Active information-seeking in support of learning extensions of novel words

Martin Zettersten (martincz@princeton.edu), Molly Cutler, & Casey Lew-Williams
Princeton University, Department of Psychology, Peretsman Scully Hall
Princeton, NJ 08540 USA

Abstract

A key debate in language learning centers on how people successfully learn the extension of a novel word, despite inherent ambiguity in the input. Across two studies, we tested whether learners reduce ambiguity about a word's extension by actively sampling the environment. Adult participants were first shown ambiguous learning situations in which novel words were presented with a set of exemplars that were drawn from a subordinate-level category (e.g., Dalmatians), a basic-level category (e.g., dogs), or a superordinate-level category (e.g., animals). Learners then had the opportunity to sample the label of additional exemplars. Participants systematically adapted their sampling choices as a function of training. Moreover, participants varied in their sampling strategies, pursuing both confirmatory strategies (selecting exemplars similar to the training set) and constraining strategies (selecting exemplars that constrain the word's extension). Overall, these findings show that learners spontaneously pursue sampling strategies that support generalizing novel word meanings.

Keywords: active learning; information-seeking; categories; generalization; word learning; word extension; word meaning

Introduction

When learners first encounter a word (like "dog") together with a novel referent (e.g., a Dalmatian), it is often ambiguous what category the meaning of the new word may generalize to (Quine, 1960; Xu & Tenenbaum, 2007b). For example, "dog" might refer to Dalmatians (a subordinate category level), dogs in general (the basic level), all animals (a superordinate category level), or a general category like "things that move" (a more general hypernym). How do learners determine the extension of a novel word?

One potentially powerful solution is that learners actively seek new information to clarify the boundaries of a word's meaning (Gottlieb et al., 2013; Gureckis & Markant, 2012). For example, they could sample new referents at different category levels, such as testing whether the new word also applies to Golden Retrievers, rabbits, or even more generally to any moving object. Past evidence both with children and adults suggests that learners will actively sample information that can aid in reducing uncertainty about novel words and categories (Kachergis et al., 2013; Markant & Gureckis, 2014; Zettersten & Saffran, 2021). However, these results have often focused on mapping one-to-one object-label associations. When learners are faced with a novel word, they are often confronted with more thorny challenges (Quine, 1960). Categories have hierarchical structure at multiple levels of abstraction, requiring learners to determine not only how the word applies to the current referent, but which of many possible classes of referents the word may extend to. How do learners seek information when learning words for complex, hierarchically structured categories?

In the present studies, we investigate how learners sample information about novel word meanings when they must disambiguate the extension of novel words across multiple possible category levels. Based on an experimental paradigm testing learners' inferences about word meanings (Xu & Tenenbaum, 2007b), we manipulate the initial sample of exemplars presented as referents of a novel word. Past work has shown that participants will systematically generalize word meanings differently depending on the composition of an initial sample of exemplars: for example, if a label is initially experienced together with three exemplars at the subordinate level (e.g., three Dalmatians), participants will tend to generalize the word meaning more narrowly than if the initial sample is composed of a broader set of exemplars (e.g., three dogs of differing breeds).

The theoretical interpretation and implications of these findings have engendered wide debate: Xu and Tenenbaum attribute the phenomenon to participants' sensitivity to how learning experiences are sampled (so-called "strong sampling"), while others have argued for alternative mechanisms grounded in general cognitive processes such as attention and memory (Spencer et al., 2011; Jenkins et al., 2021), pragmatic considerations (Lewis & Frank, 2018), and processes of comparing and contrasting exemplars (Wang & Trueswell, 2022). Though theoretical accounts of the underlying mechanisms differ, all accounts agree that learners' representations of a novel word's extension are robustly modulated by the training manipulation introduced by Xu and Tenenbaum. In the current study, we capitalize on this training paradigm to investigate whether and how participants seek additional information about a novel word's extension when presented with ambiguous input.

The present experiments examined (1) adults' sampling strategies, and (2) the consequences of their sampling strategies for learning the extension of novel words. We predict that participants will tune their information-seeking to their past learning experience: depending on their initial training condition, participants will seek information at different category levels. However, there are multiple possible strategies that participants may use as they tune their information-seeking. The main goal in the experiment is to conduct an initial exploration of these specific strategies.

