COGSCI 2018

changing / minds

40th annual cognitive science society meeting
madison wisconsin usa july 25-28

learning.development.evolution.emergence.plasticity.aging.disorder.cross-culture.experience.persuasion.propaganda.discovery.context.control
emergence.plasticity.aging.disorder.cross-culture.experience.persuasion.propaganda.discovery.context.control.learning.development.evolution
disorder.cross-culture.experience.persuasion.propaganda.discovery.context.control.learning.development.evolution.emergence.plasticity.aging

KEYNOTE SPEAKERS

Susan Gelman
University of Michigan

Matthew Botvinick
DeepMind

Michael Kearns
University of Pennsylvania

INVITED SYMPOSIA

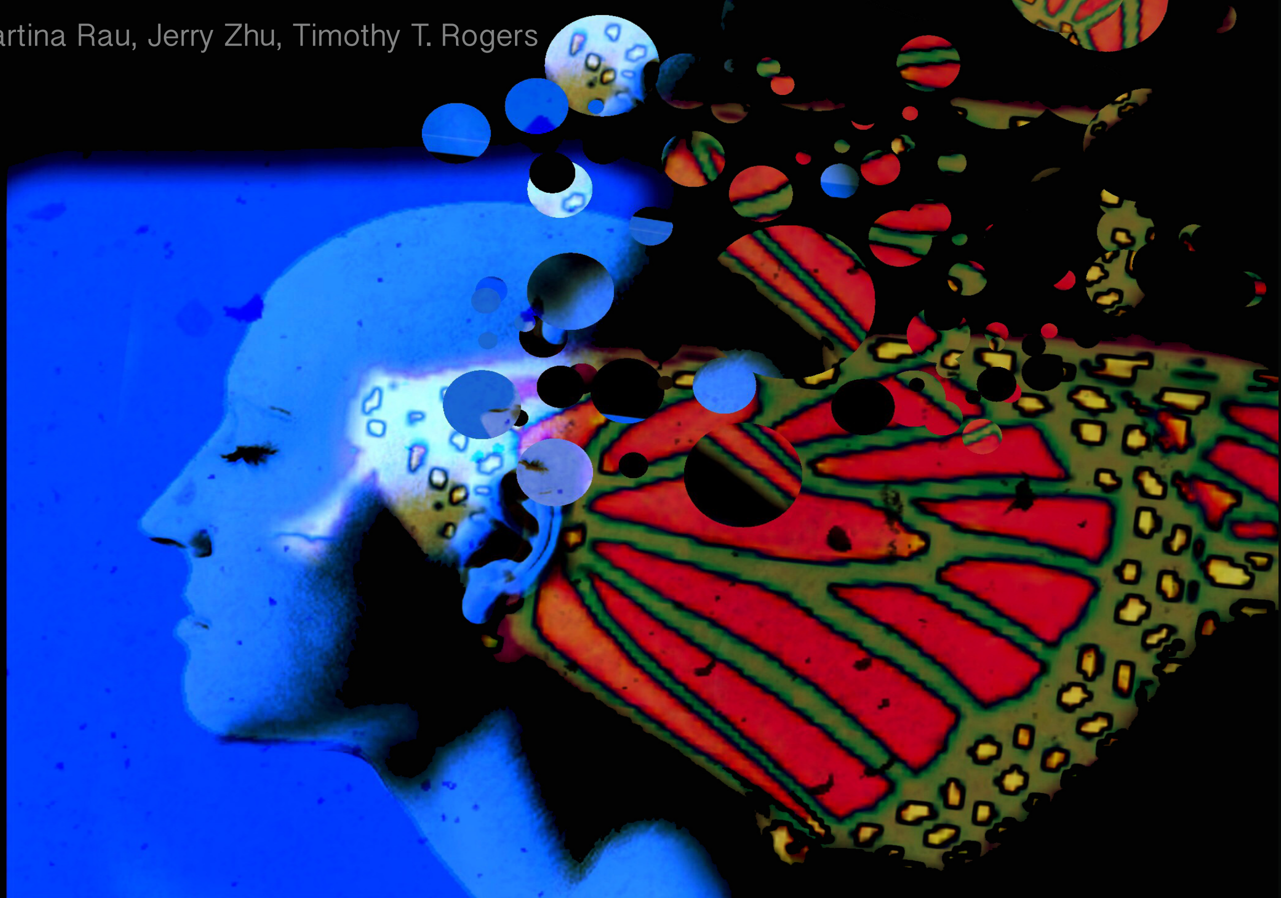
changing education
big data goes to school

changing society
persuasion, propaganda, politics

changing science
Rumelhart prize symposium

ORGANIZERS

Charles Kalish, Martina Rau, Jerry Zhu, Timothy T. Rogers



Enhancing Adaptive Learning through Strategic Scheduling of Passive and Active Learning Modes

