Schrödinger's Category: Active Learning in the Face of Label Ambiguity

John D. Patterson (jpttrsn@psu.edu)
Elisabeth A. Karuza (ekaruza@psu.edu)

Department of Psychology, The Pennsylvania State University
University Park, PA 16802 USA

Abstract

Research on active category learning—i.e., where the learner manipulates continuous feature dimensions of novel objects and receives labels for their self-generated exemplars—has routinely shown that people prefer to sample from regions of the space with high class uncertainty (near category boundaries). Prevailing accounts suggest that this strategy facilitates an understanding of the subtle distinctions between categories. However, prior work has focused on situations where category boundaries are rigid. In actuality, the boundaries between natural categories are often fuzzy or unclear. Here, we ask: do learners pursue uncertainty sampling when labels at the boundary are themselves uncertain? To answer this question, we introduce a fuzzy buffer around a target category where conflicting labels are returned from two 'teachers,' then we evaluate how sampling and representation are affected. We find that, relative to the rigid boundary case, learners avoid uncertainty, opting to sample densely from highly certain regions of the target category as opposed to its boundary. Subsequent generalization tests reveal that the sampling strategies encouraged by the fuzzy boundary negatively affected participants' grasp of category structure, even outside the fuzzy buffer zone.

Keywords: active learning; category learning; fuzzy boundary

Introduction

Category learning occupies an important role in higher-order cognition. An understanding of what does or does not constitute a member of a category can speed up processing, facilitate reasoning and decision making, and, of course, is inextricably tied to lexical development. The study of category learning has yielded an array of findings about how humans acquire categories under a variety of task conditions (e.g., *classification*: Shepard, Hovland, & Jenkins, 1961; *inference*: Yamauchi & Markman, 1998). However, despite the range of tasks typically associated with the literature, a clear majority of investigations target passive, 'reception' learning, or scenarios in which learners develop an understanding via passively received observations. In other words, the learner has no agency in the observations they experience. Reception learning can be contrasted with active or 'selection' learning, in which learners develop an understanding by actively selecting or generating observations they anticipate will be most informative (Bruner, Goodnow, & Austin, 1956; MacDonald & Frank, 2016; Markant & Gureckis, 2014; Markant, Settles, & Gureckis, 2016). Because active learning affords learners the agency to determine the distribution of items they experience, we henceforth refer to it as 'sampling'.

The study of active learning may not only elucidate the process through which humans come to select information, but may also speak to how selection patterns relate to what is ultimately learned (Markant & Gureckis, 2014). On an applied level, this line of work carries additional implications for how sampling can be optimized in pedagogical environments (e.g., Yang, Vong, Yu, & Shafto, 2019). Additionally, active learning is an important area of research in machine learning—using a model's representation to select the most informative training examples can expedite learning, lighten the human burden of providing label supervision, and improve model generalization (Cohn, Atlas, & Ladner, 1994). In sum, active learning research holds the potential to advance basic and applied interests across human and machine domains.

Though relatively nascent, the human active category learning literature has yielded several key results. For instance, it has been shown that active learners may, in some circumstances, outperform passive learners, even when a passive learner's observations are yoked to an active learner's selections (e.g., Markant & Gureckis, 2014). Furthermore, active learning appears to be particularly effective when it is preceded by passive learning, presumably because the latter enables enhanced hypothesis generation (MacDonald & Frank, 2016).

Perhaps the most consistent finding, however, is that people are driven to sample heavily from regions of the stimulus space with high label uncertainty: the regions nearest to category boundaries (MacDonald & Frank, 2016; Markant & Gureckis, 2014; Markant et al., 2016). This pattern of behavior accords with hypothesis-dependent sampling; learners select high-uncertainty samples that are most informative given their current hypothesis. As their hypotheses become more refined, sampling begins to approximate the category boundary. Not only do learners prefer to sample from regions of high label uncertainty, but they are also sensitive to *kinds* of uncertainty. Markant et al. (2016) used a three-category paradigm to assess whether people prefer to sample from regions with high 'global' label uncertainty (i.e., near the intersection of all three categories) or high 'local' label uncertainty (i.e., near the boundary between two categories). Results indicate that people prefer to sample boundary regions that isolate two categories.

