Active Sampling in Perceptual Category Learning

51st Annual Meeting of the Psychonomic Society, 2010 - St. Louis, MO

computation and cognition lab // new york university

Doug Markant and Todd M. Gureckis

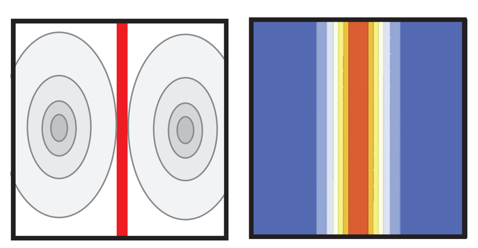
Department of Psychology, New York University

How does the ability to search for information affect category learning?

While most experimental approaches to category learning emphasize passive or observational learning, people often actively sample data to help them learn the difference between categories in the world. Our study examines how this process of active information search affects learning of basic perceptual categories. We consider two major hypotheses as to why active sampling may improve learning:

1. DATA-DRIVEN FACILITATION

Machine learning research shows that allowing an agent to select its own training data can make learning more efficient. One common method is **sampling by uncertainty**, where the agent samples the data point it is least confident about how to classify *(Settles, 2009)*.



Category Representation

Classification Uncertainty

People who can sample effectively should create **more informative data**, *leading to faster acquisition of the category boundary.*

2. DECISION-DRIVEN FACILITATION

Independent of differences in training data, prior research on discovery or inquiry-based learning suggests that allowing people to choose their own training experience leads to greater engagement and more successful learning *(Bruner, 1961)*.

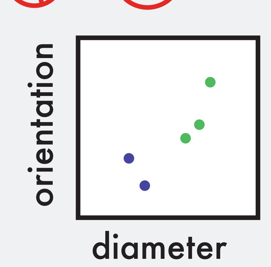
People who actively sample items should perform better than others that are **"yoked"** *to their training data,* **but do not make the sampling decisions.**

Antenna Learning Experiment

We tested perceptual category learning with "loop antennas" that varied along two dimensions: (1) the diameter of an outer loop, and (2) the orientation of the central bar:

Each antenna received one of two possible TV stations:

CH1 or **CH2**

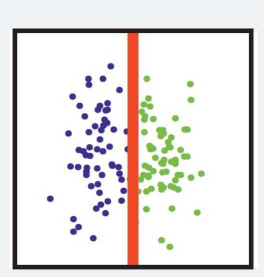


Two category types:

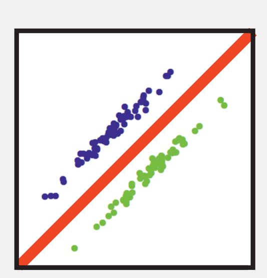
RULE BASED (RB)

Easily learned
Relies on explicit reasoning

Predictions for active learners:
Sampling close to the category boundary
Faster learning of the categories



INFORMATION INTEGRATION (II)



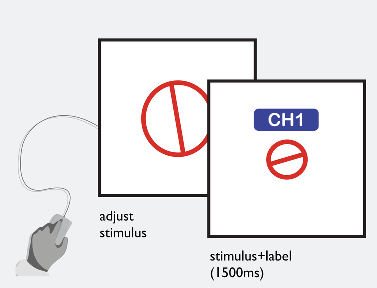
More difficult to learn than RB task
Relies on implicit or procedural learning

Predictions for active learners:
Less effective at sampling near boundary
No effect on learning speed

Four training conditions:

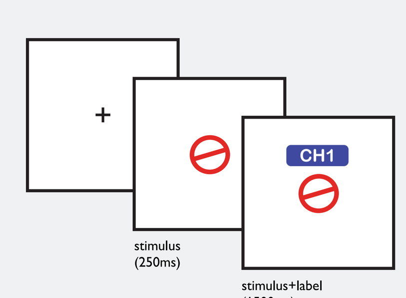
1. ACTIVE (A)

"Design" a series of antennas to help you learn the difference between CH1 and CH2 antennas.