Everett Mettler (mettler@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

Recent work suggests that optimal spacing in learning requires adaptive procedures (Mettler, Massey & Kellman, 2016). Here, we studied how adaptive techniques might be further enhanced by combining active and passive learning modes. Participants learned geography facts that were scheduled using the ARTS (Adaptive Reaction-Time-based Scheduling) system under four conditions involving passive and/or active trials. Conditions included: a) Passive Only presentations of learning items, b) Passive Initial Blocks followed by active adaptive scheduling, c) Passive Initial Items followed by active adaptive scheduling for each item introduced, or d) Active Only learning with no passive presentations. We found an advantage for combinations of active and passive presentation (by blocks or items) over Passive Only or Active Only presentation. Passive trials presented in blocks at the beginning of learning showed best performance. We discuss possible explanations for these differences and suggest principles underlying optimal combinations of active and passive modes in adaptive learning.

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

Introduction

A large body of research has demonstrated the importance of the *spacing effect*: a boost in long-term retention that results when recurrent learning episodes are spaced across gaps in time (Cepeda, Pashler, Vul, Wixted & Rohrer, 2006; Delaney, Verkoeijen & Spirgel, 2010). Spacing effects have been shown to apply to a wide variety of learning domains and learners, are robust to changes in learning conditions and test durations, and have been recommended by panels of experts seeking to improve pedagogical practice and learning outcomes in classrooms (Dunlosky, Rawson, Marsh, Nathan & Willingham, 2013; Pashler, Bain, Bottge, Graesser, Koedinger, McDaniel & Metcalfe, 2007).

Recent research has shown that spacing effects can be especially 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 & Kellman, 2014; Mettler, Massey & Kellman, 2016). In ARTS, spacing delays are updated to match changes in learning strength, as learning progresses, for individual learners and items. Evidence indicates that response time (RT) is a useful indicator of retrieval difficulty, and thus of an item's current learning strength (Pyc & Rawson, 2009; Benjamin & Bjork, 1996; Karpicke & Bauernschmidt, 2011). ARTS updates spacing by tracking underlying learning strength using an individual's accuracy and RT for learning items. It produces highly efficient learning and compares favorably with classic adaptive approaches that do not use RT information (Atkinson, 1972; Mettler, Massey & Kellman, 2011).

Adaptive learning techniques offer new answers to persistent questions about the mechanisms underlying spacing effects and optimal spacing. A longstanding issue has been whether expanding schedules of practice, where spacing interval sizes increase across subsequent presentations, are better than equal interval schedules (e.g., Bjork & Allen, 1970; Karpicke & Roediger, 2007). Mettler, Massey & Kellman (2016) suggested that the question of whether predetermined equal or expanding schedules produce better learning has no ultimate answer. In experiments with factual learning, they first showed that adaptive scheduling outperformed both equal and expanding schedules. They then used a yoking procedure to show that the advantages of adaptive learning did not derive from the particular types or distributions of spacing intervals, but depended crucially on interactions between learners and items. Their data are consistent with a *successful effort hypothesis*: the ideal time for a new learning trial for an item is the longest interval at which the learner can still respond correctly. This interval depends on underlying learning strength for each item for a given learner, which may be affected by many variables and changes throughout the course of learning. Many of these effects are difficult or impossible to predict from *a priori* models; thus, predetermined spacing arrangements cannot be optimal.

Optimal spacing may be possible with adaptive systems that use ongoing performance measures (specifically, tracking accuracy and speed of response for individual items) to gauge learning strength and determine spacing (Mettler, Massey & Kellman, 2016).

