Title
Interleaving Benefits Category Learning But Not Item Memory

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Yüksel, Ezgi Melisa
Knabe, Melina Lauryn
Green, C. Shawn
et al.

Publication Date
2024

Peer reviewed

Interleaving Benefits Category Learning But Not Item Memory

Ezgi Melisa Yüksel (yksel@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Melina L. Knabe (mknabe@wisc.edu)
Department of Educational Psychology, 1025 W. Johnson Street
Madison, WI 53706 USA

C. Shawn Green (csgreen2@wisc.edu)
Department of Psychology, 1202 W. Johnson Street
Madison, WI 53706 USA

Haley A. Vlach (hvlach@wisc.edu)
Department of Educational Psychology, 1025 W. Johnson Street
Madison, WI 53706 USA

Abstract

Interleaving, as opposed to blocking, information improves learning of categories, such as artists' painting styles. The current study examined whether presentation schedules also impact memory for specific items. 179 participants studied paintings from 12 different artists on either a blocked or interleaved schedule. In Study 1 (N = 84), participants were then asked to either identify the artists of a series of paintings (style recognition task) or determine whether they had previously seen a specific painting (item recognition task). In Study 2 (N = 93), participants completed both tasks. Results showed that the interleaved schedule led to better learning of the painting styles, but did not impact item memory. However, when participants had to recognize the style and the painting for an artist on the interleaved schedule, they incorrectly thought that they had previously seen the painting. This finding illustrates the dynamic relationship between item memory and category learning.

Keywords: interleaving; learning; categorization; induction; retrieval

Introduction

Interleaved Learning

Learning and generalizing categories is a critical cognitive process. Humans must learn that chairs are for sitting, that labradors are a type of dog, and that a triangular-shaped head belongs to a venomous snake. The process of sorting information into categories underlies various educational abilities such as acquiring knowledge about biological species through the examination of pictures of exemplars (Birnbaum et al., 2013), understanding painting styles by studying paintings from specific artists (Kang & Pashler, 2012), and identifying mathematical procedures for math problems (Higgins & Ross, 2011; Rau, Aleven, & Rummel, 2010). For that reason, cognitive and educational psychologists have spent considerable time studying category learning in the laboratory and the classroom. The goal of these research efforts has been to understand how category learning arises and how to support this critical cognitive process.

Consistent with this, research on category learning has identified several factors that influence the degree to which individuals learn and generalize category structure, including the number, perceptual similarity, and variability of individual category items (e.g., Lawson & Fisher, 2011; Perry et al., 2010; Sloutsky et al., 2007; Son et al., 2008; Spencer et al., 2011; Twomey et al., 2014). One particular vein of research relevant here has focused on how the order in which items are presented impacts category learning (e.g., Birnbaum et al., 2013; Rohrer, 2012; Pan, 2015). For example, prior studies have shown that mixing, or interleaving, the presentation of paintings from different artists (e.g., A, B, C, A, B, C) results in an improved ability to place previously unseen paintings into the correct artist category in comparison to showing paintings from the same artist back-to-back in a blocked fashion (e.g., A, A, B, B, C, C) (Kornell & Bjork, 2008). This interleaving effect has also been observed when participants are asked to learn other visual categories, such as birds and butterflies (Birnbaum et al., 2013; Higgin & Ross, 2011; Kang & Pashler, 2011; Kornell et al., 2010; Rohrer & Taylor, 2007; Wahlheim et al., 2011), and in other domains, including motor learning (Shea & Morgan, 1979), language learning (Healy & Bourne, 2013; Nakata & Suzuki, 2019), and math learning (Rau et al., 2013; Taylor & Rohrer, 2010).

Why does interleaving lead to significant differences in category learning? The precise mechanism is not fully known, but several leading theories propose that different learning schedules influence how learners allocate their visual attention to items during learning. Implicit in this claim is that differences in visual attention lead to differences in what learners encode and remember about these items. For instance, the discriminative contrast theory (e.g., Carvalho & Goldstone, 2014; 2019; Kang & Pashler, 2012; Wahlheim et

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

al., 2011) posits that learners engage in item-by-item comparisons in category learning tasks. When consecutive items belong to the same category (as in the blocked schedule), attention shifts towards the similarities within this group of items. However, if consecutive items belong to different categories (as in the interleaved schedule), attention is directed towards the differences between the items of different categories. Interleaving is therefore most effective for categories that are highly confusable, like different bird or butterfly species (e.g., Birnbaum et al., 2013).