We expected that learners would sample new information about ambiguous word meanings using one of two different strategies. One possibility is that participants pursue a confirmatory sampling strategy. For example, if presented with three referents at the subordinate level (e.g., three Dalmatians), they will prefer to sample another referent at the subordinate level (i.e., select another Dalmatian) to confirm their hypothesis that the word refers to the subordinate level. This would be consistent with evidence for a "positive test"

strategy documented across a number of other learning tasks (Austerweil & Griffiths et al., 2011; Coenen et al., 2015; Klayman & Ha, 1987, 1989; Navarro & Perfors, 2011; Wason, 1960). A second possibility is that participants attempt to constrain word meanings by sampling a potential referent "one level up" from the category implied by the training experience. For example, if presented with three referents at the subordinate level (e.g., three Dalmatians), they will prefer to sample a referent at the basic level (i.e., select a dog that is not a Dalmatian) to constrain the possible category level of the word.

Experiment 1

We investigated how learners sample new information when learning words that are ambiguous with respect to how broadly they generalize to different category levels. Participants completed a word learning task closely modeled on past studies of word learning at different levels of categorical abstraction (Xu & Tenenbaum, 2007b; Lewis & Frank, 2018), in which participants learn a new word that is associated with a set of exemplars that belong to the same subordinate, basic, or superordinate category. Learners then had the opportunity to sample a novel exemplar before being asked to generalize the word to a new set of exemplars. We predicted that participants would flexibly shift their sampling behavior in response to the categorical level of the training exemplars. The main aim of the experiment was to explore the types of information-seeking strategies people pursue in ambiguous word learning situations and to probe the consequences of different information-seeking strategies for how people subsequently generalize word meanings.

Method

The experiment was preregistered (https://osf.io/sk8zw). All materials, data, and analyses are openly available on OSF: https://osf.io/p38g9.

Participants

We recruited 200 participants (75 female, 123 male, 1 non-binary; mean age: 41.4 years, *SD* = 11.6) from Amazon Mechanical Turk using Cloud Research tools for improving data quality (Litman et al., 2017). 14 additional participants were excluded based on three preregistered exclusion criteria: (a) failing to correctly enter the novel label they saw during training on at least two of three trials (n=8), (b) always choosing the same location on all sampling or all test trials (n=1), and (c) entering nonsensical text on open response questions (n=5). Participants were paid $0.80 for completing the study. Participants were randomly assigned to three counterbalanced, within-subjects conditions: Training Condition, Correct Label Level, and Category Type (see Design & Procedure for details).

Stimuli

The image stimuli were three sets of 15 images of exemplars from three overarching categories (animals, vegetables, and vehicles). These stimuli were taken from a past study investigating how learners generalize word meanings (Lewis & Frank, 2018) that replicated the task introduced by Xu and Tenenbaum (2007b). Within each category's image set, 7 images were used during the Training Phase (see Figure 1): 3 subordinate-level exemplars (e.g., images of Dalmatians), 2 basic-level exemplars (e.g., images of dogs other than Dalmatians), and 2 superordinate-level exemplars (e.g., images of animals other than dogs). The remaining 8 images were used during the Test Phase: 2 subordinate, 2 basic, and 4 superordinate exemplars. Of the test images, 3 images were selected to appear as options in the Sampling Phase: one subordinate, basic, and superordinate exemplar each (see Figure 1). The linguistic stimuli for the experiment were six nonce words (*sibu*, *kita*, *beppo*, *tibble*, *roozer*, *guffy*). The first three words were introduced during the Training Phase as labeling the three training exemplars. The final three words were presented as the name for any image that a participant chose during the Sampling Phase that did not belong to the correct category, as determined by the condition design. Words were randomly assigned to the three training conditions for each participant.

Design & Procedure

Participants were asked to imagine they were learning a new language and that they were being told new words for the first time. Their job was to figure out what these new words mean. Participants then completed three trials, each consisting of a Training Phase, a Sampling Phase, and a Test Phase.

Training Phase. On each trial, participants were presented with three exemplars, together with a prompt labeling the images (e.g., "These are three sibus"). The nonce label also appeared as text under each individual image. The Training Condition varied whether participants saw three exemplars on the subordinate level (Narrow condition; e.g., three Dalmatians), three exemplars on the basic level (Intermediate condition; e.g., three dogs of different breeds), or three exemplars on the superordinate level (Broad condition; e.g., three animals from different basic-level categories). Training Condition varied within-participants, such that each participant was presented with each Training Condition once. The order of the training conditions was randomized, as well as the particular nonce word used to label the exemplars for each Training Condition. Each training trial involved exemplars from one of three different general categories (animals, vehicles, vegetables; Category Type). Each participant was presented with one trial involving each Category Type. The assignment of Category Type to Training Condition was counterbalanced across participants.

