Sampling to learn words: Adults and children sample words that reduce referential ambiguity

Martin Zettersten* & Jenny R. Saffran

Department of Psychology
University of Wisconsin-Madison
1202 W Johnson Street
Madison, WI 53706 U.S.A.

Corresponding author: Martin Zettersten, Department of Psychology, 1202 W Johnson St, Madison, WI 53706, USA, zettersten@wisc.edu

Data availability statement:
The data that support the findings of this study are openly available in a publicly accessible repository on the Open Science Framework (OSF) at https://osf.io/udmvh/ and on GitHub at https://github.com/mzettersten/crossAct.

Acknowledgements
This work was supported by the NSF-GRFP DGE-1747503 to MZ and grants from the NICHD awarded to JS (R37HD037466) and the Waisman Center (U54 HD090256). We would like to thank Joe Austerweil, Viridiana Benitez, and Erica Wojcik for valuable discussion at various points during the project. Special thanks to Louisa Forrest, Tess Gapinski, and Andrew Kressin for aiding in data collection, the Madison Children's Museum and all the families who participated. We would also like to thank Celeste Kidd for providing the visual stimuli used in the experiments. An earlier version of this work appeared in the Proceedings of the 41st Annual Meeting of the Cognitive Science Society, 2019.

Word count: 5,613

Research Highlights

- We tested whether learners would sample words that reduced referential uncertainty by manipulating the ambiguity of novel items in a cross-situational word learning task.
- Adults actively sampled items that reduced referential uncertainty.
- Children (3-8 years of age) also sampled items that reduced uncertainty, once the referential ambiguity of learning trials was made salient.
- The propensity to sample items that reduced referential uncertainty increased with age.

Abstract

How do learners gather new information during word learning? One possibility is that learners selectively sample items that help them reduce uncertainty about new word meanings. In a series of cross-situational word learning tasks with adults and children, we manipulated the referential ambiguity of label-object pairs experienced during training and subsequently investigated which words participants chose to sample additional information about. In the first experiment, adult learners chose to receive additional training on object-label associations that reduce referential ambiguity during cross-situational word learning. This ambiguity-reduction strategy was related to improved test performance. In two subsequent experiments, we found that, at least in some contexts, children (3-8 years of age) show a similar preference to seek information about words experienced in ambiguous word learning situations. In Experiment 2, children did not preferentially select object-label associations that remained ambiguous during cross-situational word learning. However, in a third experiment that increased the relative ambiguity of two sets of novel object-label associations, we found evidence that children preferentially make selections that reduce ambiguity about novel word meanings. These results carry implications for understanding how children actively contribute to their own language development by seeking information that supports learning.

Word count: 194

Keywords: cross-situational word learning; active learning; self-directed learning; sampling; uncertainty reduction; mutual exclusivity

Sampling to learn words:
Adults and children sample words that reduce referential ambiguity

Why do we seek out new information during learning? One proposal is that information-seeking behavior is driven by uncertainty reduction (e.g., Kidd & Hayden, 2015). At least in some contexts, children may be motivated to gather information to reduce uncertainty after ambiguous or surprising events (Lapidow & Walker, 2020; Schulz & Bonawitz, 2007; Stahl & Feigenson, 2015). For instance, infants and children preferentially seek out information from social partners when they are more uncertain (Goupil, Romand-Monnier, & Kouider, 2016) or when confronted with ambiguous or incomplete information (Bazhydai, Westermann, & Parise, 2020; Hembacher, DeMayo, & Frank, 2020; Vaish, Demir, & Baldwin, 2011). Understanding the nature of children's information-seeking strategies may provide key insights concerning how children are able to rapidly solve complex learning problems across development (Gopnik et al., 2017; Oudeyer & Smith, 2016).

A classic problem in word learning is how learners determine the meanings of words in potentially ambiguous situations (Quine, 1960). One solution is that children disambiguate word meanings by tracking co-occurrences of object-label pairs across multiple ambiguous situations (Smith & Yu, 2008; Suanda, Mugwanya, & Namy, 2014; Yurovsky & Frank, 2015). This proposal would be particularly powerful when combined with a drive to seek information that reduces uncertainty: if learners are motivated to sample object-label associations that remain ambiguous over the course of learning, this may substantially improve word learning outcomes (Hidaka, Torii, & Kachergis, 2017; Keijser, Gelderloos, & Alishahi, 2019).

Computational analyses suggest that self-directed sampling can vastly simplify the task of learning new words and categories under uncertainty, as long as learners employ strategies that increase the frequency of word meanings that have higher uncertainty (Hidaka et al., 2017; Oudeyer, Kachergis, & Schueller, 2019; Settles, 2012). For example, in one computational model of object-label association learning (Hidaka, Torii, & Kachergis, 2017), successfully learning an adult-sized vocabulary (around 60,000 words) required an unrealistic number of sampling trials when learning events were assumed to be drawn randomly from a Zipfian word distribution. However, when the model implemented an active learner that preferentially selected less frequently encountered object-label associations, learning sped up by several orders of magnitude. Other models have implemented active word learning in terms of curiosity mechanisms that preferentially sample objects that are expected to lead to the largest increases in accuracy (Twomey & Westermann, 2017). Sampling learning events based on active selection mechanisms leads to large increases in word learning speed and accuracy compared to randomly selected learning events, e.g. when learning to map labels to objects in visual scenes with multiple possible referents (Gelderloos, Mahmoudi Kamelabad, & Alishahi, 2020; Keijser et al., 2019). These models suggest that at least in principle, mechanisms that make selections based on potential learning gain or uncertainty reduction can speed up and simplify the problem of tracking object-label associations.

What sampling strategies do learners use when faced with uncertainty about novel word meanings? Children are sensitive to referentially ambiguous situations, preferentially seeking information from social partners when confronted with referential ambiguity (Hembacher et al., 2020; Hembacher & Frank, 2017; Vaish et al., 2011). Children also show stronger word learning outcomes for objects they express more interest in (Ackermann, Hepach, & Mani, 2019; Lucca

& Wilbourn, 2018). Among adult learners, active selection of label-object pairs during cross-situational word learning increases accuracy compared to a passive condition in which random sets of objects are presented (Kachergis, Yu, & Shiffrin, 2013). However, little is known about children's and adults' sampling strategies when given the opportunity to directly control their learning environment. What kind of word learning input do children actively seek? Investigating the types of information that learners sample during word learning, and how these information-seeking strategies emerge in development, can help us understand when and how active learning plays a critical role in learning novel words.

In the current experiments, we investigated whether adult and child learners actively seek information that will serve to reduce ambiguity about the meanings of novel words. We manipulated the ambiguity of novel object-label mappings by varying the degree to which object-label pairs co-occurred with one another during cross-situational word learning (Experiments 1 and 2) or whether children could use mutual exclusivity to disambiguate the referents of novel words (Experiment 3). The central question is whether adults and children preferentially sample the items that are most helpful in reducing uncertainty about novel object-label associations.

Experiment 1

Experiment 1 was designed to determine whether adults seek information that aids in disambiguating reference. Participants completed a cross-situational learning task in which their goal was to learn a set of object-label associations by determining the referent of each label across training. Participants were then given the opportunity to select which object-label association they would observe on each subsequent learning trial. The central question was whether adults make selections that reduce referential ambiguity.

Method

Participants

We recruited 28 participants through Amazon Mechanical Turk (8 female; mean age: 31.4 years, SD=7.25; all native speakers of English). Three additional participants were excluded for not passing an initial auditory attention check (2) or for restarting the experiment (1). Participants were paid $0.75 for completing the study.

Stimuli

The objects were 8 images of novel 'alien' creatures used in previous studies (Partridge, McGovern, Yung, & Kidd, 2015). Eight novel words (*beppo*, *finna*, *guffi*, *kita*, *noopy*, *manu*, *sibu*, *tesser*) were recorded by a female native speaker of English and normalized in duration and average loudness. The association between each label and its target referent and the roles of the stimuli within a condition were randomized across participants. The stimuli were presented using jsPsych (de Leeuw, 2015). All stimuli and experimental scripts for Experiment 1 and all subsequent experiments can be viewed on the Open Science Framework (OSF) [https://osf.io/udmvh/].

Design & Procedure

The experiment consisted of Training, Sampling, and Test Phases.

Training Phase. Participants completed 24 cross-situational learning trials (two blocks of 12 trials), presented in random order (Figure 1). Participants were instructed that their goal was to learn the association between eight novel labels and their referents. On each trial, participants were presented with two referents and two labels. The labels appeared sequentially in random order, both visually and auditorily. Consequently, the association between a particular label and its referent remained ambiguous on any single trial, but could be disambiguated by aggregating information across trials. Each object and its label occurred six times across the 24 training trials.

We manipulated whether or not the object-label associations became disambiguated across training trials. Half of the object-label pairs remained ambiguous: two sets of two items were yoked together such that they were never disambiguated across training (ambiguous items; Figure 2, left panel). The other half of the object-label pairs were disambiguated across the training trials; each occurred with three different object-label pairs across trials (disambiguated items; Figure 2, right panel). Note that regardless of item type, each individual object and label appeared equally frequently. Due to this manipulation, half of the label-object pairs remained highly uncertain at the onset of the Sampling phase, while the other half were (potentially) disambiguated.

Sampling Phase. Participants next completed four sampling trials. On each trial, all 8 objects appeared in randomized locations. Participants were instructed to select which of the 8 items they wanted to hear on the next cross-situational learning trial. After participants made a selection, a second object was chosen at random from the remaining objects. The two objects and their labels then appeared together in a cross-situational word learning trial with the same structure as in the training phase.

Test Phase. Participants' knowledge of the object-label associations was probed in an 8-AFC recognition test. On each test trial, all eight objects appeared in randomized locations on the screen, along with one of the eight labels. Participants were asked to select the object that went with the label. No feedback was provided. Participants were tested on each label in random order, for a total of eight test trials.

Results

The data and scripts documenting the data analysis for all experiments are openly available on the Open Science Framework [https://osf.io/udmvh/]. These materials include a walkthrough of all analyses reported in the manuscript, including further modeling details and supplementary analyses described in the Results section (accessible through a web browser at the following link: [https://mzettersten.github.io/crossAct/analysis/Crossact.html]).

Sampling choices

We used the lme4 package version 1.1-21 in R to fit a logistic mixed-effects model testing participants' likelihood of making an ambiguous selection against a chance level of 0.5 (chance level was 0.5, since the probability of selecting an ambiguous item by chance was 4/8 = 0.5; note that since logit(0.5)=0, the test of our main hypothesis is represented by whether the intercept in the logistic mixed-effects model differs from zero), including by-participant and by-item random intercepts (Bates, Mächler, Bolker, & Walker, 2015; R Development Core Team, 2019). The model was specified as follows:

$$\text{ambiguous selection} \sim 1 + (1|\text{participant}) + (1|\text{item})$$

Participants were more likely to choose ambiguous items than disambiguated items, $b = 0.62$, Wald 95% CI = [0.06, 1.17], $z=2.16$, $p=.03$. Participants chose an object from the ambiguous set on 62.5% (95% CI=[50.8%, 74.2%]) of trials (Figure 3A).

To test the robustness of participants' preference for sampling ambiguous items, we also tested subjects' average proportion of ambiguous selections against the chance level of 0.5 in a non-parametric statistical test. A Wilcoxon signed-rank test found that the distribution of ambiguous selections significantly diverged from chance, $V=152$, $p=.02$.

Test performance

Overall test accuracy. Overall, participants showed learning of the object-label pairs, accurately selecting the correct referent (M=65.6%, 95% CI=[52.7%, 78.6%], chance=12.5%). Participants' accuracy was significantly greater than chance in a logistic-mixed effects model including by-participant and by-item random intercepts and specifying an offset corresponding to the logit of chance performance, $z = 6.26$, $p < .001$. To compare accuracy for items that remained ambiguous during training to accuracy for items disambiguated during training (Figure 2), we fit a logistic mixed-effects model predicting trial-by-trial accuracy from item type (centered; ambiguous=0.5; disambiguated=-0.5), including by-participant and by-item random intercepts and a by-participant random slope for item type. We specified the model as follows:

$$\text{correct choice} \sim 1 + \text{item type} + (1 + \text{item type}|\text{participant}) + (1|\text{item})$$

Accuracy for ambiguous items (M=63.4%, 95% CI=[54.6%, 72.2%]; corrected within-participants; Morey, 2008) was marginally lower than accuracy for items that were disambiguated during training (M=67.9%, 95% CI=[59.1%, 76.6%]), z=-1.65, p=.10.

