Testing the interaction between active and passive learning

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

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

Abstract

Self-directed ("active") learning is often pitted against passive learning to see which training leads to better learning outcomes. But real-world learning unfolds over time and often contains both active and passive learning *within* the same context. In the current work, we explore the interaction between active and passive learning in three category learrning experiments with adults. First, we replicate the active over passive learning advantage found in Markant & Gureckis (2014) (Experiments 1a and 1b). Then, in Experiments 2 and 3 we provide direct tests of how different sequences of active/passive training modulate the effectiveness of active learning in two markedly different concept learning tasks. Across all three experiments, active training lead to better learning of the target concept compared to passive training. Passive-first learners performed better than Active-first in both Experiments 2 and 3. Our data provide evidence that active learning can be more effective after building a better representation of the learning task.

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

Introduction

Active learning describes a process where people can select what information they will learn next.[1] This learning stands in contrast to passive learning, where people receive information presented to them from the world. For example, consider a child playing with a new toy, testing functions to see how it works (active) compared to a child watching someone else test the toy's functions (passive). In both scenarios, the child sees some of the toy's functions, but in the active context, the child has control over the sequence of incoming information.

The potential benefits of active learning have been the focus of research in education (Grabinger & Dunlap, 1995), machine learning (Settles, 2012), and cognitive science (Castro et al., 2009). In their synthesis of this literature, Gureckis & Markant (2012) present four "cognitive" explanations for why selecting information could improve learning outcomes: (1) selection of the most informative examples, (2) increased encoding of selected examples, (3) the learner has direct knowledge of the sampling process, and (4) planning an action leads to a deeper understanding of the learning task.

In the majority of these studies, researchers isolate active and passive training, and test which regime leads to better learning outcomes. For example, Markant & Gureckis (2014) compared the effectiveness of active vs. passive training on the rate of participants' learning of two abstract category structures: a Rule-Based (RB) structure where the category boundary varied along a single dimension (e.g., size), and an Information-Integration (II) structure where the category boundary was defined by two dimensions (e.g., size and angle). In the active learning condition, people selected examples from the category to test their beliefs; whereas in passive learning condition, the examples were generated randomly. Markant & Gureckis (2014) found that participants in the active condition learned faster and achieved a higher overall accuracy, but this advantage only held for the less complex, RB category. They concluded that active training can benefit learning, but it's effectiveness depends on the complexity of the target concept.

While direct comparisons of active and passive learning are important for understanding where and when active learning might be appropriate, real-world learning unfolds over time and often involves a mix of both types of contexts. Returning to the example of a child learning about a new toy – the child might see some functions demonstrated by another person, and then be given the opportunity to play with the toy (or vice versa) *within* the same learning context. Moreover, there are markedly different costs associated with active compared to passive learning, with active learning requiring more effort and time on the part of the learner. Thus, understanding how different sequences of active/passive learning interact to affect learning outcomes is an important question for both theoretical and applied reasons.

Research in education has asked a related question: How does varying the order of different teaching methods affect students' uptake of information? One illustrative example is the work on *Productive Failure*, where researchers find that allowing students to struggle with a task before a lecture (typically in the form of self-directed problem solving), leads to better uptake of the subsequent instruction (Westermann & Rummel, 2012). The takeaway from this work is that order of presentation matters for maximizing students' learning.

In comparison, there has been relatively little basic cognitive science research testing different sequences of active and passive learning. One example is Kachergis, Yu, & Shiffrin (2013), where learners were given either passive-first or active-first training in a cross-situational word learning task. People who received active learning first preformed better overall when asked to recall newly learned words. Kachergis et al. (2013) suggest that learners developed better attentional and memory strategies during the active training, which transferred to the passive training block, and boosted their overall acccuracy. But we do not yet know whether this "active-first" advantage would generalize to other types of learning tasks.

In the current set of studies, we test the hypothesis that ac-

[1] Here we focus on deliberate decisions about what to learn, as opposed to other uses of the term "active" learning (e.g., being actively engaged with learning materials).

