COGSCI'19

Creativity+Cognition+Computation

24 - 27 JULY 2019 MONTREAL, CANADA

Invited Speakers

Elizabeth Churchill | Mary Lou Maher | Takeshi Okada

Co-Chairs

Ashok Goel | Colleen Seifert | Christian Freksa

The Synergy of Passive and Active Learning Modes in Adaptive Perceptual Learning

Everett Mettler (mettler@ucla.edu)[1]
Austin S. Phillips (asphillips@ucla.edu)[1]
Christine M. Massey (cmassey@psych.ucla.edu)[1]
Timothy Burke (mizerai@ucla.edu)[1]
Patrick Garrigan (pgarriga@sju.edu)[2]
Philip J. Kellman (kellman@cognet.ucla.edu)[1]

[1]Department of Psychology, University of California, Los Angeles
Los Angeles, CA 90095 USA
[2]Department of Psychology, St. Joseph's University
Philadelphia, PA 19131 USA

Abstract

Adaptive learning systems that generate spacing intervals based on learner performance enhance learning efficiency and retention (Mettler, Massey & Kellman, 2016). Recent research in factual learning suggests that initial blocks of passive trials, where learners observe correct answers without overtly responding, produce greater learning than passive or active trials alone (Mettler, Massey, Burke, Garrigan & Kellman, 2018). Here we tested whether this passive + active advantage generalizes beyond factual learning to perceptual learning. Participants studied and classified images of butterfly genera using either: 1) *Passive Only* presentations, 2) *Passive Initial Blocks* followed by active, adaptive scheduling, 3) *Passive Initial Category Exemplar* followed by active, adaptive scheduling, or 4) *Active Only* learning. We found an advantage for combinations of active and passive presentations over *Passive Only* or *Active Only* presentations. Passive trials presented in initial blocks showed the best performance, paralleling earlier findings in factual learning. Combining active and passive learning produces greater learning gains than either alone, and these effects occur for diverse forms of learning, including perceptual learning.

Keywords: adaptive learning; perceptual learning; spacing effect; memory; active learning; passive learning

Introduction

The well-known *spacing effect* is a boost in long-term retention that results when recurrent learning episodes are spaced across gaps in time (Carpenter, 2017; Cepeda, Pashler, Vul, Wixted & Rohrer, 2006; Delaney, Verkoeijen & Spirgel, 2010). Spacing effects apply to a wide variety of learning domains and learners, and also influence diverse learning modes such as perceptual learning (Mettler & Kellman, 2014).

Recent research has shown that spacing effects can be enhanced by dynamically adjusting the size of spacing intervals during a learning session using an adaptive algorithm, Adaptive Response-Time-based Scheduling (ARTS; Mettler, Massey & Kellman, 2011; Mettler, Massey & Kellman, 2016). In ARTS, spacing delays are updated to match changes in learning strength as learning progresses for individual learners and items. Learning strength can be reliably estimated from response time (RT), with slower response times indicating retrieval difficulty and correspondingly lower learning strengths (Pyc & Rawson, 2009; Benjamin & Bjork, 1996; Karpicke & Bauernschmidt, 2011). ARTS updates the spacing among items in real time, by tracking the underlying learning strengths using an individual's accuracy and RT for learning items or for categories, producing highly efficient learning (Mettler, Massey & Kellman, 2011, 2016). In perceptual learning and other category learning domains, the same adaptive learning approach is applied to categories, such that learning strength for each category influences the priority of a learning trial involving a new exemplar of that category. Such adaptive spacing, and the interleaving of exemplars of different categories, also produces strong learning benefits relative to other arrangements (Mettler & Kellman, 2014).

Achieving the benefits of adaptive spacing requires interactive learning trials from which performance data are obtained. Recent work, however, suggests that the benefits of adaptive spacing may be further enhanced by combining active trials with passive presentations during learning. In a study investigating the learning of geography facts, Mettler, Massey, Burke, Garrigan & Kellman (2018) compared delayed retention rates following passive learning, active learning, and combinations of passive and active learning. Combinations of passive and active learning resulted in better performance than active learning alone. Passive presentations alone fared worst. In addition, the specific manner of combining passive and active modes mattered: learning which began with multiple blocks of passive trials followed by active, adaptive learning resulted in the best performance.