Relationship between sampling and test accuracy. Participants who chose more objects from the ambiguous set during the sampling phase accurately identified more words at test, $r(26)$=.58, 95% CI=[.27, .78], p=.001 (Figure 3B).

Discussion

In a cross-situational learning task, adult learners chose to learn more about object-label pairs that remained ambiguous throughout training. In the Supplementary Materials (section S1), we report an additional in-lab experiment replicating this result: adult participants preferentially sampled object-label associations that remained ambiguous at the end of the Training Phase.

These studies provide evidence that adults seek to reduce ambiguity about object-label associations when given the opportunity to control their learning experience.

Intriguingly, we found that participants' sampling behavior was correlated with their test accuracy: participants who chose more ambiguous items during the Sampling Phase also more accurately identified object-label associations. However, the directionality of this correlation is unclear. One possibility is that learners were more accurate on the test because they sampled more ambiguous items. Another possibility is that successfully learning novel words during training makes it more likely that learners will subsequently focus on ambiguous items during the Sampling Phase. In the supplementary in-lab replication experiment (see S1 for further details), we found preliminary evidence supporting the latter explanation– as participants become more successful at learning the novel label-object associations, they also become more likely to seek out information about ambiguous items. This finding underscores the point that uncertainty is dependent on past learning: which items are perceived as most uncertain will vary depending on how well participants have learned the new words. In the current experiment, adult participants were generally successful in learning the underlying label-object associations and actively sought out information about items that remained ambiguous during training.

Experiment 2

Next, we asked whether children (4-8 years of age) also seek out information that would reduce ambiguity during cross-situational learning. As in Experiment 1, one set of novel object-label associations could be inferred based on the object-label pairs they co-occurred with during training, while another set of words remained ambiguous. Children were then given the opportunity to sample object-label associations presented in isolation. The central question was whether children would prefer to select object-label associations that remained ambiguous at the end of the Training Phase.

Method

Participants

We recruited 38 participants (M=5.9 years, SD=1.19, range=4.1–8.1 years; 19 female) at a local children's museum. Two additional participants were excluded due to inattention (e.g., not attending during task instructions and/ or requiring a high degree of experimenter support; results including these individuals follow the same pattern).

Stimuli

The object stimuli included the eight 'aliens' from Experiment 1 (Partridge et al., 2015) and two cartoon images of familiar animals (penguin, dog). Eight novel words (*biffer*, *deela*, *guffi*, *sibu*, *tibble*, *leemu*, *zeevo*, *pahvy*) and two familiar words (*penguin*, *dog*) were recorded by a female native speaker of English and normalized in duration and average loudness. The association between the novel labels and target referents, as well as the particular roles of the novel label-referent stimuli, were randomized across participants. The stimuli were presented using jsPsych (de Leeuw, 2015).

Design & Procedure

Children were tested individually in a quiet room in the local children's museum on a 10.1" Samsung Galaxy Note tablet. An experimenter guided children by giving instructions at the beginning of each phase. The experiment was presented as a game in which a cartoon bear named Teddy would first teach children the names of new alien friends, and then ask children to help her find her friends. The experiment began with a Practice Phase (see supplementary materials S2 for details), followed by the main experiment consisting of three phases: Training, Sampling, and Test.

Training Phase. Participants completed nine cross-situational learning trials (three blocks of three trials each). On each training trial, participants saw two referents appear on the screen on either side of the Teddy character and heard the labels of the two objects presented sequentially in random order. Next, the objects switched locations in a brief animation, and participants heard the same two labels presented in the same order. We included this trial repetition with flipped locations in order to reduce children's tendency to interpret the labeling event as moving from left to right on the screen.

As in Experiment 1, we manipulated whether the object-label associations could be disambiguated across training trials (Figure 4). Each object-label pair occurred on three cross-situational training trials. Four of the objects occurred with three different object label pairs (disambiguated items). The remaining two object-label pairs always occurred with one another (ambiguous items), such that children never saw evidence allowing them to link the two words unambiguously with their respective referents.

Sampling Phase. After completing the training phase, participants completed four sampling trials. On each sampling trial, all six referents appeared in randomized locations on the screen. Participants were instructed to select which of the six items they wanted to learn about next (Figure 4). When participants tapped one of the six referents, a brief animation moved the item to the center of the screen while the remaining items disappeared, and the referent was subsequently labeled in isolation.

Test Phase. Participants' knowledge of the object-label associations was probed in a 6-AFC recognition test. On each test trial, all six referents appeared in randomized locations on the screen surrounding the Teddy character. When participants tapped Teddy in the center of the screen, they heard one of the six labels. Participants were instructed to help Teddy by selecting the friend she was looking for. No feedback was provided after a choice. Participants were tested on each label in random order, for a total of six recognition test trials.

Results

Sampling choices

We fit a logistic-mixed effects model testing whether children's likelihood of selecting an ambiguous item differed from chance (0.33, since the probability of randomly selecting an ambiguous item was 2/6 = 0.33), including by-participant and by-item random intercepts. Contrary to our prediction, children did not select ambiguous object-label associations more frequently than would be expected by chance during the Sampling phase, b=-0.02, Wald 95% CI = [-0.36, 0.32], z=-0.12, p=.91. Participants chose an object from the ambiguous set on 32.9% of trials (95% CI=[27.1%, 38.7%]; Figure 5). A Wilcoxon signed-rank test conducted on children's proportion of ambiguous selections yielded comparable results (V=409, p=.57). In order to investigate whether children's sampling preferences changed with age, we fit a logistic mixed-effects model predicting the likelihood of selecting ambiguous items from age, including by-participant and by-item random intercepts. Children's propensity to select an ambiguous item was unrelated to age, b=0.004, Wald 95% CI=[-0.28, 0.29], z=0.03, p=.98.

Test performance

To investigate participants' test performance, we fit the same logistic mixed-effects models as in Experiment 1 (including the identical random effects structure). Overall, participants showed significant learning of the object-label pairs, choosing the correct object to go with a label at above-chance levels (chance = 0.167; the probability of selecting a target object if choosing completely at random), M=38.6%, 95% CI=[30.7%, 46.5%], z=6.46, p<.001. Surprisingly, children performed more accurately on the ambiguous items (M=48.7%, 95% CI=[40.0%, 57.4%]) than on the disambiguated items (M=33.6%, 95% CI=[24.8%, 42.3%]), b=0.68, Wald 95% CI=[0.08, 1.28], z=2.23, p=.026. When tested on ambiguous items, children had a strong preference to select one of the two ambiguous objects (61.8% of trials, 95% CI=[51.4%, 72.3%]) rather than the four disambiguated objects (chance=0.33). When tested on disambiguated items, children tended not to choose the two ambiguous objects, selecting them on only 18.4% of trials (95% CI=[8.0%, 28.9%]).

Relationship between sampling selections and test performance

We investigated the relationship between children's selections during the Sampling Phase and their subsequent test accuracy on sampled (vs. non-sampled) items. To do so, we fit a logistic mixed-effects model predicting children's test accuracy from item type (centered; ambiguous = 0.5; disambiguated = -0.5), sampling choice, i.e. whether or not the item was chosen by a participant during the Sampling Phase (centered; sampled = 0.5; not sampled = -0.5), and their interaction, including by-participant and by-item random intercepts, and a by-participant random slope for item type. There was a significant effect of item type, b=0.83, 95% Wald CI = [0.15, 1.50], z=2.41, p=.016. There was also a significant effect of sampling choice, b=0.89, 95% Wald CI=[0.23, 1.55], z=2.66, p=.008 (see Figure S3.3 in the supplementary materials for additional details). This indicates that participants performed more accurately on test items that they had previously selected during the Sampling Phase (M=45.2%, within-participant 95% CI=[34.6%, 55.7%]) than on items they did not sample (M=26.8%, within-participant 95% CI=[16.2%, 37.3%]). There was no significant interaction between item type and sampling choice, p=.48.

As in Experiment 1, we also tested whether children's tendency to select items from the ambiguous set predicted their test accuracy, and found no evidence that higher number of ambiguous items selected was associated with better learning, $r(36) = .22$, 95% CI = [-.11, .50], $p = .18$.

Discussion

Unlike adult learners, children did not show a tendency to select object-label associations that remained ambiguous during training. While children did not exhibit a bias towards sampling ambiguous items, children tended to have higher test accuracy for items that they selected during the Sampling Phase, suggesting that sampling behavior was linked to subsequent learning. Surprisingly, children performed better on ambiguous object-label associations than on object-label associations that were disambiguated across training trials. There are likely two reasons why children showed higher accuracy on the ambiguous items. First, since the two ambiguous items always co-occurred with one another, the training could help learners constrain the set of possible competitors for a given ambiguous label to two objects (compared to four possible objects for the disambiguated items). Indeed, children appeared to constrain their choices to the two objects that co-occurred on ambiguous trials when tested on their respective labels and rarely chose these objects when tested on the labels that occurred with the disambiguated objects.

Second, anecdotally, we observed that many children explicitly pointed to specific objects during training while listening to each label and even repeated the respective label for each object. This behavior may indicate that some children were generating explicit hypotheses about each word mapping (Trueswell, Medina, Hafri, & Gleitman, 2013). If a child formed a specific hypothesis about the mapping between the two labels and objects on the first ambiguous trial, they would subsequently hear evidence that would appear to confirm their hypothesis: the two labels and the two objects would occur together again on the subsequent two training trials. “Hypothesis-testers” would never experience evidence disconfirming their initial hypotheses. Crucially, one consequence of learners approaching the task in this manner is that the two object-label associations deemed “ambiguous” according to the experimental design may not have actually appeared notably less ambiguous to children performing the task than the putatively disambiguated items. Thus, in our next study, we adapted the task to create a learning situation in which one set of object-label associations would be more clearly ambiguous from the perspective of the child learner.

Experiment 3

In Experiment 3, we sought to increase the likelihood that children would perceive some novel object-label associations as more ambiguous than others. We used mutual exclusivity to increase the ease with which children could infer word-referent pairs for one set of novel objects (Lewis, Cristiano, Lake, Kwan, & Frank, 2020; Markman & Wachtel, 1988) while maintaining the ambiguity of a second set of novel word-referent pairs (as in the previous experiments). By giving children the opportunity to infer the referents for novel objects occurring on mutual exclusivity trials, we aimed to make it easier for children to recognize the ambiguity of consistently co-occurring object-label associations.

Method

Participants

We recruited 56 new participants (M=5.5 years, SD=1.18, range=3.3–7.9, 33 female) at a local children’s museum.[1]

---

[1] The original target age range for the study was 4.0 years-8.0 years. Three 3-year-olds were recruited and run in the experiment. Given that all three children completed the experiment without issue, we opted for an inclusive data

Two additional participants were excluded due to interruptions to the experiment (n=1) or not completing the study (n=1).

Stimuli

The novel stimuli consisted of six images and recordings composed of a subset of the items used in Experiment 2. In addition, 4 cartoon images of familiar animals (cow, dog, monkey, pig) along with audio recordings of their respective labels were used. Stimuli were recorded by the same female native speaker of English and normalized in duration and average loudness.

Design & Procedure.

The procedure and testing conditions were identical to Experiment 2.

Training Phase. Participants completed 9 cross-situational learning trials (3 blocks of 3 trials each) with 6 object-label pairs, two familiar object-label pairs (e.g., pig and dog) and four novel object-label pairs chosen randomly from the set of novel stimuli. As in Experiment 2, two referents appeared on the screen on each trial paired with two labels presented in random order. Two novel object-label associations always occurred with one another (ambiguous items), mirroring the ambiguity manipulation from Experiments 1 and 2. The two remaining novel object-label associations served as mutual exclusivity items; each novel object-label pair was yoked to a familiar object-label pair (i.e., one alien always occurred with the dog image, while the other always occurred with the pig image). We reasoned that children should be able to disambiguate reference for the mutual exclusivity items (i.e., when seeing an image of a dog and a novel “alien”, on hearing the words *leemu* and *dog*, children would successfully infer that

---

policy and included these participants in the analyses. All analytic results and conclusions are qualitatively similar if these three participants are excluded.

leemu referred to the novel alien). This would make it more likely that the ambiguous items would be perceived by child learners as having relatively high referential uncertainty. As in previous experiments, all novel objects and their labels occurred equally frequently across the training phase.