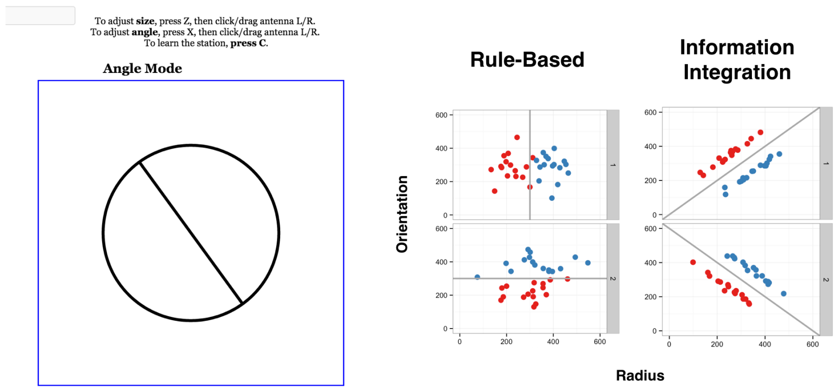

Figure 1: The left panel shows a screenshot of the stimuli used in Experiments 1 and 2. The right panel shows examples of distributions of training stimuli shown to participants passive learning condition from the Rule-Based category and the Information-Integration category.

tive learning is more effective after people gain experience with the learning task. In Experiments 1a and 1b, we directly replicate the active learning advantage found in Markant & Gureckis (2014). In Experiment 2, we build on our replication data to show that passive-first training is more effective comprared to active-first. In Experiment 3, we find the same passive-first advantage in a novel paradigm where participants learn a more complex, heirarchical category structure. Together, the data suggest that active learning can provide an advantage over passive learning, but this advantage depends on the learners' representation of the task, which can be improved by receiving more active learning or an initial bout of passive learning.

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

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. Approximately 25 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 (1 HITs). The final sample consisted of 52 participants.

Stimuli The left panel of Figure 1 shows a screenshot of the stimuli used in Experiments 1a, 1b, and 2. 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. The minimum radius and angle values were randomized for each participant, such that each participant was assigned a unique optimal decision boundary. Finally, we used a Rule-Based category structure where the category boundary is defined along a single dimension: either the antenna's size or central angle (see the right panel of Figure 1).

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 the right panel of 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 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

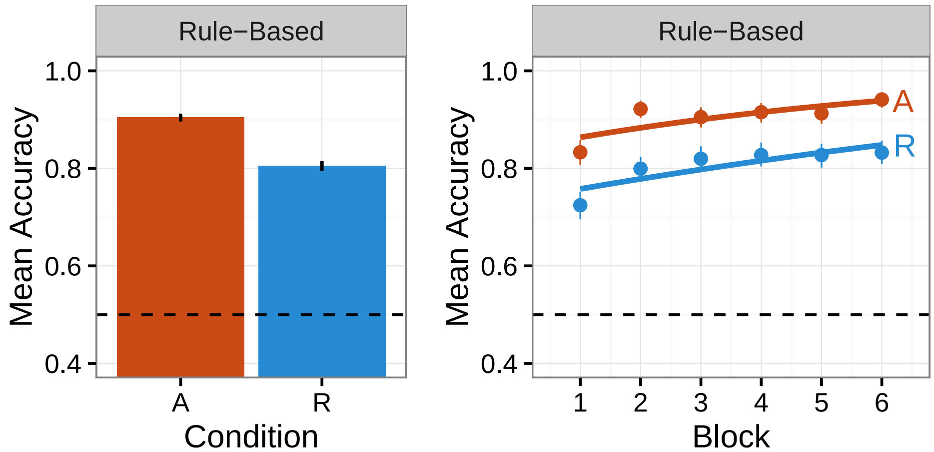

Figure 2: The left panel shows overall accuracy performance for the Active and Passive training conditions. The right panel shows participants' accuracy across each of the six blocks in the experiment. Colored lines are generated by a binomial smoother and error bars indicate 95% confidence intervals computed by non-parametric bootstrap.

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 conditions 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. At the end of each block of test trials, participants saw a summary of their accuracy on the preceding block.

Results and Discussion

Overall classification accuracy We directly followed the analysis plan of Markant & Gureckis (2014), using a t-test to compare overall test performance for participants in the active and the passive learning conditions. [2] The left panel of Figure 2 shows overall test performance, with active learners being more accurate than passive learners, $t(51) = 2.52$, $p = 0.015$.

Classification accuracy across blocks The right panel of Figure 2 shows participants' accuracies across blocks in the experiment. To quantify participants' behavior, we use mixed effects regression models with the maximal random effects structure justified by our experimental design: by-subject intercepts. We fit a logistic regression predicting test performance based on condition (active/passive) and block. The model was specified as `Correct ~ 1 + Condition * Block + (1 | subject)`. We found a significant main effect of condition (β = -0.7, p $<$ .001) with better performance for active learners, and a significant main effect of block (β = 0.2, p $<$ .001) such that responses were more accurate later in the experiment.

Relationship between sampling behavior and learning We were also interested in the relationship between participants' overall sampling behavior and learning outcomes. We follow Markant & Gureckis (2014) and quantify the quality of a sample based on it's orthogonal distance from the true category boundary, with samples closer to the boundary being of higher quality. For each participant, we computed a mean accuracy score and a mean sample distance score, and fit a linear model using sample distance to predict accuracy. We found a signficant effect of sample distance (β = -.0003, p $<$ .001) with accuracy increasing as mean sample distance decreased.