After the training trial, we included a brief attention check that asked participants to enter the label they were just taught. Neither the label prompts nor the images were visible during the attention check. Participants were accurate in recalling the label (M = 97.2%; i.e., misspellings or additions that involved a single character were permissible). Trials on which participants failed the attention check were excluded in subsequent analyses (n = 17) and participants who failed

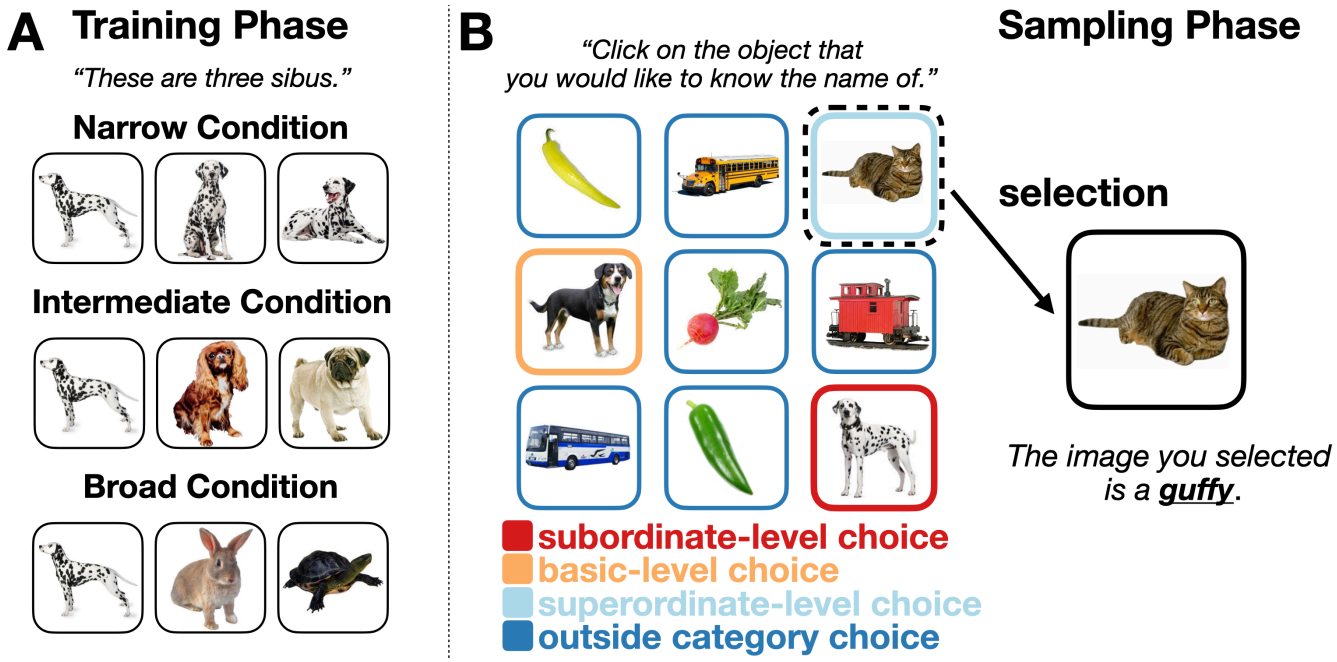

Figure 1: (A) The Training Phase (animal exemplars) and (B) the Sampling Phase. Colors indicate different choice options (subordinate-level, basic-level, superordinate-level and outside category choices) available to participants. Colored frames are used in the figure only to illustrate the trial design and choice options were not highlighted in the experiment.

more than one attention check were excluded altogether (see Participants section).

Sampling Phase. Participants subsequently entered the Sampling Phase, in which they had the opportunity to select one novel image and learn its label. On each trial, participants were presented with 9 novel images: one for each category level (subordinate, basic, superordinate) for each Category Type (see Stimuli section). Participants were prompted to choose which object they would like to learn the name of and were advised that they could only make a single selection.

