Category structure modulates interleaving and blocking advantage in inductive category acquisition

Paulo F. Carvalho (pcarvalh@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E 10th St
Bloomington, IN 47405 USA

Robert L. Goldstone (rgoldsto@indiana.edu)
Department of Psychological and Brain Sciences, 1101 E 10th St
Bloomington, IN 47405 USA

Abstract

Research in inductive category learning has demonstrated that interleaving exemplars of categories results in better performance than presenting each category in a separate block. Two experiments indicate that the advantage of interleaved over blocked presentation is modulated by the structure of the categories being presented. More specifically, interleaved presentation results in better performance for categories with high within- and between-category similarity while blocked presentation results in better performance for categories with low within- and between-category similarity. This interaction is predicted by accounts in which blocking promotes discovery of features shared by the members of a category whereas interleaving promotes discovery of features that discriminate between categories.

Keywords: category learning; order effects; interleaving;

Introduction

How to present information so that learning and memory are optimized is an important issue in teaching and training contexts (Rohrer & Pashler, 2010). It has long been demonstrated that spacing repeated presentations of the same information results in better memory than repeating the same information at a single occasion, even when time and number of presentations are equated (Ebbinghaus, 1885). This memory phenomenon, known as the "Spacing Effect," is a highly robust finding (Delaney, Verkoeijen, & Spirgel, 2010; Proctor, 1980) that has been shown both in experimental situations with words and pictures and more applied situations such as flashcard studying (Kornell, 2009). Although demonstrating the critical importance of carefully considering how to present information, the importance of maximizing memory for specific concepts or problems might not be as relevant as learning general concepts. Indeed, in educational contexts, often times inferring what characterizes or defines a concept or problem is more relevant than memorizing a single instance of that concept or fact. In this sense, a more interesting question might be to know whether the way instances are presented influences inductive learning and subsequent generalization of the acquired knowledge.

The question of how to present information in order to optimize category learning and generalization has been raised before and several proposals have been put forward. Some of these proposals are related to the categories being taught. For example, Elio and Anderson (1984; see also, Sandhofer & Doumas, 2008) have proposed that learning should start with low variability items and later introduce items with greater variability. Another proposal is that items that present the same generalization should be presented close together in temporal sequence (e.g., Elio & Anderson, 1981; Mathy & Feldman, 2009).

More recently, researchers have proposed interleaving items from the categories being taught (i.e., presented in alternating fashion), rather than grouping items together from the same category. The advantage of alternating categories has been observed in different kinds of concepts such as artists' styles (Kornell & Bjork, 2008), bird species (Kornell, Castel, Eich, & Bjork, 2010; Wahlheim, Dunlosky, & Jacoby, 2011), novel category learning in children (Vlach, Sandhofer, & Kornell, 2008) and also mathematical operations in primary school students (Taylor & Rohrer, 2010).

Initial accounts of this advantage for interleaving related the interleaved presentation with spacing of exemplars. However, Kang and Pashler, (2012) used a procedure similar to the one used by Kornell and Bjork (2008), but with added presentation conditions. In one experiment, the authors compared categorization performance in a generalization test preceded by one of four conditions: (1) blocked, (2) interleaved, (3) blocking in which every presentation of a painting was followed by an unrelated filler task (Temporal Spaced Condition), and (4) when all exemplars from the same painter were presented simultaneously (Blocked Simultaneous Condition). The results showed that only the interleaved condition resulted in better performance than the blocked condition, thus providing evidence that greater temporal spacing of presentations is not the critical factor in the interleaved advantage. The authors argue that the real advantage of interleaving might be a result of the greater opportunity to contrast and compare examples, making the differences between the artists' styles more salient (see also, Goldstone, 2003; Goldstone & Steyvers, 2001).

Further evidence for this proposal was provided in a second experiment in which the interleaved and blocked conditions were compared to a simultaneous presentation of two paintings by different artists. This latter condition resulted in similar performance to interleaved presentation and better performance than the blocked condition.

Interleaving examples from different categories, thus, may improve inductive learning because it promotes discrimination between the categories. Rapid alternation between examples of different categories, as well as simultaneous presentation of multiple categories, might lead to enhanced attention to the features that discriminate between the categories and differentiation of the dimensions on which the categories vary (Goldstone & Steyvers, 2001; Kang & Pashler, 2012; Nosofsky, 1986). However, there are also situations in which these conditions are not ideal and previous studies have also demonstrated an advantage of blocked presentation (Goldstone, 1996; Kurtz & Hovland, 1956).