An important qualification of the boundary/uncertainty preference is that it has only been observed under circumstances where category boundaries are rigidly defined. That is, everything on one side is 'in' and everything on the other is 'out.' Thus, boundary uncertainty is *resolvable*; as learning progresses and hypotheses are refined, that which was previously uncertain can become certain. However, boundaries of natural categories are frequently far from rigidly defined. Instead, they tend to be fuzzy or graded (McCloskey & Glucksberg, 1978; Rosch & Mervis, 1975). For example, questions such as, *is a hotdog a sandwich* or *are chess or cheerleading sports*, are likely to spur substantial disagreement between individuals. More often than not, natural categories involve *unresolvable* label uncertainty around their boundaries.

In the present work, we investigate how unresolvable label uncertainty at the category boundary: (1) affects how people sample during active learning; and (2) influences what is learned or represented. To this end, we compare two conditions in an A/not-A active learning paradigm where the goal is to learn to differentiate members of the target category from non-members. For each sample drawn by the learner, two labels are provided—one from each of two teachers. In the Label Certain condition, the boundary is rigid, (i.e., teachers unanimously return 'non-member' for any item beyond the boundary). However, in the Label Uncertain group, we introduce a fuzzy buffer between the member and non-member regions in which class information is ambiguous (i.e., teachers provide conflicting labels). We then evaluate how unresolvable label uncertainty at the membership horizon affects sampling behavior and representation, via a generalization test.

In light of previous work, we test the prediction that learners in the Label Certain group will favor sampling from regions of the category space with high label uncertainty (e.g., MacDonald & Frank, 2016; Markant & Gureckis, 2014; Markant et al., 2016). However, it is less clear precisely how sampling patterns will differ in the Label Uncertain group. To the best of our knowledge, the present experiment represents the first investigation of active learning in the context of boundary fuzziness (via simultaneous label disagreement). By one account, people may tolerate label uncertainty and continue to sample near the member horizon. However, Markant et al. (2016) demonstrated that people are choosy about the type of uncertainty they seek to resolve with their sampling. Further, a prevailing insight from studies of decision making is that individuals prefer to avoid ambiguity when possible (e.g., Rode, Cosmides, Hell, & Tooby, 1999). As such, Label Uncertain learners may opt to sample regions of the category with low uncertainty, which could hold negative consequences for learning.

Method

Participants

Eighty-two undergraduates (ages 18-30) recruited from Pennsylvania State University participated in the 30-minute experiment in exchange for partial course credit. There were 41 participants each in the Label Certain and Label Uncertain conditions. As the first group to study ambiguity-related boundary fuzziness, we were left to estimate an appropriate sample size. Large effect sizes have been shown for ambiguity manipulations ($d > 1.0$; Rode et al., 1999); effect sizes for manipulations within active learning paradigms are generally also moderate to large (ds > .5; Markant & Gureckis, 2014). As such, we used a sample size that would be appropriate for 80% power at an effect size of $d \approx .6$.

Stimuli

To make contact with previous work (MacDonald & Frank, 2016; Markant & Gureckis, 2014), we used stimuli consisting of black circles with a red center bar that spanned the diameter (Figure 1). The stimulus space contained two continuous dimensions of variation: size (the diameter of the circle) and orientation (the angle of the center bar). Size ranged from 100 to 500 pixels while orientation ranged from 0 to 150 degrees. The stimuli could assume any point within the space during learner-controlled trials (e.g., active learning). The test stimuli consisted of a 12x12 grid that uniformly spanned the entire space, yielding 144 items. The grid was split into quadrants. For each 36-trial test block, one quarter of the items from each quadrant was drawn randomly without replacement, such that each participant was tested on all 144 items by the fourth and final testing block. Stimuli for the final boundary test consisted of a subset of the full test grid—the two levels that straddled the middle of each dimension.

The target category was defined by a conjunctive rule. The category structure was created by partitioning the stimulus space into three regions: Member space, Non-Member space, and Fuzzy space. Member space defined the set of items that satisfied the conjunctive membership rule and occupied one quadrant of the stimulus space. Fuzzy space was adjacent to the Member boundary and created a buffer (20% of the range on each dimension) between the Member and Non-Member regions. The rest of the space was designated Non-Member space. Member and Fuzzy regions each took up ~25% of the total stimulus space while Non-Member space took up ~50%. As in previous work (e.g., Markant & Gureckis, 2014), we counterbalanced across all four 90 degree rotations of this category structure within the stimulus space and collapsed across them for each learning condition.