Sampling Phase. Participants next completed two sampling trials. On each trial, the four novel objects appeared on the screen and children were instructed to choose which object they wanted to learn more about. The procedure was otherwise identical to Experiment 2.

Test Phase. Participants' knowledge of the six words from the training phase (4 novel, 2 familiar words) was tested in a 6-AFC recognition task in the same procedure as in Experiment 2.

Results

Sampling choices

We fit a logistic-mixed effects model testing whether children's likelihood of selecting an ambiguous item differed from chance (chance level was 0.5, since the probability of selecting an ambiguous item by chance was 2/4 = 0.5), including by-participant and by-item random intercepts. As predicted, children preferentially selected ambiguous object-label associations during the Sampling phase, b=0.55, Wald 95% CI=[0.15, 0.95], z=2.71, p=.007 (Figure 6A). Participants chose an object from the ambiguous set on 63.4% of trials (95% CI=[54.4%, 72.4%]) (chance level=0.5). A Wilcoxon signed-rank test yielded similar results (V=330, p=.006). In order to investigate whether the propensity for making ambiguous selections increased with age, we fit a logistic mixed-effects model predicting the likelihood of an ambiguous selection from Age, including by-participant and by-item random intercepts. Unlike

Experiment 2, there was a significant effect of Age, b=0.46, Wald 95% CI=[0.10, 0.82], z=2.48, p=.013 (Figure 6B).

Test performance

We fit logistic mixed-effects models analogously to Experiments 1 and 2 (including the identical random effects structure). Overall, participants showed significant learning of the object-label pairs, choosing the correct object to go with a label at above-chance levels (chance selection of novel object = 0.25), M=57.6%, 95% CI=[48.4%, 66.8%], z=5.08, $p < .001$. Accuracy for mutual exclusivity items (M=61.6%, 95% CI=[52.8%, 70.4%]) and for the ambiguous items (M=53.6%, 95% CI=[44.8%, 62.4%]) was not significantly different, b=-.44, Wald 95% CI=[-1.15, 0.27], z=-1.22, p=.22 (see supplementary material S4.1 for a graphical representation of the data).

Relationship between sampling selections and test performance

As in Experiment 2, we investigated the relationship between children's selections during the Sampling Phase and their subsequent accuracy on sampled (vs. non-sampled) items. We fit the same logistic mixed-effects model predicting children's test accuracy from Item Type (centered; ambiguous = 0.5; mutual exclusivity = -0.5), sampling choice, i.e. whether or not the item was chosen by a participant during the Sampling Phase (centered; sampled=0.5; not sampled=-0.5), and their interaction, including by-participant and by-item random intercepts, and a by-participant random slope for Item Type. There were no significant effects of sampling choice (p=.33) or item type (p=.15), and no significant interaction between the two (p=.77) (see S4.2. in the supplementary materials for further information). In addition, children's tendency to sample items from the ambiguous set was not related to test accuracy (r(54)=.07, 95% CI=[-.20, .32], p=.62).

Discussion

When given the opportunity to select which object-label pairs they wanted to learn more about, 3-8-year-olds preferentially selected object-label pairs that remained ambiguous during training over object-label pairs that could be disambiguated through mutual exclusivity. These findings demonstrate that – at least in some word learning situations – children preferentially select learning events that aid in reducing referential uncertainty. The tendency to make ambiguity-reducing selections began to emerge around 5 years of age in our sample.

While children's tendency to sample ambiguity-reducing items has the potential to be a powerful driver in word learning (Hidaka et al., 2017; Keijser et al., 2019), the impact of this sampling strategy on learning outcomes remains open in the current work. Children learned the ambiguous and the mutual exclusivity items at similar rates, consistent with the fact that children used the sampling phase to aid in disentangling the reference of novel words. However, there was no evidence that children performed more accurately during the test on label-object pairs that they had previously selected during the Sampling Phase. Since the experiment was principally designed to answer questions about sampling strategy, we had limited power to trace the impact of children's selections on subsequent learning. In particular, the small number of sampling trials provided to children (n=2) limits our ability to measure correlations between sampling preference and learning. Future work will delve deeper into questions concerning the impact of children's sampling preferences by systematically manipulating the input children select and receive prior to test (e.g., Kachergis et al., 2013; Markant & Gureckis, 2014) and by increasing power to measure stable individual differences in children's sampling strategies.

General Discussion

When learning the referents of novel labels in ambiguous contexts, adult learners chose to learn more about object-label associations that remained ambiguous at the end of training (Experiment 1). Children also spontaneously sampled object-label associations that reduce ambiguity, though only when the task was simplified to emphasize referential ambiguity. When presented with a similar task as adults, children did not choose to learn about object-label associations that remained ambiguous during training (Experiment 2). However, this result is likely at least partially explained by the fact that children – contrary to our expectation - did not link novel words to their target objects more readily for disambiguated items than for ambiguous items in Experiment 2. In a simplified design that highlighted the ambiguous nature of the trials in which two referents always occurred together (Experiment 3), children chose to learn about items that reduced uncertainty about the words' referents.

The preference for selecting ambiguous items was strongly related to age, with children beginning to reliably select the ambiguous items around 5 years of age in our sample. Past work on social referencing suggests that children as young as 2 years of age (Hembacher et al., 2020) and even infants as young as 12 months are sensitive to referential uncertainty (Bazhydai et al., 2020; Vaish et al., 2011). Our studies go beyond measuring sensitivity to uncertainty by asking whether child learners choose to sample new words based on referential ambiguity. Proactively making sampling decisions based on uncertainty may require more sophisticated skills in metacognition (Ghetti, Hembacher, & Coughlin, 2013; Lyons & Ghetti, 2011) and cognitive control (Munakata, Snyder, & Chatham, 2012) that undergo substantial development during early childhood. Similar to Experiment 2, the absence of a sampling preference among younger children in Experiment 3 may also be due to younger participants being less likely to encode the learning input as providing ambiguous information for some sets of words. Limits on the extent to which younger children spontaneously make ambiguity-reducing selections raise important questions for future research on when children encode input as providing ambiguous information about word reference, the contexts in which children can effectively employ sampling strategies that reduce this ambiguity, and how children's sampling strategies develop and interact with their cognitive development more generally.

While the current experiments focused on learners' active sampling strategies, an important direction for future work will be investigating how children's active selections influence learning outcomes. In Experiments 1 and 2, we found preliminary evidence that sampling choices were associated with learning. In Experiment 1, adults who tended to select ambiguous items during the sampling phase had higher test accuracy. However, this correlation may be at least partially explained by better learners preferring to select ambiguous items, rather than ambiguous selections driving better learning alone (see S1 in the supplementary materials for further discussion). In Experiment 2, children were more accurate for test items they had selected during the sampling phase. However, we found no relationship between children's sampling selections and their subsequent test accuracy in Experiment 3.

There are several possible explanations for why we find mixed evidence for a relationship between sampling and subsequent test performance in the current experiments. First, our results highlight that information-seeking strategies and past learning are mutually dependent. What information is most relevant to the learner depends on what they have already learned. The reciprocal relationship between sampling preference and learning is clearly visible in the adult experiments – adults who preferred to sample ambiguous items were the most successful word learners, but successfully learning words during the training phase was likely a prerequisite for recognizing and subsequently sampling ambiguous object-label associations. The mutual dependence of sampling and learning highlights the need for experimental manipulations geared towards systematically manipulating sampling experiences to understand the influence of sampling selections on subsequent learning, particularly in future research with children (Markant & Gureckis, 2014; Partridge et al., 2015; Sim, Tanner, & Alpert, 2015). Second, it is important to consider how different sampling strategies may be associated with tradeoffs as learners encode new words, and how these may interact with developments in working memory (Vlach, 2019; Wojcik, 2013). While selecting an item for further study that has been associated with ambiguity in the past reduces referential uncertainty, it also comes at the cost of opportunities for studying other items - perhaps hampering successful maintenance of memory for items not sampled. Whether and how learners manage these tradeoffs in a manner that supports overall word learning is an important question for future research, for example by comparing overall learning when children actively construct learning events and when learning events are randomly generated (e.g., Kachergis et al., 2013).

Children have considerable control over their "curriculum" as they learn new words (Mani & Ackermann, 2018; Smith, Jayaraman, Clerkin, & Yu, 2018), with potentially immense consequences for the difficulty of the learning problems they face (Hidaka et al., 2017). While children are confronted with substantial referential uncertainty, active sampling strategies have the potential to structure and simplify the complex problem of linking words with their meanings in ambiguous contexts (Keijser et al., 2019; Yu & Smith, 2012). The present results demonstrate that, at least in some circumstances, children preferentially sample new words that reduce referential ambiguity. These studies contribute to a growing literature demonstrating that children are curious learners who actively contribute to their own language development.

References

Ackermann, L., Hepach, R., & Mani, N. (2019). Children learn words easier when they are interested in the category to which the word belongs. *Developmental Science*, 1–12. https://doi.org/10.1111/desc.12915

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bazhydai, M., Westermann, G., & Parise, E. (2020). "I don't know but I know who to ask": 12-month-olds actively seek information from knowledgeable adults. *Developmental Science*, 1–10. https://doi.org/10.1111/desc.12938

de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*, 1–12. https://doi.org/10.3758/s13428-014-0458-y

Gelderloos, L., Mahmoudi Kamelabad, A., & Alishahi, A. (2020). Active word learning through self-supervision. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings for the 42nd Annual Meeting of the Cognitive Science Society* (pp. 1050–1056). Cognitive Science Society.

Ghetti, S., Hembacher, E., & Coughlin, C. A. (2013). Feeling uncertain and acting on it during the preschool years: A metacognitive approach. *Child Development Perspectives*, *7*(3), 160–165. https://doi.org/10.1111/cdep.12035

Gopnik, A., O'Grady, S., Lucas, C. G., Griffiths, T. L., Wente, A., Bridgers, S., … Dahl, R. E. (2017). Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood. *Proceedings of the National Academy of*

Sciences, *114*(30), 7892–7899. https://doi.org/10.1073/pnas.1700811114

Goupil, L., Romand-Monnier, M., & Kouider, S. (2016). Infants ask for help when they know they don't know. *Proceedings of the National Academy of Sciences*, *113*(13), 3492–3496. https://doi.org/10.1073/pnas.1515129113

Hembacher, E., DeMayo, B., & Frank, M. C. (2020). Children's social information seeking is sensitive to referential ambiguity. *Child Development*, 1–16. https://doi.org/10.1111/cdev.13427

Hembacher, E., & Frank, M. C. (2017). Children's social referencing reflects sensitivity to graded uncertainty. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society* (pp. 495–500). Austin, TX: Cognitive Science Society.

Hidaka, S., Torii, T., & Kachergis, G. (2017). Quantifying the impact of active choice in word learning. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 519–525). Austion, TX: Cognitive Science Society.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2013). Actively learning object names across ambiguous situations. *Topics in Cognitive Science*, *5*(1), 200–213. https://doi.org/10.1111/tops.12008

Keijser, D., Gelderloos, L., & Alishahi, A. (2019). Curious topics: A curiosity-based model of first language word learning. In *Proceedings of the 41st Annual Conference of the Cognitive Science Society* (pp. 1991–1997). Montreal, QB: Cognitive Science Society.