For each sampling trial, we specified a "ground truth" for the meaning of each word, in order to supply participants with feedback on their sampling selections. By varying the meaning of each novel word within Training Condition, we also ensured that participants could discover new information about a word's meaning through their active selections. The "ground truth" category level for each training condition was randomly selected from among three possibilities for each training condition (e.g., in the subordinate training condition where participants see three subordinate exemplars, the "true" meaning of the label may refer to (a) the subordinate level, (b) the basic level, or (c) the superordinate level). The combination of correct category level with training condition was counterbalanced across participants.

Test Phase. After the Sampling Phase, we tested how participants generalized the word meaning to novel exemplars. Participants were presented with 24 images (2 subordinate, 2 basic, and 4 superordinate for each Category Type) and instructed to select all of the other instances of the novel word from among the test images. Note that the labeling of the items as subordinate (i.e., Dalmatians for the animal Category Type), basic (i.e., other dog images), and superordinate (i.e., other animals) is to allow for comparison of selections across the three training conditions and does not necessarily reflect how participants perceived items at test. For example, in the case of the Category Type animals, we term the Dalmatian images as the subordinate items to compare selections across training conditions, but from the perspective of participants in the Intermediate Condition, all dog items represent exemplars of different subordinate dog categories. Test trials were untimed, such that participants were free to take as much time as needed to make their selections from the test array.

Results

Sampling Choices

Learners flexibly shifted their sampling choices depending on the training condition. To test whether training condition affects sampling choices in general, we fit a multinomial logit model using the mlogit package (Croissant, 2020) in R (version 4.2.2; R Development Core Team, 2022). The model predicted participants' category-based sampling choice type (with four options: choosing the within-category subordinate exemplar, the within-category basic exemplar, the within-category superordinate exemplar, or an outside-category exemplar) from training condition (Narrow, Intermediate, Broad; dummy coded). A likelihood-ratio test indicated a significant effect of training condition on participants' sampling choices, $\chi^2(6) = 42.12$, $p < .001$ (Figure 2). The effect of training condition was robust across alternate model specifications, including fitting a logistic multinomial model with by-subject random effects

(intercepts); controlling for Category Type; and specifying the dependent measure as all 9 distinct sampling images.

Participants made both confirming and constraining sampling choices. To further investigate participants' specific sampling strategies, we used the lme4 package (Bates et al., 2015; version 1.1-31) in R to fit a logistic mixed-effects models testing participants' likelihood of (a) confirming choices and (b) constraining choices.

