However, more importantly, we found that adding a gradual component to the statistical model accounted for variance in accuracy over and beyond performance on the previous learning instance with that word. This thus indicates that participants do not simply retain hypotheses in an all-or-nothing fashion – they are also sensitive to the gradual accumulation of co-occurrence statistics. This was found even considering only trials in which participants responded incorrectly on their last encounter with a word. This also provides converging evidence for the question of multiple hypotheses (Question 1) - even when participants did not respond correctly on a prior encounter (they had the wrong hypothesis under a propose-but-verify account), they were still accumulating information relevant to the correct mappings. Thus, learning appears to be shaped both by last-encounter performance as well as overall exposure. Moreover, the fact that the effect of last-encounter performance improves with more learning suggests that the effect of last-encounter performance – argued to be the hallmark of a propose-but-verify strategy - may be a *product* of learning, not a mechanism of it (echoing the description of classic fast-mapping offered by McMurray et al., 2012).

The combined influence of both gradual learning and last-encounter performance is quite clear when examining individual subject data. Figure 8 shows the trial-by-trial accuracy for three participants for each word. Here we see that some words seem to start correct and stay that way (e.g., *pacho* for S1, *goba* for S10), seemingly confirming a propose-but-verify strategy. However, other words show a much more sporadic pattern (e.g., *goba*, for S1, *zati* for S5) and a substantial number of words show continued oscillation even out to repetition 30 or 40. Yet other words show a fairly robust run of accurate responding (suggesting a correct proposal) that is then lost for a time (e.g., *mefa* for S10 around repetitions 12-20; *pacho* for S5 from blocks 1-6). Thus, while the individual learning profiles appear to fit a range of descriptions, there appears to be a strong gradual or probabilistic component to learning.

## Experiment 2

Experiment 2 investigated one way in which context may influence cross-situational word learning (Question 3). As highlighted before, Dautriche and Chemla (2014) found that providing a stable context at the beginning of training helped learning on subsequent encounters. However, with no exposure to the context after these initial trials, it was impossible to tell if the spurious correlations created by consistent contexts may have hindered learning (since they were never encountered later). In Experiment 1, the HC competitors also offered a form of context, but one that was present probabilistically throughout the experiment, and may therefore be more likely to form spurious associations that impede learning. Thus, to determine whether this aided or hindered learning, Experiment 2 included only RO trials as a comparison.

Moreover, given the apparently conflicting results of our autocorrelation analysis and Trueswell et al.'s (2013), it was important to replicate Experiment 1 under experimental conditions that were more similar to their study; this was the case in Experiment 2 (there were no competitors with enhanced co-occurrence statistics). Thus, Experiment 2 was carried out both to establish a baseline of learning to compare accuracy of Experiment 1 against, and to replicate our autocorrelation analyses.

## 3. Experiment 2

### 3. 1 Method

**3.1.1 Participants—**Nineteen native English speakers took part in this experiment. All were students at the University of Iowa and received course credit as compensation. Participants were consented in accord with an IRB approved protocol.

**3.1.2 Design and materials—**The same stimuli were used as in Experiment 1. However, the design differed, as only RO trials were included. Thus, every trial included a target referent and two randomly chosen objects. Again, the co-occurrence rate of a word and its target was 100%. All the other seven objects were randomly selected from trial-to-trial without replacement and therefore co-occurred with the target word approximately with a co-occurrence rate of 28%. As in Experiment 1, the total number of 480 trials was separated into four blocks of 120 trials.

**3.1.3 Procedure—**The same procedure was used as in Experiment 1 with the exception that no eye movements were recorded as our primary hypotheses concerned accuracy.

### 3.2 Results Overview

Our analysis was conducted in two parts. First, we examined overall accuracy and compared it to Experiment 1 to determine how the co-occurring competitors, i.e. a constant context, affected learning. Second, we turn to our autocorrelation analysis to replicate the prior results with a study design that is closer to that of Trueswell et al.'s (2013) as well as of Dautriche and Chemla's (2014) experiments.

**3.2.1 Accuracy—**One participant was excluded from the analysis due to a computer error in the experimental script. Note that as the purpose of this was to compare accuracy in Experiment 1 and 2, the three participants who were excluded in Experiment 1 were included again to correctly reflect overall performance.

Figure 9 shows the average accuracy across for both Experiments 1 and 2. It suggests that participants performed better in the *absence* of high and low co-occurrence competitors. This pattern of results remains if one only compares RO trials of both experiments, suggesting that the presence of HC and LC competitors affected overall performance, not just accuracy in trials that included competitors with enhanced statistics. To evaluate this difference statistically, a binomial mixed model was conducted with participant, target (the object that was the correct referent) and stimulus as random effects. Experiment and block were the fixed effects (coded similarly to the prior analysis). The dependent variable was accuracy. Model selection suggested that a model with random intercepts for stimulus, target and participant as well as random slopes of block on participant, target and participant offered the best fit to the data. Using this, the effect of block was highly significant (B= 1.55, SE= .165, Z= 9.425, $p$<.001), suggesting improvement over time. Crucially, the effect of experiment was significant (B=.990, SE=.503, Z= 1.969, $p$ =.0489), as performance in Experiment 2 was better than in Experiment 1. The interaction of experiment and block was marginally significant (B= .468, SE= .274, Z= 1.709, $p$ = .087), and may indicate that participants did not simply learn better in Experiment 2 but also more quickly. Thus, highly co-occurring competitors, a probabilistic context, appear to impede learning.

One concern is that these differences may not reflect differences in the overall quality of learning, but the fact that some of the trials in Experiment 1 - those with an HC or LC competitor - were simply harder (though they would have been harder only because people had formed spurious associations). Thus, we controlled for this in a second analysis, by comparing only the RO trials of Experiment 1 to all of the trials from Experiment 2 (which were all RO). We found that a model that included random intercepts for stimulus, target and participant with a random slope of block on participant was best supported by the data. Using this model, the effect of block was still highly significant (B= 1.57, SE= .140, Z= 11.30, $p$<.001). Interestingly, the effect of experiment was still marginally significant (B=

0.90, SE= .50, Z= 1.81, $p$=.07), suggesting tentatively that the difference in accuracy across the two experiments was not simply driven by a lower accuracy in HC, HCLC and LC trials in Experiment 1, but that the inclusion of high-occurrence competitors impeded overall learning. However, given this marginal significance, this result needs to be confirmed in further research. The interaction between experiment and block did not reach significance, $p$ = .191.

**3.2.2 Autocorrelation Analysis—**Our autocorrelation analyses used the general statistical approach as in Experiment 1. This model included *last-target-correct* (centered), *log-target-count* (centered) and their interaction as fixed effects. In addition, a random intercept for participant, word and target object was added. We also included a random slope for *last-target correct*, *log-target-count* and their interaction term on both participant and word.

We found that (as in Experiment 1) the model which included *log-target-count* and its interaction with *last-target-correct* (both were added as random slopes on participant and word) offered a significantly better fit than a model with *last-target-correct* alone ($\chi^2(16)$ = 544.85, $p$< .001), replicating the previously reported influence of a gradual element to unsupervised learning. As in the previous analyses, it was found that there was a significant effect of *last-target-correct* (B = 2.16, SE = 0.25, Z = 8.49, $p$ < .001) and *log-target-count* (B = 1.33, SE = 0.17, Z = 8.21, $p$ < .001), indicating that both accuracy on a previous learning instance as well as experience (time) positively predicted accuracy on a current trial. Similar to Experiment 1, the interaction between *last-target-correct* and *log-target-count* reached significance (B = .45, SE = 0.19, Z = 2.39, $p$ = .017), suggesting that there was more of an impact of *last-target-correct* performance later on in the experiment.

### 3.3 Discussion

This experiment offered a clear answer to our third question, showing that participants learned more poorly in the presence of competitors that co-occur with a target word than in the presence of purely random competitors. This suggests an important caveat to our understanding of context. In natural situations, contexts like a kitchen or a park create sets of objects that frequently co-occur with each other and with their names. If such contexts are repeated across exposure (not just in the beginning of the experiment), this may create spurious associations between words and incorrect referents that can impede learning. It is likely that the statistical difference between Experiment 1 and 2 would have been more pronounced if learning in general had been more difficult, as most participants in Experiment 2 reached ceiling by block 2 (Figure 7).