Interactive learning provides information to guide spacing, but it may have certain drawbacks, especially at the start of learning. With interactive "test" trials from the start, the learner must initially guess and receive feedback. There are at least two potential drawbacks in such a situation, one cognitive and one motivational. The cognitive issue is that wrong answers generated by guessing may persist later in learning. Motivationally, being tested on material one has not learned is deflating, perhaps more so for some learner groups (e.g., middle school students learning mathematics). In this paper, we seek to enhance adaptive learning systems by considering possible modifications of initial learning trials. We examined two ways of including initial passive learning trials in which learners are introduced to correct information but not required to respond. We compared these combinations to conditions in which learners received either active or passive trials alone throughout learning.

Passive presentations in learning have been studied previously in a variety of contexts. Recent memory studies have compared the effectiveness of study vs. test trials during learning (Roediger & Karpicke, 2006) and other studies have explored the role of passive learning when the learning task consists of category learning (Carvalho & Goldstone, 2015), relational concept learning (McDonald & Frank, 2016), hypothesis testing (Markant & Gureckis, 2014), and specific domains such as physical simulation (Bramley, Gerstenberg & Tenenbaum, 2016). Research examining the difference between passive study and active testing has identified a powerful effect of testing trials over study trials, such that testing trials result in greater long-term retention (Carpenter, Pashler & Cepeda, 2009; Halamish & Bjork, 2011; Roediger & Karpicke, 2006).

As in studies of memory, research in category learning and perceptual learning have found that active presentations generally contribute more to learning, however, passive presentations sometimes play a beneficial role. Carvalho and Goldstone (2015) found that passive presentation with massed trials of exemplars from the same category produced better category learning than interleaved passive trials and also better than massed active learning trials. Active, spaced trials, however, were as good as passive, massed trials in one experiment, and clearly better in another. Most relevant to the present study, Thai, Krasne & Kellman (2015) studied learning efficiency in a perceptual-adaptive learning module (PALM) for training interpretation of electrocardiograms and found that an initial block of passive trials, followed by adaptive category learning, enhanced the efficiency of learning relative to passive only or active only conditions.

In the present work, we examined the use of passive modes of learning in the early phases of learning for factual material. One goal was to understand whether and how the spacing benefits of adaptive learning might be affected or enhanced by initial passive experience. The second goal was to compare two approaches to the integration of passive and active modes. We approached these goals through experiments and learning analytics aimed at revealing specific interactions between passive presentations and spacing dynamics.

One method for use of introductory passive trials is to present one or more blocks of passive trials, with each block containing one presentation of each learning item (as used by Thai, Krasne & Kellman, 2015). In the *Passive Initial Block* condition in the present study, two initial trial blocks of passive trials were followed by active, adaptive learning in ARTS. We also tested a second approach. In the *Passive Initial Item* condition, the initial presentation of each learning item was a passive trial. That trial served to "unlock" that item for subsequent adaptive learning. The passive trial was treated by ARTS in the same way as an active error trial: the item was given a high priority for reappearance as an active learning trial, recurring on average two trials later. The dynamics of this "unlocking" procedure resulted, after a few trials, in a mix of passive and active trials – new items presented passively on their initial appearance intermixed with active learning items whose reappearance depended on learner performance. This *Passive Initial Item* condition had the potential advantage of mixing modes of learning and sustaining interest. We also tested *Active Only* and *Passive Only* conditions as controls. We hypothesized that one or both combinations of passive with active, adaptive learning would produce enhancements in learning efficiency.

Method

Participants

Participants were 120 undergraduate psychology students who participated for course credit.

Materials

All materials were presented on a computer within a web-based application. Participants were to identify 24 African countries on a 500 x 800-pixel map of Africa presented on the left side of the screen. Responses were indicated by mouse clicking on a two-column list of African countries alphabetically organized by column then row, presented on the right side of the screen.

Design

Each participant was assigned to one of four scheduling conditions consisting of passive trials only, active trials only, or one of two variations combining passive and active trials. Passive trials consisted of a four-second presentation of the target country highlighted on a map, accompanied only by the correct country label. After four seconds a continue button appeared, which participants clicked to advance to the next trial. On active trials, the target country was highlighted on a map and the learner selected the name from the full set of country labels, then correct / incorrect and response time feedback was provided while the correct country label was highlighted. All active trials were adaptively scheduled according to ARTS (see below).