Design and Procedure

The experiment was programmed using the PsychoPy suite for Python (Peirce et al., 2019). All participants engaged in four blocks of active learning, where each block consisted of 15 samples. Following each active learning block, learners completed a test block (36 trials). In this report we focus only on primary measures related to active learning and test blocks. Ancillary measures were collected following the last test block. In the interest of methodological completeness, we describe them here; however, that set of results is beyond the scope of the present report. Immediately following the last test block was a final border test set, consisting of items at the border between Member and Fuzzy space. After the final border test set, participants then were asked to create a 'most average' example of the target category in the make-one task. Example trials for these tasks can be seen in Figure 2. Lastly, learners rated how much they trusted each of the teachers.

Active Learning Participants were informed that they would see objects that varied in size and angle of their center bar, that some of them were known as 'Lunqs', and that their goal was to learn about the characteristics of a Lunq. On each active learning trial, a randomly initialized item within the stimulus space was presented and learners were instructed to modify the item however they wanted. Participants were required to make a modification to at least one dimension before submitting it in order to learn whether it was a Lunq. On-screen instructions told participants how to modify the item: they could hold down either the 'x' or 'z' key and move the mouse to the left or right. The mapping between key (x, y) and dimension (size, angle) was randomly determined for each participant. The same was true for the mapping between mouse movement (left, right) and dimension change (decrease, increase). Upon submitting their item, participants were asked how likely it was the item they created was a Lunq and registered their response on a scale that ranged from 'Definitely Not' to 'Definitely'. After submitting their rating, participants were provided label feedback.

Our primary manipulation was the nature of feedback provided for Fuzzy region samples. The goal was to create unresolvable label uncertainty at the category boundary for the Label Uncertain group. We considered probabilistically returning different labels for samples in this region, an approach taken in studies of probability learning (e.g., Kruschke & Johansen, 1999). However, substantial exposure would be required before the learner came to appreciate the uncertainty of the region. Instead, we opted to return two labels for each sample, one per teacher. As can be seen in Figure 2, feedback from both teachers, 'Biv' and 'Zup,' appeared on separate lines below the stimulus. While participants in both conditions received unanimous 'Lunq' and 'NOT a Lunq' feedback for Member and Non-Member samples respectively, the Label Uncertain group received one of each label for Fuzzy region samples while the Label Certain group received unanimous non-member feedback. Thus, in the Label Uncertain condition, each teacher has a slightly different conception of the stimulus space which they maintain throughout training—one has a more inclusive category while the other's is more exclusive—and participants have been given no reason to believe that one teacher is more reliable than the other.

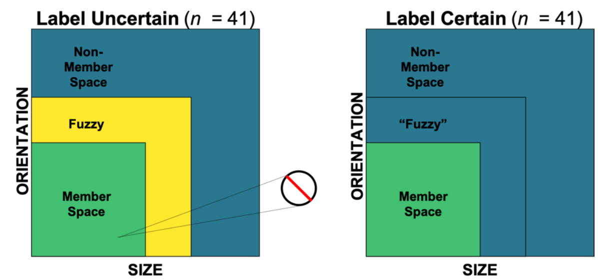

Figure 1: Category structure for each condition and example stimulus. Note: the 'Fuzzy' space in the Label Certain group was Non-Member space.

Assessment On each block test and final border test trial, an item was presented and participants were asked if the item was a member of the category 'Lunq.' They used a slider scale to simultaneously classify the item and register their confidence. The scale ranged from 'Definitely Not' to 'Definitely,' where right (left) of the middle point indicated class inclusion (exclusion) and distance from the middle point indicated confidence. Text below the slider dynamically indicated the current classification and degree of confidence based on the position of the slider, and participants clicked submit when both reflected their opinion.

In the make-one task that followed the final border test, participants were asked to create a member of the category they considered to be most average, using the same controls from the active learning phase. After the make-one task, participants were asked to rate their trust of each teacher. Two sliders appeared in the middle of the screen, with the name of each teacher beside them. The scales ranged from 'No Trust' to 'Full Trust.' Participants registered their responses and clicked submit for each.