Indirectly, this offers further evidence that people track more than one object-referent mapping hypothesis at the time too, even if this approach to learning has the potential to impair overall performance. Finally, we replicated the autocorrelation results from Experiment 1 (in the absence of HC and LC competitors) demonstrating that a gradual element significantly increased the fit of the model to the data, even in the presence of a strong effect of the last response. This highlights the need for a theory of word learning that assumes that both factors influence learning, i.e. that participants must bring more information forward from prior trials (Question 2 and 4).

# 4. Further Effects on Learning

## 4.1 Overview

Our previous autocorrelation analyses indicated that at least two factors influence learning on a trial by trial basis (the prior choice behavior and the gradual effect of number of exposures). We next extended our investigation to determine what other variables from prior trials may influence performance (Question 4). In part this is motivated by the previously described animal learning work (Wasserman et al., 2015) that suggests a rich tapestry of information on prior encounters with a word can shape responding. This includes the distance between the last encounter with an object and the present trial, the learners' knowledge of the foils, and the spatial arrangement of the items on prior trials (relative to that on the present trial). In this analysis, we asked whether such variables play a role in human cross-situational word learning over and above the previously examined number of exposures and/ or prior accuracy.

For this purpose, we combined the data from Experiment 1 and 2 in order to obtain more power to detect small effects. As our baseline, we started with the more complex model including *last-target-correct*, *log-target-count* (both centered) and the interaction term as fixed effects, and the same random effect structure (random intercept for participant, word and target object; random slopes for *last-target correct*, *log-target-count* and their interaction on participant and word). During initial explorations, we also examined models which added experiment and its interactions (with the other terms) as fixed effects. However, this did not improve the fit of the model to the data over the baseline model ($p = .104$), so these terms were not included in the analyses presented here. It should be noted that the baseline effects of *last-target-correct* and *log-target-count* were similar in all cases to the prior analyses. Thus, we do not discuss them here and focus on the most important new findings.

We investigated the following factors: The *delay between repetitions* (*last-trial-distance*) may distinguish unsupervised from supervised learning (Carvalho & Goldstone, 2014), or indicate the involvement of a decaying working memory. We also assessed learners' knowledge of the foils on the last encounter with the target (*last-foil-accuracy*). This could implicate some form of mutual-exclusivity or competition process that helps in correctly identifying the target (McMurray et al., 2012; Yurovsky, Yu, & Smith, 2013). Finally, we examined the degree to which visual-spatial factors may be involved by assessing whether the target object appeared in the same location on the last encounter (*last-target-same-location*). This may implicate a spatially organized working memory (Samuelson et al., 2011), a fairly naïve associative learner that had not yet determined the relevant features (Wasserman et al., 2015), or some kind of episodic memory. Each was examined in a separate model, adding one of these three factors along with its interactions to the baseline model (last-target-correct × log-target-count).

We started by considering the number of trials between the prior encounter with the word and the current trial. Figure 10A shows a small effect of *last-trial-distance* such that a short distance between the current and prior encounter led to better learning, though this was most pronounced early in training. A model including *last-trial-distance* (and its interactions) significantly improved the fit to the data ($\chi^2(4) = 36.77$, $p < .001$) over the baseline model. This was due to a significant main effect of *last-trial distance* on accuracy (B = −.013, SE = .004, Z = −3.02, $p = .002$). This indicates that an increase in target distance decreases accuracy (regardless of whether participants selected the target or not), potentially suggesting some form of recent memory that may influence performance. *Last-trial-distance* also interacted with *last-target-correct* (B = −.03, SE = .01, Z = −3.45, $p < .001$) as well as with *log-target-count* (B =.01, SE = .005, Z = 2.51, $p = .01$), and the three-way interaction was also significant (B =.02, SE = .008, Z = 2.19, $p = .03$). Earlier in the experiment, the effect of *last-target-correct* depended on *last-target-distance*, with a higher distance increasing the probability of an accurate trial if one was incorrect beforehand. This may indicate that the participants' response on a current trial was influenced by what they remember to have selected when they last encountered that target: This memory effect may decrease if distance is higher or when participants have already encountered a majority of mappings (late in the experiment). However, the main effect (which persisted with the interaction) also suggests some kind of momentum – learning was better if items were repeated nearby, regardless of the response.

We next examined the effect of foil accuracy (Figure 10B). This effect is very pronounced as accuracy is substantially higher if participants had previously responded correctly to the foil objects of a current trial. When we added *last-foil-accuracy* and its interactions to the model, this significantly improved the fit of the model over and above the baseline ($\chi^2(3) = 12.785$, $p = .005$). This was due to a significant main effect of *last-foil-accuracy*: Learners were more likely to be correct if they had been correct on a trial on which either of the foils had been the target *(*B =.34, SE = .122, Z = 2.76, $p = .006$). This also significantly interacted with *log-target-count (*B =.34, SE = .11, Z = 3.09, $p = .002$), as participants were more likely to be accurate if they knew one or both foils late in the experiment. This would seem to implicate some sort of eliminative processes by which increasing knowledge of the foils can be used to rule them out for a target word. This also parallels the *last-target-correct* × *log-target-count* interaction suggesting that the influence of foil knowledge is also a product of learning.

Finally, we examined whether performance differed when the target appeared in the same location on two subsequent trials helps learning (*last-target-same-location*, Figure 10C). This prediction easily falls out of associative or exemplar learning accounts in which words are not just associated with the objects, but perhaps with the whole context, or with co-occurring (irrelevant) factors such as spatial location. As Figure 10C shows, there was a small benefit when the target reappeared in the same spatial location as on the last encounter, though like target distance, this effect waned over training. We found that adding the main effect of *last-target-same-location* (but not its interactions) improved the fit of the model ($\chi^2(1) = 4.1$, $p = .04$). Accuracy was significantly increased if the target appeared in the same location the trial before (independently of whether one was correct on that trial; B

=.15, SE = .07, Z = 2.03, $p$ = .042). The fact that participants' learning is influenced by a target's position even if they did not click on it within that trial is strong evidence that they must be sensitive to variables beyond their current response (or hypothesis). This did not interact with any of our measures (as adding the interaction did not increase the fit of the model), suggesting a locus in the dynamics of learning and/or memory.

### 4.2 Discussion

These analyses showed that learners' performance was influenced by a variety of factors over and above of whether one chose the right referent on the last trial or one's position in the learning curve. As it was not possible to evaluate these factors simultaneously (the statistical models were too complex to fit), we cannot really evaluate the relative importance or size of these effects. It is also important to point out that these factors were not experimentally manipulated, rather the analyses are in some ways correlational. Moreover, these factors obviously do not exhaust all of the possibilities; it is very likely that there are other trial-by-trial factors that also influence learning significantly. Consequently, these analyses should be treated as exploratory.

Nonetheless, the results are quite compelling suggesting effects of nearby repetition, of spatial location, and of knowledge of the foils. In general, these analyses show that a wide variety of factors from prior encounters with a word shape behavior in the moment. It is not clear whether such information is carried forward by associative mechanisms, a short- or long-term memory mechanism or by some sort of propositional inference (or all three). However, these results point into the direction of a more complex learning mechanism that takes a number of sources of information into account.

First, learning appears to be characterized both by some associative effects. The significance of *log-target-count*, for example, suggests that there is some accumulation of information across multiple trials over and above any hypothesis formed on the immediate prior encounter. Similarly, the main effect of spatial location, suggests information that is not strictly necessary for word-object mappings is nonetheless preserved, consistent with an associative mechanism that is not specifically geared for word learning.