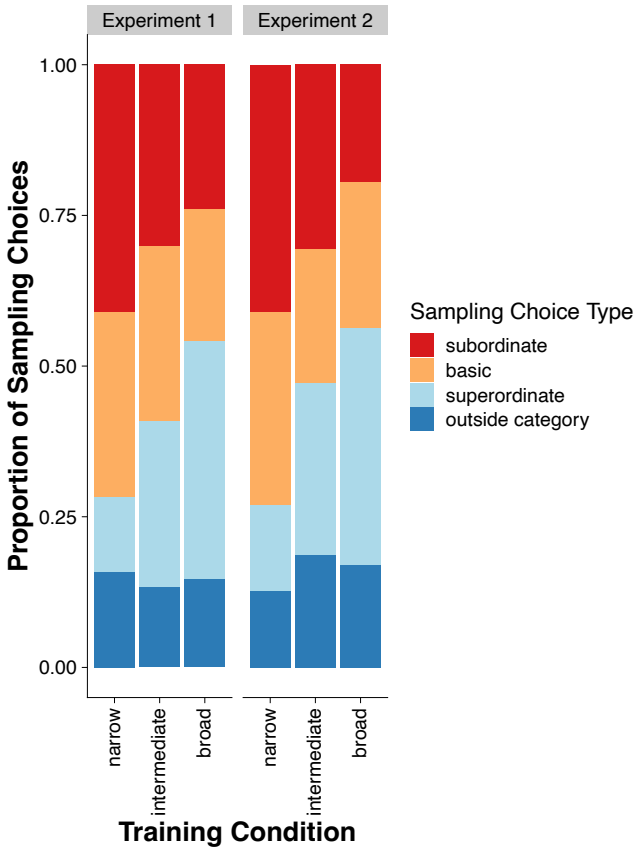

Figure 2: Sampling Choices in Exp 1 and Exp 2

Confirming sampling choices. We first fit an intercept-only logistic mixed-effects model predicting whether participants were more likely than would be predicted by chance to make confirming choices on each trial, including a by-participant random intercept. Since the chance level of making a confirming choice changed across trials depending on training condition (e.g., chance = 1/9 for the subordinate training condition and chance = 1/3 for the superordinate training condition), we adjusted chance on a trial-by-trial basis. Participants were more likely than would be expected by chance to make confirming sampling choices, $b = 2.04$, $z = 14.1$, $p < .001$. To test whether the likelihood of confirming choices differed across conditions, we included a fixed effect of Training Condition (dummy coded) in the otherwise identical logistic mixed-effects model. There was a significant overall effect of condition, $\chi^2(2) = 28.23$, $p < .001$. There was no evidence that the proportion of confirming sampling choices changed across trials ($p = .33$).

Constraining sampling choices. Using the same analytic approach as for confirming choices, we found a significant overall effect of Training Condition on the likelihood of making constraining choices, $\chi^2(2) = 231.31$, $p < .001$. Follow-up analyses showed that participants were more likely than would be expected by chance to make constraining sampling choices in the Narrow condition ($b = 1.16$, $z = 6.32$, $p < .001$) and in the Intermediate condition ($b = 0.99$, $z = 5.20$, $p < .001$). However, in the Broad condition, participants were *less* likely than would be expected by chance to make constraining choices ($b = -2.66$, $z = -10.93$, $p < .001$), likely because participants showed a preference against making choices outside of the overarching Category Type across all conditions (see Figure 2). The number of constraining sampling choices increased across trials ($b = 1.58$, $z = 2.63$, $p = .008$).

Confirming sampling choices were more prevalent than constraining sampling choices. We computed the average percent confirming choices and constraining choices for each participant, subtracting trial-level chance to account for variation in chance-levels. We then conducted a paired *t*-test between the adjusted confirming and constraining averages. Confirming sampling choices (unadjusted $M = 62.0\%$, 95% CI = [57.1%,66.9%]) were substantially more prevalent than constraining sampling choices (unadjusted $M = 24.3\%$, 95% CI = [20.4%, 28.1%]), $t(199) = 10.84$, $p < .001$.

Test Performance

Training condition modulated participants' choices at test. We investigated how the training manipulation affected participants' likelihood of selecting within-category images at the subordinate, basic, and superordinate category level. We computed the proportion of within-category images selected at each category level (subordinate, basic, superordinate) for each participant in the Test Phase (as in Xu & Tenenbaum, 2007b; Lewis & Frank, 2018) and investigated whether each of these proportions was predicted by training condition using mixed-effects regression models. Each model included a by-participant random intercept and random slope. Degrees of freedom for significance tests were approximated using the Satterthwaite approximation, implemented using the lmerTest package in R (Kuznetsova et al., 2017).

Training condition was a significant predictor for basic-level ($\chi^2(2) = 126.48$, $p < .001$) and superordinate-level choices ($\chi^2(2) = 374.32$, $p < .001$), but not for subordinate-level choices ($\chi^2(2) = 4.52$, $p = .10$). Basic-level choices were significantly higher in the Intermediate ($t(388.9) = 10.65$, $p < .001$; Figure 3A) and Broad condition ($t(391.0) = 8.45$, $p < .001$) than in the Narrow condition; superordinate-level choices were higher in the Broad condition than in the Narrow ($t(389.1) = 18.52$, $p < .001$) and Intermediate conditions ($t(388.5) = 14.18$, $p < .001$). These findings suggest that learners systematically shifted their generalizations of a novel word's extension from a narrow, subordinate-level interpretation to a broader, superordinate-level interpretation across the three training conditions.