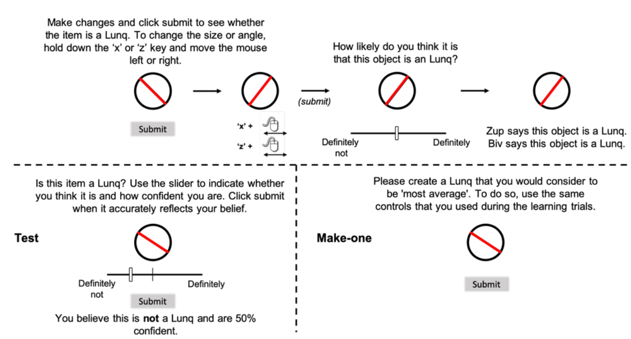

Figure 2: Learning and assessment tasks.

Results

Active Learning

Area-wise Analysis Of principal interest in the present work was assessing how unresolvable uncertainty at the boundary affected sampling behavior. To this end, we first tested whether condition (Label Uncertain, Label Certain) predicted the extent to which learners sampled from each region (i.e., Member, Fuzzy, Non-Member; Figure 3).

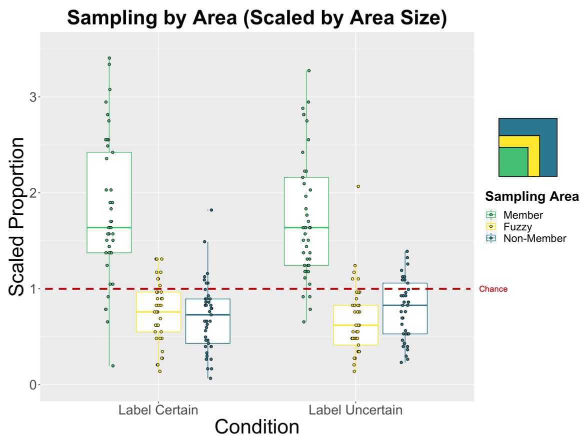

Figure 3: Sampling by area and condition. A value of 1 on the y-axis means sampling was at 'chance' or exactly proportional to the size of the region.

We used condition and region to predict scaled sampling proportion (proportion of samples within a given area, divided by proportion of the stimulus space the region took up) in a linear mixed effect regression with subject as a random intercept. The model yielded a main effect of sampling area ($F(2, 240) = 122.05$, $p < .001$), where learners sampled reliably more from the Member region, relative to both Fuzzy and Non-Member regions, but Fuzzy and Non-Member regions did not differ (*Member-Fuzzy*: $t(160) = 13.74$, $p < .001$; *Member-Non-Member*: $t(160) = 13.31$, $p < .001$). Somewhat surprisingly, however, neither condition, nor the interaction showed significant differences.

Member-region Analysis We next conducted a post-hoc analysis based on our qualitative observations of Member space sampling. Despite the very obvious limitations of this data-driven approach (which we do not intend to understate), it enabled a more fine-grained portrayal of exactly *how* learners in the two conditions sampled the Member space, given that participants in both conditions sampled most heavily from this region relative to the other two. Qualitatively, the two conditions exhibited strikingly different sampling patterns within Member space (Figure 4). Label Certain learners showed more evenly-distributed sampling across Member space, while Label Uncertain learners focused their sampling in regions of Member space that: (1) minimize uncertainty along one or both category-relevant dimensions (i.e., close to axis lines in Figure 4); and (2) are primarily distant from Fuzzy region label ambiguity. This pattern of results is consistent with ambiguity avoidance shown in the decision making literature.

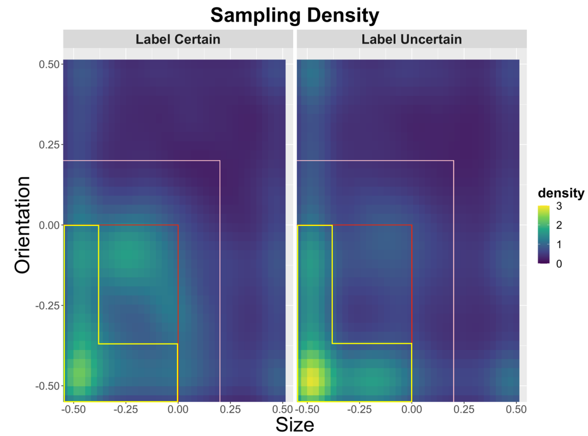

Figure 4: Sampling density by condition. Axes show scaled size and orientation dimensions. The red box corresponds to the boundary of Member space while the area between the red and pink boxes reflects Fuzzy space. The yellow 'L' shows the idealized sampling region.