Second, we also see the simultaneous influence of potentially inferential processes (knowledge of the target, knowledge of the foils). However, it is important to note that these processes do not need to be conscious. A number of these seemingly inferential effects could be characterized in either way: For example, foil knowledge could reflect some sort of basic competition process (e.g. McMurray et al., 2012; Yurovsky et al., 2013), but they may also reflect something more like mutual-exclusivity (Halberda, 2003). Other effects may reflect contributions of both sorts of mechanisms. The main effect of target distance, for example, seems to implicate basic learning or memory which operates in a graded manner, while its interaction with *last-target-correct* suggests this memory may be crucial for allowing inference from prior trials. Crucially, these possibly inferential effects (particularly *last-foil-accuracy* and *last-target-accuracy*) both interact with the amount of exposure (they increase over time) suggesting that they are a product of learning, not a mechanism of it.

Finally, effects like that of spatial location are hard to view in any kind of propositional or inferential framework – they may represent potentially erroneous memory or associative biases. Thus, neither a simple propose-but-verify, memory-based account nor an associative model that only considers overall statistics (but not the behavior of the participant) can explain the richness of people's learning behavior.

## 5. General Discussion

This study was motivated by four unresolved questions that may help refine theoretical frameworks. First, we asked whether learners maintain multiple hypotheses for a word and whether these are simultaneously activated in-the-moment while listeners decide the referent of a novel word. Second, we asked whether learners gradually accumulate knowledge about a word above and beyond the effects of any hypotheses they may have from prior encounters. Third, we investigated whether a consistent context can exert a cost on word learning. Fourth and finally, we asked what other information is carried forward from previous encounters with a word to shape performance. Before we address each of these questions, we start by discussing some general limitations of the methodology we used. We close with a discussion about the nature of observational learning.

### 5.1 Limitations

Cross-situational word learning in general has been criticized as an experimental paradigm that is not representative of real-life vocabulary acquisition. It simplifies the problem of referential ambiguity by using isolated objects rather than embedding them in a cluttered scene. This could significantly facilitate learning (e.g. Medina et al., 2011; Trueswell et al., 2013). These concerns are important, though in many ways, cross-situational learning is a step toward better distilling the problem of referential ambiguity – many word learning paradigms feature no ambiguity (ostensive naming/ teaching), give strong cues to the correct referent (e.g., eye gaze), or explicit feedback. In this way, cross-situational learning may allow us to isolate and study the key aspects of unsupervised learning in the face of moderate referential ambiguity in a paradigm in which learning that can unfold in the space of a few laboratory sessions.

Nonetheless, if one takes these concerns at face value, one might argue that our design is particularly guilty of these simplifications. Our task featured a relatively small number of words, and they were repeated a large number of times. Consequently, performance was very high by the end of both experiments. This leaves open the possibility that this characteristic of our experiments significantly influenced our results and may have encouraged strategies that children may not use during more naturalistic word learning. Our intuition was that if anything these design choices should have encouraged a more propositional approach to learning, so the fact that we observed multiple hypotheses, gradual learning, and effects like that of spatial location may be quite telling. However, it is also possible that they encouraged a more statistical strategy, as there is little data (of this level of detail) on how children learning words cross-situationally, and on learning with more naturalistic levels of ambiguity. We do know that animal learning (which is clearly not propositional) shows virtually all of the hallmarks demonstrated here (Wasserman et al., 2015), but it remains an empirical question whether these findings generalize to more naturalistic settings.

Our own choices in how to distill the paradigm were motivated by a desire to investigate aspects of observational learning that may have been missed. The use of more (rather than less) repetitions allowed us to investigate longer-term acquisition. This revealed an effect of gradual statistical learning that could not be seen in shorter experiments, and it is worth noting that outside of the lab children learn words quite slowly. More importantly, this additional training gave us experimental power to ask and isolate questions that have not been addressed within (observational) word learning before, such as what other factors contribute to word learning or how gradual information across different contexts contribute to word learning. Moreover, this expanded training along with the somewhat easier nature of the learning problem (8 word-object linkages) yielded high accuracies that were necessary for the logic of our eye-tracking design which offered a much more direct measure of whether people were tracking multiple hypotheses for a word and activating them in the moment of naming.

This trade-off may restrict the ability of our studies to generalize to vocabulary acquisition of natural language (in children), though this is still an empirical question. However, our goal in this study was not to capture word learning as a whole but rather to isolate mechanisms within observational learning that had been underrepresented by previous studies and that are likely to impact how words are learned.

With this important caveat, we now turn to our four questions.

### 5.1 Do learners maintain multiple hypotheses in parallel?

This question was addressed by our analysis of the eye movements in Experiment 1 which revealed that participants are more likely to fixate the high co-occurrence competitor than a randomly chosen object (the baseline), even as they are clicking on the target. This indicates that they retain multiple hypotheses about the word and are actively considering both the HC competitor and the target simultaneously *on the same trial*, even in the absence of disconfirming evidence or memory failure. This difference in looks to the RO/ LC and HC competitor was driven by eye movements made late in the time course of processing; this may be an indication that people had some small amount of difficulty disengaging from the HC competitor when it was present. Though this effect was numerically quite small, this was expected given that structure of the task: instructions made it clear that one word ultimately maps onto one object only, there was little bottom-up perceptual ambiguity, and we excluded the trials when consideration of this competitor was high enough to trigger an incorrect response to the HC object. More importantly, however, this effect was statistically robust and carried a moderate effect size. Given all of this, it is clear that multiple hypotheses were considered at least to some degree spontaneously and simultaneously.

The marginally significant interaction of object-type and block on the HCLC trials, as well as the overall effect of block in our analysis of the log odds ratio offers additional insight into this issue, by suggesting it may change over the course of learning. This indicates that these erroneous associations may grow over time, as listeners acquire more co-occurrence data to support them, though this is clearly not as robust an effect as the overall main effects of object-type.

It is important to consider an alternative account of these results. It is possible that our repetitive design by itself led participants to be rather bored, and this in turn led participants to encode other aspects of the experiment that are not strictly relevant to the correct answer (e.g., they started learning about the high-co-occurrence competitors, but only due to boredom). While we cannot completely rule this out as an explanation of our eye movement data, there are several reasons why it is unlikely to be true. First, this predicts the strongest effects of object-type late in the experiment. However, the interaction between block and object-type were inconsistent across the two trial-types, and not reliable for the HC trials – the strongest evidence for multiple hypotheses. More importantly, while this hypothesis may address our eye movement results, it fails to account for the accuracy and autocorrelation results: It does not explain why participants' performance was consistently better in Experiment 2 (and not just at the end when the eye movement effect becomes more evident) or why the effect of prior accuracy, presumably a marker of inferential processing, in fact becomes stronger as the experiment progresses. Moreover, the effect of number of exposures - which can be seen over and above learner's prior hypothesis - also offers converging evidence that learners are building multiple hypotheses over time. Even when we only consider trials on which the learner was incorrect on the prior encounter, there is still a strong effect of number of exposures, implying that they were encoding multiple hypotheses. Finally, as similar learning patterns have been observed in pigeons using exactly the same autocorrelation analyses, this again indicates that this alternative hypothesis cannot account for all behavioral results (Wasserman et al., 2015). Thus, while we cannot fully rule out this account for our eye-movement results, it is not clear that it can explain the rest of our results. Thus, these arguments offer converging support for the idea that our eye-movement results are not a spurious result from a boring task, but rather serve as evidence of a more basic learning mechanism.

So what kind of learning mechanism? The maintenance of multiple hypotheses for a word's referent is a clear property of associative accounts of cross-situational word learning in which a given word can be partially associated with multiple objects (Kachergis, Yu, & Shiffrin, 2012; McMurray, Horst, & Samuelson, 2012; Ramscar et al., 2013; Yu & Smith, 2007), and also of Bayesian accounts (Frank et al., 2009) which capture the likelihhod of each referent for a given word. The indication that looks to the HC competitor increased over training (as seen in the HCLC trials) is also a clear prediction of both associative and statistical accounts, in which these associations gradually build as a result of the accumulation of co-occurrence statistics.