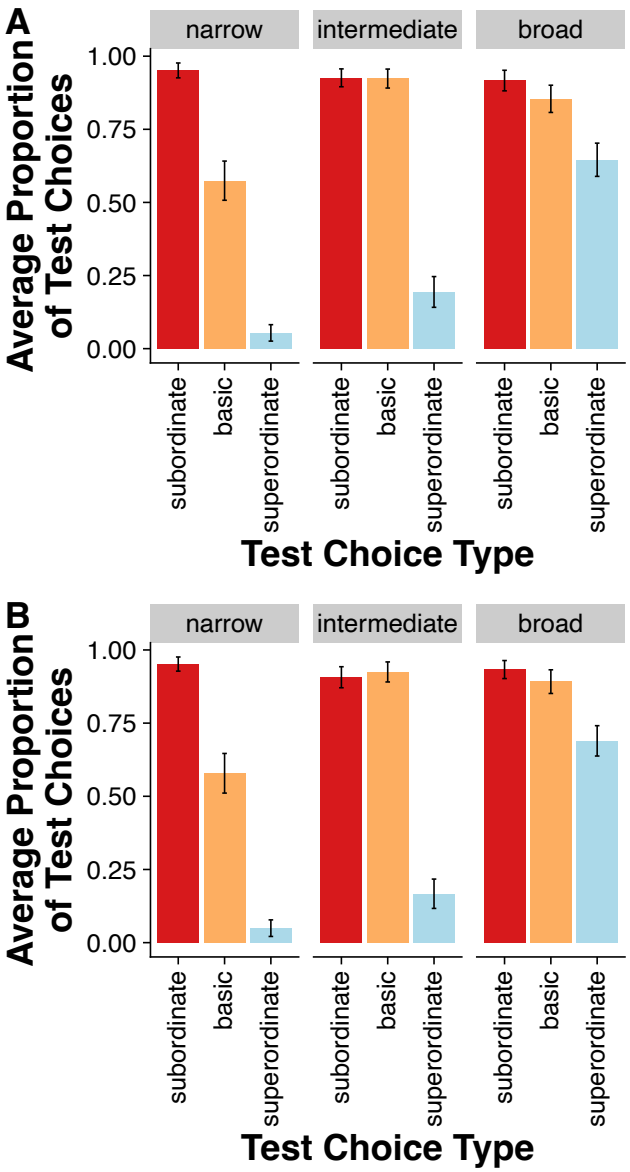

Figure 3: Proportion choices at different category levels during the Test Phase depending on Training Condition (facetted) in (A) Experiment 1 and (B) Experiment 2.

Relationship between Sampling and Test

Participants who make more confirmatory choices during sampling are *less* accurate at test. We fit a linear mixed-effects model predicting adults' d prime score for each test procedure (where hits and false alarms are determined based on the ground truth correct choices for each label; note that the same pattern of results holds when using overall test accuracy as the dependent measure) from their average proportion confirming choices across sampling trials. The model included a by-participant random intercept and slope. The more participants made confirming choices, the less accurate they were overall at test, b = -0.52, $t(148.6) = -4.88$, $p < .001$ (Figure 4A). There was no interaction with Training Condition ($p = .71$).

Participants who make more constraining choices during sampling are *more* accurate at test. We used the same analytic approach to predict participants' d prime score for each test procedure from their average proportion constraining choices across training trials. Participants' accuracy at identifying a novel word's extension increased if they had adopted a constraining sampling strategy, b = 0.67, $t(116.5) = 5.26$, $p < .001$ (Figure 4C). There was no interaction with Training Condition ($p = .11$).

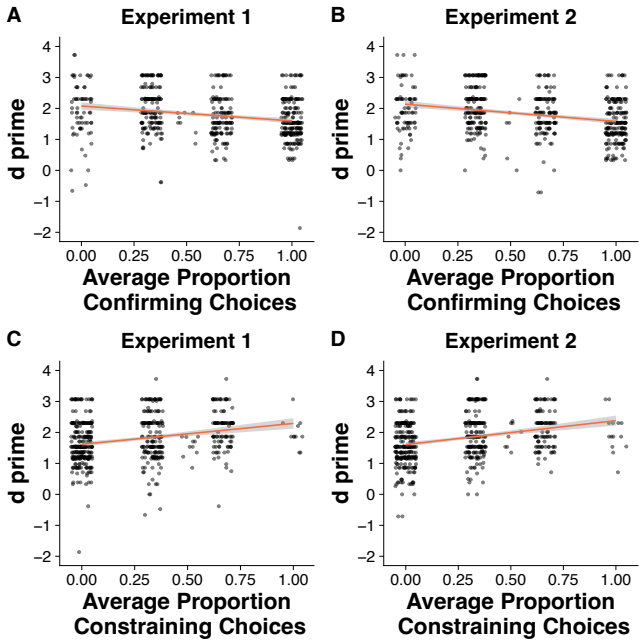

Figure 4: Relationship between participants' sampling choices and test accuracy (d prime). Accuracy decreases with a higher proportion of confirming choices in (A) Exp 1 and (B) Exp 2. Accuracy increases when participants make more constraining choices in (C) Exp 1 and (D) Exp 2.