In the current study, we investigated whether the same learning advantages for passive combined with active learning might exist for perceptual learning (PL), which presumably rests on different mechanisms (changes in information selection and encoding vs. explicit storage of memory items). For factual information, spacing was manipulated among individual factual items. Here spacing was manipulated among categories of perceptual stimuli, but with each re-presentation of a category, a new exemplar was shown. Some earlier work suggested that combining passive and active modes might benefit PL (Thai, Krasne & Kellman, 2015); however, no work has explored different modes of combining active and passive trials.

Why might including some passive learning trials among active learning trials result in better PL than active trials alone? One benefit of passive trials may be to prevent the negative cognitive and motivational consequences of asking learners to generate answers in initial interactive learning trials - similar to the hypothesized benefits of initial passive trials in factual learning. Specific to PL, passive trials might focus attention on some characteristics of categories, and active trials might complement this learning by targeting other characteristics. For example, Carvalho & Goldstone (2015) suggested that passive trials can increase attention to commonalities between members of the same category when certain between-category and within-category similarity relations hold, but that active trials provide greater benefits to learning when the inverse similarity relations hold. Combining passive and active trials could be a strategy then to increase overall learning due to the complementary strengths of active and passive presentations in the learning of categories that possess a variety of internal structures. In the current study, we systematically compared learning schedules that included passive and active trials alone, and two different combinations of passive and active trials. We analyzed subsequent retention of perceptual classification after a delay, and we examined whether passive and active training was affected by internal category structures such as between and within-category similarity.

We compared four conditions: a) *Passive Only* presentations of learning items, b) *Passive Initial Blocks* followed by active, adaptive scheduling, c) *Passive Initial Category Exemplar* followed by active, adaptive scheduling for each category introduced, and d) *Active Only* learning with no passive presentations. We hypothesized that introductory presentations of passive trials, followed by active learning would fare the best, however, the effect of passive learning might be better if passive trials were limited to single presentations rather than blocks.

Method

Participants One hundred twenty undergraduate psychology students participated to partially fulfill course requirements.

Materials 12 categories (genera) of butterflies (lepidoptera) were used, where each genus contained images of 9 exemplars. On each learning trial, an image of one category exemplar was presented on the left side of the screen. In Active trials, the 12 possible category name responses were shown in a two-column list organized alphabetically on the right side of the screen. In Passive trials, only the correct category label was shown and the alternate category names were omitted.

Figure 1: Images of 2 butterfly genera with 3 exemplars from each genus. Danaus (top) and Neptis (bottom).

Design A 4x3x2x2 mixed factorial design was used. There were four between-subject passive/active conditions (*Passive Only, Passive Initial Block, Passive Initial Category Exemplar,* and *Active Only*). A pretest/posttest design consisted of three test phases (Pretest, Immediate posttest, and 1 week delayed posttest). In addition there was a within-subject factor of Familiarity (Familiar vs Unfamiliar); that is, at each test, each category was tested twice with both new and previously seen exemplars. Finally, there was a between-subject factor of Assessment List, such that the familiar and unfamiliar exemplars for each category were randomly selected differently for each of the two lists.

Procedure Participants completed two sessions separated by one week. The initial session consisted of a pretest, training phase and immediate posttest. The second session consisted of a delayed posttest only. In all tests and training, participants were shown a genus exemplar and were asked to identify the matching genus name from a list of all 12 category names. No feedback was provided. Tests consisted of two presentations of each genus: one presentation was a 'familiar' exemplar shown during training, and the other exemplar was an 'unfamiliar' exemplar withheld from training. There were two assessment lists and each participant was randomly assigned one of the versions. Each participant saw the same test version, and thus the same familiar and unfamiliar exemplars for each category, across pre, post and delayed tests.