Instead of focusing on what is encoded, another theory focuses on what learners remember (and forget) during and after learning (Knabe et al., 2023; Vlach et al., 2008; 2012; 2022; Vlach & Kalish, 2014). This forgetting-as-abstraction account (for a review, see Vlach, 2014)—derived from study-phase retrieval theories—suggests that when learners are shown items on a spaced or interleaved schedule, they begin to forget features that do not appear in each item and might not be relevant for category membership. Features that reappear in each item, however, are reactivated in memory across time. This reactivation supports categorization because features that are important for category membership become more retrievable for the learner, and they are better able to abstract the categories from the individual items.

The discriminative contrast theory and forgetting-as-abstraction account differ in their emphasis on encoding versus forgetting during category learning, but share a key assumption: that learners are differentially attending to and remembering the item features that comprise a category on blocked and interleaved schedules, and this difference impacts their category learning. To date, no studies have extended these theories to test learners' memory of the items that comprise a category. This is an important gap, as prior work on categorization has shown that there are bidirectional relationships between memory and categorization processes (e.g., Knabe et al., 2023; Vlach, 2016). Indeed, the idea that individual items and categories interact is undisputed. The degree to which learners *remember* individual items that comprise a category, however, is debated in categorization research (see exemplar-based vs. prototype theories; Brooks, 1978; Posner & Keele, 1968).

In sum, the interleaving effect is thought to arise from differences in what learners attend to and remember on blocked versus interleaved schedules. Interleaved schedules lead to better category learning, but how does this type of schedule impact learners' memory for the individual items that comprise a category? On the one hand, the discriminative contrast theory is agnostic about the downstream consequences of attentional differences in blocked and interleaved schedules on overall item memory. On the other hand, the forgetting-as-abstraction account would predict that learners might demonstrate poorer memory for specific items on the interleaved schedule because they must forget features of the items to better abstract the categories.

The Current Research

The aim of the present research was to determine to what degree learners remember the specific items that comprise a category depending on whether they encountered the items in a blocked or interleaved schedule (Study 1) and how their item memory impacts category learning (Study 2). We addressed this aim by showing learners the Kornell and Bjork (2008) painting styles task, a commonly used visual category learning task. In Study 1, participants were assigned to a typical category learning test (Study 1a) that included new and old paintings at the test, or to an item recognition task (Study 1b) where they had to identify whether a painting was new or old. In Study 2, participants completed the item recognition task prior to the category learning test.

Study 1.

Methods

Participants We recruited 88 undergraduate students (Study 1a: $N = 46$; Study 1b: $N = 42$) from a public university in the Midwestern United States who participated in exchange for course credit. Four participants were excluded from Study 1a due to noncompliance with the instructions: they either pressed the same key 90% of the time during the test or did not press any key. The participants were between 18 and 28 years (*M*: 20.0 years, *SD* = 1.7) and consisted of 80.2% females.

Design Participants completed the Kornell and Bjork (2008) painting styles task, where they saw 72 paintings from 12 artists. Half of the artists appeared on a blocked schedule and half of the artists appeared on an interleaved schedule (Figure 1). Participants then engaged in a category learning task (Study 1a) or in an item recognition task (Study 1b). The category learning task in Study 1a was adapted from Kornell and Bjork (2008) with a slight modification: Rather than presenting participants with 4 new paintings and asking them to identify the corresponding artist, here participants had to select the artist for 3 new paintings and 3 old paintings. This was done to ensure that the interleaving effect persisted when participants were presented with paintings seen during the learning phase (as opposed to only new paintings). The item recognition task in Study 1b involved showing participants a combination of new and old paintings at test, and was not included in the original paradigm. Specifically, participants saw 3 old paintings and 3 new paintings for all 12 artists. After each item, they had to indicate whether they remembered seeing the specific painting during the learning phase.

Materials The materials were the same for Study 1a and Study 1b. Participants completed the Kornell and Bjork (2008) painting styles task, using materials sourced from the same study. A total of ten paintings from the following twelve artists were chosen: Georges Braque, Henri-Edmond Cross, Judy Hawkins, Philip Juras, Ryan Lewis, Marilyn Mylrea, Bruno Pessani, Ron Schlorff, Georges Seurat, Ciprian Stratulat, George Wexler, and YieMei. Six paintings from each artist were randomly assigned to the blocked or interleaved schedule, and 3 paintings were randomly assigned to the test phase.