In the *Passive Only* condition, countries were presented in 10 blocks of 24 passive trials. Each country appeared once per block, in random order. In the *Passive Initial Block* condition, participants first completed 2 blocks of passive trials, as in the *Passive Only* condition, followed by adaptive scheduling. In the *Passive Initial Item* condition, the first presentation of each country was a passive trial followed by a fixed spacing interval of at least one intervening trial, 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 country 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 previous work (Mettler, Massey & Kellman, 2011, 2016). In this study, the enforced delay D was set to 1 trial, the incorrect penalty W was set to 20 and parameters a, b, r were set to 0.1, 1.1, and 3.0 respectively. In general, the priority score system results in an item reappearing soon after an error; however, an enforced delay prohibits reappearance while the answer still resides in working memory. Spacing for correct responses is an inverse function of log response time, such that faster responses (indicating higher learning strength) produce lower priority scores (resulting in longer recurrence intervals).

Procedure

Participants were shown a map of Africa featuring an outlined country and were asked to select its name from a list of all country names by clicking with a computer mouse. Participants attended two sessions, separated by 1 week. In the first session, participants initially took a pretest on all 24 items, presented in random order, without feedback. The pretest was followed by a learning phase in one of the four experimental conditions, which took up the majority of the first session. Accuracy feedback followed every trial, and, after every 10 trials, participants received block feedback indicating their average response accuracy and speed for the previous block of 10 trials and every previous block up to 10 prior blocks. After the learning phase, participants took a posttest that was identical to the pretest. (Due to a programming error, 17 participants in the Passive Initial Item condition received their first posttest trial as a passive presentation, with accuracy marked as incorrect. To adjust for this error the posttest score for those 17 participants was scored out of 23 items.) The entire first session took no more than 1 hour for each participant. After the posttest, participants were instructed not to study or reflect on the information learned and to return in 1 week to complete a delayed posttest, which was identical to the immediate posttest. No feedback was given on either posttest.

Participants were automatically assigned to a scheduling condition using an algorithm designed to simultaneously randomly assign participants to condition as well as balance average pretest scores across conditions. Participants were assigned to the condition for which their pretest score would minimize the differences between average pretest scores across conditions. The algorithm did not allow any condition to get more than one participant ahead of any other condition. The balancing algorithm also acted as a filter to screen out participants deemed unsuitable for the study, due to pretest scores with either accuracy > 35% (n= 4 participants) or average response times < 1 second, (n = 1) participant).

To ensure reduction of noise and comparability of conditions, any learner in adaptive conditions who did not successfully retire all 24 items was removed from our primary analysis (n = 5 participants). Learning criteria were enforced for the three conditions that had adaptive scheduling. Learning criteria encompassed both speed and accuracy and ensured that items were well learned before removal from the active learning set. Specifically, items were retired if correctly responded to on 4 out of the last 4 presentations with RT less than 7 seconds, similar to prior studies (Mettler, Massey & Kellman 2016). There were no learning criteria for the *Passive Only* condition and the learning session ended after exactly 240 trials for every participant in that condition.

Dependent Measures and Data Analysis

Because we 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. We also examined accuracy gain and trials to criterion separately. 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 counted for trial and efficiency calculations. 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 and Discussion

Primary Results

Efficiency. Learning efficiency results are shown in Figure 1, which plots efficiency at posttest and delayed posttest separately for the 4 learning conditions. The *Passive Only* condition showed lower efficiency in both posttests. The *Passive Initial Block* condition appeared somewhat better than the other conditions. *Passive Initial Item* and *Active Only* conditions showed little difference. These observations were confirmed by the analyses. A 4x2 mixed factorial ANOVA on Passive/Active Scheduling Condition and Test phase (Immediate vs. Delayed) was conducted on the Efficiency scores. There was a significant main effect of

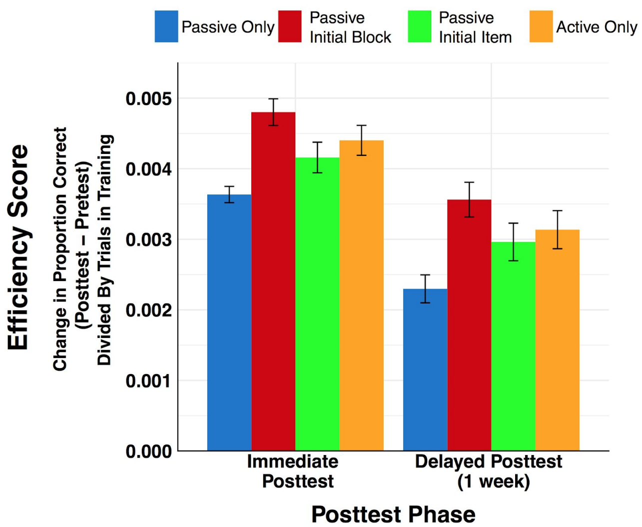


Figure 1. Learning Efficiency in Posttest and Delayed Posttest.