This qualitative observation prompted us to conduct a post-hoc analysis to evaluate whether the conditions differed quantitatively in the extent to which they sampled from these high-certainty regions of Member space. In the analysis, we operationalized 'high-certainty' as being in the third of Member space closest to the extremes on *either* dimension (i.e., next to axis lines in Figure 4). We included *either* dimension, as this would include the region of maximal certainty across *both* dimensions (the corner), but would also capture high-certainty samples if the learner had only reached a unidimensional understanding of the category (at a given stage of learning or overall).

The resulting L-shaped region of interest (Figure 4), which we refer to as the 'idealized' Member region, was used to score each Member sample based on whether it was in or out, yielding a binomial outcome. An important consequence of this coding scheme is that Member samples outside the idealized region are closer to the area of maximal uncertainty, or the intersection of each dimension's Member boundary. This dependent variable was regressed onto condition, block, and their interaction using a generalized linear mixed effect model. An intercept for participant and a random slope for block were included in the random effects structure. The analysis showed a main effect of condition ($\chi^2_{condition}(1) = 10.43$, $p < .01$), where the Label Uncertain group was more likely to sample from the idealized region than the Label Certain group ($\beta = -1.02$, $SE = 0.32$, $z = 3.23$, $p < .001$). There was also a main effect of block and a reliable block-by-condition interaction ($\chi^2_{block}(1) = 15.16$, $p < .01$; $\chi^2_{interaction}(1) = 5.70$, $p < .05$). The rate of idealized sampling decreased across blocks, though the interaction revealed that this was underpinned by a reliable decrease across blocks in the Label Uncertain group ($\beta = -0.59$, $SE = 0.15$, $z = -3.89$, $p < .001$), but not the Label Certain group. As Figure 5 shows, this effect appears to reflect a 'late start' to more fully exploring

Member space for the Label Uncertain group. Given our method of defining the idealized region [1 = idealized region, 0 = rest of Member space], this analysis also indicates that Label Certain learners sampled more from regions that were closer to the boundaries of both dimensions (areas with higher global uncertainty).

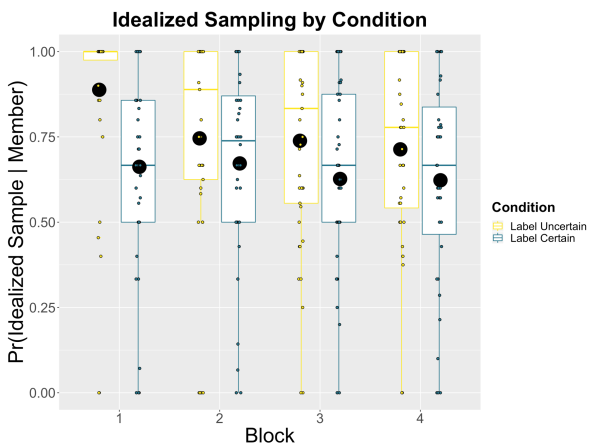

Figure 5: Idealized sampling by block and condition. Black dots represent condition means.

Assessment

As noted above, we focus the assessment analyses on our core measure: the test grid of the whole stimulus space. However, as a manipulation check, we first compared confidence for Fuzzy region test items between the two conditions. As expected, Label Uncertain learners were less confident than Label Certain learners for Fuzzy region items, but the two conditions did not differ in their confidence for the rest of the space (*Fuzzy*: $t(79.98) = 2.37$, $p < .05$; *Rest*: $t(79.98) = 1.54$, *ns*). These results suggest that condition differences cannot be attributable to globally reduced confidence.

Next, we turn to how participants in the Label Certain and Uncertain conditions compare in terms of how well they learned the unambiguous regions of the space. We compare the conditions using *d'*—a measure of learners' ability to accurately endorse Members without inappropriately endorsing Non-Members—for each participant. Recall, the two conditions received different feedback for Fuzzy region samples during learning. As such, it would be unreasonable to include Fuzzy region test items in the *d'* calculation. However, the conditions were perfectly matched in the training feedback received for Member and Non-Member regions (as seen in Figure 1). Thus, only test items that fell within the Member and Non-Member regions were included in the *d'* calculation. We regressed *d'* scores onto condition, block, and their interaction in a linear mixed effect regression, with a random intercept for participant and a random slope for block.