Experiment 2

Experiment 1 represented a first step toward understanding learners' information-seeking when tasked with inferring a novel word's extension. In Experiment 2, we replicated the findings from Experiment 1 in a new sample, in order to conduct a confirmatory test of key findings. The design of the experiment was identical to Experiment 1, with one exception: we also introduced a performance-based bonus to further incentivize participants to learn the novel word meanings. This allowed us to probe whether variation in sampling strategies persisted in the presence of increased incentives to learn the novel word extensions accurately.

Method

The experiment was preregistered (https://osf.io/86vu4).

Participants

We recruited 200 participants (75 female, 123 male, 1 non-binary, 1 alternative classification; mean age: 40.3 years, *SD*

= 11.5) from Amazon Mechanical Turk using identical recruitment methods. 16 additional participants were excluded based on the same three preregistered exclusion criteria (failing to correctly enter the novel training label on at least two of three, n=7; choosing the same sampling locations on all trials, n=2; entering nonsensical text on open response questions, n=7). Participants were paid $0.80 for completing the study and received a $0.20 bonus if they finished in the top quartile in word learning success.

Stimuli, Design, & Procedure.

The stimuli, design, and procedure were identical to Experiment 1.

Results

Sampling Choices

Using the same analytic approach as in Experiment 1, we again found that participants flexibly shifted their sampling choices depending on the training condition, $\chi^2(6) = 45.38$, $p < .001$ (Figure 2). Participants were more likely than chance to make confirming sampling choices ($b = 1.88$, $z = 13.02$, $p < .001$) overall, but showed an above-chance preference for constraining sampling choices only in the Narrow ($b = 1.25$, $z = 7.26$, $p < .001$) and Intermediate ($b = 1.07$, $z = 5.89$, $p < .001$) conditions. As in Experiment 1, confirming sampling choices (unadjusted M = 59.1%, 95% CI = [54.1%,64.0%]) were substantially more prevalent than constraining sampling choices (unadjusted M = 25.7%, 95% CI = [21.8%, 29.5%]), $t(199) = 9.72$, $p < .001$. There were no effects of trial number on the proportion of confirming ($p = .69$) or constraining ($p = .57$) sampling choices.

Test Performance

Unlike in Experiment 1, there was a significant effect of training condition on participants' subordinate-level choices at test ($\chi^2(2) = 12.05$, $p = .002$; Figure 3B). All other patterns of results for the Test Phase mirrored the results from Experiment 1. Training condition had a strong effect on basic-level ($\chi^2(2) = 147.95$, $p < .001$) and superordinate-level choices ($\chi^2(2) = 554.45$, $p < .001$). Pairwise comparisons across conditions yielded similar results to Experiment 1.

Relationship between Sampling and Test

As in Experiment 1, the more participants made confirming choices during the Sampling Phase, the less accurate they were overall at test, b = -0.58, $t(128.3) = -5.91$, $p < .001$ (Figure 4B). Conversely, the more participants made constraining choices, the more accurate they were at identifying the novel word's extension at test, b = 0.76, $t(100.9) = 6.24$, $p < .001$ (Figure 4D).

General Discussion

Across two experiments, adult learners selectively sampled information about novel words based on their past learning experience. Participants were given the opportunity to control their input, and they systematically made both confirmatory choices – sampling the label for exemplars that occurred within the extension implied by the training sample – and constraining choices – sampling the label for exemplars at the boundary of the extension implied by the training sample. However, learners showed a strong preference for sampling category-confirming exemplars, with consequences for how well they correctly generalized novel word meanings. Participants who showed a stronger tendency towards constraining choices performed better at test. Conversely, participants who showed a stronger tendency to make confirmatory sampling choices were worse at test.

These studies build on prior research on learning the extension of novel words by allowing participants to select their own learning curriculum. Overall, participants pursued information-seeking strategies that in principle can aid in successfully disambiguating a novel word's extension, consistent with past work on how people sample information about novel object-label mappings in ambiguous word learning situations (Kachergis et al., 2013; Zettersten & Saffran, 2021). These findings thus further extend our understanding of how active sampling strategies may help learners rapidly acquire novel words (Hidaka et al., 2017; Keijser et al., 2019) by showing how learners may seek information that aids in correctly generalizing word meanings at multiple levels of abstraction.