Simultaneously, it may be possible to integrate these findings in a weak propose-but-verify- or memory-based framework (e.g. Koehne, Trueswell, & Gleitman, 2014). For instance, Koehne et al. (2014) suggested that participants may retain hypotheses that have been considered before. Given the design of this particular study (high number of HC trials, small number of words), it is likely that participants clicked the HC competitor at some point during learning. Thus, such a weak propose-but-verify account may be in agreement with our data. However, such an account would predict activation of multiple hypotheses only under certain circumstances, i.e. in the light of disconfirming evidence and/ or memory failure. In contrast, we find evidence of such parallelism even when the learner is correct. Moreover, if such processes were to explain our results, activation of alternative hypotheses should have decreased over time (as a result of higher levels of certainty), where we see evidence for either a stable property of learning, or an increase with training. Thus, whilst the overall finding of participants maintaining multiple hypotheses is consistent with a weak propose-but-verify account, a closer analysis of its predictions does not concur with the particular pattern of effects we observed. More broadly, while we do not think our results are inconsistent with all propositional learning accounts, it is clear that a careful consideration of what information is available to learners in real-time may constrain theoretical notions of what learners may use multiple hypotheses for.

### 5.2 What do learners carry forward from prior trials?

Our second goal was to determine whether there also is a gradual element that contributes to cross-situational word learning. For that purpose, we enhanced the autocorrelation analysis introduced by Trueswell et al. (2013) by adding variables beyond whether participants were correct on their last encounter with the target word. We found that the number of encounters with an object was a significant predictor of accuracy over and above the participants' prior accuracy with that word. More precisely, participants were more likely to be correct on a current trial if they had encountered it more often, *over and beyond* the effect of last-trial performance. This indicates that last-encounter performance, despite capturing an important aspect of learning, is only one of several factors that influences learning (and thus need to be accounted for in any theory of word learning). Crucially, while the specific hypothesis that learners may arrive at on a previous trial may indeed shape learning, their gradual accumulation of evidence for this hypothesis (as well as others) appears to be just as important. This is underscored by the fact that this gradual learning effect was observed even when we consider only trials on which learners were entertaining the wrong hypothesis.

This does not negate the prior responding effect – both were robust in our analysis. Thus, there are clear influences of prior behavior (not just statistics) on learning. Trueswell et al. (2013) attribute such effects to an explicit propose-but-verify strategy within a propositional framework. However, we are hesitant to claim that such last-encounter effects are unique markers of such an approach. First, Wasserman et al. (2015) showed that pigeons behave similarly in a supervised cross-situational learning paradigm, making it unlikely that an effect of last-encounter performance is uniquely an indicator of hypothesis testing. Second, there is a range of possible processes that could give rise to these effects. For example, participants could be biased by short-term memory of prior responses (whether or not they think those are correct for that word). This hypothesis may be supported by the interactions of last-encounter performance with the distance to the last trial as well as the significant three-way interaction between last-encounter performance, target distance and target count. These indicated that incorrect last-encounter performance may have less of a negative impact on accuracy if the period between presentations was larger. Alternatively, the last-encounter performance effect on learning may reflect a sort of biased competition or hysteresis, where once the system has settled into a response for one word, it is more likely to settle there again. It could even be due to something as simple as fluctuations in performance around a gradually increasing mean (e.g., if the participant is on a "run" of good performance), or a statistical artifact of not fully characterizing a non-linear learning curve. Thus, while it is clear that this prior-responding effect is a very important driver of learning behavior, future work must clearly disentangle what it means. This is particularly clear given that the here described effects seem to be consistent with either theory of observational learning.

### 5.3 Can a consistent context impair learning?

This question was addressed by comparing accuracy in Experiment 1, which manipulated context in a probabilistic manner, and Experiment 2, in which the co-occurrence of individual objects and a target word was kept constant across the seven foils (and thus overall lower than for the HC and LC competitors of Experiment 1). We found that participants' accuracy was decreased by the presence of HC and LC competitors and that this difference does not appear to be driven by HC and HCLC trials only. Interestingly, learners appeared to consider more than one hypothesis, despite the fact that this impedes learning: If participants had treated trials in Experiment 1 as RO trials (i.e. ignored the higher co-occurrence of the HC and LC competitors with the target), learning would have been faster (as in Experiment 2).

This finding contrasts with the benefit of context reported by Dautriche and Chemla (2014). This is likely due to the fact that their study implemented the co-occurrence only in the beginning of the experiment and not throughout. Thus, this indicates that context may help situate a word (e.g. by restricting the possible interpretations of *spoon* to a kitchen item), but impedes learning when the co-occurrence of members of one context is held high during learning. More specifically, the word *spoon* would be more difficult to learn if it were repeatedly experienced with a fork and knife, but would receive a boost if it were seen in a completely different context and thus different competitors (e.g. in a sandbox). For real word learning, this may indicate that encouraging use of words outside their typical contexts may promote vocabulary acquisition, particularly after a word has been experienced in constant surroundings beforehand. This is consistent with the notion of contextual interference from motor skill learning (Wulf & Shea, 2002), and with the idea that variation in irrelevant factors can improve learning (Apfelbaum & McMurray, 2011; Gómez, 2002; Rost & McMurray, 2009). Targeted variation - in this case, in the co-occurring foils - can help learners to form more robust mappings between words and the correct foils.

This result may be explained within the propose-but-verify account: Accuracy may be lower in Experiment 1, as participants might consider the HC competitor as the target for longer (as this hypothesis is more likely to be confirmed in a later trial than in Experiment 2), thus leading to lower performance. This pattern of performance, however, should lead to the biggest differences between the two experiments early in learning (unless there is a lot of noise in how well people remember the hypotheses), suggesting this is not the whole story. Our results are also consistent with an associative account. As people track multiple hypotheses, this may result in spurious associations which favor the HC competitor, thus reducing overall accuracy.

However, this account by itself may not be able to account for both the benefits of context observed by Dautriche and Chemla (2014) and the cost shown here. Indeed, our work is an important complement to their study highlighting the richness of contextual manipulation. The relative frequency of these two effects (e.g. benefit of learning vs. negative impact of context on learning) is hard to estimate, making it difficult to make clear predictions outside of the lab. But the presence of both a benefit and a cost implies something about the learning mechanism that is quite novel: it must involve both long-term tracking of statistics and associations (to show the cost) as well as short term competition or inference mechanisms (to get the benefit). This seems to suggest some kind of hybrid model in which active competition or inference processes are built on top of more basic associative mechanisms (McMurray et al., 2012; D Yurovsky et al., 2013).

### 5.4 What other information is retained from prior encounters during learning?

Our autocorrelation analysis identified a range of other factors from prior encounters with a word (or object) that may influence learning. While these analyses were somewhat exploratory, they considerably go beyond the issues of gradual learning (Question 2) and hypothesis testing by building on recent work in animal learning. In these analyses, target distance emerged as a significant additional predictor. When the target word had been encountered recently, learners were more likely to respond correctly, regardless of choice accuracy. This effect is quite interesting, as a recent study by Carvalho and Goldstone (2014) suggests that unsupervised learning is much more likely to benefit from nearby repetitions of a stimulus than supervised learning. This finding generally reinforces the possibility of a largely associative or general mechanism for this behavior, though it is also consistent with a memory-based account, in which short-term memory facilitates performance if target distance is shorter.

We also observed an effect of spatial location in which participants were more accurate when the target object was in the same spatial location on a prior trial. This is consistent with Wasserman et al.'s (2015) animal model. In pigeons, this effect appeared to be driven by the animals not knowing early in training that space is not relevant for predicting reward. Thus, consistently with error factor theory (Harlow, 1959), pigeons needed to learn to suppress responding on the basis of location. Here we showed that also humans benefit from a target being presented in the same spot as on a previous trial, independently of whether they clicked on it at that prior encounter. This happened in an unsupervised learning paradigm (no reward); nevertheless, the explanation is likely to be similar as in the pigeons: words and objects are not just associated with each other, but also with irrelevant factors such as spatial location. This benefits learners if the target object occurred in the same location subsequently (and hurt them if was not). Moreover, this effect could also reflect a more exemplar approach to memory in which all aspects of a context are recorded as part of the memory trace for a word or a reflection of space being used as a cue to bind the novel word label to the object (Samuelson et al., 2011). (Neither of these explanations is necessarily inconsistent with each other.) At this point, it is not clear how this finding could be integrated in a more propositional account which seem to assume strong functional goals on the part of the learner as to what are the relevant aspects of this task.