Kidd, C., & Hayden, B. Y. (2015). The psychology and neuroscience of curiosity. *Neuron*, *88*, 449–460. https://doi.org/10.1016/j.neuron.2015.09.010

Lapidow, E., & Walker, C. M. (2020). Informative experimentation in intuitive science: Children select and learn from their own causal interventions. *Cognition*, *201*, 104315. https://doi.org/10.1016/j.cognition.2020.104315

Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., & Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, *198*, 104191. https://doi.org/10.1016/j.cognition.2020.104191

Lucca, K., & Wilbourn, M. P. (2018). The what and the how: Information-seeking pointing gestures facilitate learning labels and functions. *Journal of Experimental Child Psychology*, *178*, 417–436. https://doi.org/10.1016/j.jecp.2018.08.003

Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development*, *82*(6), 1778–1787. https://doi.org/10.1111/j.1467-8624.2011.01649.x

Mani, N., & Ackermann, L. (2018). Why do children learn the words they do? *Child Development Perspectives*, *12*(4), 253–257. https://doi.org/10.1111/cdep.12295

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology. General*, *143*(1), 94–122. https://doi.org/10.1037/a0032108

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Morey, R. D. (2008). Confidence Intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Munakata, Y., Snyder, H. R., & Chatham, C. H. (2012). Developing cognitive control: Three key transitions. *Current Directions in Psychological Science*, *21*(2), 71–77. https://doi.org/10.1177/0963721412436807

Oudeyer, P.-Y., Kachergis, G., & Schueller, W. (2019). *Computational and robotic models of early language development: A review*. Retrieved from http://arxiv.org/abs/1903.10246

Oudeyer, P.-Y., & Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, *8*(2), 492–502. https://doi.org/10.1111/tops.12196

Partridge, E., McGovern, M. G., Yung, A., & Kidd, C. (2015). Young children's self-directed information gathering on touchscreens. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (Eds.), *Proceedings of the Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

R Development Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Schulz, L. E., & Bonawitz, E. B. (2007). Serious fun: Preschoolers engage in more exploratory play when evidence is confounded. *Developmental Psychology*, *43*(4), 1045–1050. https://doi.org/10.1037/0012-1649.43.4.1045

Settles, B. (2012). *Active Learning*. *Synthesis Lectures on Artificial Intelligence and Machine Learning* (Vol. 6). https://doi.org/10.2200/S00429ED1V01Y201207AIM018

Sim, Z. L., Tanner, M., & Alpert, N. Y. (2015). Children learn better when they select their own data. In P. P. Noelle, D. C., Dale, R., Warlaumont, A. S., Yoshimi, J., Matlock, T., Jennings, C. D., & Maglio (Ed.), *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (pp. 2194–2199). Austin, TX: Cognitive Science Society.

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in Cognitive Sciences*, *22*(4), 325–336. https://doi.org/10.1016/j.tics.2018.02.004

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*, *348*(6230), 91–94. https://doi.org/10.1126/science.aaa3799

Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, *126*, 395–411. https://doi.org/10.1016/j.jecp.2014.06.003

Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, *66*(1), 126–156. https://doi.org/10.1016/j.cogpsych.2012.10.001

Twomey, K. E., & Westermann, G. (2017). Curiosity-based learning in infants: A neurocomputational approach. *Developmental Science*, 1–13. https://doi.org/10.1111/desc.12629

Vaish, A., Demir, Ö. E., & Baldwin, D. (2011). Thirteen- and 18-month-old infants recognize when they need referential information. *Social Development*, *20*(3), 431–449. https://doi.org/10.1111/j.1467-9507.2010.00601.x

Vlach, H. A. (2019). Learning to remember words: Memory constraints as double-edged sword mechanisms of language development. *Child Development Perspectives*, *13*(3), cdep.12337.

https://doi.org/10.1111/cdep.12337

Wojcik, E. H. (2013). Remembering new words: integrating early memory development into word learning. *Frontiers in Psychology*, *4*, 151. https://doi.org/10.3389/fpsyg.2013.00151

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. https://doi.org/10.1016/j.cognition.2012.06.016

Yurovsky, D., & Frank, M. C. (2015). An integrative account of constraints on cross-situational learning. *Cognition*, *145*, 53–62. https://doi.org/10.1016/j.cognition.2015.07.013

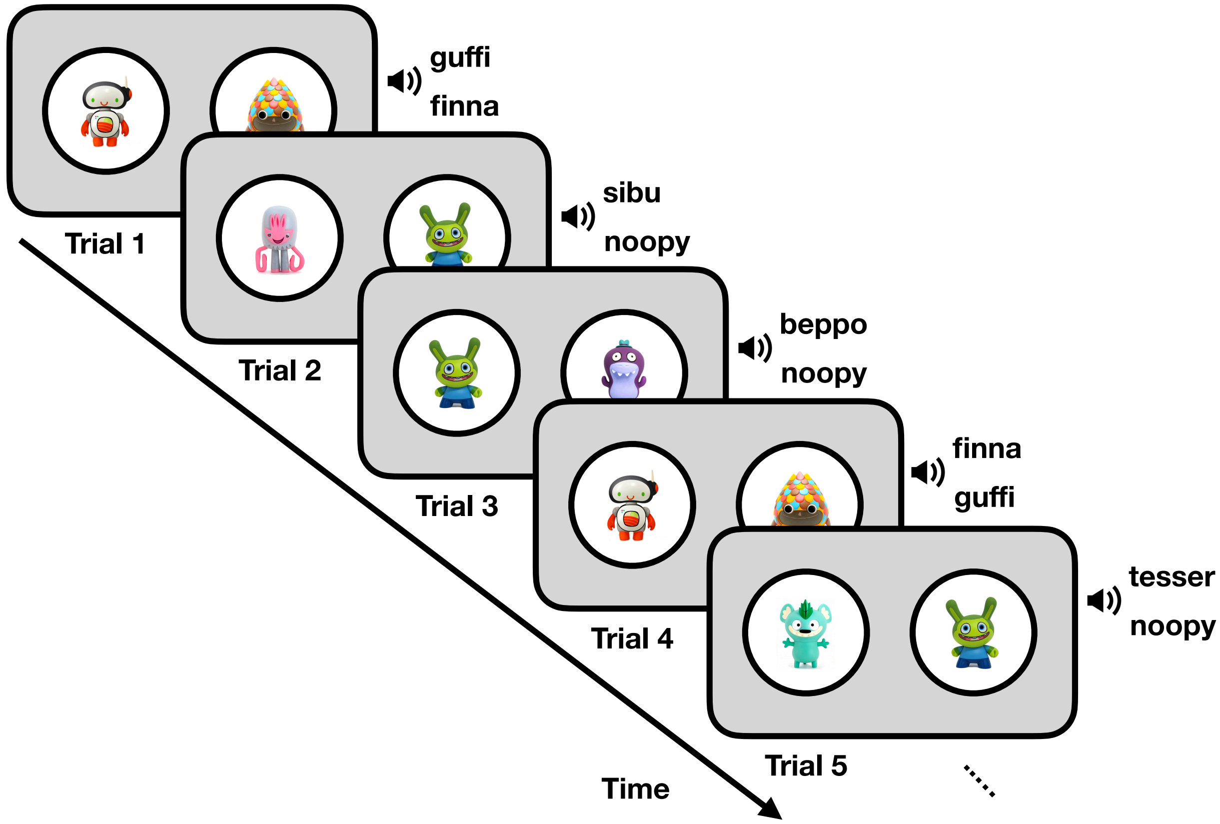

Figure 1. Overview over the training procedure.

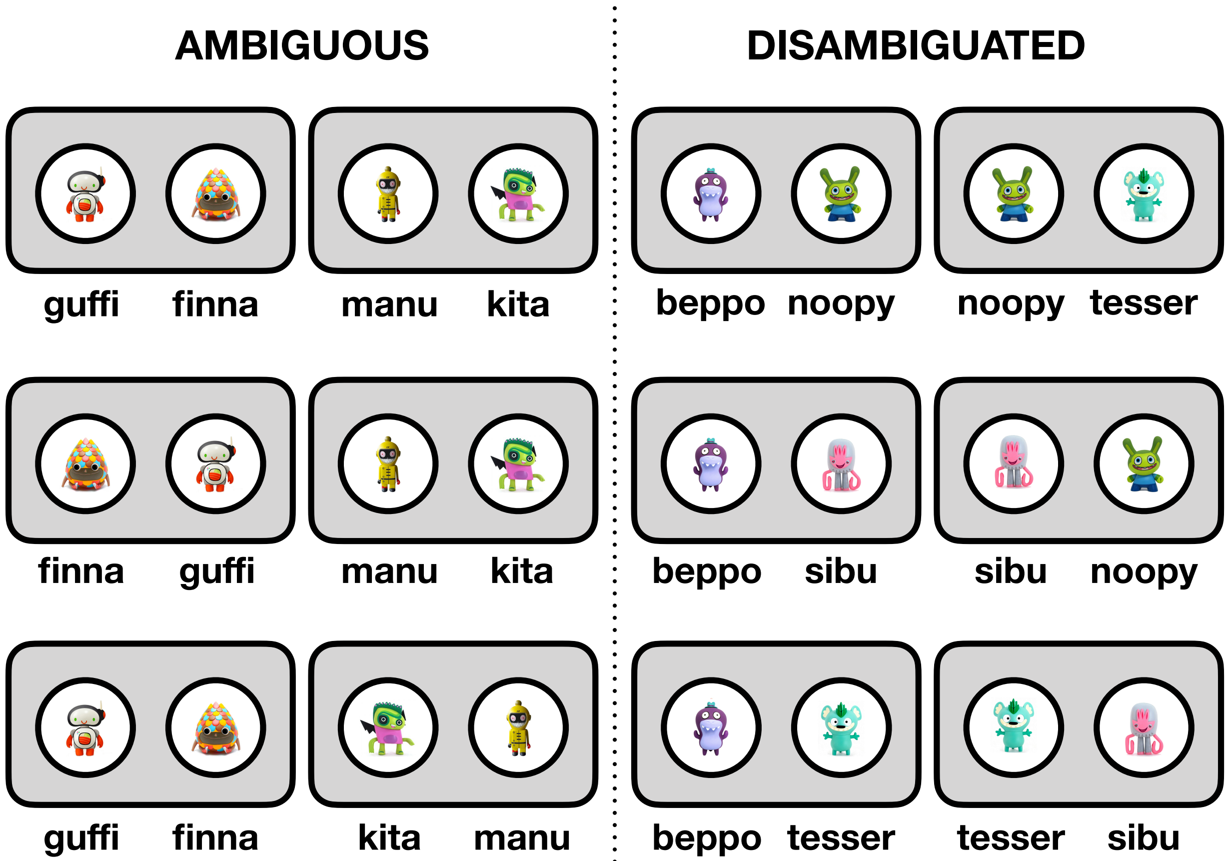

Figure 2. Overview over one block of the Training Phase.

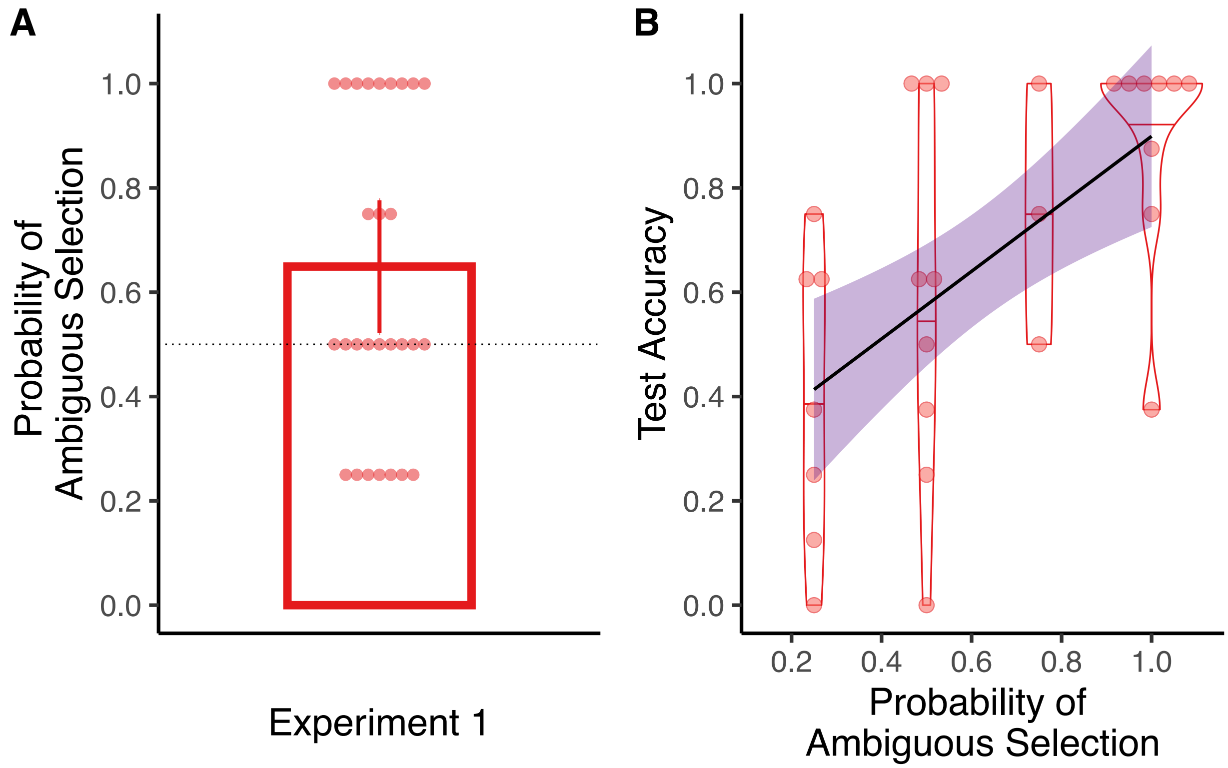

Figure 3. (A) Proportion of more ambiguous items selected in Experiment 1. Error bar represents the 95% CI of the model predictions. (B) Relationship between choosing more ambiguous items and test accuracy in Experiment. Dots and violin plots represent the distribution of individual participants' test accuracy. Error bands represent 95% CIs.

