Title
How to present exemplars of several categories? Interleave during active learning and block during passive learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 35(35)

ISSN
1069-7977

Authors
Carvalho, Paulo
Goldstone, Robert

Publication Date
2013

Peer reviewed

How to present exemplars of several categories? Interleave during active learning and block during passive learning

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

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

Abstract

Research on how information should be presented during inductive category learning has identified both interleaving of categories and blocking by category as beneficial for learning. Previous work suggests that this mixed evidence can be reconciled by taking into account within- and between-category similarity relations. In this paper we present a new moderating factor. One group of participants studied categories actively, either interleaved or blocked. Another group studied the same categories passively. Results from a subsequent generalization task show that active learning benefits from interleaved presentation while passive learning benefits from blocked presentation.

Keywords: interleaving; blocking; learning; comparison

Introduction

Can the method with which information is presented substantially affect learning? The answer seems to be "yes." Changing the way with which information is presented not only changes what is learned (Schyns, Goldstone, & Thibaut, 1998) but also how well it is learned (Goldstone, 1996). One example is the order in which instances are presented in a study session and the effect this has for inductive learning. Kornell and Bjork (2008) demonstrated that if participants are given study examples of paintings from several artists' interleaved, participants' later memory and generalization is substantially improved when compared to presenting each artist in a separate block.

The advantage of interleaving over blocking for inductive learning has been repeatedly shown in recent years. Interleaving of categories has been shown to improve learning of naturalistic materials for both young and older adults (Kornell, Castel, Eich, & Bjork, 2010; Wahlheim, Dunlosky, & Jacoby, 2011), as well as flashcard learning (Kornell, 2009). It has also been demonstrated in children (Vlach, Sandhofer, & Kornell, 2008; Vlach & Sandhofer, 2012).

Notwithstanding the clear benefit of interleaving in some situations, there have also been demonstrations of the advantage of blocking for category learning. For example, Goldstone (1996) presented participants with complex images consisting of 20 line segments. There were two conditions: frequent alternation of categories (interleaving) and infrequent alternation (blocking). The results showed that participants were better at learning the categories in the infrequent alternation condition. The author associates this advantage with the relative difficulty in finding the common features shared by members of each category (for further evidence of blocking advantages using different kinds of tasks and stimuli see Kurtz & Hovland, 1956; Whitman & Garner, 1963).

Given this mixed evidence about the best way to sequentially present information for optimal learning, an important question is: what conditions yield an advantage for interleaving compared to blocking?

This question has received some attention in recent years. For instance, Carvalho and Goldstone (2012) showed that when studying low similarity categories, blocked presentation resulted in improved subsequent generalization performance. This pattern was reversed for high similarity categories (for similar results with category discriminability, see Zulkiply & Burt, 2013).

Carvalho and Goldstone (2012; see also Goldstone, 1996) have proposed that interleaving categories allows participants to identify the features that *distinguish* between the categories, while blocked presentation promotes the identification of features that are *characteristic* among stimuli within a single category. This dichotomy is the result of the same principle: the selective emphasis of categorization-relevant features during comparison of sequentially presented objects.

In the case of interleaved presentation, differences between objects belonging to different categories will be emphasized while for blocked presentation, similarities among objects belong to the same category will be emphasized. This same process will result in improved or hindered learning depending on whether similarities or differences need to be learned. One possible way in which category learning could move from an emphasis on differences towards an emphasis on similarities is by changing the similarity relations within and between categories (Carvalho & Goldstone, 2012; Zulkiply and Burt, 2013).

However, the characteristics of the categories being studied are not the only factors that can have an influence on how sequential comparison impacts learning. In theory, any property of the learning situation that changes attentional constraints could have similar impact by changing the task demands from an emphasis on similarities to an emphasis on differences.

Inductive learning can take place in several ways. One such way is active category learning. In this kind of learning task, participants actively try to categorize never-before-seen stimuli into one of the categories provided. Participants are then given feedback on the accuracy of their responses. Learning takes place via feedback-informed update of perceptual, attentional and decisional processes.

Another kind of inductive learning task can be referred to as passive learning. In this kind of situation, participants study category exemplars along with their correct category assignment. This task can be considered 'passive' in the sense that participants do not actively make responses during learning, and no feedback is provided via which a participant could adjust their judgments.