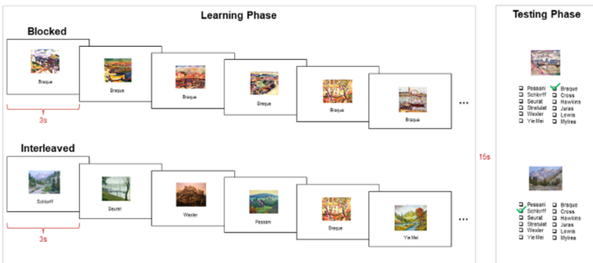

Figure 1: Example of Kornell and Bjork (2008) painting styles task trial with eye-tracking paradigm.

Procedure The study was administered in-person. Participants were told that they would study a series of paintings before answering questions and completing a demographics survey. During the study, participants viewed a total of 72 paintings: six from each of the 12 artists. The allocation of artists to interleaved (I) and blocked (B) conditions was randomized for each participant. The sequence of interleaved and blocked conditions followed the pattern BIIBBIIBBIIB, which was identical to the original Kornell and Bjork (2008) paradigm. Each painting was displayed on-screen for 3 seconds with the corresponding artist's name presented below the painting. For half of the artists, paintings were presented on a blocked schedule where six paintings by the same artist were shown consecutively (e.g., Braque, Braque, Braque, Braque, Braque, Braque). The other half of the artists were displayed on an interleaved schedule, alternating the six paintings with paintings from other artists (e.g., Pessani, Schlorff, Stratulat, Seurat, Cross, Lewis).

After the learning phase, participants completed a 15-second distractor phase where they counted backwards from 547 in intervals of 3. Participants then entered the testing phase, which contained 4 test blocks. In contrast to the original paradigm, participants in Study 1a were presented with 3 new and 3 old (i.e., seen in the study phase) paintings from the 12 artists in each block. For each painting, they had to select who painted the painting from a list of 12 artist names. No feedback was provided. Participants in Study 1b were presented 3 new and 3 old paintings and were asked to indicate if they had seen the painting during the learning phase ("Yes" or "No"). 'Yes' was the correct answer for previously seen paintings, whereas 'No' was the correct answer for new paintings.

Results

All analyses were conducted in R (R Core Team, 2022). We conducted three analyses: First, we compared category learning performance on the blocked and interleaved schedules (Study 1a) to replicate the interleaving effect in this stimulus set. Next, we compared item recognition performance on blocked and interleaved schedules (Study 1b). Finally, we conducted a detailed analysis of participants' item recognition performance in Study 1b by examining their ability to discriminate old and new items with measures used in signal detection theory (i.e., d' prime, hit rates, false alarms, correct rejections, and misses).

Study 1a: Category Learning Performance We first assessed whether we could replicate the interleaving effect in learning artists' painting styles. We calculated participants' accuracy by determining the proportion of test trials where they chose the correct artist from among 12 artist names (Figure 1). A linear mixed effect model was constructed where accuracy at test was regressed on study schedule (contrast coded: "-0.5" for blocked and "0.5" for interleaved), controlling for participant random effects. The results revealed that participants were significantly more accurate in recognizing the painting style when the artists were presented on the interleaved schedule (M = .48, SD = .50) compared to the blocked schedule (M = .30, SD = .46), (b = 0.18, SE = 0.02, t = 11.06, p < .001). This replicated prior findings (Figure 2).

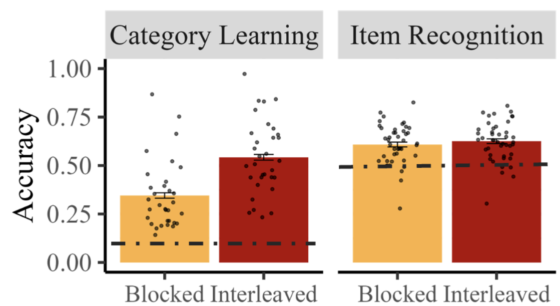

Figure 2. Accuracy on category learning (Study 1a) and item recognition (Study 1b) tasks by condition (blocked or interleaved)