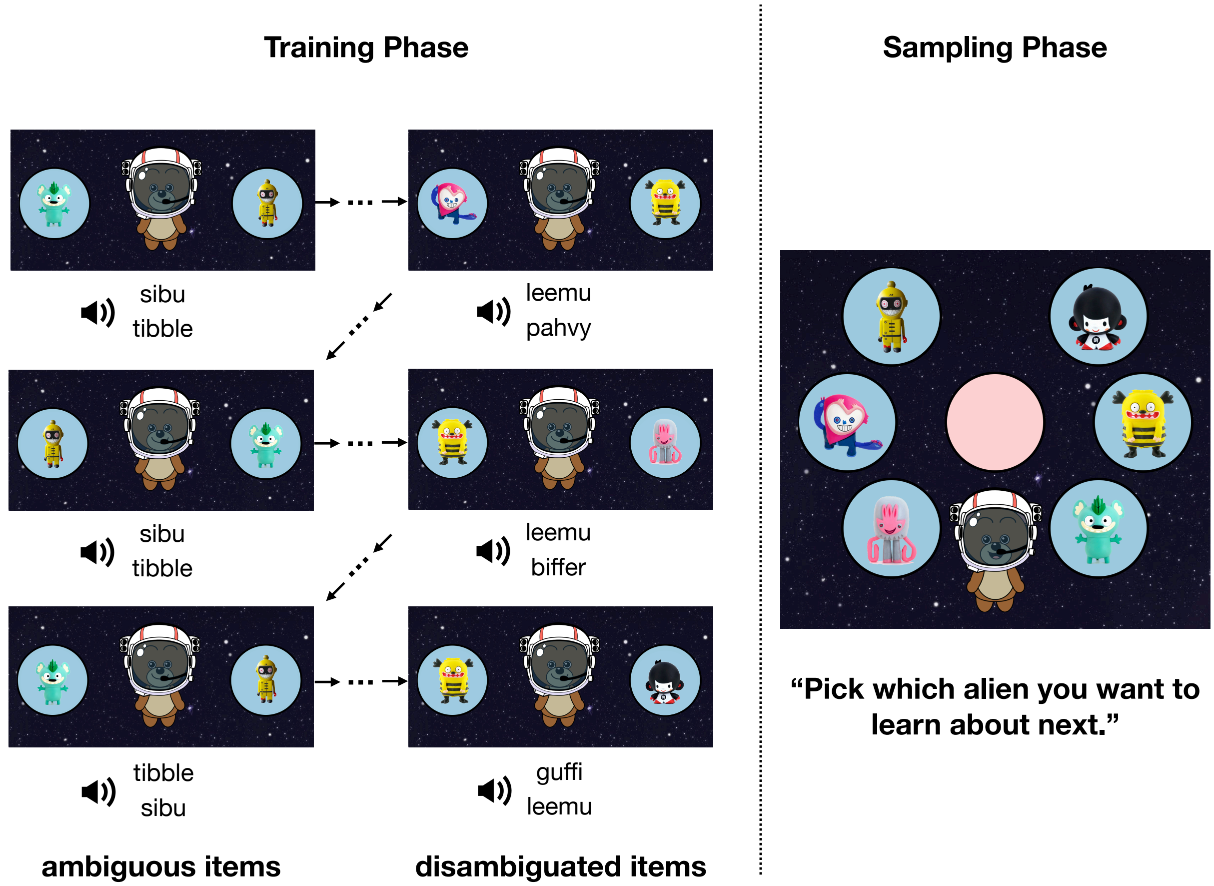

Figure 4. Overview over the design of the Training and Sampling Phase in Experiment 2.

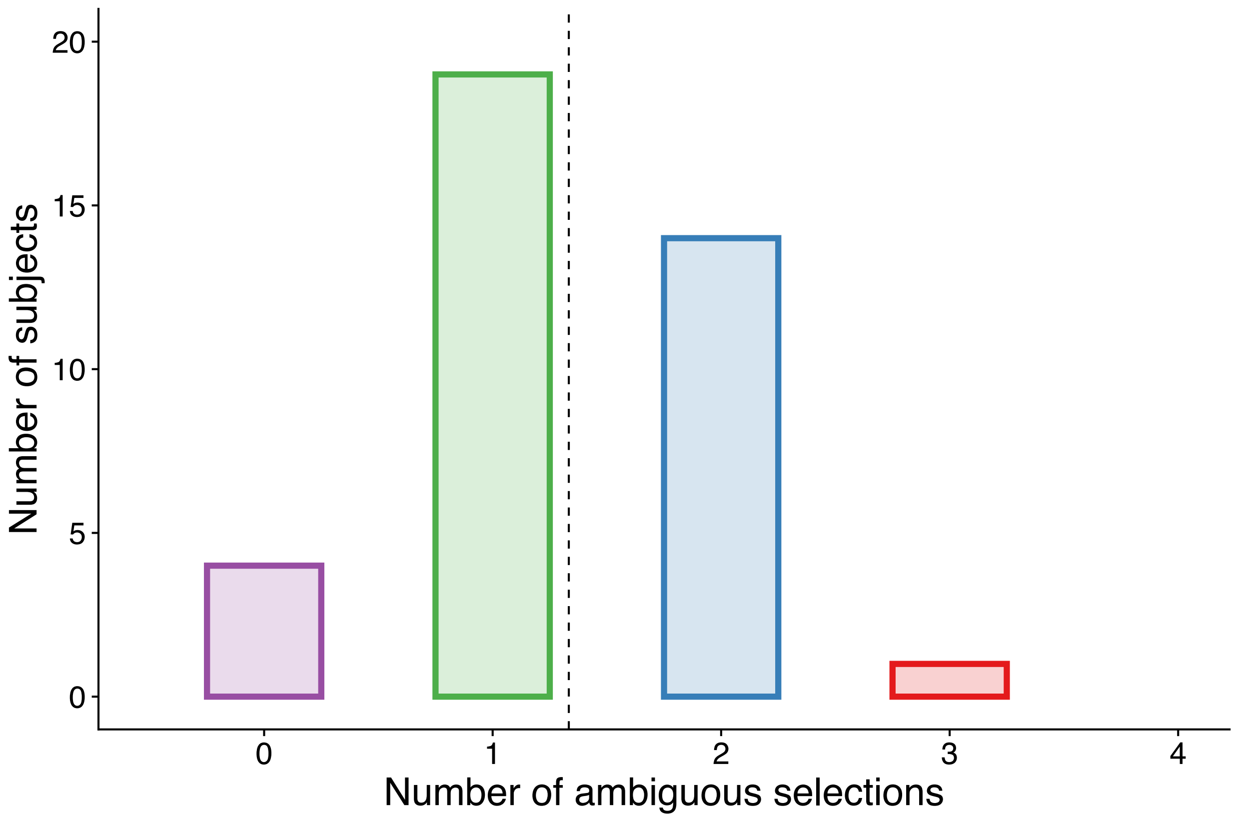

Figure 5. Children's sampling choices in Experiment 2. The plot depicts the number of subjects (out of 38) selecting 0, 1, 2, 3, or 4 ambiguous items across the four sampling trials. The dashed line represents the expected value if items are sampled randomly.

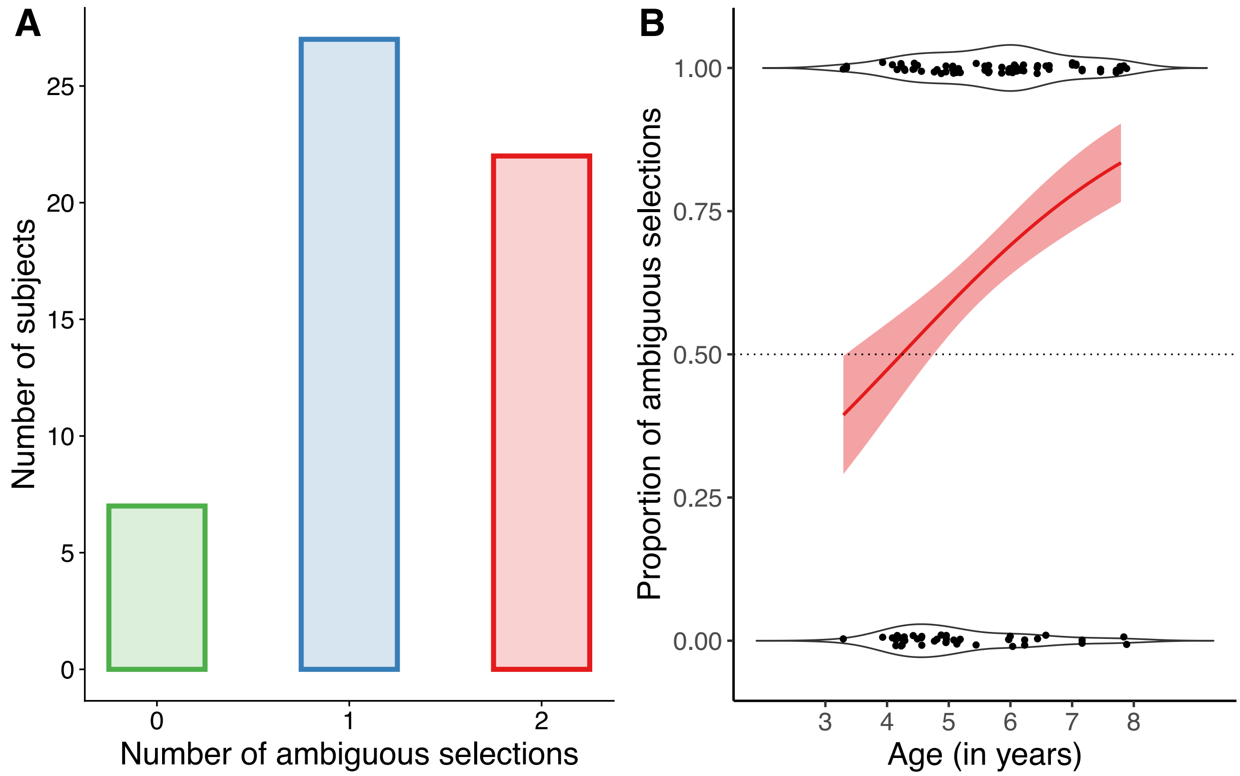

Figure 6. Distribution of ambiguous item selections in Experiment 3 overall (A) and across age (B). Error bands are +1/-1 SEs based on model estimates.

Supplementary Materials

S1: Replication of Experiment 1

We sought to replicate Experiment 1 using an in-lab task. As in Experiment 1, the central question was whether adults would make selections that reduced referential ambiguity. In addition to replicating the result from Experiment 1, we explored the degree to which adults' sampling decisions were sensitive in a graded fashion to the extent of ambiguity in their options. We manipulated the degree of ambiguity across two between-subjects conditions. One condition (the Fully Ambiguous condition) matched the design of Experiment 1, such that one set of items remained fully ambiguous throughout the training phase, while object-label associations were disambiguated for the remaining items. In the second condition (the Partially Ambiguous condition), one set of items was manipulated to remain moderately ambiguous across training, while the remaining items were disambiguated as in the Fully Ambiguous condition.

As in Experiment 1, we predicted that participants would be more likely to choose to learn about the ambiguous items than about the disambiguated items in the sampling phase. For the Partially Ambiguous condition, we expected participants to have a weaker preference for ambiguous items over the disambiguated items, since adults accurately tracking the co-occurrence evidence could successfully learn all word-referent pairs.