Condition ($F(3,116) = 5.95$, $p=.001$, $\eta_p^2=.133$) a significant main effect of Test phase ($F(1,116)=320.9$, $p<.001$, $\eta_p^2=.73$), and no significant Condition by Test phase interaction ($F(3,116)=0.17$, $p=.914$, $\eta_p^2=.004$). Paired comparisons revealed reliable differences between conditions at immediate test (*Passive Only* vs. *Passive Initial Block*, $t(58)=5.28$, $p<.001$, $D=1.40$; *Passive Only* vs. *Passive Initial Item*, $t(58)=2.14$, $p=.036$, $D=0.58$; *Passive Initial Block* vs. *Passive Initial Item*, $t(58)=2.24$, $p=.029$, $D=0.58$; *Passive Only* vs. *Active Only*, $t(58)=3.17$, $p=.002$, $D=0.86$). *Passive Initial Block* vs. *Active Only* did not reach significance ($t(58)=1.40$, $p=.165$, $D=0.364$). Other differences between conditions at immediate test were not significant (ps>.165). Paired comparisons at delayed posttest showed significant differences between *Passive Only* and *Passive Initial Block*, $t(58)=4.01$, $p<.001$, $D=1.04$, and *Passive Only* vs. *Active Only*, $t(58)=2.51$, $p=.015$, $D=0.65$). There was a marginally significant difference between *Passive Only* and *Passive Initial Item* ($t(58)=2.00$, $p=.050$, $D=0.52$). The remaining comparisons did not reach significance: *Passive Initial Block* vs. *Active Only* ($t(58)=1.17$, $p=.247$, $D=0.30$), *Passive Initial Block* vs. *Passive Initial Item* ($t(58)=1.65$, $p=.102$, $D=0.429$), and *Passive Initial Item vs. Active Only* ($t(58)=0.46$, $p=.65$, $D=0.119$).

Accuracy Change. Accuracy change score measured posttest minus pretest accuracy at immediate and delayed posttest. Pretest accuracies were comparable across conditions (Passive Only, M: 0.07, SD: 0.04; Passive Initial Block, M: 0.07, SD: 0.07; Passive Initial Item, M: 0.07, SD: 0.08; Active Only: M:0.06, SD: 0.06; $F(3,116)=0.80$, $p=.99$). A 4x2 mixed factorial ANOVA on Passive/Active Scheduling Condition and Test phase (Immediate vs. Delayed) was conducted on the Accuracy change score. The ANOVA found no significant main effect of Condition ($F(3,116)=1.87$, $p=.138$, $\eta_p^2=.046$), a significant main effect of test ($F(1,116)=269.33$, $p<.001$, $\eta_p^2=.693$) and no significant interaction of Condition by Test ($F(3,116)=0.609$, $p=.611$, $\eta_p^2=.015$). None of the paired comparisons between conditions at immediate or delayed test showed reliable differences, except for a marginally significant difference at delayed posttest between *Passive Only* vs. *Passive Initial Block* ($t(58)=1.97$, $p=.053$, $D=0.53$).

Trials. Trials taken to reach learning criteria or the end of the session are shown in Figure 2. A between subjects ANOVA was conducted on Trials comparing the conditions. There was a reliable effect of condition ($F(3,116)=3.94$, $p=.010$, $\eta_p^2=.092$). Paired comparisons showed a significant difference between *Passive Only* and *Passive Initial Block* ($t(58)=5.50$, $p<.001$, $D=2.01$) and between *Passive Initial Block* and *Passive Initial Item* ($t(58)=2.49$, $p=.016$, $D=0.66$). There was a marginally significant difference between *Passive Initial Block* and *Active Only* ($t(58)=1.75$, $p=.09$, $D=0.46$). No other comparisons were significant (ps>.125).