Finally, we also found an effect of foil-responding on previous trials such that accurate identification of the foils (on a prior trial) led to better accuracy on the current trial. This would seem to suggest a form of mutual exclusivity that may come into play as foils are learned; however, this is consistent with both inferential versions of mutual exclusivity (e.g. Carey & Bartlett, 1978; Halberda, 2003) in which participants explicitly use their 'knowledge', or with more basic mechanisms such as some form of real-time competition (McMurray et al., 2012; Yurovsky et al., 2013). As with the effect of *last-target-correct*, this effect also grew with more exposures to a word. This implies that it is a product of learning, not a core mechanism of it.

In summary, what is clear is that a substantial amount of information about prior trials (and prior experience) shapes learning on any given trial – both what the participant did and his or her experience with an object, as well as potentially more subtle factors (particularly given that our list is likely to not be exhaustive given the exploratory nature of these analyses). Thus, it would be incorrect to oversimplify learning into a process in which only one particular type of information is considered.

### 5.4 The nature of observational learning

While individually, our various findings are consistent with several accounts, our conclusions do not derive from any one finding (e.g., the eye movement results) or from the answer to any question in isolation. Rather the converging evidence across multiple findings is perhaps our most novel contribution. This highlights the complexity of processes that take place during word learning, processes that include the online activation of multiple hypothesized referents, potentially inferential or competitive strategies to rule out particular hypotheses in the moment on the basis of both knowledge of the target word, and of the target and foils objects, short-term memory processes that track prior responding to both targets and foils, and very basic associative phenomena that capture the cross-situational statistics.

Whilst the goal of this study was not to definitively test any particular theoretical accounts, it is possible to draw some conclusions. The richness of the results described here makes it clear that any oversimplified account is unlikely to be true: This applies both to simple co-occurrence counting as well as a strong propose-but-verify account. More specifically, a strong propose-but-verify account cannot account for our eye-movements data (as people should not have tracked more than one hypothesis). Similarly, pure co-occurrence counting is not consistent with the subtle effects observed in the autocorrelation analyses. Thus, we need something more complex.

So do these two experiments appear to favor any particular elements of a theory of learning? On the one hand, the evidence for multiple hypotheses from the eye movements, and the evidence for gradual learning in the autocorrelation analyses, as well as the effect of spatial location in the same analyses all point toward the idea that associative learning (or something like it) plays at least some role in the learning. Under associative accounts, connections between words and objects that co-occur together are strengthened during learning; this process is gradual and accumulative to avoid over-commitment to one mapping on the basis of a few exposures, accounting for the rather protracted effect of exposure (in the autocorrelation analyses). This effect could not be seen over five trials but did emerge over the whole course of training. This gradual and accumulative nature also means that participants would have formed stronger connections with a word and its respective HC competitor than any random word (as they were seen more often together), thus leading to increased levels of activation during trials (as reflected in a higher number of looks to the HC competitors in Experiment 1). Moreover, one would predict that this effect may increase over time as a result of the gradual nature of associative learning as these spurious connections are strengthened by more data. Indeed, we found some evidence for this in the eye movements. Moreover, the marked enhancement of learning in Experiment 2 when there were no spuriously correlated objects suggests that these additional associations interfere with learning. Finally, as associative learning may not know a priori which aspects of the stimulus are relevant for learning (i.e. its appearance), this may lead to the observed effects of spatial location, as spatial locations are associated with words (c.f., Wasserman et al., 2015). While alternative explanations for any of these findings are possible (as we describe above), all told, this converging pieces of evidence paint a picture of some form of associative or statistical core to word learning, even as this must be embedded in a much richer processing system of some kind.

This is not likely the whole story however. There were strong effects of accuracy on prior trials including effects of accuracy for both the target (as seen in prior studies) and foil objects (seen for the first time in humans here). This appears to be in line with propositional accounts. In a propose-but-verify framework, the effect of prior accuracy is thought to be a reflection of retaining a single hypothesis (actively) from a prior encounter with a word. Thus, if participants chose the target on a prior trial, they should be very likely to be correct on the current trial, too. (This may not be perfect, however, because of memory limitations.) At the same time, an incorrect choice on a prior trial means that the person's hypothesis had to be rejected and the participant is at chance when being confronted with the same word. Similarly, while propose-but-verify does not explicitly address knowledge of the foils, one could imagine propositional accounts in which learners use a mutual exclusivity strategy to choose an object for a novel word by ruling out objects for which they know the labels (as in mutual exclusivity accounts of child word learning). This would thus increase the probability of responding correctly if at least one known foil is present. These findings seem quite consistent with a more inferential or propositional logic.

But such findings are not unambiguous evidence for propositional learning. First, the fact that we see potential evidence for inferential processes (*last-target-correct* and *last-foil-correct* effects) does not negate an associative core to how memories are formed. Indeed, we know of few other ways by which information is stored in the brain. However, such effects suggest that such a core must be embedded in a richer system which not only learns but also infers and acts. That is, learners may engage in propositional reasoning in the moment, even as the results of this reasoning gradually accumulate in the association matrix. Second, the effects of prior accuracy and foil responding do not appear to be unique markers of an inferential account: Wasserman et al. (2015) reported extremely similar effects of prior trial effects in pigeons. These animals do not have a developed prefrontal cortex, but are rather biological associative models. Of course, it is possible that these behaviors indicate different mechanisms in varying species. However, right now, we simply do not know what representations and/ or mechanisms lead to these effects in humans and we have clear evidence that they can arise in an animal without higher-level inferential processes. Indeed, there may be simpler, domain-general mechanisms that could give rise to such effects; for example, competition between possible referents may instantiate a form of mutual exclusivity (McMurray et al., 2012).

This would appear to leave us in a situation where we need some kind of framework that is associative in core, but can also account for processes that appear inferential. McMurray et al. (2012) offer a conceptual (and computational) model that suggests a way to rectify some of these more inferential-looking processes with associative learning in the context of observational word learning. It argues that associative learning must be embedded within a system that is capable of real-time decisions (a form of constraint satisfaction or competition); in the specific simulations, they suggest that simple Hebbian learning embedded in a system of real-time competition (to choose the best referent) can learn words under high degrees of uncertainty and account for a range of developmental effects.

To broaden this framework, we might suggest that proposing and/or verifying may be primarily real-time processes (that are either propositional or perhaps based in lower level mechanisms like competition), even as the underlying associative learning system can build multiple alternatives more slowly. That is, in real-time participants only choose a single object (even as they evaluate multiple), and may engage in a variety of decision making processes (including, for example, verification processes), even as the underlying learning is gradual and tracks multiple co-occurrence statistics, which are still implicitly activated during response choice.

It is important to note that the mechanisms that lead to effects such as mutual exclusivity in this model are not actually propositional in nature; in contrast, they emerge out of simple real-time competition between referents. However, whatever the nature of the real-time processes, this model makes the point that such interactions across timescales can be quite powerful, allowing the system to act intelligently, even if the associative weights are not fully formed[6]. It is unlikely that the specific computational instantiation of McMurray et al.'s (2012) model can account for all of the specific data right now. For example, it does not include any units to encode context (e.g., Dautriche & Chemla, 2014), nor does it account for effects of spatial location. Nevertheless, it makes the important point that a combination of real-time processes and associative learning has the power and richness to account for the different types of behavior we are observing in Experiment 1 and 2.