The analysis showed main effects of block and condition, but the interaction was not significant ($\chi^2_{block}(1) = 40.56$, $p < .001$; $\chi^2_{condition}(1) = 4.76$, $p < .05$; see Figure 6). Though learners' understanding of the space improved as training progressed ($\beta = 0.33$, $SE = 0.04$, $t = 7.67$, $p < .001$), learning in the context of boundary fuzziness led to poorer overall understanding (*Label Certain* > *Label Uncertain*: $\beta = -0.41$, $SE = 0.19$, $t = -2.18$, $p < .05$).

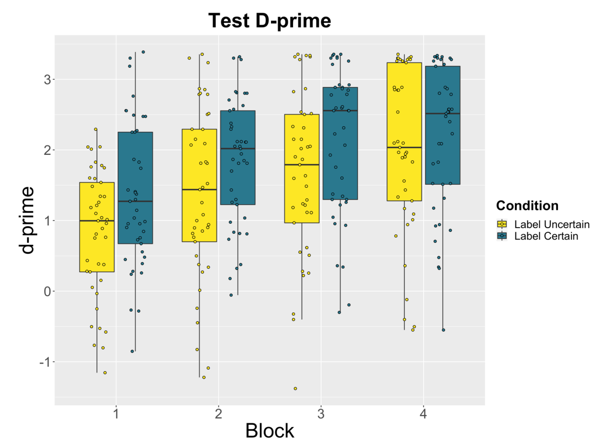

Figure 6: Test *d'* by block and condition.

To ascertain if lower *d'* in the Label Uncertain condition was attributable to poor accuracy on Members, Non-Members or both, we regressed trial-by-trial accuracy on condition, region (Member, Non-Member), and their interaction in a generalized linear mixed effect regression. The random effects structure included a slope for region and an intercept in the participant term and a slope for condition and an intercept for item. The analysis revealed a reliable main effect of region and an interaction ($\chi^2_{region}(1) = 21,89$, $p < .001$; $\chi^2_{interaction}(1) = 4.30$, $p < .05$). Learners were more accurate for Non-Members than they were for Members ($\beta = 1.73$, $SE = 0.21$, $z = 8.39$, $p < .001$), likely due to the small size of Member space. Of central importance, the interaction showed that although the two conditions did not differ in their accuracy for Member items, Label Uncertain learners were reliably worse at accurately rejecting Non-Members than the Label Certain group ($\beta = -1.03$, $SE = 0.31$, $z = -3.34$, $p < .001$). This finding suggests that the lower *d'* for the Label Uncertain group is driven by an increased false alarm rate (as can be seen in Figure 7).

Relating Sampling & Assessment The analyses above show that fuzzy, uncertain boundaries alter: (1) how people actively sample; and (2) how well they learn the overall category structure, even outside the fuzzy buffer. But how do the two relate? Specifically, how does the amount that learners sampled from the Fuzzy region affect their understanding of the *rest* of the space? We addressed this by regressing overall *d'* on condition, Fuzzy sampling (area-scaled proportion), and their interaction in a linear regression. The model returned only a significant interaction ($F(1, 78) = 5.42$, $p < .05$). As Fuzzy region sampling increased, so too did *d'* for the Label Certain group ($\beta = 1.31$, $SE = 0.47$, $z = 2.80$, $p < .01$; Figure 8), but the Label Uncertain group saw no benefit. Recall, the Fuzzy region was not 'fuzzy' for the

Label Certain group, and returned non-member feedback. This finding reaffirms the benefits of boundary/uncertainty sampling shown in previous work (e.g., Markant & Gureckis, 2014), yet crucially, in the case of unresolvable boundary uncertainty, boundary sampling conferred no benefit.

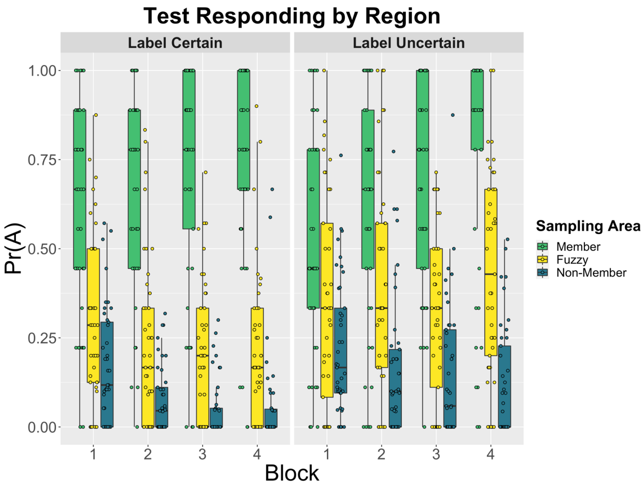

Figure 7: Probability of responding 'Member' at test by region and condition.