2. PASSIVE (P)

Passively observe training data drawn from category distributions (cf. Ashby et al. 2002)



3. NAIVE PASSIVE-YOKED (Y1)

Passively observe training data previously selected by an **ACTIVE** participant, *without any knowledge of the source of the data.*

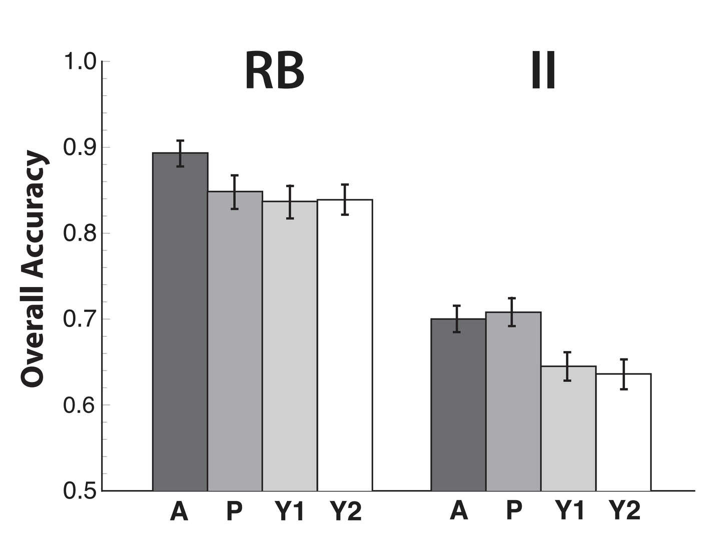
4. AWARE PASSIVE-YOKED (Y2)

Passively observe training data previously selected by an **ACTIVE** participant, *after being told that the antennas were designed by another person.*

Accuracy

In the RB task, **ACTIVE** training led to greater learning of the category boundary than **ALL** passive groups.

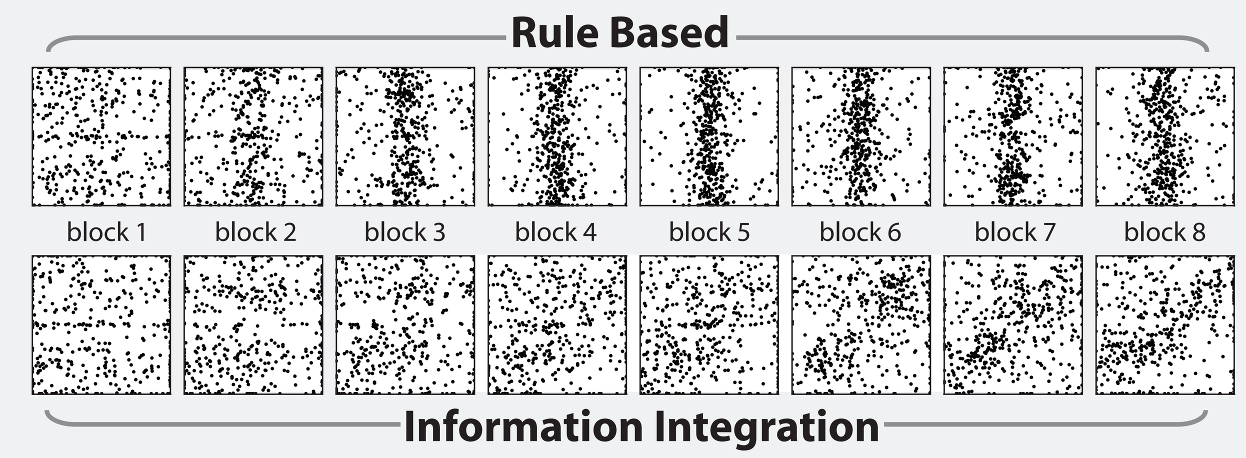
In the II task, accuracy was lower overall, and **ACTIVE** and standard **PASSIVE** training led to equal performance.