Finally, observational learning is suggested as a developmental account, one of the mechanisms by which children learn words. Thus, it would be worth considering whether our results may differ in children. Recent data by Ramscar, Dye, and Klein (2013) suggest that children's behavior is more consistent with an associative account whilst adults seem to behave more propositionally and to at least sometimes apply explicit strategies. In contrast,

[6]Similarly, Bayesian models that track the distribution of possible mappings and their likelihoods should give rise to comparable effects, (Frank, Goodman, & Tenenbaum, 2009).

we show clear hallmarks of associative learning in adults here. Part of the discrepancy may be due to the fact that Ramscar et al. (2013) only used two exposures per word (nine word-object mappings in total) – clearly favoring an inferential approach that may have been out of reach for children, and this may also apply to cross-situational word learning experiments that only used a low number of trials. However, it may be possible that the factors that emerged during the autocorrelation analysis only correspond partly to the variables that may significantly predict performance in children. That is, children may show different effects of last-trial performance, gradual learning, spatial location and the like when confronted with a similar learning paradigm.

In summary, this study shows clearly that people are sensitive to not only complex co-occurrence matrices but also other trial-to-trial information (such as the distance between two target trials or the location of an object). This indicates that observational learning must be supported by a much richer learning mechanisms than previously assumed, even as they are built on a deeply parallel and gradual core. Given that different frameworks in which cross-situational learning could be explained are still developing, it is unclear at this point what gives rise to this behavior. However, any future account of observational learning will need to acknowledge that word learning is a dynamic product of both gradual and real-time effects.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The authors would like to thank Keith Apfelbaum for his highly appreciated input and assistance with various aspects of this project, Toby Mordkoff who suggested the log-likelihood analysis, Teresa Treat who assisted with advice on mixed model analyses, Dan McEchron for technical assistance and for recruiting participants, and the other members of MAClab team for further technical assistance. This research was supported by NIH grant DC0008089 to BM, and by a Fulbright Scholarship which funded TCR's visit to the University of Iowa.

## 7. References

Apfelbaum KS, McMurray B. Using variability to guide dimensional weighting: associative mechanisms in early word learning. Cognitive Science. 2011; 35(6):1105–38. http://doi.org/10.1111/j.1551-6709.2011.01181.x. [PubMed: 21609356]

Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language. 2013; 68(3) http://doi.org/10.1016/j.jml.2012.11.001.

Bates D, Maechler M. lme4: Linear mixed-effects models using S4 classes. 2009 Retrieved from http://cran.r-project.org/package=lme4.

Carey S, Bartlett E. Acquiring a single new word. Papers and Reports on Child Language Development. Aug.1978 15:17–29. Retrieved from http://www.mendeley.com/research/acquiring-single-new-word-1/.

Carvalho PF, Goldstone RL. The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. Psychonomic Bulletin & Review. 2014 http://doi.org/10.3758/s13423-014-0676-4.

Dautriche I, Chemla E. Cross-situational word learning in the right situations. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2014; 40(3):892–903. http://doi.org/10.1037/a0035657.

Ebbinghaus, H. Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie. Neue unveränd und ungek Ausgabe nach der 1 Aufl 1885 (Vol. Neue, unve). 1885.

Frank MC, Goodman ND, Tenenbaum JB. Using speakers' referential intentions to model early cross-situational word learning. Psychological Science. 2009; 20(5):578–585. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19389131. [PubMed: 19389131]

Gleitman L. The Structural Sources of Verb Meanings. Language Acquisition. 1990; 1(1):3–55.

Golinkoff RM, Hirsh-Pasek K, Bailey LM, Wenger NR. Young children and adults use lexical principles to learn new nouns. Developmental Psychology. 1992; 28(1):99–108. http://doi.org/10.1037/0012-1649.28.1.99.

Gómez RL. Variability and detection of invariant structure. Psychological Science : A Journal of the American Psychological Society / APS. 2002; 13(5):431–436.

Halberda J. The development of a word-learning strategy. Cognition. 2003; 87(1):B23–B34. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0010027702001865. [PubMed: 12499109]

Kachergis G, Yu C, Shiffrin RM. An associative model of adaptive inference for learning word-referent mappings. Psychonomic Bulletin & Review. 2012; 19(2):317–24. http://doi.org/10.3758/s13423-011-0194-6. [PubMed: 22215466]

Koehne, J.; Trueswell, JC.; Gleitman, LR. Multiple Proposal Memory in Observational Word Learning Learning based on Co-occurrence Frequencies. In: Knauff, M.; Pauen, M.; Sebanz, N.; Wachsmuth, I., editors. Proceedings of the 35th Annual Con- ference of the Cognitive Science Society; Austin, TX: Cognitive Science Society; 2014. p. 805-810.

Markman EM. Constraints children place on word meanings. Cognitive Science. 1990; 14(1):57–77. http://doi.org/10.1016/0364-0213(90)90026-S.

McMurray B, Aslin RN, Tanenhaus MK, Spivey MJ, Subik D. Gradient sensitivity to within-category variation in words and syllables. Journal of Experimental Psychology: Human Perception and Performance. 2008; 34(6):1609–1631. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3011988&tool=pmcentrez&rendertype=abstract. [PubMed: 19045996]

McMurray B, Horst JS, Samuelson LK. Word learning emerges from the interaction of online referent selection and slow associative learning. Psychological Review. 2012; 119(4):831–77. http://doi.org/10.1037/a0029872. [PubMed: 23088341]

McMurray B, Samelson VM, Lee SH, Tomblin JB. Individual differences in online spoken word recognition: Implications for SLI. Cognitive Psychology. 2010; 60(1):1–39. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19836014. [PubMed: 19836014]

McMurray B, Zhao L, Kucker SC, Samuelson LK. Gogate L, Hollich G. IGI Global. Theoretical and Computational Models of Word Learning. 2013 http://doi.org/10.4018/978-1-4666-2973-8.

Medina TN, Snedeker J, Trueswell JC, Gleitman LR. How words can and cannot be learned by observation. Proceedings of the National Academy of Sciences of the United States of America. 2011; 108(22):9014–9. http://doi.org/10.1073/pnas.1105040108. [PubMed: 21576483]

Pavlik PI Jr, Anderson JR. Practice and Forgetting Effects on Vocabulary Memory : An Activation-Based Model of the Spacing Effect. Cognitive Science. 2005; 29:559–586. [PubMed: 21702785]

Quine, WVO. Word and object: An inquiry into the linguistic mechanisms of objective reference. The MIT Press; Cambridge, MA: 1960.

R Development Core Team. R: A Language and Environment for Statistical Computing. the R Foundation for Statistical Computing; Vienna, Austria: 2011. ISBN: 3-900051-07-0. Available online at http://www.R-project.org/

Ramscar M, Dye M, Klein J. Children value informativity over logic in word learning. Psychological Science. 2013; 24(6):1017–23. http://doi.org/10.1177/0956797612460691. [PubMed: 23610135]

Rost GC, McMurray B. Speaker variability augments phonological processing in early word learning. Developmental Science. 2009; 12(2):339–349. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3011987&tool=pmcentrez&rendertype=abstract. [PubMed: 19143806]

Salverda AP, Brown M, Tanenhaus MK. A goal-based perspective on eye movements in visual world studies. Acta Psychologica. 2011; 137(2):172–80. http://doi.org/10.1016/j.actpsy.2010.09.010. [PubMed: 21067708]

Samuelson LK, Smith LB, Perry LK, Spencer JP. Grounding word learning in space. PLoS ONE. 2011; 6(12)

Siskind JM. A computational study of cross-situational techniques for learning word-to-meaning mappings. Cognition. 1996; 61(1-2):39–91. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8990968. [PubMed: 8990968]

Smith K, Smith ADM, Blythe R. a. Cross-Situational Learning: An Experimental Study of Word-Learning Mechanisms. Cognitive Science. 2011; 35(3):480–498. http://doi.org/10.1111/j.1551-6709.2010.01158.x.