Goldstone (1996) proposed that category learning might be difficult for two different reasons: high between-category similarity or low within-category similarity. High between-category similarity refers to category structures in which both categories share most of their features, making discriminating the categories a matter of finding subtle differences between exemplars (as is the case of distinguishing between alligators versus crocodiles, for example). Low within-category similarity, on the other hand, refers to category structures in which the exemplars of the same category share very few features (as is the case for the category "animal," for example).

Each one of these two kinds of categories requires different mechanisms for efficient category learning. In the case of high between-category similarity, one has to identify subtle differences between categories, which might be facilitated by frequent alternation. However, rapid alternation between two categories with low within-category similarity will not allow for the identification of the relevant properties characteristic of a category. In this case it might be more beneficial to block exemplars of each category separately so that the learner can identify the shared features among members of a category hidden within their diversity.

By this analysis, the best presentation schedule would depend upon the learning situation at hand. In this paper we approach this question directly by manipulating only the characteristics of the stimuli presented during learning. We aim to demonstrate that interleaving or blocking can both be beneficial for inductive category learning. In Experiment 1 participants were taught three categories with high within- and between-category similarity, in both an interleaved and blocked schedule. In Experiment 2 we employ the same learning procedure but using three categories that possess low within- and between- category similarity. Performance during category learning and in a subsequent generalization task are measured.

Experiment 1: High Similarity

In this experiment we aim to replicate previous results showing an advantage for interleaved presentation of categories for category learning. Participants were presented with 3 categories (either blocked or interleaved). These 3 categories had high within- and between- category similarity. In this sense, every stimulus from the same category shared several features, some category relevant and others not, while they also shared almost every feature with every stimulus from the other categories. We believe this is a situation close to previous work on sequencing effects in category learning (e.g., Kornell & Bjork, 2008) and literature on perceptual discrimination learning showing an interleaving advantage (Mitchell, Kadib, Nash, Lavis, & Hall, 2008; Mundy, Honey, & Dwyer, 2007). Under this situation, rapid alternation between categories will lead to the direction of attention to the diagnostic features because identifying what is changing from category to category is relatively easy given the high overall similarity.

Method

Participants Forty-four Indiana University undergraduate students participated in this experiment in return for partial course credit. All participants completed every condition. Fifteen participants did not reach the criterion of 34% correct responses during categorization learning and were excluded from further analyses.

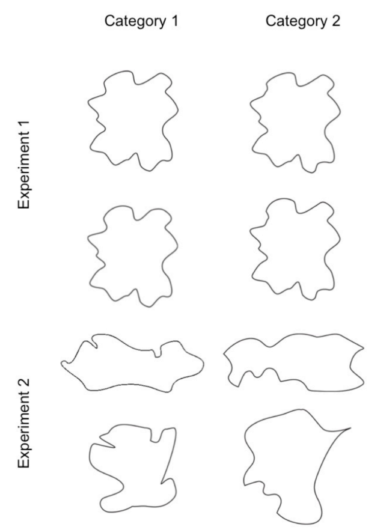

Fig. 1: Examples of stimuli used in Experiments 1 and 2.

Apparatus and stimuli In this and the subsequent experiment, blob figures were used as stimuli (see Fig. 1). All blobs were created by randomly generating curvilinear segments. The blobs used were very similar, sharing almost every feature. Variation within each category was exactly the same for all categories, so that a difference that could exist between two exemplars in Category 1 would also exist between 2 exemplars of Categories 2 and 3. As a cover story, participants were told that a recent expedition to Mars recovered several cells of alien organisms. Each cell could be categorized into one of 3 species based solely on its perceptual features. Stimuli were presented on a computer screen and participants responded using keys on the keyboard with a consistent mapping to the category assignment.