At the same time, it is notable that participants tended to prefer choices that confirmed training evidence, at the expense of constraining possible word meanings. These results are broadly consistent with findings from other cognitive domains that learners often pursue positive-test strategies (Coenen et al., 2015; Klayman & Ha, 1987; Wason, 1960). A key question for future work will be to investigate how flexibly participants pursue different information-seeking strategies across learning contexts and to what extent there are stable individual differences in learners' information-seeking. Understanding consistency and flexibility in individuals' information-seeking approaches could aid in clarifying the underlying mechanisms shaping how learners sample information about new words.

In future work, we plan to adapt the current design to investigate how children approach the task of actively constraining word meanings (Xu & Tenenbaum, 2007a). Past work has found that infants (Bazhydai et al., 2020; Lucca & Wilbourn, 2019) and children (Hembacher et al., 2020; Zettersten & Saffran, 2021) actively seek novel information about words. To date, work on children's active information-seeking about word meanings has typically been limited to learning one-to-one object-label mappings. Investigating how children actively sample information about the extension of novel words could advance our understanding of how children solve the task of learning word meanings at multiple levels of abstraction.

In sum, our findings suggest that adult learners adapt their sampling in response to differing word learning experience, and active information-seeking may play an important role in successfully disambiguating a novel word's extension.

Acknowledgements

This work was supported by a grant from the NIH NICHD under Award Number F32HD110174, awarded to Martin Zettersten, and a stipend from Princeton University's Rematch+ summer program, awarded to Molly Cutler. We would like to thank Molly Lewis and Michael C. Frank for making the stimuli used in this study publicly available.

References

Austerweil, J. L., & Griffiths, T. L. (2011). Seeking confirmation is rational for deterministic hypotheses. *Cognitive Science*, *35*, 499–526.

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.

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, 23*(5), 1–10.

Coenen, A., Rehder, B., & Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, *79*, 102–133.

Croissant, Y. (2020). Estimation of random utility models in R: The mlogit Package. *Journal of Statistical Software, 95*(11), 1–41.

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information-seeking, curiosity, and attention: computational and neural mechanisms. *Trends in Cognitive Sciences*, *17*, 585–93.

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, *7*, 464–481.

Hembacher, E., DeMayo, B., & Frank, M. C. (2020). Children's social information seeking is sensitive to referential ambiguity. *Child Development, 91*(6), 1–16.

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). Cognitive Science Society.

Jenkins, G. W., Samuelson, L. K., Penny, W., & Spencer, J. R. (2021). Learning words in space and time: Contrasting models of the suspicious coincidence effect. *Cognition, 210*, 104576.

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

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

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211–228.

Klayman, J., & Ha, Y. (1989). Hypothesis testing in rule discovery: Strategy, structure, and content. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 596–604.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26.

Lewis, M., & Frank, M. (2018). Still suspicious: The suspicious-coincidence effect revisited. *Psychological Science, 29*(12), 2039-2047.

Litman, L., Robinson, J., & Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods, 49*(2), 433-442.

Lucca, K., & Wilbourn, M. P. (2019). The what and the how: Information-seeking pointing gestures facilitate learning labels and functions. *Journal of Experimental Child Psychology, 178*, 417–436.

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*, 94–122.

Navarro, D. J., & Perfors, A. F. (2011). Hypothesis generation, sparse categories, and the positive test strategy. *Psychological Review*, *118*, 120–134.

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

R Development Core Team. (2022). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Spencer, J. R., Perone, S., Smith, L. B., & Samuelson, L. K. (2011). Learning words in space and time: Probing the mechanisms behind the suspicious-coincidence effect. *Psychological Science, 22*(8), 1049-1057.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140.

Xu, F., & Tenenbaum, J. B. (2007a). Sensitivity to sampling in Bayesian word learning. *Developmental Science, 10*, 288–297.

Xu, F., & Tenenbaum, J. B. (2007b). Word learning as Bayesian inference. *Psychological Review, 114*, 245–72.

Wang, F. H., & Trueswell, J. (2022). Being suspicious of suspicious coincidences: The case of learning subordinate word meanings. *Cognition, 224*, 105028.

Zettersten, M., & Saffran, J. (2021). Sampling to learn words: Adults and children sample words that reduce referential ambiguity. *Developmental Science, 24*(3), e13064