These two tasks differ in a number of educationally relevant aspects (e.g., motivation, engagement, etc.) but also in their cognitive aspects. It could be argued that if participants actively study the categories, emphasis will be placed on finding the differences between the categories. This is perhaps the most obvious way one can learn to discriminate As from Bs and achieve good performance. By contrast, in passive learning, participants are not tasked with learning how to discriminate between the categories. They are explicitly given the category assignments for the stimuli. Instead participants may search for features that characterize each category or the similarity amongst objects in each category.

Put another way, if participants passively study the stimuli along with their correct category assignments, their self-imposed task may be to identify the features that characterize that category (e.g., that all the 'Zups' have a similar nose shape). By contrast, if participants are not given the category assignment but instead have to try to categorize the stimulus and only then receive feedback, they may focus on finding differences between objects of different categories (e.g., that 'Zups' are round and 'Rikes' are squares; see Markman & Ross, 2003; Yamauchi & Markman, 2000).

Following Carvalho and Goldstone's (2012) proposal, combining interleaved study with active study will be beneficial for category learning because they are both compatible with focusing on features that differentiate between the categories being acquired. Likewise, combining blocked study with passive learning should be beneficial because they are consistent in leading participants to find similarities that are useful for successfully learning each category in isolation.

In this paper we aim to extend previous evidence for a comparison and attentional mechanism as the unifying processes behind both blocked and interleaved study, by manipulating the properties of the study session that affect attention allocation. One group of participants completed a passive study session associated with both interleaved and blocked presentations. The other group of participants completed an active study session, while keeping all other aspects of the task constant between the two groups.

An Experiment

In this experiment, participants studied a set of six categories, three presented interleaved and the remaining three presented blocked. Critically, for one group of participants, the study session was set up as a passive learning task. Participants studied each object for a short period of time during which the correct category assignment was also presented on the screen. For the other group of participants, the categories were studied in an active learning task. Both groups completed the same generalization task afterwards.

Method

Participants Eighty-one undergraduate students at Indiana University volunteered to participate in return for partial course credit. Data from seven participants in the passive learning group were excluded from analyses due to failure to repeat the label of the object just presented on more than half of the total number of study trials (see bellow for details). All participants in the active learning group reached the criterion of 34% correct responses during study and their data were kept for analyses.

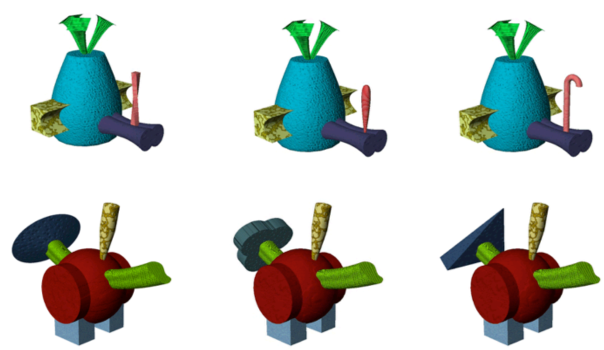

Figure 1: Examples of one exemplar of each of the 6 categories used. The top row constitutes one group of categories and the bottom row another group.

Apparatus and Stimuli In this experiment, stimuli were "Fribble" objects (Williams, 1998). Three of the categories were composed of very similar objects differing diagnostically only in one of their parts (see top panel in Figure 1). The other three categories were also very similar and differed diagnostically from each other only in one of their parts, however, they were substantially different from the other three categories (see bottom panel in Figure 1). Random variation existed in each of the categories but was the same across the three categories in each group.

Each category was given a unique label that perfectly predicted the presence of the unique feature that defined that category. At the start of the experiment, one label was randomly picked for each category from the pool: beme, kipe, vune, coge, zade, and tyfe (Hendrickson, et al., 2012).

Design and procedure This experiment had two conditions manipulated within-subjects (schedule of presentation: blocked category learning and interleaved category learning), and two conditions manipulated between-subjects (study type: active vs. passive). Each of these four conditions was composed of a study and test phase.

Study phase For the passive study group, during this study phase, participants were presented with a stimulus in the center of the screen along with the correct category assignment above the object for 2.5 s. Immediately after the presentation of the stimulus, three buttons were shown on the screen corresponding to the three possible category names for that study session. The participants' task was to press the button corresponding to the category of the object they just saw. This task was introduced to ensure participants' attention to the task and to equate the active and passive learning situations for the presence of a motor response. However, note that in this condition, participants simply needed to repeat the category they had just seen. There was no need for participants to learn a categorization rule. The mapping between the position of the buttons on the screen and the label was randomly shifted each trial.