Method

Participants

62 University undergraduates at a large public university in the Midwestern United States (27 female; mean age: 19.1 years, SD = 1.01; 56 native speakers of English) participated for course credit and were randomly assigned to the Fully Ambiguous Condition (n = 28) or the Partially Ambiguous Condition (n = 34).

Stimuli

The stimuli were identical to Experiment 1. All stimuli and the experimental scripts are openly available on the Open Science Framework [https://osf.io/udmvh/].

Design & Procedure

The experiment consisted of a Training Phase, a Sampling Phase, and a Test Phase. In Experiment S1 (and unlike Experiment 1), participants completed the Test Phase twice, once after the Training Phase and once after the Sampling Phase (i.e., the order was Training-Test-Sampling-Test).

Training Phase. As in Experiment 1, participants viewed 24 cross-situational learning trials (2 blocks of 12 trials), presented in random order. The design and procedure were identical to Experiment 1, with the following exceptions.

Participants were assigned to one of two conditions: Fully Ambiguous or Partially Ambiguous. The Fully Ambiguous condition was identical to the design of Experiment 1, where half of the object-label pairs were never disambiguated (ambiguous items; Figure S1.1, top left) and half of the object-label pairs were disambiguated across training trials (disambiguated items; Figure S1.1, right panel). In the Partially Ambiguous condition, half of the object-label pairs were grouped such that two specific objects co-occurred on 4 out of their 6 occurrences, but each object occurred with a different object from the ambiguous object set on the remaining 2 trials (partially ambiguous items; Figure S1.1, bottom left). The other four objects were disambiguated as in the Fully Ambiguous condition. In both conditions, participants saw each individual object and label equally frequently.

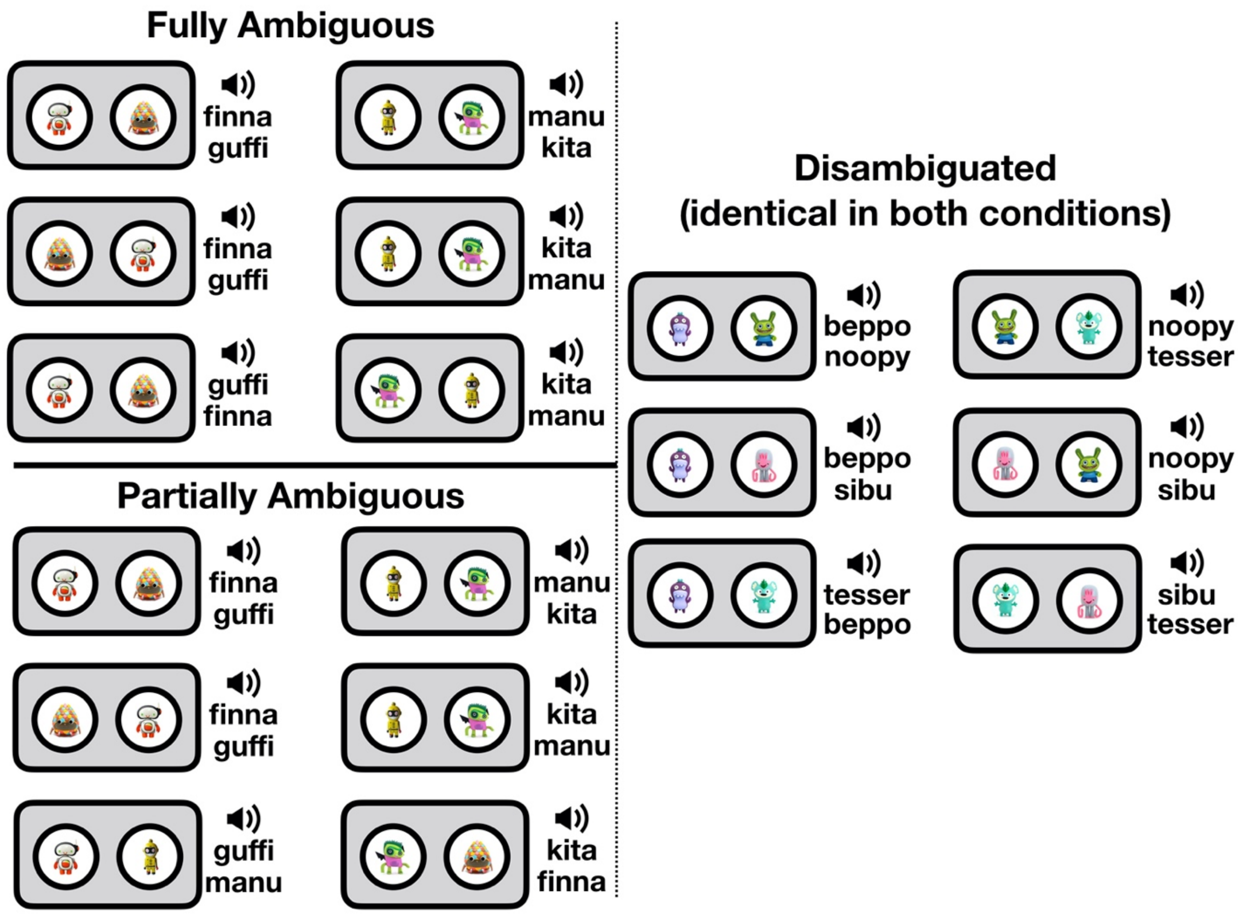

Figure S1.1. Overview over one block of the Training Phase for the Fully Ambiguous Condition and the Partially Ambiguous Condition

Sampling Phase. The procedure for the Sampling Phase was identical to the Sampling Phase in Experiment 1 (and identical in both conditions).

Test Phase. Knowledge of the object-label associations was tested in the same manner as in Experiment 1, with the exception that the 8-AFC test phase was presented twice: once immediately following the Training Phase (Test Block 1) and once immediately following the Sampling Phase (Test Block 2), as in Experiment 1. Our rationale in testing participants twice was to further assess the consequences of participants' sampling behavior for word learning. By testing object-label knowledge before and after the Sampling Phase, we hoped to investigate whether participants' choices led to gains in test accuracy. While adding a test phase immediately following the Training Phase provided participants with additional exposure to the labels and the objects prior to the Sampling Phase, we reasoned that the additional test phase (Test Block 1) should not help participants learn the particular object-label associations, since all eight objects were presented with each label (i.e., no cross-situational learning could occur during the test phase) and participants did not receive feedback on their test choices.

Results

Sampling choices

We fit the same logistic mixed-effects model as in Experiment 1, with the addition of a fixed effect for condition (centered; fully ambiguous = 0.5, partially ambiguous = -0.5). To test whether participants' likelihood of selecting an ambiguous item differed from chance within each condition, we refit the model while recoding the condition variable with each condition coded as zero. Participants were significantly more likely to select a (fully or partially) ambiguous item in the Fully Ambiguous condition (M = 64.3%, 95% CI = [53.6%, 75.0%]) than in the Partially Ambiguous condition (M = 47.8%, 95% CI = [39.1%, 56.5%]), $b = 0.70$, 95% Wald CI = [0.14, 1.25], $z = 2.47$, $p = .01$ (Figure S1.2). Participants in the Fully Ambiguous condition selected ambiguous items more frequently than disambiguated items ($b = 0.60$, 95% CI = [0.19, 1.02], $z = 2.85$, $p = .004$), replicating the sampling preference from Experiment 1. Participants in the Partially Ambiguous condition did not select (partially) ambiguous items more frequently than disambiguated items ($b = -0.09$, 95% CI = [-0.45, 0.27], $z = -0.50$, $p = .62$).

We conducted non-parametric statistical analyses to test the robustness of the results in the Sampling Phase. A Wilcoxon rank-sum test comparing subjects' proportion of ambiguous selections between the Fully Ambiguous condition and the Partially Ambiguous condition was significant, $W = 632$, $p = .02$. Participants' average proportion of ambiguous selections significantly diverged from chance in the Fully Ambiguous condition ($V = 140$, $p = .015$), but not in the Partially Ambiguous condition ($V = 101$, $p = .59$).

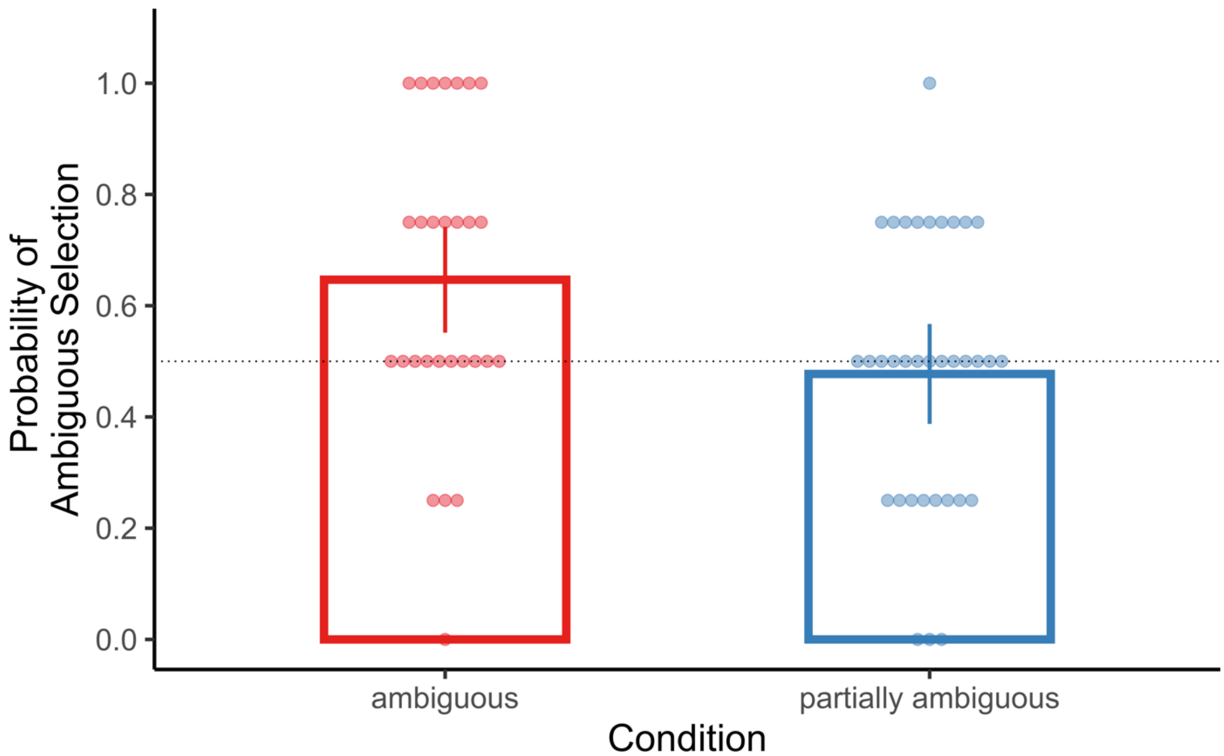

Figure S1.2. Proportion of more ambiguous items selected in each condition. Error bars in represent the 95% CI of the model predictions.

Test performance

Overall test accuracy. Participants demonstrated successful word learning in the 8-AFC recognition test that immediately followed the Training Phase (Fully Ambiguous condition: M = 65.2%, 95% CI = [61.2%, 69.2%]; Partially Ambiguous condition: M = 76.8%, 95% CI = [72.5%, 81.2%]) and in the (identical) test phase following the Sampling Phase (Fully Ambiguous condition: M = 72.8%, 95% CI = [68.8%, 76.8%]; Partially Ambiguous condition: M = 77.6%, 95% CI = [73.2%, 81.9%]).