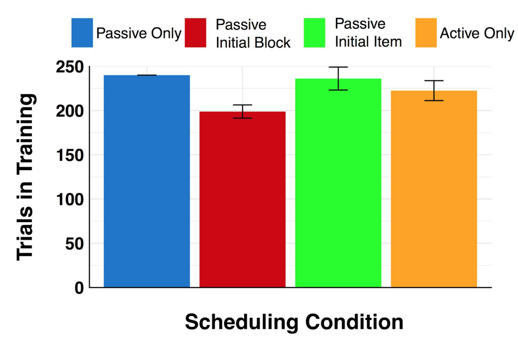


Figure 2. Learning trials by Scheduling Condition.

Learning Analytics

We carried out more detailed analyses to understand the impact initial trials had on learning performance. As mentioned, a grounding hypothesis of the ARTS system, and a key to optimal spacing, is the *successful effort hypothesis*: extending recurrence intervals improves learning benefits, so long as the learner can still respond successfully (c.f., Pyc & Rawson, 2009). This hypothesis predicts that effective adaptive learning regimens should minimize what we call "snaps" – occasions where the learner fails to answer correctly as the retention interval is stretched. From data in purely active, adaptive learning, we noticed an interesting pattern: After missing an item and successfully answering it on its next appearance, learners had a relatively low proportion of accurate responding on the next appearance of that item. Moreover, this pattern of a sequential error and correct response (designated '0,1') occurred often: 1673 times in the data we examined. This frequency includes many occasions at the start of learning where the first trial in active learning requires a guess and produces an error.

In the present study, we examined the frequency of 0,1 sequences and accuracy of subsequent responses, in an attempt to understand the benefits of initial passive trials. We discovered that the proportion of success after 0,1 trials varied with condition, with higher success rates in the combined *passive-active* conditions than in *Active Only.* Statistical analyses revealed a significant difference between the *Passive Initial Block* (M=0.68, SD=0.17) and *Active Only* (M=0.56 , SD=0.14) conditions (t(52) = 3.16, p=.003, D=0.86) and a marginally significant difference between the *Passive Initial Item* (M=0.63, SD=0.18) and *Active Only* conditions: t(57)=1.86, p=.068, D=0.486). The difference between *Passive Initial Block* and *Passive Initial Item* conditions was not significant (t(51)=1.163, p=.25).

A second finding was that the use of initial passive trials vastly reduced the frequency of 0,1 sequences: Compared to the 1673 occurrences in the *Active Only* condition, there were 1158 in the *Passive Initial Item* condition, and 631 in the *Passive Initial Block* condition. We also examined occurrences relative to the initial active trial in each condition, shown in Figure 3. Trials 1 & 2 reflect initial active trials in the *Active Only* condition — likely due to initial guessing. In trials 2 & 3, 0,1's appear in the Passive Initial Item condition at a lower rate than Active Only. In trials 3 & 4, the occurrences of 0,1 sequences are lowest in the Passive Initial Block condition. The rate of occurrences remains lowest in the Passive Initial Block condition for all remaining trials in the learning session suggesting that the advantages of initial passive presentations extend into the learning session.

General Discussion

We tested combinations of passive and active learning in adaptive learning to see whether and how the addition of passive presentations early in learning can enhance learning efficiency. Consistent with prior work, all three conditions with active adaptive scheduling were better than passive only presentations. A combination of active and passive trials did enhance learning: presenting two blocks of passive presentations prior to all active learning was (numerically) the fastest method of learning, and exceeded other

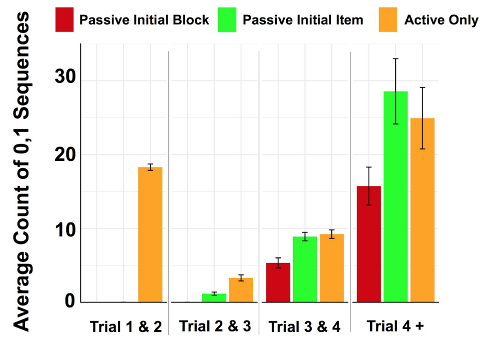


Figure 3. '0,1' Sequences by Condition and Trial.

conditions in terms of the rate of deleterious trial sequences.

Learning analytics were used to probe the source of this effect. We found that in active learning, 0,1 sequences for an item (an error, followed by a correct response) occurred frequently and were often followed by an error on the next occurrence of that item. This pattern is suboptimal, as efficiency benefits in adaptive learning may derive largely from keeping learning as nearly errorless as possible (Mettler, Massey & Kellman, 2016).