For the active learning group, participants were presented with a stimulus for 500 ms. without its label. After the stimulus was removed, three buttons were shown on the screen and the participant had to choose the category assignment for that stimulus. After the participant's response, the stimulus along with the correct category assignment was shown on the screen for 2 s. The mapping between the position of the buttons on the screen and the label was also randomly shifted each trial.

For both groups, a 1000 ms inter-trial interval followed the trial and then the next 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. That is, in the interleaved condition, the probability of an object being followed by an object 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, 2012; Goldstone, 1996).

Each study phase was composed of 4 blocks for both groups of participants and the entire study phase took approximately the same amount of time for each group. Each block had 24 trials (4 exemplars of each category repeated 2 times each). After the 4th block of study a new set of instructions was presented on the screen and the second phase began. Each participant completed two sets of study and test phases (one for each schedule of presentation).

The two schedule conditions (blocked vs. interleaved) differed only in the frequency of category change during study and the species labels. Which condition was presented first was counterbalanced across participants and the allocation of the stimuli to each category and condition was randomized across participants.

Test Phase This second phase was a generalization task during which 36 stimuli were shown in random order – the 12 blobs participants studied during the learning task and 24 new stimuli. The new stimuli were similar to the studied 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, by pressing a key on the screen. After a 1000 ms inter-trial interval, a new trial would begin. No feedback was provided during this phase. Each test phase followed the respective study phase.

Results

We started by analyzing participants' performance during the study phase in the active learning group. As can be seen from Figure 2, participants' performance improves across the task for both the interleaved and blocked conditions.

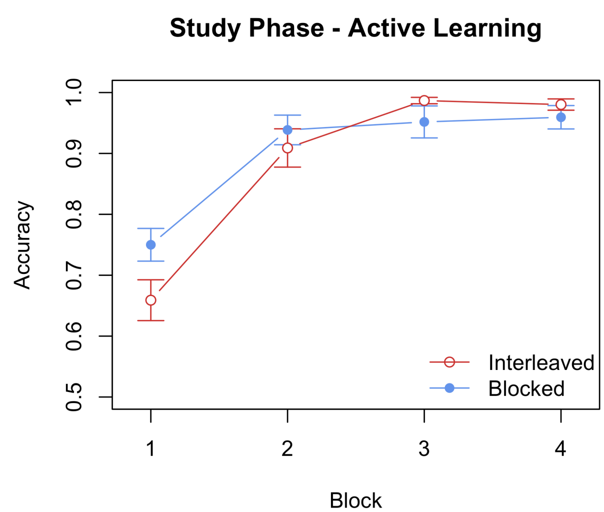

Figure 2: Results from the Study Phase for the Active Learning group. Error bars indicate standard errors of the mean.

A within-subjects ANOVA with Block (1 vs. 2 vs. 3 vs. 4) and schedule of presentation (interleaved vs. blocked) as factors confirmed this interpretation. There is a main effect of Block, F (3, 111) = 139.93, p < .0001, η_G^2 = 0.36. No main effect of schedule of presentation, F (1, 37) < 1 was found but the interaction between the two variables was also significant, F (3, 111) = 4.15, p = 0.008, η_G^2 = 0.03, indicating a larger improvement for the interleaved condition compared to the blocked condition.

However, the result of greater interest is how well participants are able to generalize this learning to new stimuli. These analyses will not only allow us to test the effect of interleaving vs. blocking and of passive vs. active learning but, more importantly, the interaction between the two.

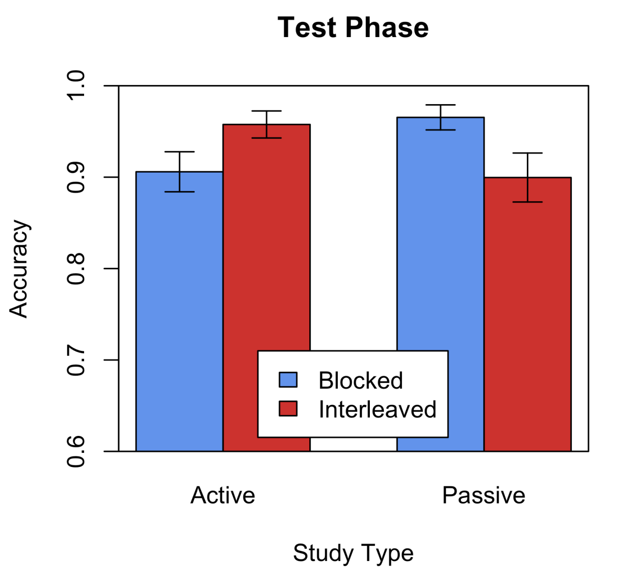

Figure 3: Main results of the Test Phase for both the Active and Passive groups and for each of the schedule of presentation conditions. Error bars indicate standard errors of the mean