Effects of Test Half and Item Type. To investigate the relationship between Test Half, Item Type (ambiguous vs. disambiguated), and Condition, we fit a logistic mixed-effects model predicting trial-by-trial accuracy from the three-way interaction between these three predictors (centered), including all lower-order effects. We included by-participant and by-item random intercepts, and by-participant random slopes for Test Block, Item Type, and their interaction. There was a significant effect of Item Type (b = -1.33, Wald 95% CI = [-2.10, -0.56], z = -3.37, p < .001), indicating that participants generally performed better on items disambiguated during training across conditions, and a significant effect of Test Block (b = 0.82, Wald 95% CI = [0.11, 1.53], z = 2.27, p = .02), indicating that participants performed better on the second test block than the first. This effect of Test Block appeared to be driven mainly by a significant increase in accuracy for ambiguous items in the Fully Ambiguous condition from Test Block 1 M = 46.4%, 95% CI = [37.8%, 55.1%]) to Test Block 2 (M = 59.8%, 95% CI = [51.6%, 68.1%]), b = 1.04, Wald 95% CI = [0.18, 1.89], z = 2.38, p = .017 (see Figure S1.3A). There was also a significant interaction between Item Type and Condition, indicating that the difference in accuracy between disambiguated and ambiguous items was greater in the Fully Ambiguous condition than in the Partially Ambiguous condition (averaging across test block), b = -2.75, Wald 95% CI = [-3.87, -1.64], z = -4.83, p < .001 (Figure S1.3). All other effects, including the three-way interaction, were non-significant, all ps > .25.

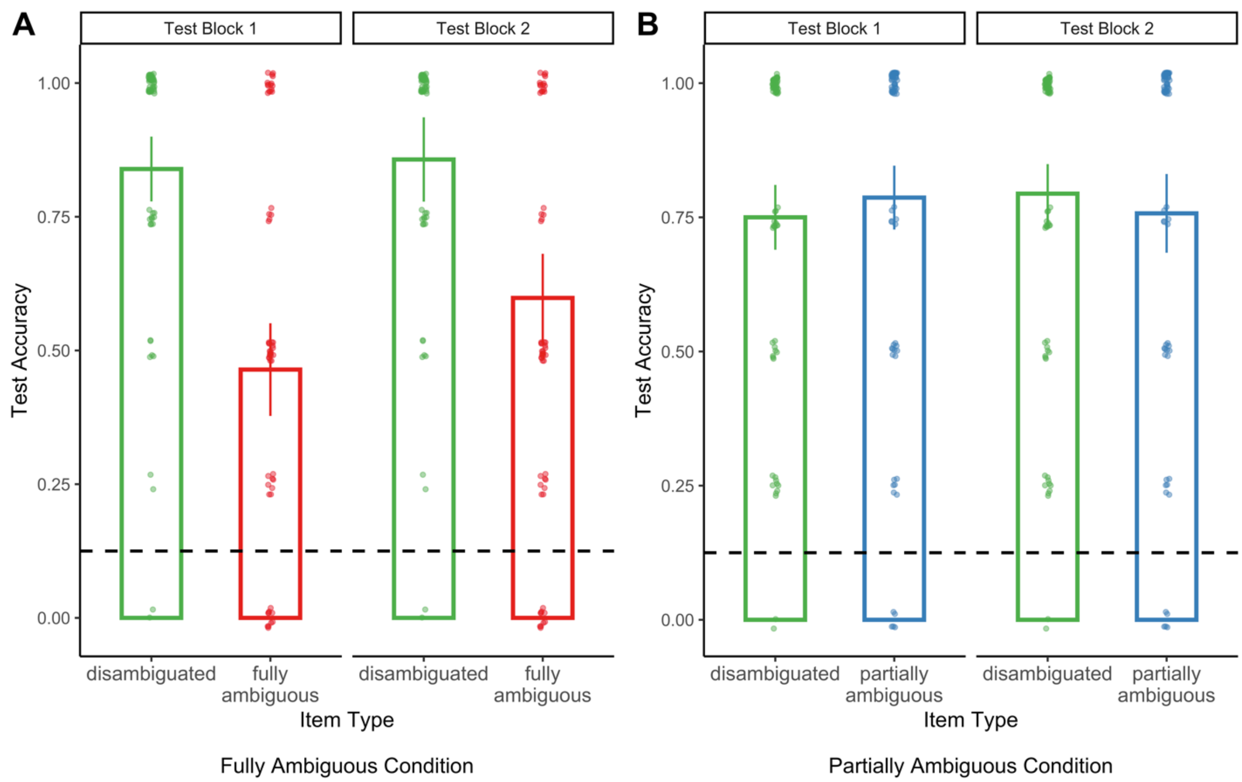

Figure S1.3. Test accuracy by item for (A) the Fully Ambiguous condition and (B) the Partially Ambiguous condition in Experiment S1. The dashed line represents chance-level performance. Error bars represent within-participant 95% CIs (Morey, 2008).

Relationship between sampling and test accuracy. In the Fully Ambiguous condition, participants' proportion of ambiguous selections was correlated with their test accuracy before the Sampling Phase (Test Block 1), $r(26) = .45$, 95% CI = [.10, .71], $p = .015$. The proportion of ambiguous selections was marginally correlated with participants' test accuracy after the Sampling Phase (Test Block 2), $r(26) = .35$, 95% CI = [-.03, .64], $p = .07$ (Figure S1.4, panel A). In the Partially Ambiguous condition, participants' proportion of (partially) ambiguous items selected was not significantly correlated with their test accuracy before (Test Block 1: $r(32) = .10$, 95% CI = [-.25, .42], $p = .58$) or after the Sampling Phase (Test Block 2: $r(32) = -.12$, 95% CI = [-.44, .23], $p = .50$; Figure S1.4, panel B).

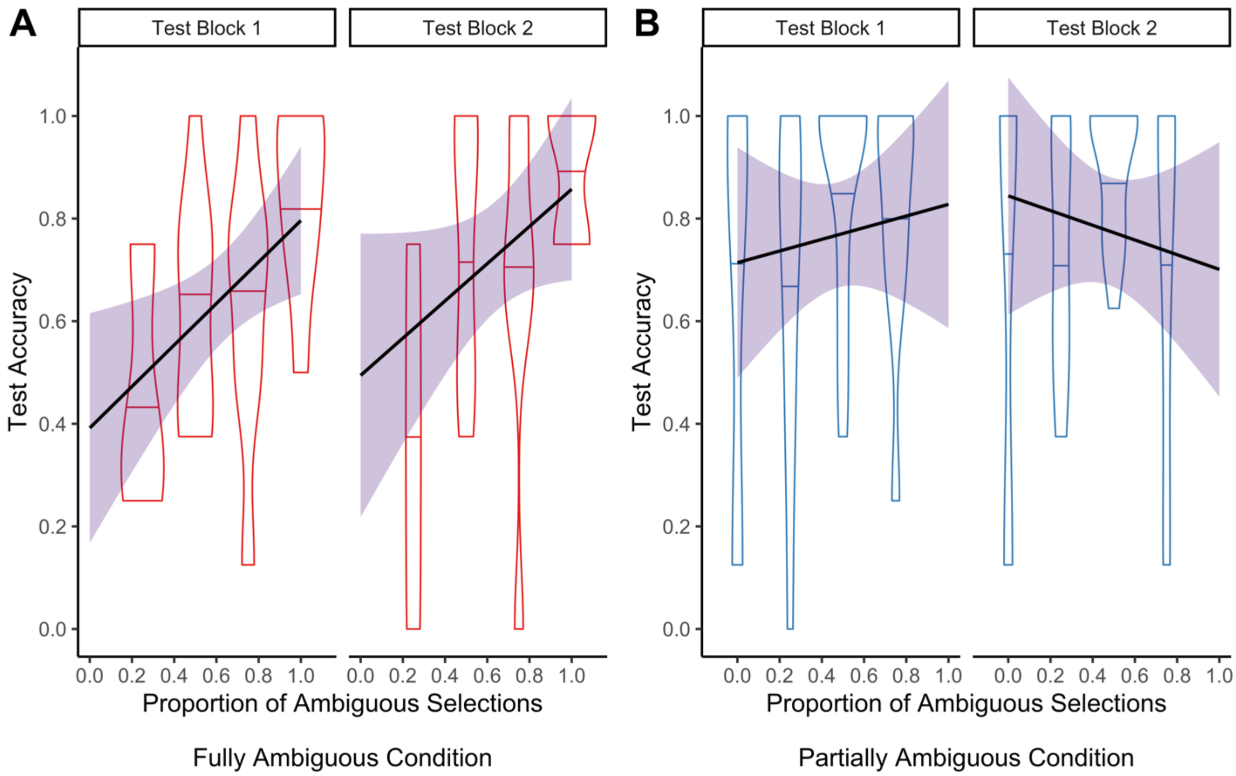

Figure S1.4. Relationship between choosing more ambiguous items and test accuracy for (A) the Fully Ambiguous condition and (B) the Partially Ambiguous condition in Experiment S1, split by Test Block. The distribution of individual participants' test accuracy for different proportions of ambiguous choices is represented with violin plots. Error bands represent 95% CIs.

By including a test phase immediately preceding the Sampling Phase, we aimed to further understand the correlation between test accuracy and preference for selecting ambiguous items observed in Experiment 1. Specifically, do participants have higher test accuracy at the conclusion of the experiment because they preferentially selected ambiguous items, or do participants who are more successful at learning the object-label associations show a stronger preference for selecting ambiguous items?

To address this question, we correlated participants' proportion of ambiguous selections with their increase in accuracy from Test Block 1 to Test Block 2. If participants' selections of ambiguous items are driving higher accuracy, then participants who show a preference for sampling ambiguous items should show the largest increases in accuracy from Test Block 1 to

Test Block 2. However, proportion of ambiguous items selected was not significantly correlated with an increase in test accuracy in the Fully Ambiguous condition ($r(26) = -.08$, 95% CI = [-.44, .30], $p = .69$) and was negatively correlated in the Partially Ambiguous condition ($r(32) = -.36$, 95% CI = [-.62, -.03], $p = .04$), i.e. participants who were more likely to select the (partially) ambiguous items showed a lesser increase in test accuracy. Similar relationships were found with test accuracy for both ambiguous and disambiguated items.

Discussion

In a cross-situational learning task, adult learners chose to learn more about those object-label pairs that remained ambiguous throughout training. Adults showed this tendency when the object-label pairings were truly ambiguous based on the training evidence (Fully Ambiguous condition), but not when the object-label pairs became disambiguated at any point during training (Partially Ambiguous condition). This experiment replicates the result from Experiment 1 (in the Fully Ambiguous condition) and provides further evidence that adult learners will seek to reduce ambiguity about object-label associations, primarily when word referents remain entirely ambiguous from past experience.

At test, participants tended to show poorer overall learning of the ambiguous object-label pairs in the Fully Ambiguous condition, though not in the Partially Ambiguous condition. This finding is consistent with the fact that the fully ambiguous items were (by definition) more difficult object-label mappings to learn (though note that we only found a marginal difference in accuracy between item types in Experiment 1). Intriguingly, we also found that participants' sampling behavior was correlated with their test accuracy in the Fully Ambiguous condition: participants who chose more ambiguous items during the Sampling Phase also more accurately identified object-label associations. By testing participants before and after the Sampling Phase in Experiment S1, we were able to partially disentangle the directionality of this effect. Participants who had learned the novel words better following the Training Phase were more likely to sample ambiguous items. In other words, participants' learning success appeared to predict their likelihood of targeting ambiguous items during the Sampling Phase, rather than participants' sampling choices (solely) driving their test accuracy. Perhaps surprisingly, participants' likelihood of selecting ambiguous items did not appear to lead to larger increases in accuracy from the first to the second testing (though accuracy for ambiguous items showed the largest increase in general).

There are a few potential reasons why we did not observe greater increases in accuracy following ambiguous item selections in the present study. First, participants who showed a strong preference for selecting ambiguous items were already performing quite well on the test, suggesting that the lack of an effect may in part be due to a ceiling effect. Second, the Sampling Phase was designed such that participants continued to receive ambiguous, cross-situational learning trials: even after selecting ambiguous items, participants would view a cross-situational learning trial involving an additional, randomly selected item. While these trials in principle provided the opportunity to disambiguate object-label mappings, they also required participants to simultaneously continue to update and maintain the object-label relationships from past training, which may in turn have "washed out" some of the learning benefits that participants might have accrued from targeting the ambiguous items during the Sampling Phase. Future research could specifically target what types of sampling behavior leads to better learning outcomes, relative to randomly selected training events.