In the *Passive Only* condition, butterflies were presented in 12 blocks of 12 passive trials. Each category appeared once per block, in random order, and a random exemplar from the category was chosen for each presentation. In the *Passive Initial Blocks* condition, participants first completed 2 blocks of passive trials, with blocks having the same structure as the *Passive Only* condition, followed by adaptive scheduling. In the *Passive Initial Category Exemplar* condition, the first presentation of each category was a passive trial followed by a fixed spacing interval of two intervening trials, so that the correct response was not still in working memory. All trials in this condition that did not involve the first presentation of a category were adaptively scheduled. In the *Active Only* condition, all trials were adaptively scheduled.

The ARTS algorithm determined the adaptive scheduling for active trials. After every response, ARTS calculates a priority score for each learning item and compares scores across items to determine which item will be presented next. Equation 1 shows the priority score calculation.

$$P_i = a(N_i - D)[b(1 - \alpha_i)\, Log(RT_i / r) + \alpha_i W] \qquad (1)$$

Detailed description of the ARTS algorithm can be found in Mettler, Massey & Kellman (2011, 2016). ARTS parameters were the following: the enforced delay D was set to 2 trials, the incorrect penalty W was set to 20, parameters a, b, r were set to 0.1, 1.1, and 1.7 respectively, and the timeout was 30 seconds.

Learning for each category continued until 5 out of the last 6 presentations were correctly answered with all correct response times less than 7 seconds. Learning criteria, adopted from previous studies, included both speed and accuracy, where speedy responses also ensured that final presentations were widely spaced.

Participants were assigned to Condition using a pretest balancing algorithm (similar to a procedure called Minimization; Pocock & Simon, 1975; Mettler et al., 2018). The condition balancing algorithm was constrained so that, across conditions, the largest difference in number of assigned participants never exceeded one. There were exactly 30 participants in each of the 4 conditions.

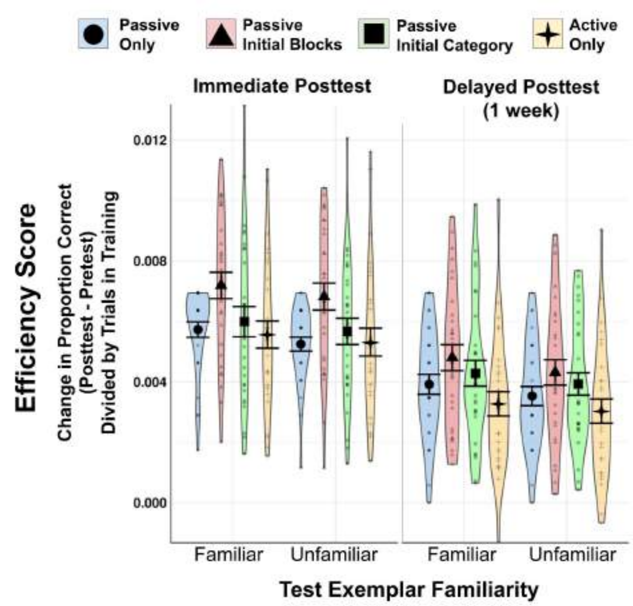

Figure 2: Learning Efficiency in Immediate and Delayed Posttest by Test Item Familiarity. (Violin plot shows mean, +/- 1 standard error of the mean, density estimate and individual data points).

Dependent Measures and Data Analysis

Because all adaptive conditions used learning to criterion, our primary measure was learning *efficiency*, defined as accuracy gain from pretest to posttest divided by the number of trials invested in learning. Efficiency gives a way of measuring learning that incorporates both variations in posttest performance, and variations in the number of learning trials required to reach the learning criteria. It may be thought of as a rate measure, indicating performance improvement per trial. The number of passive trials was determined based on pilot work to be roughly equal to the number of trials needed to reach mastery in active conditions. In the two conditions combining passive and active trials, all trials were included in trial and efficiency calculations.