Design and procedure This experiment had two conditions (blocked category learning and interleaved category learning), manipulated within-subject. Each condition was composed by 2 phases. The first phase was a category learning task. During this task participants were presented with a stimulus in the center of the screen for 500 ms. After the blob was removed, the participant was asked to classify the blob into one of 3 species (Q, Y or P) by pressing the corresponding key on the keyboard. After a participant's response, the blob was presented again for 2000 ms together with the presentation of feedback on the accuracy of their response and the correct species of the blob (e.g, "CORRECT! This cell belongs to species Q" or "Sorry, that is INCORRECT! This cell belongs to species Y").

A 1000 ms inter-trial interval followed and then a new trial began. In the blocked condition, the categories presented alternated 25% of the time while in the interleaved condition they alternated 75% of the time. Thus, in the interleaved condition, the probability of a blob being followed by a blob of the same category was low, whereas for the blocked condition this probability was high. We used this probabilistic approach rather than creating purely interleaved or blocked conditions in order to diminish the possibility that participants noticed the pattern of alternation in responses, which would affect categorization accuracy. Furthermore, if a purely blocked condition had been used there would be no way to guarantee participants' attention to the task, as there would be no uncertainty as to the correct categorization. This approach has been used before in similar tasks with successful results (Carvalho & Goldstone, 2011; Goldstone, 1996).

The learning phase was composed of 4 blocks. Each block had 48 trials (8 exemplars of each category repeated 2 times each). After the 4th block of categorization learning a new set of instructions was presented on the screen and the second phase began. This second phase was a generalization task during which 48 stimuli were shown in random order – the 24 blobs participants studied during the learning task and 24 new stimuli. The new stimuli were generated in the same manner as training stimuli, with new instantiations of the unique features. Each stimulus was presented in the center of the screen for 500 ms, after which participants were asked to classify it into one of the species just learned. After a 1000 ms inter-trial interval, a new trial would begin. No feedback was provided during this phase.

The two learning conditions (blocked vs. interleaved) differed only in the frequency of category change and the species labels. In one of the conditions Q, Y and P labels/keys were used, while in the other, A, G and L were used, by random assignment. Which condition was presented first was counterbalanced across participants and the allocation of the stimuli to each category and condition was randomized across participants.

Results and Discussion

The top panel of Fig. 2 depicts the main results of the categorization task. A 2 x 4 repeated measures ANOVA with learning condition (blocked vs. interleaved) and learning block (1 vs. 2 vs. 3 vs. 4) as factors revealed main effects of learning block, $F(3,81) = 59.42$, $p < .00001$ and condition, $F(1,27) = 18.58$, $p = .0002$, but no interaction between the two variables, $F(3,81) = 1.43$, $p = .24$. Thus, there is an improvement in categorization accuracy during the task, regardless of presentation schedule, and blocked presentation results in the overall best accuracy rates.

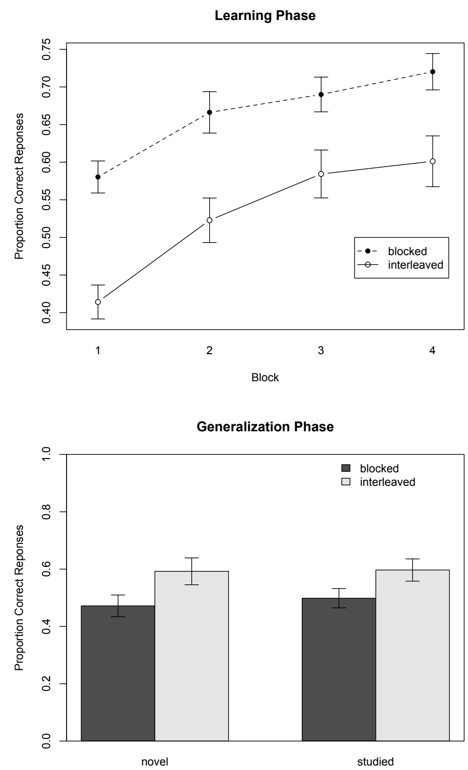

Fig. 2: Results from Experiment 1. The top graph represents accuracy during the learning phase for both conditions across the four blocks of trials. The bottom graph represents accuracy during the generalization phase for novel and studied stimuli for each condition. Error bars represent standard errors.

The results from the generalization task are depicted in the bottom panel of Fig. 2. A 2 x 2 repeated measures ANOVA with learning condition (blocked vs. interleaved) and stimulus presented (novel vs. studied) as factors, revealed a main effect of condition, $F\ (1,27) = 4.50$, $p = .04$, with higher accuracy for the interleaved condition, but no main effect of stimulus presented, $F\ (1,27) < 1$, or interaction between the two variables, $F\ (1,27) < 1$.