Smith L, Yu C. Infants rapidly learn word-referent mappings via cross-situational statistics. Cognition. 2008; 106(3):1558–1568. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17692305. [PubMed: 17692305]

Tanenhaus MK, Spivey-Knowlton MJ, Eberhard KM, Sedivy JC. Integration of visual and linguistic information in spoken language comprehension. Science. 1995; 268(5217):1632–1634. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7777863. [PubMed: 7777863]

Trueswell JC, Medina TN, Hafri A, Gleitman LR. Propose but verify: fast mapping meets cross-situational word learning. Cognitive Psychology. 2013; 66(1):126–56. http://doi.org/10.1016/j.cogpsych.2012.10.001. [PubMed: 23142693]

Vouloumanos A. Fine-grained sensitivity to statistical information in adult word learning. Cognition. 2008; 107(2):729–42. http://doi.org/10.1016/j.cognition.2007.08.007. [PubMed: 17950721]

Wasserman EA, Brooks DI, McMurray B. Pigeons acquire multiple categories in parallel via associative learning: a parallel to human word learning? Cognition. 2015; 136:99–122. http://doi.org/10.1016/j.cognition.2014.11.020. [PubMed: 25497520]

Wulf G, Shea CH. Principles derived from the study of simple skills do not generalize to complex skill learning. Psychonomic Bulletin & Review. 2002; 9(2):185–211. http://doi.org/10.3758/BF03196276. [PubMed: 12120783]

Yu C, Smith LB. Rapid word learning under uncertainty via cross-situational statistics. Psychological Science. 2007; 18(5):414–20. http://doi.org/10.1111/j.1467-9280.2007.01915.x. [PubMed: 17576281]

Yu C, Smith LB. Modeling cross-situational word-referent learning: prior questions. Psychological Review. 2012; 119(1):21–39. http://doi.org/10.1037/a0026182. [PubMed: 22229490]

Yurovsky D, Fricker DC, Yu C, Smith LB. The role of partial knowledge in statistical word learning. Psychonomic Bulletin & Review. 2014; 21(1):1–22. http://doi.org/10.3758/s13423-013-0443-y. [PubMed: 23702980]

Yurovsky D, Yu C, Smith L. Competitive processes in cross-situational word learning. Cognitive Science. 2013; 37:891–921. Retrieved from http://www.indiana.edu/~dll/papers/yurovsky_yu_smith_c12.pdf. [PubMed: 23607610]

## Highlights

- There is considerable uncertainty around the nature of observational word learning
- Eye movements suggested learners maintain and activate multiple hypotheses
- Accuracy was influenced by many factors, including spatial location, foil accuracy
- Probabilistic context throughout experiment hurt people's learning
- This supports an associative mechanism that is buttressed by real-time processes

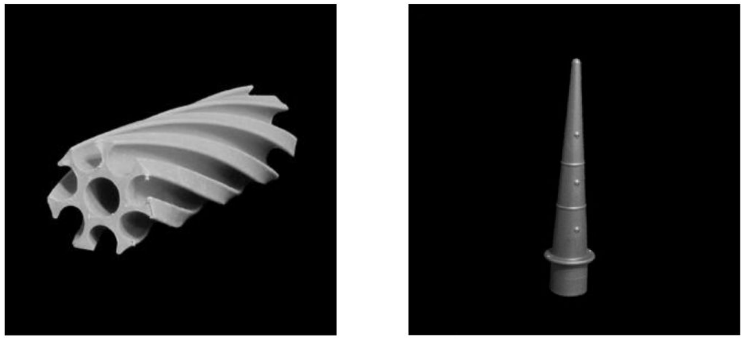

**Figure 1.**
Examples of the novel, differently colored objects used in Experiment 1.

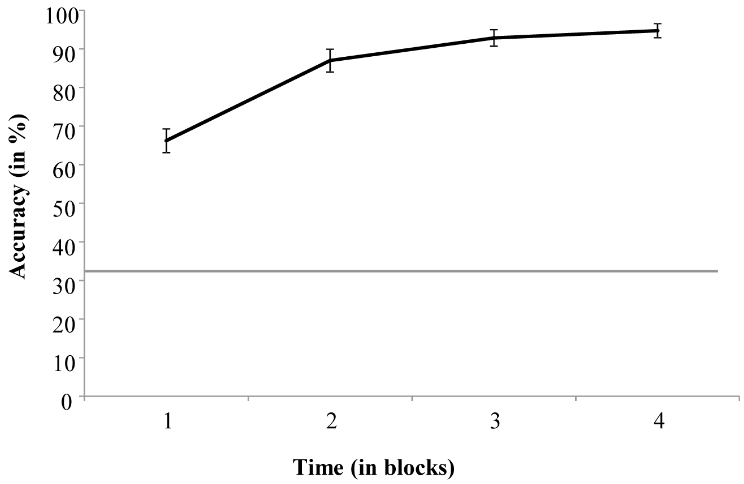


**Figure 2.**
Average accuracy across blocks. Errors bars mark the standard error of the mean.

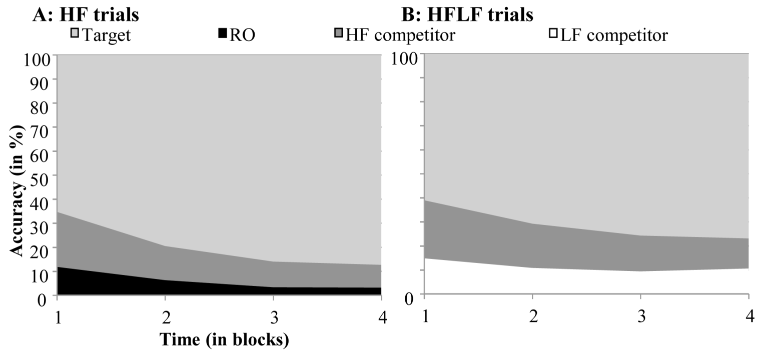


**Figure 3.**

The distribution of overt responses (indicated by shading) as a function of block. Target (light grey) is correct.

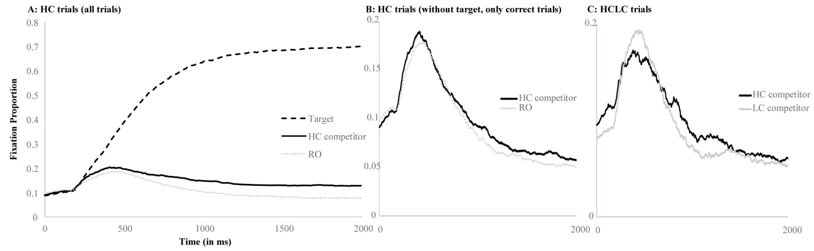


**Figure 4.**

Time course graphs for HC trials with looks to the target for all trials (Panel A) and without looks to the target (Panel B) and for HCLC trials (Panel C) including only correct trials.

Author Manuscript

Author Manuscript

Author Manuscript

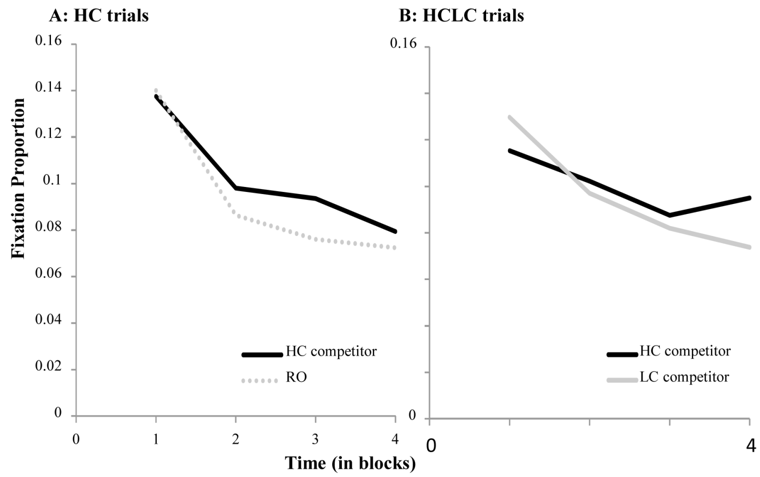


**Figure 5.**

Looks to the competitors in HC (Panel A) and HCLC trials (Panel B) (separated by block).