In addition to efficiency we measured change in accuracy and reaction time. All measures were assessed using standard parametric statistics, such as ANOVA. Because we sought to compare differences across learning conditions, we conducted planned comparisons between pairs of conditions. All statistical tests were two-tailed, with a 95% confidence level, all effect sizes d are Cohen's d, and all error bars in graphs show +/- 1 standard error of the mean.

Results

Pretests A 4x2x2 ANOVA on Condition, Assessment List and Familiarity showed no significant main effect of Condition ($F(3,112)=0.213$, $p=.887$, $\eta_p^2=.006$), Assessment List ($F(1,112)=0.457$, $p=.500$, $\eta_p^2=.004$) or Familiarity ($F(1,112)=2.395$, $p=.125$, $\eta_p^2=.021$).

Efficiency *Efficiency*, defined as posttest accuracy gain from pretest divided by learning trials to criterion, is shown in Figure 2 for each of the posttests, the 4 learning conditions and for familiar vs. unfamiliar test items. The *Passive Initial Blocks* condition appeared to have higher efficiency at immediate posttest and highest numerical efficiency at delayed posttest. A 4x2x2x2 mixed factorial ANOVA on Passive/Active Scheduling Condition, Test Phase (Immediate vs. Delayed Posttest), Item Familiarity (Test exemplar seen vs. withheld in training) and Assessment List (1 vs 2) showed a significant main effect of Condition ($F(3,112)=2.921$, $p=.037$, $\eta_p^2=.073$) a significant main effect of Test Phase ($F(1,112)=277.127$, $p<.001$, $\eta_p^2=.712$), a significant main effect of Familiarity ($F(1,112)=17.832$, $p<.001$, $\eta_p^2=.137$), and no significant main effect of Assessment List ($F(1,112)=0.018$, $p=.893$, $\eta_p^2<.001$). Interactions were not significant ($ps>.127$) but there was a marginally significant interaction between Condition and Phase ($F(3,112)=2.197$, $p=.092$, $\eta_p^2=.056$) and Assessment List and Familiarity ($F(1,112)=3.391$, $p=.068$, $\eta_p^2=.029$).

The marginally significant interaction between Condition and Test appears to be driven by the clear superiority of

Passive Initial Blocks at immediate test that is less pronounced at delayed test. Paired comparisons revealed significant differences between conditions at immediate test (*Passive Only* vs. *Passive Initial Block*, t(58)=3.12, p=.003, d=0.84; *Passive Initial Blocks* vs. *Active Only* , t(58)=2.53, p=.014, d=0.65), and a marginally significant difference between *Passive Initial Blocks* vs. *Passive Initial Category* (t(58)=1.868, p=.067, d=0.48). Other comparisons did not reach significance (ps >.51). Paired comparisons at delayed posttest showed significant differences between *Passive Initial Blocks* and *Active Only* (t(58)=2.514, p=.015, d=0.65). There was a marginally significant difference between *Passive Initial Category* and *Active Only* (t(58)=1.74, p=.088, d=0.45). The remaining comparisons did not reach significance (ps > .105). Between immediate and delayed posttests, all pairwise comparisons were significant (p<.05) except for between *Active Only* at immediate test and *Passive Initial Blocks* at delayed posttest (t(58)=1.47, p=.147, d=0.38).

Trials in training Mean trials to reach learning criteria or the end of the session are shown in Figure 3. A 3x2 mixed factorial ANOVA was conducted on Condition and Assessment List. The *Passive Only* condition was removed from the ANOVA and paired comparisons due to its fixed (preset) number of trials. There was a significant effect of condition (F(2,84)=3.448, p=.036, η_p^2=.076). Paired comparisons showed significant differences between *Passive Initial Blocks* and *Passive Initial Category* (t(58)=2.068, p=.043, d=0.554) and between *Passive Initial Blocks* and *Active Only* (t(58)=2.707, p=.009, d=0.732), but not between *Passive Initial Category* and *Active Only* (t(58)=0.623, p=.536, d=0.161). One sample t-tests were used to compare each Active condition against the *Passive Only* condition mean of 144 trials. There was a significant difference for *Active Only* (t(29)=2.69, p=.012) and a