Taken together, our data provide strong evidence for a successful replication of the original results reported in Markant & Gureckis (2014). We found a comparable advantage in overall classification accuracy for active learners over receptive learners in a web-based experiment with two fewer train-

[2] All of our data, processing, and analysis code can be viewed in the version control repository for this paper at: https://github.com/kemacdonald/act-learn.

ing/test trial blocks. Our results differ from the original study in that we found an immediate advantage for active learners after the first block that was not present in the original study. Next we attempt to replicate Markant & Gureckis (2014)'s findings for the II category structure and for the yoked passive learning condition.

Experiment 1b

The goals of Experiment 1b are to (a) replicate Markant & Gureckis (2014)'s findings for the more difficult Information-Integration (II) category structure, and (b) replicate the finding that passive learners did not benefit from being "yoked" to active learners' data.[3] They did not find an active learning advantage for the II category structure and yoked learners were worse than active learners even though they had seen the exact same learning information. We used the same stimuli and followed the exact procedures as the original study (described below). However, we reduced the length of the experiment to two blocks. We made this decision based on finding an immediate active learning advantage in Experiment 1a.

Methods

Stimuli Visual stimuli were identical to Experiment 1a.

Participants Participant recruitment and inclusionary/exclusionary criteria were identical to those of Experiment 1a (excluded No, 3 HITs). 196 HITs were posted across each of the between-subjects conditions: two category structures (II and RB) and three training conditions (Active, Passive, and Yoked).

Design and procedure Procedures were identical to those of Experiment 1a. We added a "yoked" learning condition, in which we match each passive learning participant with training data generated from an active learning participant's sampling behavior. Thus, both the active and yoked participants saw the exact same data, but the active participants were in control of the information flow.

Results and Discussion

Overall classification accuracy Figure 3 shows the overall effect of category structure and training condition on participants' accuracy performance. We fit the same logistic regression as specified in Experiment 1a. We found a significant main effect of category structure (β = -0.95, $p < .001$) with better performance in the RB category structure, and a significant main effect of condition (β = -0.7, $p < .001$) such that participants in the passive and yoked conditions performed worse than participants in the active condition. We did not find any significant interactions.

Relationship between sampling and test across blocks Since the main goal of this work was to test different sequences of active/passive learning, we were interested in exploring the relations between participants' sampling behavior and test accuracy *over time* in this task. To explore these relations, we performed an exploratory analysis where we fit a linear model predicting each participant's mean accuracy based on the quality of their sampling behavior and block. As expected participants' accuracy improved in the second block (β = 0.21, p = 0.01). There was a significant two-way interaction between sampling distance and block (β = -0.0013, p = 0.01) such that the relationship between quality of sampling and accuracy did not emerge until the second block. Figure 4 shows this interaction.

TODO: quick recap and then transition to Experiment 2.

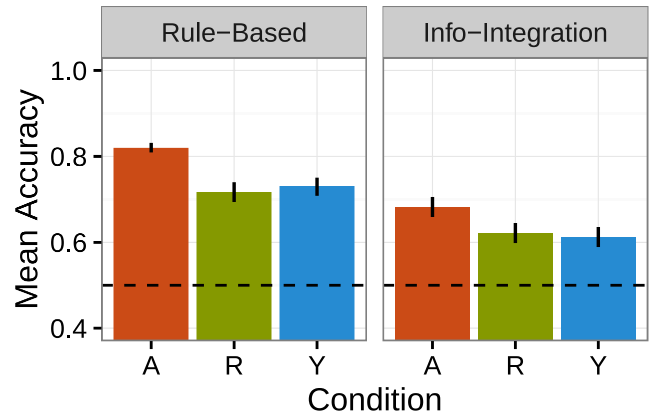

Figure 3: The left panel shows overall accuracy performance for the Active and Passive training conditions. The right panel shows participants' accuracy across all six blocks in the experiment.

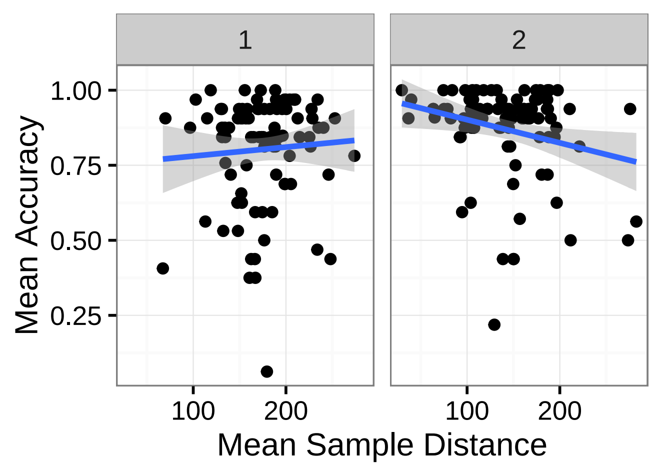

Figure 4: The relations between quality of sampling and accuracy on test trials across blocks.