The graph depicted in Figure 3 shows performance in the generalization task for participants in the active and passive study conditions and for each study presentation format. Given the overall high level of performance we began by analyzing possible differences in response time. No main effect or interaction between any of the variables was found for RT.

We then proceeded to analyze performance differences. The most obvious result is an interaction between the study condition and the presentation format. While for participants studying the stimuli actively, interleaving presentation of the objects results in better generalization performance, blocking is better for participants in the passive learning condition.

A mixed ANOVA with presentation schedule (blocked vs. interleaved) as a within-subject factor and study type (passive vs. active) as a between-subjects factor confirms this analysis. There is no main effect of presentation schedule, F (1, 73) < 1 or study type, F (1, 73) < 1. However, the interaction between the two variables was reliable, F (1, 73) = 7.30, p = .04, η_G^2 = 0.03.

Performance results for each group and schedule of presentation, sub-divided by studied and novel stimulus, are presented in Table 1. As can be seen, no differences in performance between novel and studied stimuli were found. We repeated the ANOVA analyses with stimulus type (studied vs. novel) as another within-subject factor. These analyses revealed the same critical interaction between schedule of presentation and study type, but no differences in performance for novel and studied stimuli or interaction with any of the other variables.

Table 1: Categorization accuracy in the test phase for both groups and schedule conditions, broken down by type of item (novel vs. studied). Standard deviations are presented in parentheses.

	Active		Passive	
	Interleaved	Blocked	Interleaved	Blocked
Novel	0.96 (0.13)	0.90 (0.21)	0.91 (0.20)	0.96 (0.13)
Studied	0.96 (0.13)	0.91 (0.18)	0.89 (0.26)	0.97 (0.11)

Discussion

Determining how to order information so that learners can achieve the best learning outcomes is crucial for effective training. In this work we present further evidence that the way information is ordered impacts learning and that this influence is modulated by whether learning is active or passive.

The results of the experiment presented here show that whether interleaving examples of several concepts or blocking examples by category is beneficial is a function of the training task's implicit demands. More specifically, in a task involving discrimination of the concepts being studied by identifying their differences (the active learning situation), interleaved study results in better performance in a subsequent generalization task. However, if the learning situation involves creating a positive, stand-alone representation of the concepts by identifying the similarities among the instances within each category (the passive learning situation), blocked study benefits performance in the generalization task.

Interestingly, both study conditions result in similar performance during the learning task in the active learning condition. This eliminates the possibility that one study condition is more difficult and results in greater cognitive effort, which is a known important factor contributing to improved learning (Bjork, 1994).

In the generalization task, the interaction between the type of study situation and the schedule of presentation of the exemplars had an effect for both studied stimuli and novel ones. This is an interesting result, suggesting that study benefits go beyond memorization of the whole exemplars. Very likely, participants succeeded at the categorization task by identifying the single defining parts for each category. These defining parts were instantiated identically for studied and novel objects, explaining why novel objects were not more difficult to categorize. Future research will be needed to assure that the interaction between learning activity and presentation schedule generalizes to other category structures.

Overall these results are consistent with the framework proposed by Carvalho and Goldstone (2012; see also Goldstone, 1996) hypothesizing that participants compare successive objects and update attention to stimulus features as a result of these comparisons.

The role of allocating one's attention during category learning has been highlighted before in different models (Kruschke, 1992; Love, Medin, & Gureckis, 2004; Minda & Smith, 2002; Nosofsky, 1986) and the use of eye tracking technology has made it possible to study the patterns of overt trial-by-trial, or even within-trial, attention. For example, Blair, Watson, Walshe, and Maj (2009) have demonstrated that in a categorization task, different stimuli can elicit different patterns of attention allocation to their features. Additionally, previous research has also demonstrated that during category learning, participants often take into account information from only the previous few trials to decide whether a stimulus belongs in one category or another (Jones, Love, & Maddox, 2006; Jones & Sieck, 2003; Stewart, Brown, & Chater, 2002; Stewart & Brown, 2004; Stewart & Chater, 2002).