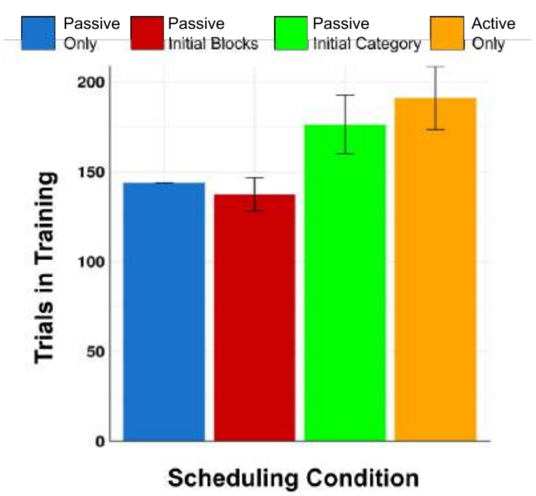

Figure 3: Trials in training session by 4 scheduling conditions.

marginally significant difference for *Passive Initial Category* (t(29)=1.97, p=.057), but no significant difference for *Passive Initial Blocks* (t(29)=0.70, p=.49).

Learning Analytics

In order to explore the reasons why performance was highest for *Passive Initial Blocks* conditions and lower for *Active Only*, we explored trial-by-trial data during learning. In prior work with learning of factual items we determined that initial blocks of passive items significantly reduced the severity of certain deleterious trial sequences. Specifically, the incidence of errors followed by correct responses (dubbed 0,1 sequences) across conditions, and these sequences followed by another error (0,1,0 sequences), were reduced in conditions that included initial passive blocks, relative to the other active conditions.

We examined 0,1 trial sequences during learning across the three adaptive scheduling conditions. First, the incidence of 0,1 sequences was highest in the *Active Only* condition and lowest in the *Passive Initial Blocks* condition, even when adjusting for the first few trials where there are necessarily errors in the *Active Only* condition due to initial guessing. The frequency of 0,1 instances across the three conditions and for groups of initial trials are shown in Figure 4. Trials 4+ are most instructive, showing that *Passive Initial Blocks* had the fewest occurrences of 0,1 among the three conditions. A 3 way ANOVA run on Condition for Trials 4+, found a significant effect of condition (F(2,87)=5.23, p=.007, η_p^2=.107) and paired comparisons showed significant differences between *Passive Initial Blocks* and *Passive Initial Category* (t(58)=2.52, p=.014, d=0.66), *Passive Initial Blocks* and *Active Only* (t(58)=3.15, p=.003, d=0.82), but not between *Passive Initial Category* and *Active Only* (t(58)=0.65, p=.519, d=0.17).

We also examined accuracy following 0,1 sequences. Again, the first 3 trials were removed to equate conditions with respect to number of prior presentations. Figure 5

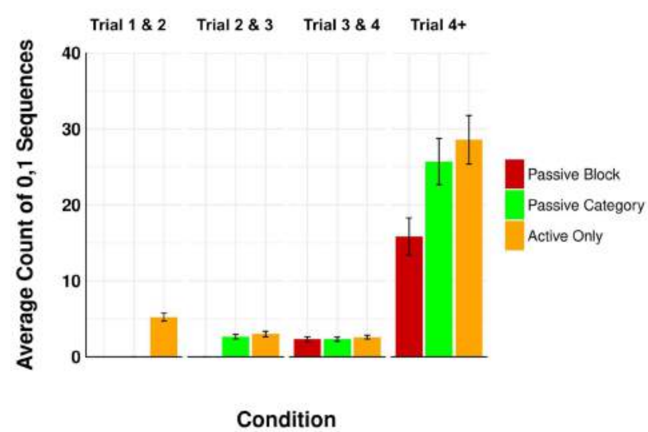

Figure 4: Frequency of 0,1 sequences by condition and by trial in learning session.