Overall these results replicate previous demonstrations that interleaving very similar categories during learning results in better memory for the presented items (as seen by the higher accuracy rates for the studied stimuli) but also better generalization to novel items. The finding that blocked presentation results in higher accuracy rates during learning has also been demonstrated before and has been explained by a relative ease in performing the training task (Taylor & Rohrer, 2010), which would make the interleaved condition a desirable difficulty in this kind of task (Kornell & Bjork, 2008).

The results are consistent with the hypothesis that interleaved presentation promotes the discovery of features that discriminate between alternative categories, which is important when categories are highly similar.

Experiment 2: Low Similarity

In this Experiment we analyze the effect of the schedule of presentation in a categorization situation different from that of Experiment 1. We tried to create a category learning task in which blocked presentation of the category exemplars would allow for the discovery of hard-to-find common features among dissimilar category members.

We employ the same procedure as in Experiment 1 but the 3 categories now have low within- and between-category similarity. In this sense, exemplars from the same category share only one feature (the category relevant feature) while differing in all other features. Moreover, exemplars from different categories are also highly dissimilar, sharing none of their features. Thus, in this condition, it is not expected that interleaving will have a beneficial effect on learning. Rapid alternation of categories would not allow for the identification of important features because all the features change from trial to trial. On the contrary, we expect blocked presentation to result in superior category learning and performance on a later generalization task, by promoting comparison of objects from the same category and thereby promoting discovery of their subtle commonality.

Method

Participants Thirty-two Indiana University undergraduate students participated in this experiment in return for partial course credit. All participants completed every condition and all reached the criterion of 34% correct responses or more during categorization learning.

Apparatus and stimuli The same kind of blob figures as in Experiment 1 were used. However, the blobs used in this experiment differed from every other blob in all the segments except the one that defined each one of the six categories (each one of the discriminating features). In this experiment, Blobs also differed in their overall shape, some being more circular and others more elongated. All other details were kept the same as in Experiment 1.

Design and procedure This experiment followed the same procedure as Experiment 1 with the exception of the stimuli used.

Results and Discussion

The graph in the top panel of Figure 3 depicts response accuracy during category learning as a function of both learning block and schedule of presentation.

A 2 x 4 repeated measures ANOVA with both learning condition (blocked vs. interleaved) and learning block (1 vs. 2 vs. 3 vs. 4) as repeated measures factors, revealed a main effect of learning block, $F(3, 93) = 169.74$, $p < .0001$, showing an overall improvement for both presentation conditions. Moreover, categorization is overall better for the blocked condition when compared to the interleaved condition, $F(1, 31) = 29.03$, $p < .0001$. Finally, a significant interaction between presentation condition and learning block was also found, $F(3, 93) = 7.06$, $p = .0002$, indicating that the difference in accuracy between the two conditions decreases as the categorization progresses.

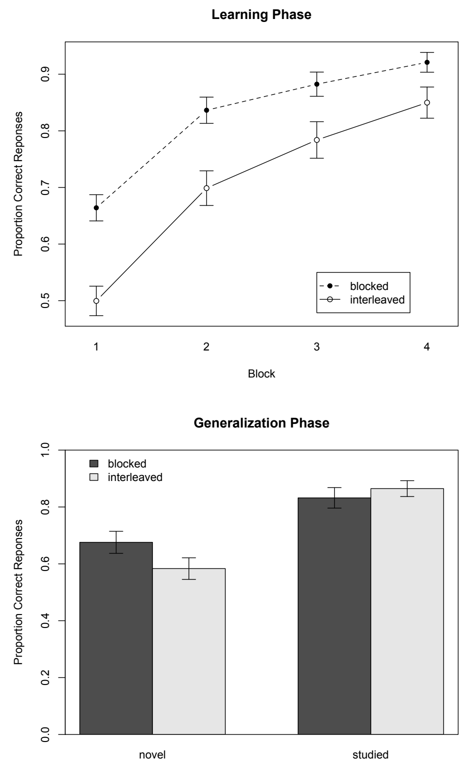

Fig. 3: Results from Experiment 2. The top graph shows accuracy during the learning phase for both conditions across the four blocks of trials. The bottom graph shows accuracy during the generalization phase for novel and studied stimuli for each condition.