Next, we examined whether the interleaving effect was present for paintings participants had seen during learning (old items) and for paintings they had not seen during learning (new items). We constructed a linear mixed effects model where accuracy at test was regressed on study schedule (contrast coded: "-0.5" for blocked and "0.5" for interleaved), item type (contrast coded: "-0.5" for old and "0.5" for new), and their interaction. There was no significant interaction between study schedule and item type, which suggests that the pattern of results did not differ across old and new items: the interleaving effect was present when participants had to retain the artist names for previously seen paintings and when they had to generalize artist names to new paintings, b = -0.005, SE = 0.03, t =-0.16, p = .87. There was a significant main effect of item type, such that participants were more accurate at selecting the corresponding artist for old items (M = .41, SD = .49) than new items (M = .37, SD = .48), b = -0.04, SE = 0.02, t = -2.77, p = .005.

Study 1b: Painting Recognition Memory Performance In Study 1b, we first determined whether interleaving impacted item recognition memory. Participants' item recognition memory was assessed by calculating the proportion of test trials where participants correctly identified the paintings as either old or new (Figure 3). That is, the proportion of trials where participants said "yes" to having seen old paintings and "no" to new paintings. We constructed a linear mixed effect model where recognition memory accuracy at test was regressed on study schedule (contrast coded: "-0.5" for blocked and "0.5" for interleaved), controlling for participant random effects.

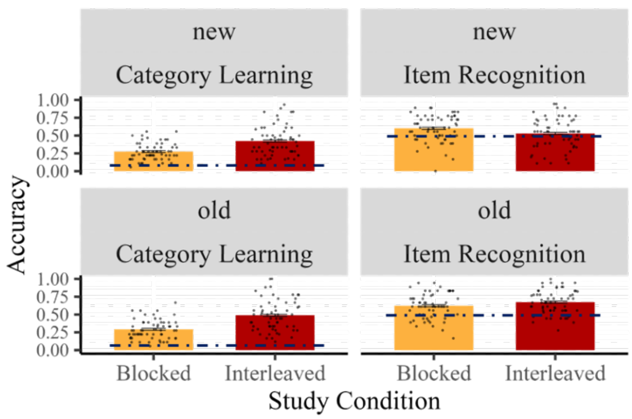

Figure 3. Accuracy on category learning (Study 1a) and item recognition (Study 1b) tasks by condition (blocked or interleaved) and item (old or new paintings). Dotted lines indicate the chance levels.

The results showed that participants' item recognition performance was comparable on the interleaved (M = .64, SD = .48) and blocked (M = .62, SD = .49) schedules, b = 0.02, SE = 0.02, t = 1.06, p = .29.

Next, we examined whether item recognition performance differed between old and new items. We constructed a linear mixed effect models where item recognition memory performance at test was regressed on study schedule (contrast coded: "-0.5" for blocked and "0.5" for interleaved), item type (contrast coded: "-0.5" for old and "0.5" for new), and their interaction. This analysis revealed that the interaction between study schedule and item type was not significant, b = -0.05, SE = 0.03, t = -1.36, p = .17.

The item memory recognition analysis only considered whether participants correctly identified the old paintings as old and the new paintings as new. To obtain a more sensitive measure of participants' ability to discriminate, or differentiate, old from new paintings, we calculated participants' d-prime (d'). D-prime is a measure derived from signal detection theory and is calculated by subtracting the z transform of false alarms rates from the hit rates. Hits refers to participants correctly identifying an old painting as "old", whereas false alarms refers to participants incorrectly identifying a new painting as "old" (Table 1).

We constructed a linear mixed effects model where we regressed d' on study schedule (contrast coded: "-0.5" for blocked and "0.5" for interleaved). The results showed that participants had a significantly higher d' on the interleaved schedule (M = .89, SD = 0.64) than the blocked schedule (M = .72, SD = 0.56), b = 0.18, SE = 0.022, t = 7.88, p < .001. This means that participants were better at discriminating between old and new paintings when they had seen the artists on an interleaved schedule.