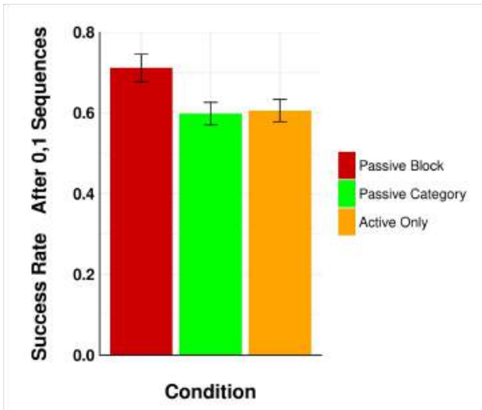

Figure 5: Success rate after 0,1 sequences, corrected for initial guessing (beginning at trial 3 for all conditions).

shows accuracy following 0,1 sequences. A 3 way ANOVA on success rate after 0,1 sequences found a significant effect of Condition ($F(2,87)=4.34$, $p=.016$, $\eta_p^2=.091$). Paired comparisons showed significant differences between *Passive Initial Blocks* and *Passive Initial Category* ($t(58)=2.71$, $p=.009$, $d=0.7$), *Passive Initial Blocks* and *Active Only* ($t(58)=2.22$, $p=.030$, $d=0.57$), but not between *Passive Initial Category* and *Active Only* ($t(58)=0.62$, $p=.539$, $d=0.16$).

Within-category and between-category similarity relations Since prior research indicates the importance of within and between category similarity for benefits from passive or active trial scheduling, we examined passive only and active only learning efficiency as a function of between and within-category similarity. Similarity relations were determined by subject ratings of each category, first for between-category relations and then again, separately for within-category relations. All 12 categories were rated on a 3 point similarity scale for between-category similarity with 3 being highest and 1 lowest. Subject ratings were averaged for each category and categories were divided into 1 of 3 between-category similarity groups based on the tertile of their averaged rating. The same procedure was repeated for within-category ratings. Thus, within and between-category similarities were estimated independently. Posttest efficiencies were compared for two scheduling conditions, *Passive Only* and *Active Only*, across the three levels of within and between-category similarity.

Average efficiency differences, plotted separately for each within and between-category similarity group are shown in Figure 6. Two 2x2x3 ANOVAs were conducted, each with training schedule (*Passive Only, Active Only*), and Test phases (Immediate vs. Delayed posttest) as factors. One ANOVA also included within-category similarity as a factor, and the other also included between-category similarity as a factor. The ANOVA with within-category similarity as a factor showed no significant effect of Condition ($F(1,176)=1.63$, $p=.204$, $\eta_p^2=.009$), a significant

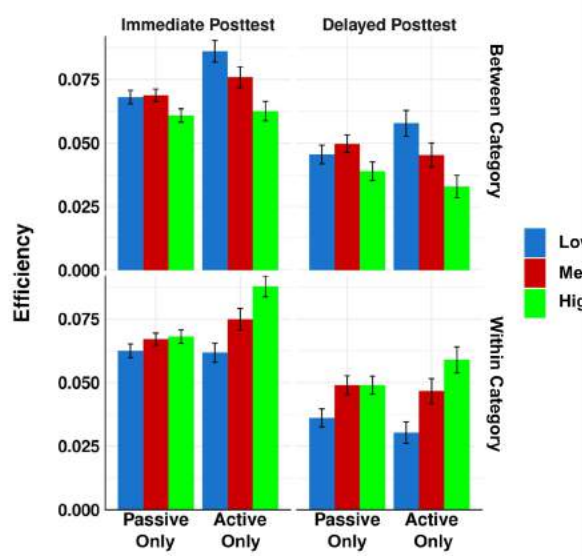

Figure 6: Efficiency for between-category similarity groups (top) and within-category similarity groups (bottom) for low, medium and high similarity, by *Passive Only* and *Active Only* conditions at immediate and delayed posttests.

effect of within-category similarity ($F(1,176)=15.92$, $p<.001$, $\eta_p^2=.083$) and an effect of Test phase ($F(1,176)=223.67$, $p<.001$, $\eta_p^2=.56$). There were two significant interactions, Condition with Similarity group ($F(1,176)=3.92$, $p=.049$, $\eta_p^2=.022$) and Condition with Test phase ($F(1,176)=6.04$, $p=.015$, $\eta_p^2=.033$).