The bottom panel of Fig. 3 shows accuracy performance in the generalization task as a function of stimulus presented (novel vs. studied) and presentation condition during learning (blocked vs. interleaved). A 2 x 2 repeated measures ANOVA with both presentation condition (blocked vs. interleaved) and stimulus presented (novel vs. studied) as factors revealed a main effect of stimulus, $F(1, 31) = 74.20$, $p < .00001$, with higher accuracy for the studied than novel stimuli. There was no significant main effect of presentation condition during learning, $F(1, 31) = 1.14$, $p = .30$. However, the interaction between the two variables was significant, $F(1, 31) = 14.61$, $p = .0006$. To further investigate this interaction, pairwise *t-test* comparisons were performed correcting for the error associated with multiple comparisons using Bonferroni correction (critical $\alpha = .025$). This analysis revealed that blocked presentation results in higher accuracy in categorizing novel stimuli, $t(31) = 2.47$, $p = .02$, but there is no difference between the two conditions for studied stimuli, $t(31) = 1.21$, $p = .24$.

Thus, the results of this experiment demonstrate an advantage of blocked presentation of category exemplars in accuracy during learning, although participants in the interleaved condition approach that performance by the last trial. Perhaps more interestingly, blocked presentation resulted in better performance when categorizing new stimuli when compared to interleaved presentation. However, for studied stimuli no difference was found.

These results directly challenge the proposal that interleaving categories during learning improves performance due to the increased spacing of exemplars. They are, however, consistent with our proposal that the interleaving advantage was due to greater opportunity for contrasting categories and discovery features that distinguish between categories. When finding feature differences between categories is easy, as in Experiment 2, then the more important task is to find features shared by members of the same category, a task promoted by blocking. This clearly demonstrates that the category structure has a modulating effect on the advantage of one sequencing over another.

General Discussion

Learning the characteristics of concepts and categories inductively takes place frequently. In learning contexts, it is also important to be able to generalize that knowledge to new exemplars. In this work we studied the interaction between object presentation during learning and the structure of the category being taught for inductive learning and generalization optimization.

Experiment 1 replicated previous results demonstrating that interleaving categories being taught results in better memory and generalization for those categories. Additionally, in Experiment 2 we presented results pointing to an advantage of blocking by category. In Experiment 2 generalization was better for categories that had their exemplars presented in adjacent temporal sequence during learning.

Critically, the only difference between these two experiments was the properties of the categories being taught. while in Experiment 1, the three categories were highly similar among each other and were constituted by exemplars that were also very similar, in Experiment 2 the similarity within and between the categories was low, resulting in categories that did not share any features and were composed of stimuli that shared only one feature with other objects in their same category.

A possible explanation to at least some of the results presented here could be that interleaving constitutes a "desirable difficulty" (Bjork, 1994). In both experiments, blocked presentation during the learning phase resulted in a higher proportion of correct responses. This could indicate that interleaved presentation involved a greater effort on the participants' part, maintaining their attention to the stimuli and leading to better performance in the subsequent task. Although this can, in fact, account for the results found in Experiment 1, that is not the case for Experiment 2. In that experiment, the condition resulting in greater accuracy during learning also resulted in better generalization.

In line with Goldstone (1996) we propose that interleaving categories allows for the identification of the features that are different between each category while blocked presentation promotes the identification of the features that are common among stimuli from the same category. This dichotomy is the result of the same principle: the opportunity to compare and contrast the properties of the categories, what will emphasize different features in different situations,

We further build on this proposal by suggesting a mechanistic account of how it takes place in a trial by trial basis. It has previously been demonstrated that during category learning participants take into account information from the previous few trials to decide whether a stimulus belongs to one category or another (Stewart & Brown, 2004; Stewart, Brown, & Chater, 2002). We further propose, that when two successive stimuli are similar and in different categories, participants' attention will be directed to the differences between exemplars by a comparison process. On the other hand, if stimuli are very dissimilar but in the same category, participants' attention will be directed towards the similarities between successive stimuli. This proposal can account for both the advantage of interleaving over blocking with high-similarity categories and the reversal with low-similarity categories.