Experiment 2

TODO: introduce experiment 2.

Methods

Stimuli Stimuli were identical to Experiment 1.

[3] Yoked designs are important because they help dissociate the effects of selection from the effects of seeing better data.

Participants Participant recruitment and inclusionary/exclusionary criteria were identical to those of Experiment 1 (No, 3 HITs). Approximately 44 HITs were posted for each condition for total of 176 paid HITs.

Design and procedure Procedures were identical to those of Experiment 1. Participants were randomly assigned to one the two between-subjects conditions: Active-Receptive (AR) vs. Receptive-Active (RA). In the AR condition, participants completed a block of active learning then proceeded to a block of passive learning. In the RA condition, the order of the blocks was flipped.

Results and Discussion

Intercept is the mean of the means (or the grand mean) of all the groups. These data are unbalanced. Active better than passive. Information integration worse than rule-based.

Experiment 3

Experiment 3 is a conceptual replication of the order effect findings using a novel paradigm where participants learn a higher dimensional concept.

Methods

Stimuli

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)	-0.07	0.09	-0.7	0.46
x	1.98	0.09	23.3	0.00

General Discussion

- Recap findings
  - Active learning advantage in a direct replication
  - 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:

Acknowledgements

We are grateful to Doug Markant and Todd Gureckis for sharing the details and code from the original experiment. We thank the members of the Language and Cognition Lab for their helpful feedback on this project. This work was supported by a National Science Foundation Graduate Research Fellowship to KM.

References

Castro, R. M., Kalish, C., Nowak, R., Qian, R., Rogers, T., & Zhu, X. (2009). Human active learning. In *Advances in neural information processing systems* (pp. 241–248).

Grabinger, R. S., & Dunlap, J. C. (1995). Rich environments for active learning: A definition. *Research in Learning Technology*, *3*(2).

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning a cognitive and computational perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Kachergis, G., Yu, C., & Shiffrin, R. M. (2013). Actively learning object names across ambiguous situations. *Topics in Cognitive Science*, *5*(1), 200–213.

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.

Settles, B. (2012). Active learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, *6*(1), 1–114.

Westermann, K., & Rummel, N. (2012). Delaying instruction: Evidence from a study in a university relearning setting. *Instructional Science*, *40*(4), 673–689.

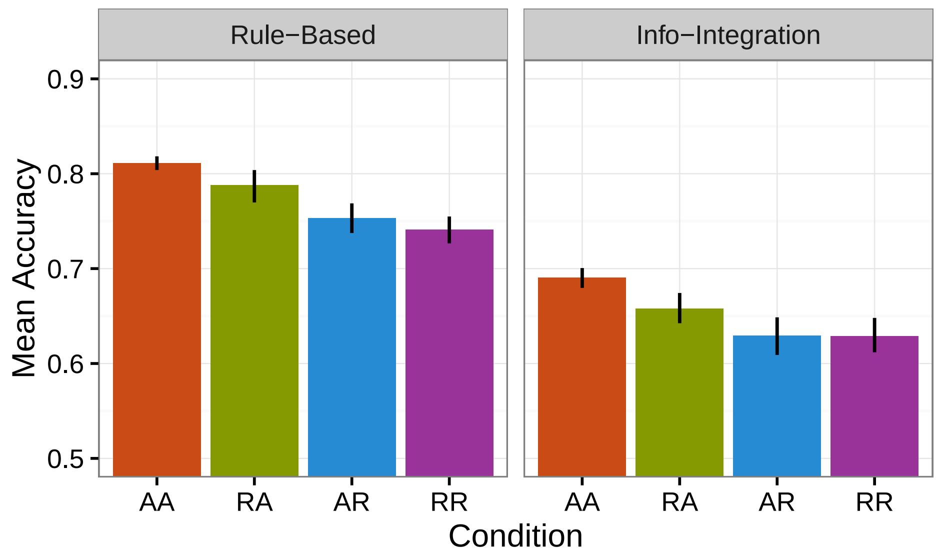

Figure 5: The left panel shows accuracy performance across both blocks for the different sequence of active/passive training. The right panel shows overall accuracy performance plotted with the active-active and receptive-receptive data from Experiment 1.