S2: Experiment 2 Practice Phase

At the beginning of the experiment, participants completed a practice phase in which they encountered the two familiar word object stimuli and two novel object-label associations. We introduced this short practice phase to give children experience with the overall structure of the main experiment under less demanding circumstances, using a smaller set of items and mixing familiar and novel items. First, children were exposed to 4 training practice trials similar in structure to the training trials in the main experiment. On each trial, two referents appeared on the screen on either side of the Teddy character and children heard two labels, one for each object, in random order. On the first trial, children always saw the two familiar items (i.e., the penguin and the dog), followed by a second trial in which children saw two novel object-label associations (i.e., an ambiguous labeling event). On the final two training practice trials, children saw each of the familiar items occur with one of the two novel items (permitting the disambiguation of the novel object-label associations). Next, children saw two sampling practice trials, in which children had the opportunity to select which of the four items they wanted to learn about next, followed by four practice test trials, in which participants' knowledge of the items was tested in a 4-AFC recognition test. The procedure for each of these practice trial types mirrored the procedure for the Sampling phase and the Test phase described in the methods section of Experiment 2.

S3: Experiment 2 Supplementary Figures

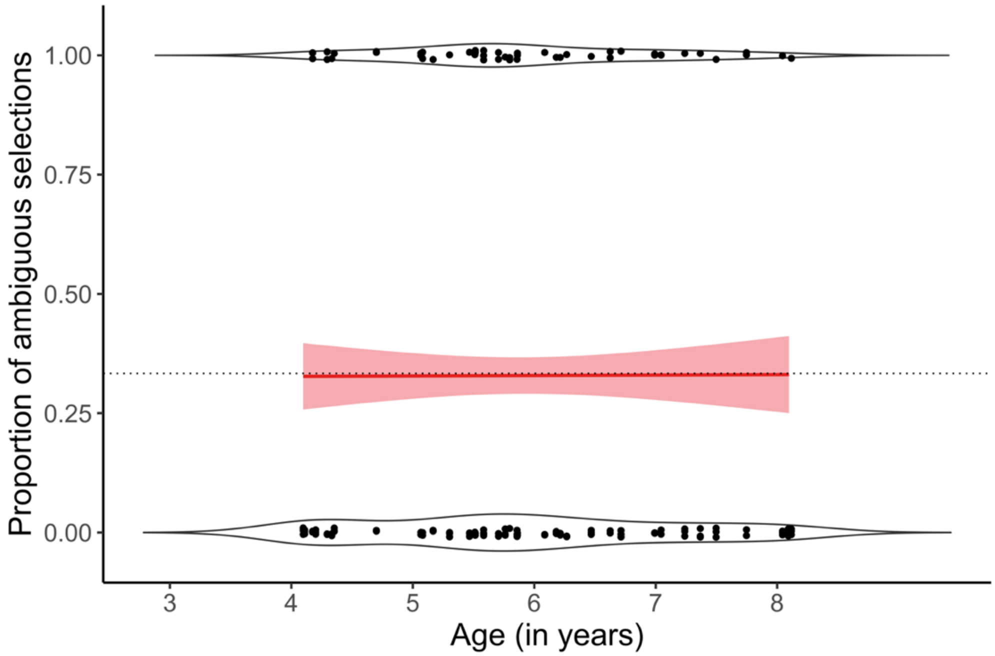

Figure S3.1. Distribution of ambiguous item selections in Experiment 2 across age. Error bands are +1/-1 SEs based on model estimates.

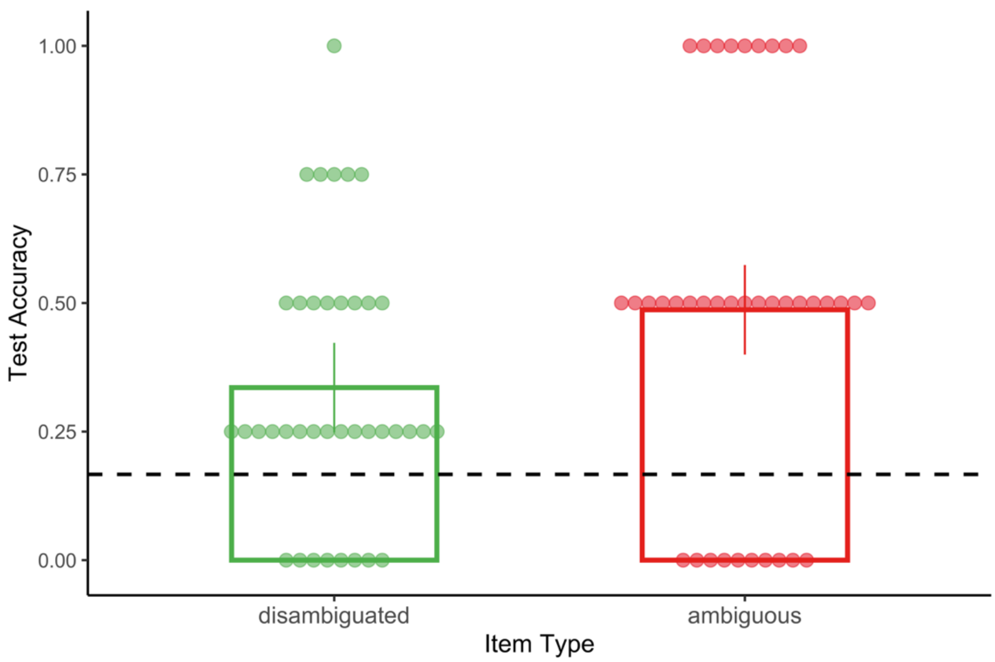

Figure S3.2. Children's test accuracy in Experiment 2, split by Item Type (disambiguated vs. ambiguous). Error bars represent within-participant 95% CIs (Morey, 2008).

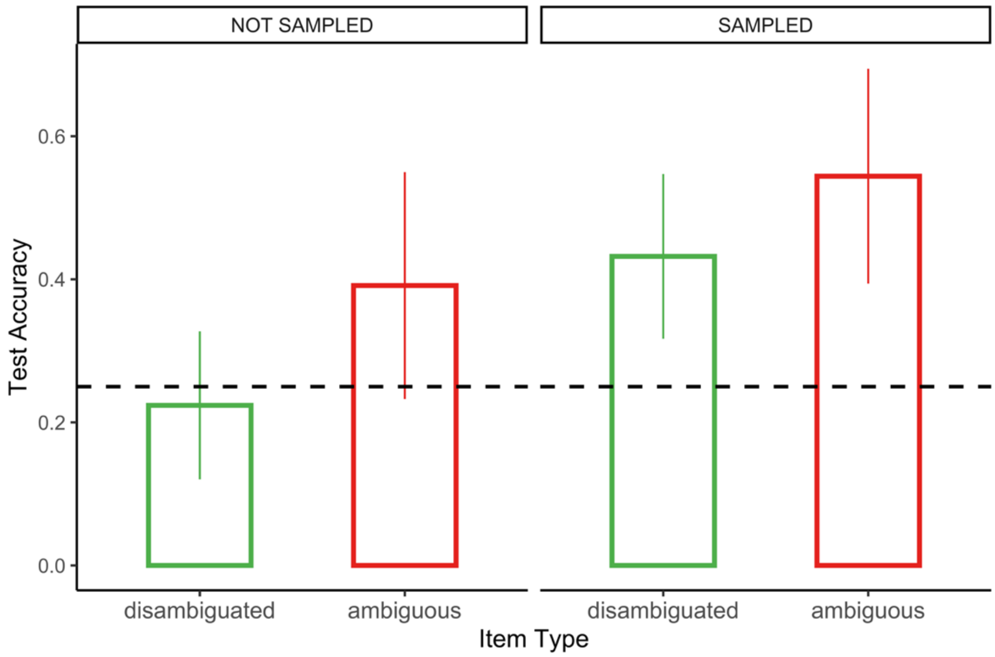

Figure S3.3. Children's test accuracy in Experiment 2, split by Item Type (disambiguated vs. ambiguous) and whether a given test item was selected during the Sampling Phase (not sampled vs. sampled). Error bars represent within-participant 95% CIs (Morey, 2008).

S4: Experiment 3 Supplementary Results

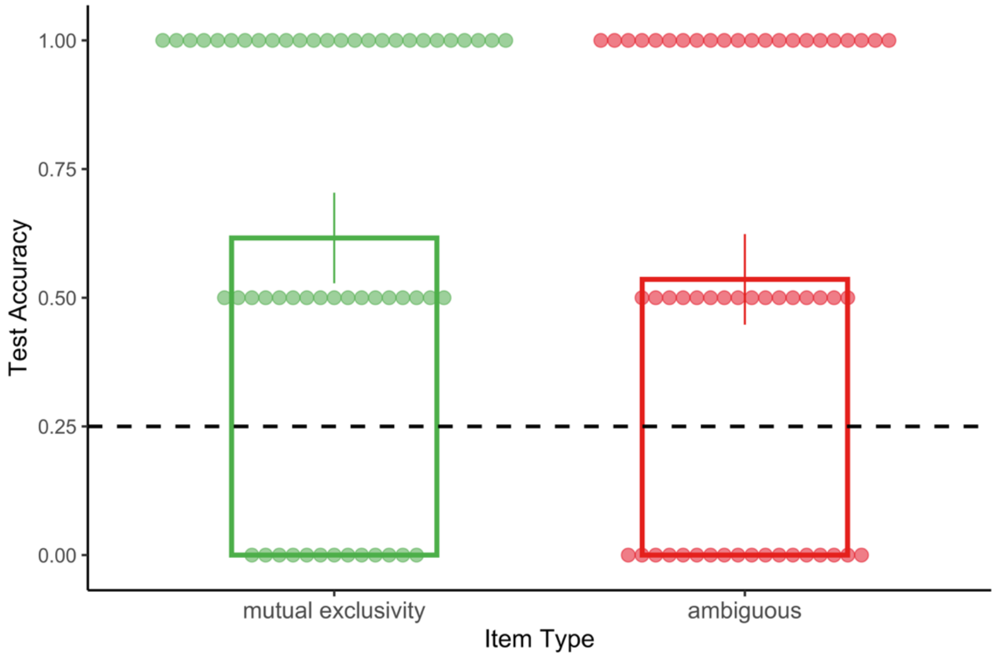

Figure S4.1. Children's test accuracy in Experiment 3, split by Item Type (disambiguated/ mutual exclusivity vs. ambiguous). Error bars represent within-participant 95% CIs (Morey, 2008).

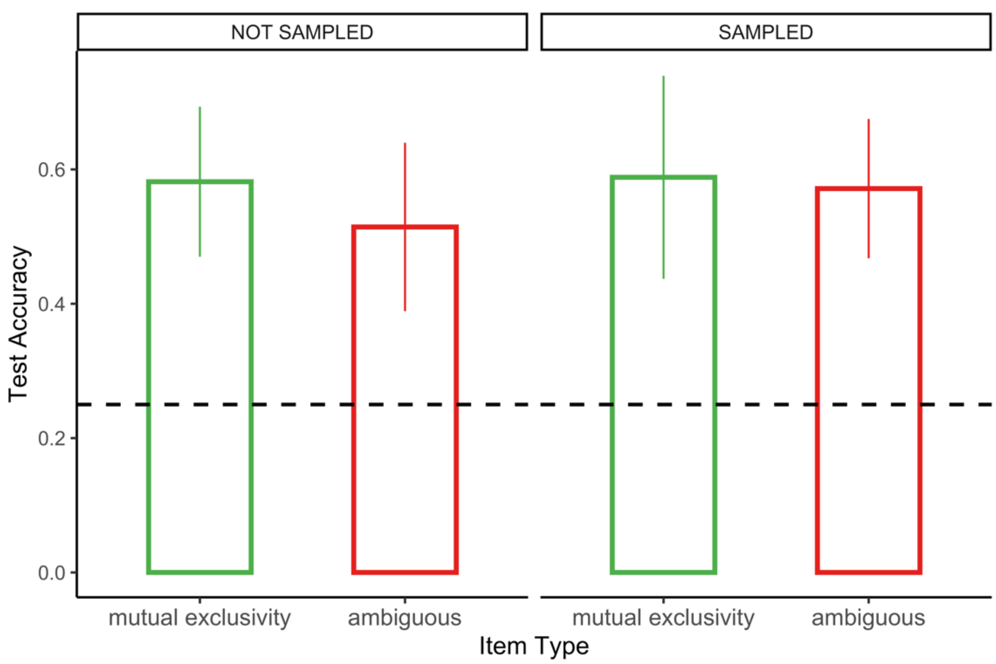

Figure S4.2. Children's test accuracy in Experiment 3, split by Item Type (mutual exclusivity vs. ambiguous) and whether a given test item was selected during the Sampling Phase (not sampled vs. sampled). Error bars represent within-participant 95% CIs (Morey, 2008).