Order effects in active and passive hypothesis testing

Kyle MacDonald
kyle.macdonald@university.edu
Department of Psychology
Stanford University

Michael C. Frank
mcfrank@university.edu
Department of Psychology
Stanford University

Abstract

Active learning can speed concept learning by allowing people to collect highly informative samples based on their current hypotheses. But real-world learning often involves both active and passive learning contexts, an interaction that we know little about. In three category learning experiments with adults, we explore the effectiveness of active learning in a well-understood category learning task. Experiments 1a and 1b are direct replications of the active learning advantage found in Markant & Gureckis (2014). Experiment 2 is an extension that tests how different sequences of active/passive training modulate the effectiveness of active learning. Experiment 3 is a conceptual replication of the order effect findings, using a novel paradigm where participants learn a higher dimensional concept. Across all experiments, active training lead to better learning of the target category boundary. Passive-then-active training was more effective compared to Active- then-passive in both Experiments 2 and 3. Our replication data provides additional support that active learning can provide an advantage over passive learning, and our extension data show that active learning can be more effective when the learner already has some passive experience with the learning task.

Keywords: active learning, hypothesis testing, category learning, replication, order effects

Introduction

- What is active learning?

- Why is active learning helpful?

  - Human active
  - Machine active learning
  - When is active learning most effective

- Current work

In their experiment, Markant & Gureckis (2014) compared the effectiveness of active vs. passive training on the rate of participants' category learning. They tested two different types of category structures: a Rule-Based (RB) category, which varied along one dimension, and an Information-Integration (II) category, which varied along two dimensions. In the active learning condition, learners were allowed to choose specific observations from the category to test their beliefs; whereas in passive learning condition, the learner had no control over what data they saw. They found that participants in the active condition learned the category structure faster and achieved a higher overall accuracy rate compared to participants in the passive learning condition, but this advantage only held for the less complex, RB category structure.

Experiment 1a

Experiment 1a is a direct replication of the advantage for active learning over passive learning found in Markant &

Gureckis (2014). We tested participants' category learning for the RB category structure after receiving either Active or Passive training. We used the same stimuli and followed the exact procedures as the original study (described below). All of the stimuli and the experiments can be viewed and downloaded at the project page for this paper: https://kemacdonald.github.io/Act-Learn/.

Methods

Power analysis We calculated Cohen's d for the t-test comparing participants' overall accuracy rate in the Active and Passive training conditions (d = 0.47). Next, we performed a post hoc power analysis and found that the original study achieved a power of 0.43. Finally, we conducted an a priori power analysis for the replication in order to achieve 80%, 90%, and 95% power to detect the effect size in the original study. The results were: 80%, n = 84 (42 in each group); 90%, n = 158 (79 in each group), 95%, n = 238 (119 in each group).

Participants We posted a set of Human Intelligence Tasks (HITs) to Amazon Mechanical Turk. Only participants with US IP addresses and a task approval rate above 85% were allowed to participate, and each HIT paid one dollar. 137 HITs were posted for each of the two between-subjects conditions. Data were excluded if participants completed the task more than once or if they reported that they did not understand the task at the end of the experiment (2 HITs). The final sample consisted of 135 participants.

Stimuli Figure 1 shows a screenshot of the stimuli used in all experiments. Visual stimuli were black "antennas" on a white background. Each antenna could vary along two continuous dimensions – radius size or central angle – and was assigned a value between 1 and 600. These values were converted to pixel values for display on a computer screen. To ensure that participants could not complete a full rotation of the antenna, the rotation of the central angle was limited to 150 degrees. Finally, the minimum radius and angle values were randomized for each participant, such that each participant was assigned a unique optimal decision boundary.

Radius and angle values for the 96 passive training trials were generated from two Gaussian distributions with identical mean and covariance parameters as Markant & Gureckis (2014) (see Figure 1). For test trials, we created a uniform grid of 192 unique test items that covered the entire feature space. We randomly sampled 8 items from each quadrant to get 32 test trials for each block. We then randomized the order of the training and test trials within each block for each
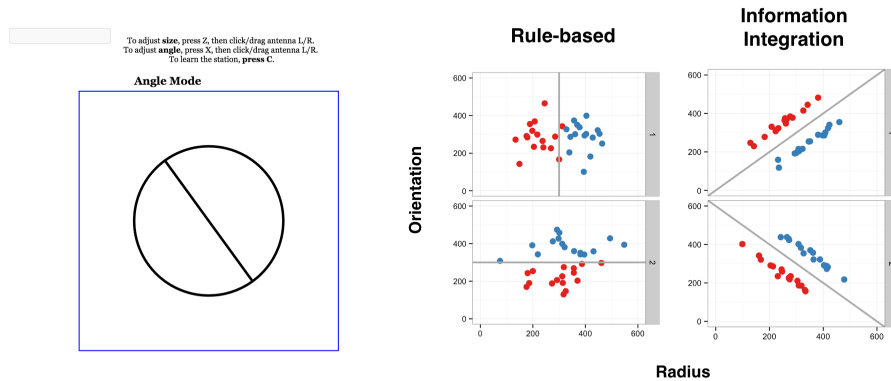
Figure 1: The left panel shows a screenshot of the stimuli used in all experiments. The right panel shows the true category distributions for the single dimension, Rule-Based category and the two dimensional, Information-Integration category.

participant.