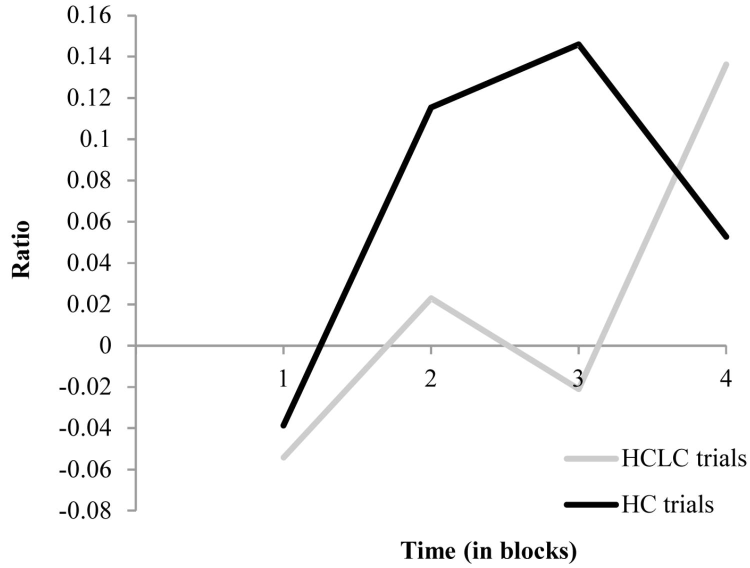


**Figure 6.**
The mean log-odds ratio as a function of block.

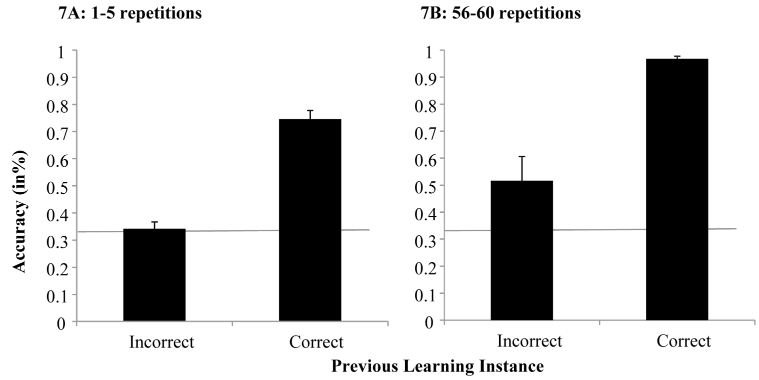


**Figure 7.**

Accuracy as a function of how participants responded on previous trials with the same target for five target replications (as used by Trueswell et al., 2013) at beginning of experiment (7A) and at the end of the experiment (7B). Error bars indicate the standard error of the mean.

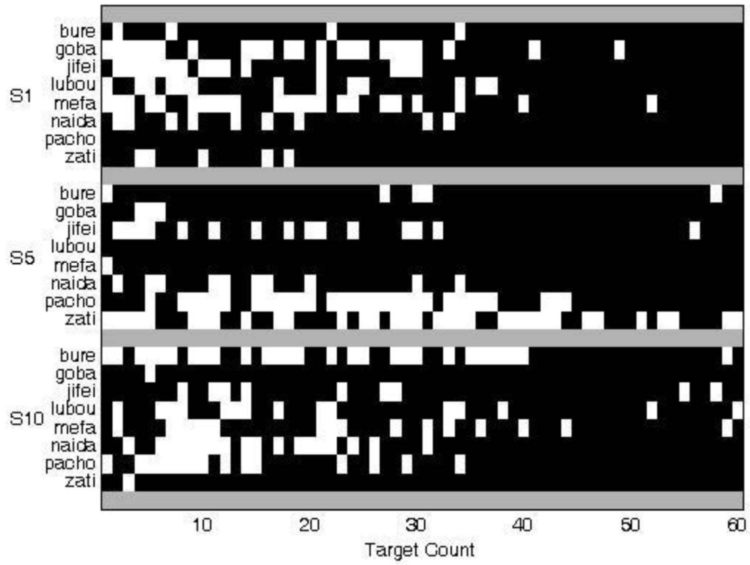


**Figure 8.**

Accuracy per item for three exemplary participants (S1, S5, S10) (black = correct trials; white = incorrect trials).

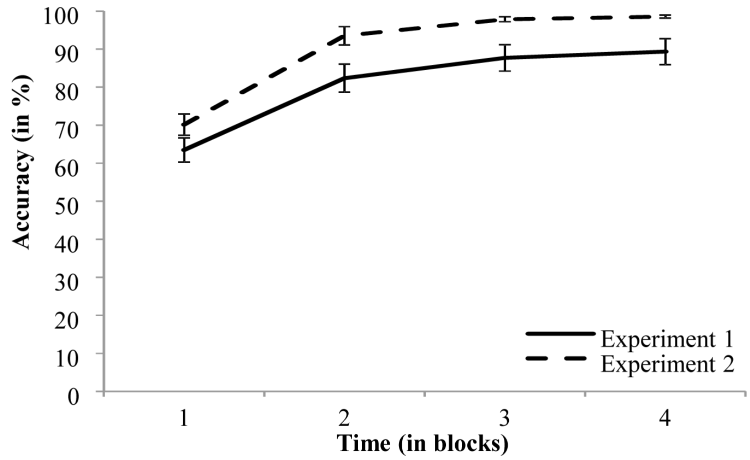


**Figure 9.**
Average accuracy across blocks for Experiment 1 and 2.

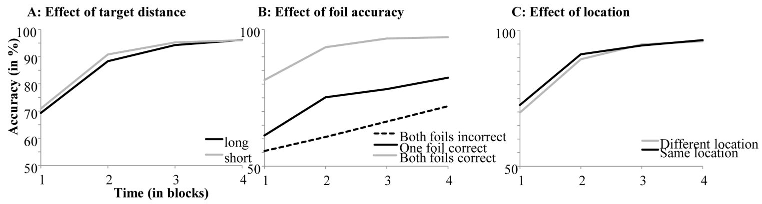


**Figure 10.**

Main effects of target distance (Panel A), foil accuracy (B) and location of target in prior trial (C). Target distance was separated by a median split, median = 6.

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1**

Novel words used

| Written form | IPA |
|---|---|
| Mefa | /meɪfɑ/ |
| Goba | /goubɑ/ |
| Jifei | /dʒifeɪ/ |
| Bure | /buɹeɪ/ |
| Naida | /nɑɪdɑ/ |
| Zati | /zæti/ |
| Lubou | /lubo/ |
| Pacho | /pɑtʃou/ |

**Table 2**

Example of co-occurrence statistics over all four blocks.

| | Object 1 | Object 2 | Object 3 | Object 4 | Object 5 | Object 6 | Object 7 | Object 8 |
|---|---|---|---|---|---|---|---|---|
| **Mefa** | 60 | 36 | 24 | 12 | 12 | 12 | 12 | 12 |
| **Goba** | 12 | 60 | 36 | 24 | 12 | 12 | 12 | 12 |
| **Jifei** | 12 | 12 | 60 | 36 | 24 | 12 | 12 | 12 |
| **Bure** | 12 | 12 | 12 | 60 | 36 | 24 | 12 | 12 |
| **Naida** | 12 | 12 | 12 | 12 | 60 | 36 | 24 | 12 |
| **Zati** | 12 | 12 | 12 | 12 | 12 | 60 | 36 | 24 |
| **Lubou** | 24 | 12 | 12 | 12 | 12 | 12 | 60 | 36 |
| **Pacho** | 36 | 24 | 12 | 12 | 12 | 12 | 12 | 60 |

**Table 3**

The four trial-types and the number of times each is repeated in a single block and over the course of the experiment.

| Trial-type | Object types on screen | Repetitions / block and word | Repetitions |
|---|---|---|---|
| HCLC | Target<br>HC competitor<br>LC competitor | 2 | 64 |
| HC | Target<br>HC competitor<br>Random object | 7 | 224 |
| LC | Target<br>LC competitor<br>Random object | 4 | 128 |
| RO | Target<br>Random object 1<br>Random object 2 | 2 | 64 |