Rapid alternation of categories allows participants to identify differences between categories, which will be particularly beneficial if those differences are hard to detect, as in the case of the stimuli used in Experiment 1 and the artists' styles or bird species used in previous studies. Infrequent alternation of categories, on the other hand, will allow participants to identify the commonalities within each category, which is particularly beneficial if categories are composed by members with high variability, like the ones used in Experiment 2.

In appreciating the benefits of blocking it is important to keep in mind that not all concept learning takes place by identifying discriminating features among categories. For example, sometimes it is possible to create an absolute characterization of a category in terms of its prevalent features, regardless of their discriminative values (Markman & Ross, 2003). Furthermore, in other situations, memorizing instances might be a highly useful strategy.

In sum, the results presented here show that there may not be a single answer to the question of how should an instructor sequence information so that the learner acquires the knowledge and is able to generalize it more efficiently. Best sequencing practices will depend on the nature of the categories being sequenced.

Acknowledgments

This research was supported in part by National Science Foundation REESE grant 0910218 and Department of

Education IES grant R305A1100060. PFC was also supported by a Fulbright Research Fellowship and Graduate Fellowship SFRH/BD/68554/2010 from the Portuguese Foundation for Science and Technology (FCT). The authors would like to thank Rachel Selonick for her help with data collection.

References

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing*. Cambridge, MA: MIT Press.

Carvalho, P. F., & Goldstone, R. L. (2011). Sequential similarity and comparison effects in category learning. In C. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Delaney, P. F., Verkoeijen, P. P. J. L., & 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*(10), 63–147.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. *City*. New York, NY: Teachers College, Columbia University.

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(6), 397-417.

Elio, R., & Anderson, J. R. (1984). The effects of information order and learning mode on schema abstraction. *Memory & cognition*, *12*(1), 20-30.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & cognition*, *24*(5), 608-28.

Goldstone, R. L. (2003). Learning to perceive while perceiving to learn. In R. Kimchi, M. Behrmann, & C. Olson (Eds.), *Perceptual organization in vision: Behavioral and neural perspectives*. New Jersey: Lawrence Erlbaum.

Goldstone, R. L., & Steyvers, M. (2001). The sensitization and differentiation of dimensions during category learning. *Journal of Experimental Psychology: General*, *130*(1), 116-139.

Kang, S. H. K., & Pashler, H. (2012). Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. *Applied Cognitive Psychology*, *26*, 97-103.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology*, *23*(9), 1297-1317.

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: is spacing the "enemy of induction"? *Psychological science*, *19*(6), 585-92.

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and aging*, *25*(2), 498-503.

Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of experimental psychology*, *51*(4), 239.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*(4), 592-613.

Mathy, F. & Feldman, J. (2009). A rule-based presentation order facilitates category learning. *Psychonomic Bulletin & Review, 16, 1050-1057.*

Mitchell, C. J., Kadib, R., Nash, S., Lavis, Y., & Hall, G. (2008). Analysis of the role of associative inhibition in perceptual learning by means of the same-different task. *Journal of experimental psychology. Animal behavior processes*, *34*(4), 475-85.

Mundy, M. E., Honey, R. C., & Dwyer, D. M. (2007). Simultaneous presentation of similar stimuli produces perceptual learning in human picture processing. *Journal of experimental psychology. Animal behavior processes*, *33*(2), 124-38.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of experimental psychology. General*, *115*(1), 39-61.

Proctor, R. W. (1980). The Influence of Intervening Tasks on the Spacing Effect for Frequency Judgments. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(3), 254-266.

Rohrer, D., & Pashler, H. (2010). Recent Research on Human Learning Challenges Conventional Instructional Strategies. *Educational Researcher*, *39*(5), 406-412.

Stewart, N., & Brown, G. D. A. (2004). Sequence effects in the categorization of tones varying in frequency. *Journal of experimental psychology. Learning, memory, and cognition*, *30*(2), 416-30.

Stewart, N., Brown, G. D. A., & Chater, N. (2002). Sequence effects in categorization of simple perceptual stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 3-11.

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology*, *24*(6), 837-848.

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163-7.

Wahlheim, C. N., Dunlosky, J., & Jacoby, L. L. (2011). Spacing enhances the learning of natural concepts: an investigation of mechanisms, metacognition, and aging. *Memory & cognition*.