The most instructive interaction, Condition x Similarity group, indicated that similarity relations modulated the effect of Condition. Paired comparisons indicated that differences in efficiency varied more across levels of similarity in the *Active* condition than in the *Passive* condition. Specifically, the greater the within group similarity, the greater the efficiency in the *Active Only* condition. In the *Active Only* condition, there were significant differences in learning efficiency between low within similarity and high within similarity ($t(238)=4.96$, $p<.001$, $d=0.64$), between medium within similarity and low within similarity ($t(238)=2.7$, $p=.007$, $d=0.35$), and between high within similarity and medium within similarity ($t(238)=2.13$, $p=.034$, $d=0.28$). In the *Passive Only* condition, the difference between low within similarity and medium within similarity was significant ($t(238)=2.226$, $p=.027$, $d=0.287$) and the difference between low within similarity and high within similarity was significant ($t(238)=2.388$, $p=.018$, $d=0.308$), but the difference between medium within similarity and high within similarity was not significant ($t(238)=0.136$, $p=.892$, $d=0.018$).

The ANOVA with between-category similarity included as a factor showed no significant effect of condition ($F(1,176)=1.73$, $p=.190$, $\eta_p^2=.01$), a significant effect of between-category similarity ($F(1,176)=12.34$, $p<.001$, $\eta_p^2=0.066$), and a significant effect of Test phase

($F(1,176)=236.08$, $p<.001$, $\eta_p^2=0.573$). There was one significant interaction, between Condition and Test phase ($F(1,176)=6.38$, $p=.012$, $\eta_p^2=.035$), and a marginally significant interaction of Condition x Similarity group ($F(1,176)=3.79$, $p=.053$, $\eta_p^2=0.021$). As with within-category relations, paired comparisons showed that between-category similarity modulated the effects of Condition. In the *Active Only* condition, there were significant differences in efficiency between high between-category similarity and low between-category similarity ($t(238)=4.26$, $p<.001$, $d=0.55$), between medium and low similarity ($t(238)=2.36$, $p=.019$, $d=0.31$), and a marginally significant difference between high similarity and medium similarity ($t(238)=1.94$, $p=.054$, $d=0.25$). In the *Passive Only* condition, there was one significant difference between the medium and low similarity conditions ($t(238)=2.43$, $p=.016$, $d=0.31$) and a marginally significant difference between high and low similarity conditions ($t(238)=1.76$, $p=.080$, $d=0.23$).

Discussion

The synergy of passive and active presentations in perceptual learning was remarkably similar to that found previously in factual learning (Mettler et al., 2018). In both studies the following conditions were compared: 1) passive presentations alone, 2) initial blocks of passive presentations followed by active, adaptive learning, 3) initial passive presentations for each category that unlocked later adaptive learning, or 4) active, adaptive learning alone with no passive presentations. In this experiment the learning consisted of perceptual learning across multiple categories (butterfly genera). We found an advantage for combining passive with active presentations such that initial passive presentations, especially when grouped into initial blocks of passive trials in which all learning categories were interleaved, resulted in the greatest efficiency of category classification at posttest. Learning persisted across time as measured by a 1-week delayed test. In addition, the benefits of passive and active combined schedules generalized to unfamiliar category exemplars that had not been shown during the learning phase. Unsurprisingly, combinations of passive and active presentations were better than passive presentations alone. More important, combinations of passive and active trials were much more effective than active, adaptive presentations alone: a few initial presentations (1 or 2 presentations for each category) was enough to generate learning gains beyond those found with purely active, adaptive schedules. Passive block and adaptive trial synergy was so strong that the *Passive Initial Blocks* condition at delayed test was not statistically different from the *Active Only* condition performance at immediate test. Further analysis of trial-by-trial learning data including sequences of correctness supported the idea that the benefits of a *Passive Initial Blocks* condition extended well into the active, adaptive learning component.