Design and procedure Participants saw a total of 288 trials (96 training trials and 192 test trials) across 6 blocks. Each block consisted of 16 training trials and 32 test trials. Before starting the task, participants were told that this was a game where they would see "loop antennas" for televisions and each antenna received one of two channels (CH1 or CH2), and their goal was to learn the difference between the two types of antennas. We introduced some uncertainty by telling participants that the antennas could pick up the wrong channel on occasion, and that they should learn what channel is most often received by a particular type of antenna.

After the instructions, participants were randomly assigned to one the two between-subjects conditions (Active vs. Passive training). In the Active training condition, participants were able to design their own antennas to test. They modified the antenna by clicking and dragging the mouse from left to right. To change the size of the antenna, they first pressed the "Z" key. To change the angle, they first pressed the "X" key. When participants were finished with their design, they pressed the spacebar to see which channel (Ch1 or Ch2) the antenna received. The channel label appeared in a text box with a green border located above the antenna.

In the Passive training condition, participants were shown antennas with size and angles generated from the underlying category distributions. After a two second delay they were told which channel the antenna received. To ensure that participants saw the channel, they had to click on the channel text in order to advance the experiment. When they clicked the channel text, a green box appeared around the text to indicate that their response had been recorded.

After completing the training, participants in both condi-

tions proceeded to the test trials. On each test trial participants saw an antenna and were asked, "Which channel does this antenna receive?" To indicate their response participants selected one of two buttons located above the antenna.

Results and Discussion

Overall classification accuracy

Classification accuracy across blocks

Relationship between evidence selection and learning

Experiment 1b

Methods

Stimuli Stimuli were identical to Experiment 1a.

Participants Participant recruitment and inclusionary/exclusionary criteria were identical to those of Experiment 1 (excluded TODO HITs). TODO HITs were posted for each condition (TODO) for total of TODO paid HITs.

Design and procedure

Results and Discussion

Experiment 2

Methods

Stimuli Stimuli were identical to Experiment 1.

Participants Participant recruitment and inclusionary/exclusionary criteria were identical to those of Experiment 1 (excluded TODO HITs). TODO HITs were posted for each condition (TODO) for total of TODO paid HITs.
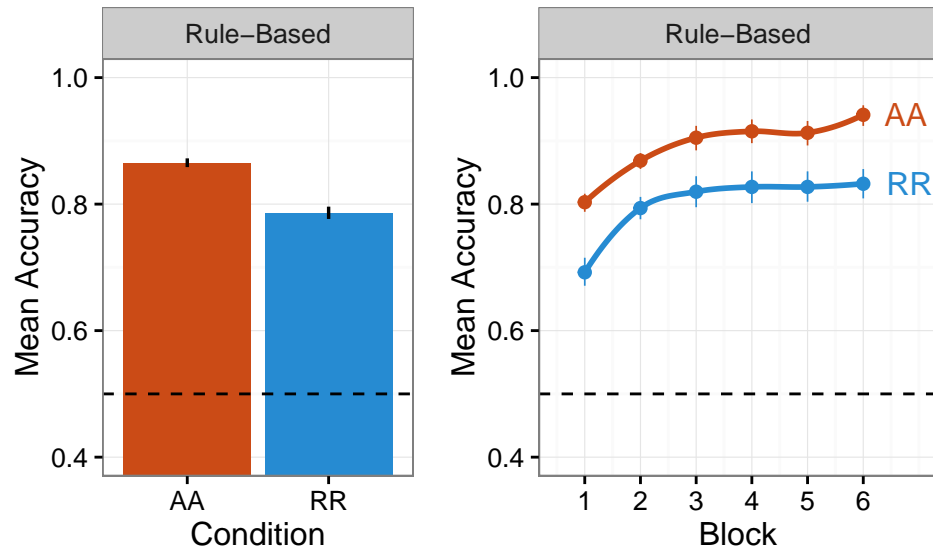
Design and procedure

Figure 2: The left panel shows a screenshot of the stimuli used in all experiments. The right panel shows the true category distributions for the single dimension, Rule-Based category and the two dimensional, Information-Integration category.

Results and Discussion

Experiment 3

Methods

Stimuli Stimuli were identical to Experiments 1 and 2.

Participants Participant recruitment, and inclusionary/exclusionary criteria were identical to those of Experiment 1 and 2 (excluded TODO HITs). 40 HITs were posted for each condition (TODO) for total of TODO paid HITs.

Design and procedure

Results and Discussion

	Estimate	Std. Error	t value	Pr($>$|t|)
(Intercept)	0.05	0.10	0.5	0.61
x	2.10	0.11	18.5	0.00

General Discussion

- Recap findings
  - Active learning advantage in a direct replication (yay science!)
  - Passive-active better than Active-passive
  - Conceptual replication
- Expand on why we see AR > RA
  - Sequential hypothesis testing model
  - Gain some understanding of task before exploring
  - RA is bad because you can't refine your current hypothesis. Can only use the data you are given to confirm/reject current hypothesis
- Limitations

- AA was always best
- task analysis
- complexity of real world learning
- Takeaway point:

References

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.