Our analyses indicated three findings regarding 0,1 sequences. First, the need for initial guessing in a purely active, adaptive condition contributes to the high frequency of 0,1 sequences, and this guessing can be reduced or eliminated by initial passive presentations. Second, initial passive trials reduced the number of 0,1 sequences far more than would be expected by eliminating initial guessing alone. This was especially true in the best condition, *Passive Initial Block*, in which 0,1 sequences were almost 1/2 as frequent as in the Active Only condition, not including initial guessing. The third finding was that the success rates after 0,1 sequences were higher in both combined conditions than in the *Active Only* condition.

Initial passive learning reduces later errors in learning, leading to more efficient mastery in an adaptive framework. What accounts for the benefits? We mentioned two possibilities at the outset. Besides producing errors on initial trials, early incorrect guesses may persist in learning and lead to later errors. A related idea is that unduly strong associations between cues and incorrect guesses can impede later recall (Knight, Ball, Brewer, DeWitt, & Marsh, 2012). Another is that motivation may be affected by having to guess. We add here a third possibility: Some work in problem solving indicates that having to solve problems before getting a foothold in learning may add cognitive load that impedes the learning itself (Paas & Merrienboer, 1994). Passive exposure may cushion learners from deleterious features of initial active learning, such as cognitive effort. It is also possible that adaptive spacing parameters, such as parameters that decide the general relationship between RTs and spacing interval size, may have positively interacted with learners who had received initial passive presentations. Passive presentations may provide learners with enough initial learning strength to deliver more accurate indications of learning strength to adaptive scheduling routines.

Our current findings are limited in several ways. Most prominently, we tested particular implementations of passive learning, including using two blocks in the *Passive Initial Block* condition. We do not know whether a single block (one presentation of each item) would suffice to attain benefits, or if more passive trials, or a different schedule of passive items, would be beneficial. Finally, it is possible that our findings are somewhat unique to the learning material used, although Thai, Krasne & Kellman (2015) found similar advantages of initial passive blocks in adaptive perceptual learning (classification of electrocardiograms).

Adaptive learning frameworks that leverage learner performance to arrange spacing and sequencing in learning provide substantial benefits to learning. The present results indicate that the benefits are further enhanced by combining active responding with passive modes of learning, especially at the start of learning. Fully understanding the mechanisms underlying these benefits, their range of applicability, and how to optimize them in adaptive learning, pose important questions for further research.

Acknowledgments

We gratefully acknowledge expert assistance from Tim Burke, Rachel Older, Kandice Tsoi and Lucy Mullin. This material is based on work supported by 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

Atkinson, R. C. (1972). Optimizing the learning of a second-language vocabulary. *Journal of Experimental Psychology*, 96, 124–129.

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.

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior, 9*, 567–572.

Bramley, N., Gerstenberg, T. & Tenenbaum, J. B. (2016). Natural science: Active learning in dynamic physical microworlds. *Proceedings of the 38th Annual Conference of the Cognitive Science Society.*

Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2009). Using tests to enhance 8th grade students' retention of U. S. history facts. *Applied Cognitive Psychology,* 23, 760-771.

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.

Dunlosky, J., Rawson, K.A., Marsh, E.J., Nathan, M.J., & Willingham, D.T. (2013) Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology, *Psychological Science in the Public Interest*, 14(1), pp.4-58.

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 801-812.

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.

Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 704–719.

Knight, J. B., Ball, B. H., Brewer, G. A., DeWitt, M. R., & Marsh, R. L. (2012). Testing unsuccessfully: A specification of the underlying mechanisms supporting its influence on retention. *Journal of Memory and Language*, *66*(4), 731-746.

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.

MacDonald, K., & Frank, M.C. (2016). When does passive learning improve the effectiveness of active learning? *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*.

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., & Kellman, P. J. (2011). Improving adaptive learning technology through the use of response-times. In *Proceedings of the 33rd Annual Meeting of the 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.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing Instruction and Study to Improve Student Learning* (NCER 2007-2004). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Paas, F., & Van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Education & Psychology*, 86, 122–133.

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.

Roediger, H.L. & Karpicke, J.D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, 17, 249-255.

Thai, K. P., Krasne, S., & Kellman, P. J. (2015). Adaptive perceptual learning in electrocardiography: The synergy of passive and active classification. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*.