To identify what was driving this difference in d', we constructed two linear mixed effects models where we regressed the number of hits or false alarms separately on study schedule (contrast coded: "-0.5" for blocked and "0.5" for interleaved). The results showed that participants had a significantly higher hit rate on the interleaved schedule (M = .65, SD = 0.19) than the blocked schedule (M = .61, SD = 0.17), b = 0.04, SE = 0.006, t = 6.79, p < .001. There was no significant difference in the false alarm rate between the interleaved (M = .37, SD = 0.18) and blocked schedule (M = .37, ms, SD = 0.18), b = -0.0008, SE = 0.007, t = -0.125, p = .9. These results indicate that the higher d' on the interleaved schedule was driven by a higher hit rate and lower false alarm rate, which means participants were better at discriminating between old and new paintings from artists presented on an interleaved schedule.

Table 1: Signal Detection Theory Measures by Presentation Schedule for Study 1

Outcome	Blocked		Interleaved		*p*
	M	*SD*	*M*	*SD*	
D-Prime	0.72	0.56	0.89	0.64	<.001**
Hits	0.61	0.17	0.65	0.19	<.001**
False Alarms	0.37	0.18	0.37	0.18	.89
Correct Rejections	0.63	0.18	0.63	0.18	.89
Misses	0.39	0.17	0.35	0.19	<.001**

Study 2

Study 1 examined differences in item memory across blocked and interleaved schedules. The next step was to assess whether memory for an artists' paintings predicted learning of the artists' style. Therefore, participants underwent tests for both item memory and style learning performance.

Methods

Study 2 followed the same procedure as Study 1, but the type of test was manipulated within-subjects as opposed to between-subjects. That is, after the learning phase, participants had to indicate if they remembered seeing the paintings in the learning phase (item recognition task) and then select the artist who painted each painting (category learning task).

Participants. Ninety-eight undergraduate students from a public university in the Midwestern United States participated in Study 2 in exchange for course credit. Six participants were excluded from the study due to noncompliance with the instructions: they either pressed the same key 90% of the time during the test or did not press any key. The participants were between 18 and 28 years (*M*: 19.81 years, *SD* = 1.23) and consisted of 63% females.

Results

Category Learning Performance We first replicated the interleaving effect in learning artists' painting styles: Participants' ability to identify the artist's name was significantly better for paintings presented on the interleaved condition (*M* = .35, *SD* = .48) compared to the blocked condition (*M* = .22, *SD* = .41), *b* = 0.13, SE = 0.01, *t* = 12.81, *p* < .001).

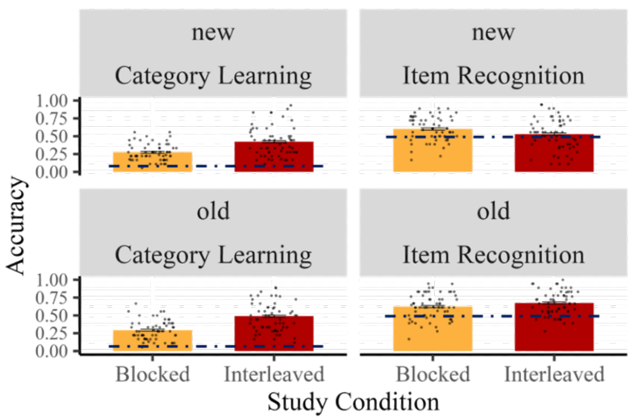

Figure 4. Accuracy on category learning and item recognition tasks by condition (blocked or interleaved) and item (old or new paintings) in Study 2. Dotted lines indicate the chance levels.

Painting Recognition Memory Performance We then examined whether the learning schedule impacted item recognition memory. Similar to Study 1, we found no significant difference in participants' ability to identify the paintings between interleaved (*M* = .58, *SD* = .49) and blocked conditions (*M* = .58, *SD* = .49), *b* = -0.003, SE = 0.01, *t* = -0.26, *p* = .79.

To determine whether correctly identifying paintings predicted category learning performance, we ran a linear mixed effect model where we regressed accuracy on the category learning task on accuracy on the item recognition task while controlling for participants' random effects. The results revealed that the accuracy in item memory was not a significant predictor of accuracy in learning the artist's style, *b* = 0.01, SE = 0.01, *t* = 0.88, *p* = 0.37.

Next, we compared participants' ability to identify old paintings as old (hit) and new paintings as old (false alarm) (Table 2). Unlike Study 1b, we found that participants' hit rates were higher for the interleaved condition (*M* = .63, *SD* = .17) than the blocked condition (*M* = .61, *SD* = .19), *b* = 0.02, SE = 0.005, *t* = 3.92, *p* < .001. We also found that false alarm rates were significantly higher in the interleaved condition (*M* = .47, *SD* = .22) than in the blocked condition (*M* = .44, *SD* = .21), *b* = 0.03, SE = 0.005, *t* = 4.94, *p* < .001.