In addition to these results, we investigated the effect of category similarity on passive + active synergies. The overall apparent lower performance in the *Active Only* condition compared to the *Passive Only* condition appears to hold only when similarity between categories is high or when within-category similarity is low. For lower levels of between-category similarity and for greater levels of within-category similarity, *Active Only* conditions fared better than passive presentations. These effects of category similarity are somewhat different than results by Carvalho & Goldstone (2015) who showed that passive presentations result in slightly worse performance when categories have relatively low within-category similarity.[1] Unlike Carvalho & Goldstone, we found that active presentations had the greatest benefit when between-category similarity was lowest and when within-category similarity was highest. By one interpretation, high similarity between categories implies greater difficulty of making category discriminations. Thus active presentations are best when categories are more discriminable from each other. A natural interpretation of the effects in adaptive category sequencing is that with low within-category similarity (and potentially with high between-category similarity) assessments of category learning strength gotten from each active trial by the adaptive algorithm are less reliable when category instances are more diverse, making learning less efficient.

To conclude, we investigated the contribution of including passive presentations with interactive, adaptive learning. We found that combining passive with active presentations such that an initial passive phase (passive blocks) in which passive presentations were given for all learning categories resulted in the greatest retention performance at posttest. In perceptual learning, the effects of passive presentations appear to temper differences in category structure across variable within and between-category relations, and to enhance active, adaptive learning with fewer errors throughout the learning session.

Adaptive learning frameworks that leverage learner performance data to arrange spacing and sequencing in learning substantially improve learning across diverse types of learning, including perceptual learning. These benefits are further enhanced by combining active responding with passive modes of learning at the start of learning. The present results may help lead to a theoretical understanding of the mechanisms that enable passive + active synergies across different types of learning, and they contribute to a practical understanding of how to optimize these effects in instructional technology.

[1] It should be noted that blocking in Carvalho and Goldstone referred to massing exemplars from the same category, whereas in our Passive Initial Blocks condition all of the passive trials were presented as a block, but we interleaved exemplars from every category consistently in all conditions.

Acknowledgements

We gratefully acknowledge support for this work from the National Science Foundation under Grant No. DRL-1644916. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. Reder (Ed.), *Implicit memory and metacognition* (pp. 309–338). Hillsdale, NJ: Erlbaum.

Carpenter, S. K. (2017). Spacing effects on learning and memory, in: J.T. Wixted (Ed.), *Cognitive Psychology of Memory*. Academic Press, Oxford.

Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review*, 22, 281–288.

Cepeda, N.J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132, 354–380.

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: a deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation*, 53, 63–147.

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 1250-1257.

Mettler, E., & Kellman, P. J. (2014). Adaptive response-time-based category sequencing in perceptual learning. *Vision Research*, 99, 111–123.

Mettler, E., Massey, C. M., Burke, T., Garrigan, P. & Kellman, P. J. (2018). Enhancing adaptive learning through strategic scheduling of passive and active learning modes. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), Proceedings of the 40th Annual Conference of the Cognitive Science Society (pp. 768-773). Austin, TX: Cognitive Science Society.

Mettler, E., Massey, C. M., & Kellman, P. J. (2011). Improving adaptive learning technology through the use of response times. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), In Proceedings of the 33rd Annual Conference of the Cognitive Science Society (pp. 2532-2537). Austin, TX: Cognitive Science Society.

Mettler, E., Massey, C. M. & Kellman, P. J. (2016). A comparison of adaptive and fixed schedules of practice. *Journal of Experimental Psychology: General*, 145(7): 897- 917.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language*, 60, 437–447.

Thai, K. P., Krasne, S., & Kellman, P. J. (2015). Adaptive perceptual learning in electrocardiography: The synergy of passive and active classification. In D. C. Noell, R. Dole, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), Proceedings of the 37th Annual Meeting of the Cognitive Science Society (pp. 2350-2355). Austin, TX: Cognitive Science Society.