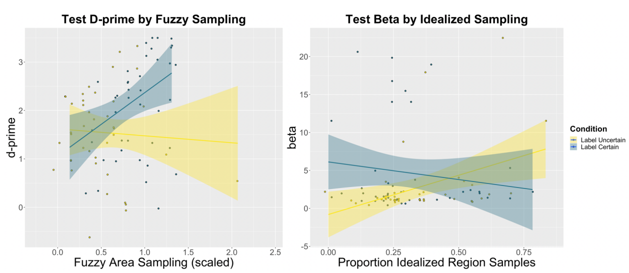

Figure 8: d' by Fuzzy Sampling (left) and β by idealized sampling (right).

Next, we evaluated how the idealized sampling of the Label Uncertain group related to test performance. We first conducted a d' analysis identical to the one above, but substituted idealized sampling for Fuzzy sampling. No effect of idealized sampling was observed. If learners focused their sampling in idealized regions of Member space, then their notion of membership might be highly constrained, which would lead to an offset in the d' measure (lower false alarm rate, but also lower hit rate). We assessed this possibility via a signal detection measure of response bias, β. Values of β less than 1 indicate an increasingly liberal notion of the category (over-extension of membership), while values greater than 1 indicate an increasingly constrained notion of the category. Using the same model, but with β instead of d', the model revealed a marginal effect of condition, a reliable effect of idealized sampling, and an interaction ($F_{condition}(1, 78) = 2.87$, $p = .094$; $F_{idealized}(1, 78) = 5.16$, $p < .05$; $F_{interaction}(1, 78) = 5.61$, $p < .05$). The Label Uncertain group exhibited a marginally more liberal representation of the category ($\beta = -0.98$, $SE = 0.58$, $t = -1.70$, $p = .094$). However, as idealized sampling increased, Label Uncertain learners' conception of the category became increasingly constrained ($\beta = 10.32$, $SE = 3.62$, $t = 2.85$, $p < .01$). Yet, idealized sampling had no relation to response bias in the Label Certain group ($p >. 3$).

Discussion

Motivated by a commonly observed property of natural categories, the experiment presented here asks, for the first time, how sampling and category knowledge is affected by a fuzzy boundary. Generally, we found that unresolvable label ambiguity near a category boundary exerts a pronounced effect on both. In the presence of a fuzzy boundary, participants were more likely to sample from highly-certain, idealized regions of the category, suggesting a tendency to avoid unresolvable uncertainty. This finding comports with previous work which has shown avoidance of ambiguity in decision making (Rode et al., 1999). The present work offers an extension of this tendency into the realm of active category learning. Further, it suggests an important qualification to a core finding in active learning research: people sample regions of uncertainty, but only when that uncertainty is *resolvable*.

We also found markedly reduced representational quality for fuzzy boundary learners, owed to a broader construal of the category and increased false alarm rate. Our analyses relating sampling and test performance highlighted that sampling the area adjacent to the member boundary (i.e., Fuzzy region) was associated with improved knowledge of the space in the rigid boundary case, a result consistent with previous work showing benefits of boundary sampling (e.g., Markant & Gureckis, 2014). However, in the context of a fuzzy boundary, where samples from this area did not afford sharpening of the boundary, no benefit was conferred. This possibility may, at least in part, explain higher false alarm rates observed for fuzzy boundary learners.

Although the fuzzy boundary led to increased sampling from idealized regions of the category, this sampling behavior was unrelated to classification accuracy at test. It was, however, related to the development of more conservative representations of the category, as evidenced by the signal detection measure of response bias, β. As poorer test performance in the fuzzy boundary condition was the result of a liberal conception of the category (i.e., high false alarm rate), idealized sampling can be seen as somewhat adaptive, serving to reduce false alarm rate at a potential cost to hit rate.