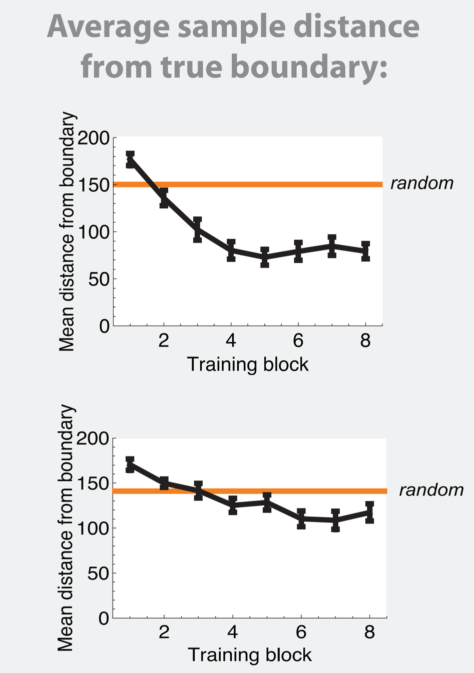


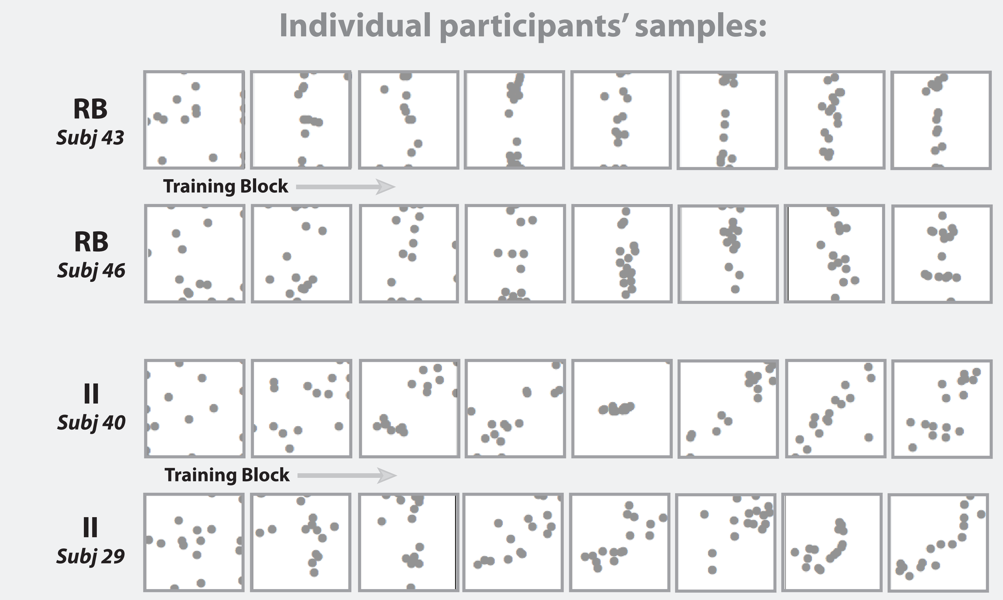
In both tasks, **YOKED** learners performed worse than the **ACTIVE** group, despite being given identical training sets.

Sampling

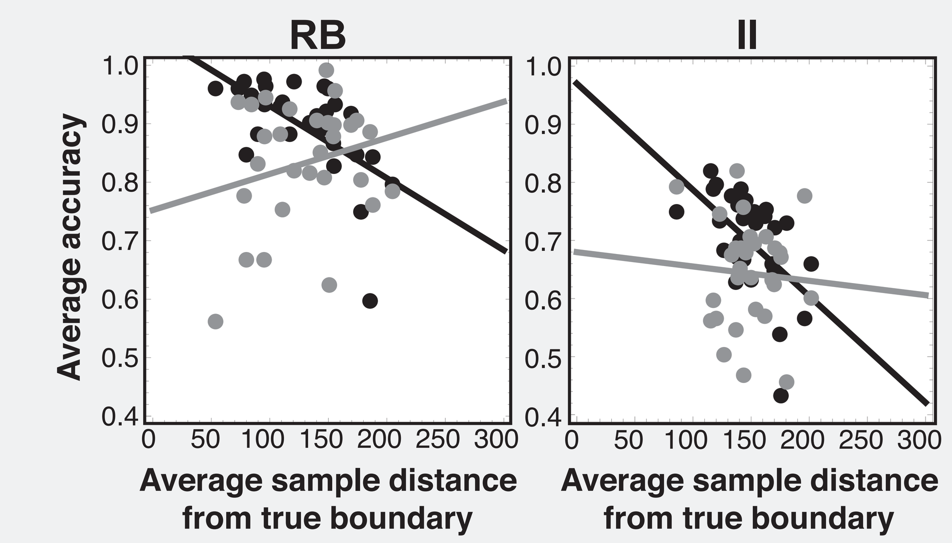
Were ACTIVE participants able to sample along the category boundary?