We then examined participants' ability to discriminate old from new paintings using d' prime. This analysis showed no significant difference between the interleaved (*M* = .50, *SD* = .59) and blocked condition (*M* = .52, *SD* = .65), *b* = -0.01, SE = 0.02, *t* = -0.92, *p* = .35.

The finding that participants had higher hit rates and false alarm rates in the interleaved condition suggests that participants had a tendency to say that they had previously seen a painting. Saying "yes" to having previously seen a painting – even if they had not – means that participants were not able to discriminate old and new paintings well, thereby decreasing d'. One explanation for why participants showed poorer discrimination in the interleaved condition in Study 2 than in Study 1b is the difference in the testing phase: In Study 2, participants selected the artist name after they had indicated whether they had seen the painting. We address this further in the Discussion section.

Table 2: Signal Detection Theory Measures by Presentation Schedule for Study 2

Outcome	Blocked		Interleaved		*p*
	M	*SD*	*M*	*SD*	
D-Prime	0.53	0.65	0.50	0.59	.35
Hits	0.61	0.19	0.63	0.17	<.001**
False Alarms	0.44	0.21	0.47	0.22	<.001**
Correct Rejections	0.56	0.21	0.53	0.22	<.001**
Misses	0.39	0.17	0.37	0.17	<.001**

Discussion

The bidirectional relationship between category exemplars and the overarching category is well acknowledged, yet there is limited empirical evidence on how presentation schedules impact this relationship. Theories explaining the interleaving effect, such as the discriminative contrast theory, highlight the importance of learners attending to and remembering different item features based on the learning schedules. This theory remains agnostic about how attentional differences in learning schedules impact overall item memory. Whereas, the forgetting-as-abstraction account proposes that interleaving influences memory for individual items. Implicit in this theory is a potential trade-off where interleaving enhances category abstraction but harms item memory by causing forgetting of individual items. We addressed this gap by examining the impact of learning schedules on item memory.

Our findings revealed that the interleaved schedule did not impact item memory differently than the blocked schedule. Instead, item memory was low across both schedules. Our results also showed that participants' item memory performance was not a significant predictor of their category learning performance.

According to the discriminative contrast theory, learners encode similarities or differences in the features of items depending on the learning schedule. Due to the complexity of naturalistic categories like paintings, we could not directly test learners' memory for specific features, which would provide a more direct test of the discriminative contrast theory. It is plausible that attention to the similarities within or differences between categories might differentially impact learning of category boundaries across the two learning schedules, which in turn could produce different item memory performance. We found comparable item memory performance across the two learning schedules, so further studies are necessary to better understand the observed pattern of results.

According to the forgetting-as-abstraction account, presenting items on an interleaved schedule should lead to forgetting of category-irrelevant features, thereby benefiting abstraction and categorization (Vlach et al., 2008, 2012; Vlach & Kalish, 2014; for a review, see Vlach, 2014). Our findings showed that participants' memory performance remained consistently low for the items, irrespective of the learning schedule. The results support the idea that participants might be forgetting the items that comprise a category, yet this forgetting is not limited to the interleaved schedule and also extends to items shown on a blocked learning schedule. An alternative explanation is that participants did not forget, but rather failed to encode, the items on both schedules during the learning phase. This might not impact their category learning as participants abstract the relevant category structure, but it might harm their item memory recognition.

Even though learners failed to encode or rapidly forgot the individual paintings, they were still able to correctly generalize the artists' style to new paintings at test. Another question we addressed was whether correctly remembering the individual items would predict category learning, irrespective of the learning schedule. We found that learners' item memory did not predict category learning, which supports the idea that participants were not relying on their item memory to generalize.

Although our study indicated that overall item memory was not a strong predictor of generalization, we did observe evidence pointing towards a nuanced relationship between item memory and categorization: First, participants were better at discriminating old from new paintings in the interleaved condition in Study 1a. When participants had to recognize the painting and then the category back-to-back in Study 2, however, their false alarm rates increased on the interleaved schedule; that is, participants might have had a greater sense of familiarity with the paintings on the interleaved schedule than the blocked schedule, leading them to incorrectly think that they had previously seen a painting when they had not (Wallis, & Bülthoff, 2001). This finding shows that the degree to which a category is learned influences the immediate recognition of individual items. This might also be related to the stimuli used in this study as visual categories like paintings are highly confusable. Thus, the nature of the task and the stimuli might be pivotal in shaping how learners remember and recognize individual items.