Carvalho and Goldstone (2012) propose that when studying a new exemplar in an inductive learning task, participants compare the properties of that object with the properties they recall from the previous ones. However, learners do not remember all the features from all the objects presented. Instead, when studying a new exemplar, learners weight more heavily the information presented in the immediately preceding instances. If the previous trial consisted of an object in one category and the current trial consists of another object in a different category, participants' attention will be directed towards the *differences* between the two objects. Conversely, if the two objects come from the same category, learners will attend to *similarities* between the objects.

This framework can aptly account for the results presented here: passive learning requires attending to similarities, while active learning requires attending to differences. When the presentation order also emphasizes those factors, learning will be facilitated.

Finally, although our results do not directly speak to the importance of active vs. passive learning, it is worth noting that this interaction should be taken into account when deciding whether learners should be given worked examples to study or not.

Indeed, we think one of the most important contributions of the present work is the proposal that when deciding how to structure learning, one needs to take into account the entire learning situation and possible interactions between situational factors. So far, we have demonstrated this for interleaving/blocking benefits relative to the training activity (passive vs. active) and the similarity structure of the categories being studied (similar vs. dissimilar; Carvalho & Goldstone, 2012).

Acknowledgements

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 Graduate Training Fellowship SFRH /BD/78083/2011 from the Portuguese Foundation for Science and Technology (FCT). The authors would like to thank the Percepts and Concepts Lab members for discussion and Spenser Benge for his help with data collection. Stimulus images courtesy of Michael J. Tarr, Center for the Neural Basis of Cognition and Department of Psychology, Carnegie Mellon University, **http://www.tarrlab.org/**

References

Blair, M., Watson, M. R., Walshe, R. C., & Maj, F. (2009). Extremely selective attention: Eye-tracking studies of the dynamic allocation of attention to stimulus features in categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(5), 1196–1206.

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., & Goldstone, R. (2012). Category structure modulates interleaving and blocking advantage in inductive category acquisition. In C. R. P. Miyake N, Peebles D (Ed.), *Proceedings of the 34th Annual Conference of the Cognitive Society*. Austin, TX: Cognitive Science Society.

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

Hendrickson, A. T., Kachergis, G., Fausey, C. M., & Goldstone, R. L. (2012). Re-learning labeled categories reveals structured representations. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Jones, M., Love, B. C., & Maddox, W. T. (2006). Recency effects as a window to generalization: separating decisional and perceptual sequential effects in category learning. *Journal of experimental psychology. Learning, memory, and cognition*, *32*(2), 316–32.

Jones, M., & Sieck, W. R. (2003). Learning myopia: An adaptive recency effect in category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 626–640.

Kornell, N. (2009). Optimizing 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.
Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological review*, *99*(1), 22.
Kurtz, K. H., & Hovland, C. I. (1956). Concept learning with differing sequences of instances. *Journal of experimental psychology*, *51*(4), 239.
Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A Network Model of Category Learning. *Psychological Review*, *111*(2), 309–332.
Markman, A.B., & Ross, B.H. (2003). Category use and category learning. *Psychological Bulletin*, *129*(4), 592-613.
Minda, J. P., & Smith, J. D. (2002). Comparing prototype-based and exemplar-based accounts of category learning and attentional allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(2), 275–292.
Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of experimental psychology. General*, *115*(1), 39–61.
Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *The Behavioral and brain sciences*, *21*(1), 1–17; discussion 17–54.
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.
Stewart, N., & Chater, N. (2002). The effect of category variability in perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(5), 893.
Vlach, H. A., & Sandhofer, C. M. (2012). Distributing Learning Over Time: The Spacing Effect in Children's Acquisition and Generalization of Science Concepts. *Child development, 83*, 1137–1144.
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*, 39(5), 750-763.
Whitman, J. R., & Garner, W. R. (1963). Concept Learning as a Function of Form of Internal Structure. *Journal Of Verbal Learning And Verbal Behavior*, *2*(2), 195–202.
Williams, P. (1997). Prototypes, exemplars, and object recognition. New Haven, CT: Yale.
Yamauchi, T., & Markman, A.B. (2000). Inference using categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(3), 776-795.
Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition, 41*(1), 16-27.