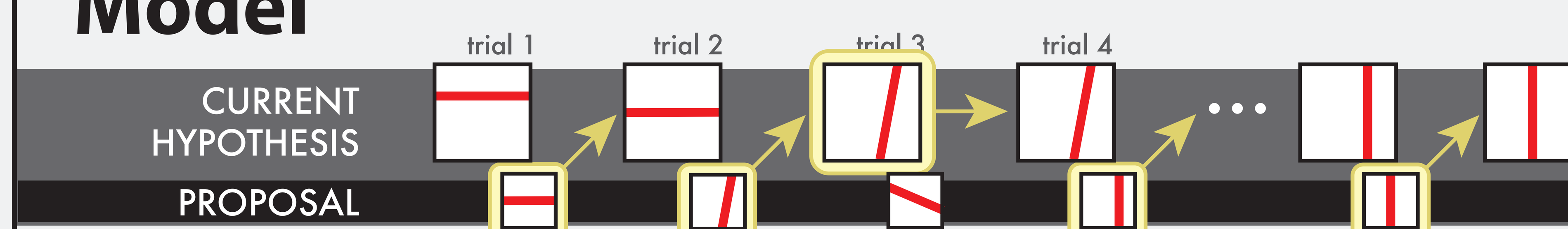
Did the ability to sample close to the boundary predict classification performance?



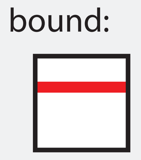
In both tasks, **ACTIVE** learners who sampled close to the boundary performed better overall.

In contrast, **YOKED** learners performed the same regardless of sample distance. Participants who received highly informative data in the RB task performed the worst overall.

Model



Model learning as **serial hypothesis testing**, in which the learner represents a single hypothesis about the category boundary in the form of a linear decision bound:

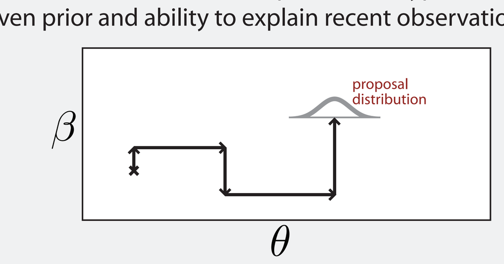


θ : angle
β : bias (offset from center)

Assume learner can evaluate the relative probability of a given decision bound using prior distribution over parameters and sigmoidal likelihood function over recent observations.

Local hypothesis generation

On each trial, sample a new hypothesis from proposal distributions centered on the current state. Adopt the new hypothesis if it is more likely given prior and ability to explain recent observations.

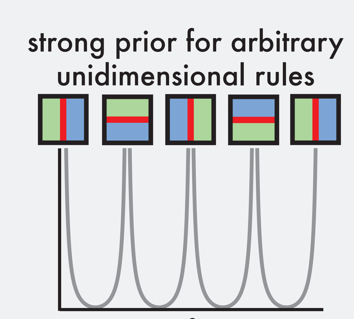


Small working memory

Participants in the behavioral experiment showed a high frequency of shifts in their response rules from block to block, suggesting they were evaluating hypotheses with respect to a small number of observations. The model captures this behavior by computing the likelihood over *n* recent observations (e.g., $n = 5$).

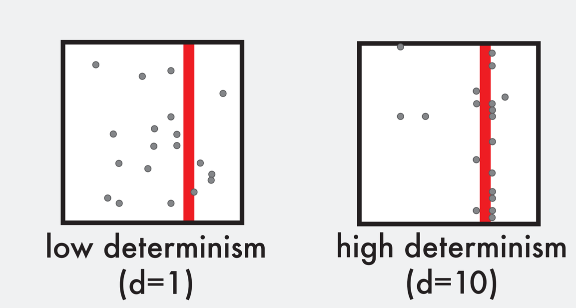
Unidimensional bias

While RB participants quickly learned the relevant feature dimension for classification, II participants tended to alternate between sub-optimal solutions on one of the two dimensions from block to block, and rarely responded according to the correct II rule. The model accounts for this behavior by assuming a strong prior bias toward unidimensional rules:



Dynamic role of sampling

To allow the model to actively sample during learning, we use **label entropy** to calculate the model's uncertainty over the stimulus space on any given trial. A **determinism** parameter **(d)** controls the randomness of sampling:



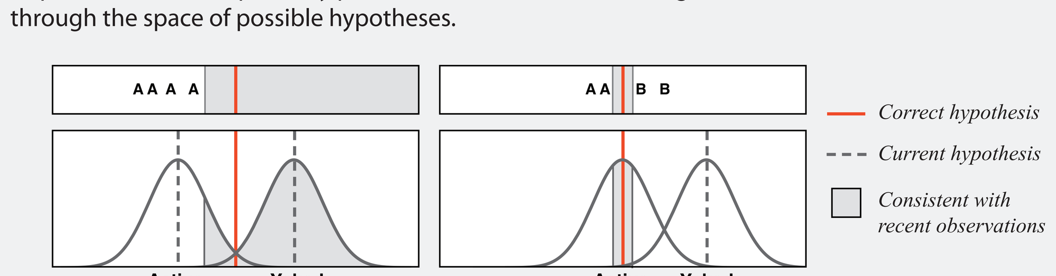
low determinism (d=1)

high determinism (d=10)

1. Due to small working memory, the determinism of recent samples affects the range of potential proposals that may be accepted on next trial.

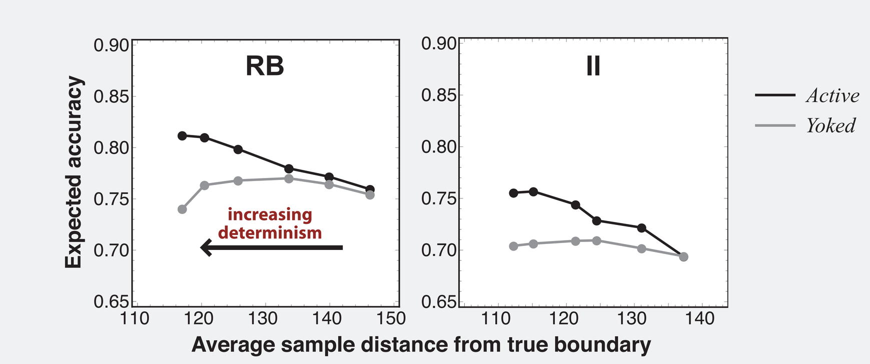
2. For **ACTIVE** models, deterministic samples lead to more efficient hypothesis search by limiting the set of new proposals that will be adopted on the next trial.

3. For the **PASSIVE-YOKED** learner whose current hypothesis differs from their **ACTIVE** counterpart, the same samples may provide weaker evidence, leading to less efficient search through the space of possible hypotheses.



Simulation

Trained 200 ACTIVE models in each task (RB/II). ACTIVE models generated a new stimulus on each trial based on current hypothesis. The simulation was run for a range of values of sampling determinism. For each ACTIVE model, we then trained a set of PASSIVE-YOKED models on the dataset generated by the ACTIVE model. Compared the expected accuracy of ACTIVE model with average expected accuracy over YOKED models.



When ACTIVE models sample randomly, there is no difference between ACTIVE and YOKED performance. **As the determinism of their samples increases, performance increases for the ACTIVE models only**, while performance stays the same for YOKED models, recovering the relationship between sample distance and accuracy.

Conclusions

1. Active sampling is effective for rule-based learning, but significantly limited for more complex tasks. Participants' ability to actively sample data was impaired in the Information Integration task in comparison to the Rule-Based task. Based on previous work with these tasks, this suggests sampling is limited by participants' ability to generate hypotheses consistent with the true category boundary.

2. Active sampling facilitates learning in the Rule-Based task but not the Information-Integration task relative to passive observation. In both tasks, overall accuracy was highly correlated with the quality of a participant's samples.

3. In both tasks, people who were "yoked" to active learners performed worse despite learning from identical training sets, regardless of whether they were aware of the source of the data. Even participants who observed highly informative data close to the category boundary (but did not generate those data themselves) performed worse overall.

4. Contrary to models that treat learning as the slow accumulation of data, our behavioral findings can be accounted for by a process of serial hypothesis testing. Our model learns through local, stochastic shifts in a single hypothesis, but evaluates the likelihood of that hypothesis using a small number of recent observations. As a result, the ability to test one's hypothesis through sampling plays an important role in guiding hypothesis search for the active learner, while the same data provides no benefit to passive learners.