Collectively, these findings convincingly show that boundary fuzziness exerts a profound impact on the kinds of exemplars people select during learning and the quality of acquired category knowledge—and this has far-reaching implications. More often than not, research in category learning has employed stimuli, learning tasks, and feedback that caricature naturalistic experience. Yet, imbuing the learning environment with more naturalistic elements (e.g., boundary ambiguity) can radically alter learning processes and representations. This highlights the importance of improving the degree of contact between paradigms of study and naturalistic categorization for understanding basic category learning phenomena.

The present work has taken important steps toward understanding active learning and the influence of more naturalistic and fuzzy categories. However, many open questions that remain pave the way for future research. Though the observed certainty-sampling effect was highly reliable, a limitation of this work is the post-hoc nature of the sampling analysis. Thus, cautious interpretation of that effect is warranted and an important next step will be confirmation via follow-up studies.

It will also be important in future work to evaluate how ambiguity-tied fuzziness fits into the broader category learning picture. What role does the learning task play in the representational deficits seen for the ambiguous condition? How did the use of a relatively complex category structure affect the pattern of results? An examination of fuzzy and rigid learning under active *and* passive learning formats—using category structures of differing complexity—will make important connections to previous active learning research (Markant & Gureckis, 2014). The active/passive manipulation would also serve to clarify whether reduced learning stems exclusively from the feedback (i.e., the absence of boundary-sharpening non-members) or if fuzziness-driven changes to sampling behavior also play a role.

Another open question pertains to the learning goal—to differentiate category 'A' from 'not-A' in the present study. We found learners sampled primarily from 'A' space and, within that region, fuzzy boundary learners selected more high-certainty samples. However, it is conceivable that an 'A' versus 'B' goal format may increase the pressure to find the boundary of 'A' and more fully explore 'not-A' space. Under these circumstances, perhaps unresolvable label ambiguity at the boundary would be tolerated or ignored, as the learning goal is incompatible with avoidance. Given the general lack of comparisons between these two goal formats in the literature, this is an important area for future work in many respects.

Finally, it will be important to consider factors that influence the degree of ambiguity that results from label conflict—and how degree of ambiguity relates to sampling and learning. In this study, we opted to return simultaneous conflicting labels from two self-consistent teachers. There are three factors embedded in this label conflict: time, source, and information type. First, the labels were juxtaposed in time. However, it is reasonable to expect temporally-offset label conflict (across trials) might mitigate degree of ambiguity in the short term. Second, there were two sources of information (teachers) and each provided deterministic feedback. However, one might expect that the number of sources, across which boundary ambiguity is instantiated, would modulate the effect of label conflict. Another potential contributor to degree of ambiguity is the self-consistency of each source's feedback; stochasticity in teacher feedback should serve to increase ambiguity and certainty sampling. Lastly, regarding information type, the feedback provided in this experiment was discrete in nature. Yet, an alternative flavor of this task would be to provide membership probabilities rather than discrete labels. Systematic manipulation of these factors may prove fruitful in future work.

References

Bruner, J., Goodnow, J., & Austin, G. (1956). A study of thinking. New York, NY: John Wiley and Sons.

Cohn, D., Atlas, L., & Ladner, R. (1994). Improving generalization with active learning. *Machine Learning*, *15*(2), 201-221.

Kruschke, J. K., & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1083.

MacDonald, K., & Frank, M. (2016). When does passive learning improve the effectiveness of active learning? In D. Grodner, D. Mirman, A. Papafragou, J. Trueswell, J. Novick, S. Arunachalam, S. Christie, & C. Norris (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp 2459-2464). Austin, TX: Cognitive Science Society.

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.

Markant, D. B., Settles, B., & Gureckis, T. M. (2016). Self-directed learning favors local, rather than global, uncertainty. *Cognitive Science*, *40*(1), 100-120.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets?. *Memory & Cognition*, *6*(4), 462-472.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods.*

Rode, C., Cosmides, L., Hell, W., & Tooby, J. (1999). When and why do people avoid unknown probabilities in decisions under uncertainty? Testing some predictions from optimal foraging theory. *Cognition*, *72*(3), 269-304.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573-605.

Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs: General and Applied*, *75*(13), 1.

Yamauchi, T., & Markman, A. B. (1998). Category learning by inference and classification. *Journal of Memory and Language*, *39*(1), 124-148.

Yang, S. C. H., Vong, W. K., Yu, Y., & Shafto, P. (2019). A unifying computational framework for teaching and active learning. *Topics in Cognitive Science*, *11*(2), 316-337.