This finding contributes to existing theories of the interleaving effect. Current theories of the interleaving effect have typically tested learners' attention to and memory for specific features of artificial categories. To understand the dynamic relation between item memory and generalization better, future studies should continue to examine naturalistic stimuli.

In sum, this study was the first to examine the impact of learning schedules (or sequencing) on item memory. The results showed that interleaving does not harm item memory, but does impact the ability to discriminate between similar paintings depending on the task structure. There are several avenues for future research that arise based on these findings. Going forward, it would be valuable to examine how interleaving affects item memory when the items and categories are dissimilar and more distinguishable. Furthermore, future studies should probe participants' item memory in other ways, such as asking them to freely recall individual features of items. This would provide insight into how learning schedules impact memory for specific details compared to overall item recognition, and would specifically address predictions made by the discriminative contrast theory. By expanding our research efforts in these directions, we will better understand the dynamic relationship between memory and category learning.

References

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2012). Why interleaving enhances inductive learning: The roles of discrimination and retrieval. Memory & Cognition, 41(3), 392-402. https://doi.org/10.3758/s13421-012-0272-7

Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological Bulletin, 145*(11), 1029-1052. https://doi.org/10.1037/bul0000209

Carpenter, S. K., & Mueller, F. E. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition, 41*(5), 671-682. https://doi.org/10.3758/s13421-012-0291-4

Carvalho, P. F., & Goldstone, R. L. (2014). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic Bulletin & Review, 22*(1), 281-288. https://doi.org/10.3758/s13423-014-0676-4

Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(11), 1699-1719. https://doi.org/10.1037/xlm0000406

Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing simultaneously promotes multiple forms of learning in children's science curriculum. *Applied Cognitive Psychology*, *28*(2), 266-273.

Goldwater, M. B., Don, H. J., Krusche, M. J., & Livesey, E. J. (2018). Relational discovery in category learning. *Journal of Experimental Psychology: General*, *147*(1), 1.
Healy, A. F., & Bourne, L. E. (2013). Empirically Valid Principles for Training in the Real World. *The American Journal of Psychology, 126*(4), 389-399. https://doi.org/10.5406/amerjpsyc.126.4.0389
Higgins, E., & Ross, B. (2011). Comparisons in category learning: How best to compare for what. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 33, No. 33).
Kang, S. H. K., & Pashler, H. (2011). Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast: Spacing promotes contrast. *Applied Cognitive Psychology, 26*(1), 97-103. https://doi.org/10.1002/acp.1801
Kirchoff, B. K., Delaney, P. F., Horton, M., & Dellinger-Johnston, R. (2014). Optimizing learning of scientific category knowledge in the classroom: The case of plant identification. *CBE—Life Sciences Education*, *13*(3), 425-436.
Knabe, M. L., Schonberg, C. C., & Vlach, H. A. (2023). When time shifts the boundaries: Isolating the role of forgetting in children's changing category representations. *Journal of Memory and Language*, *132*, 104447.
Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories: Is Spacing the "Enemy of Induction"? *Psychological Science, 19*(6), 585-592. https://doi.org/10.1111/j.1467-9280.2008.02127.x
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. https://doi.org/10.1037/a0017807
Lawson, C. A., & Fisher, A. V. (2011). It's in the sample: The effects of sample size and sample diversity on the breadth of inductive generalization. *Journal of Experimental Child Psychology*, *110*(4), 499–519. https://doi.org/10.1016/j.jecp.2011.07.001
Nakata, T., & Suzuki, Y. (2019). Mixing Grammar Exercises Facilitates Long-Term Retention: Effects of Blocking, Interleaving, and Increasing Practice. *The Modern Language Journal*. https://doi.org/10.1111/modl.12581
Pan, S. C. (2015). The interleaving effect: mixing it up boosts learning. Scientific American, 313(2).
Perry, L. K., Axelsson, E. L., & Horst, J. S. (2016). Learning what to remember: Vocabulary knowledge and children's memory for object names and features. *Infant and Child Development*, *25*(4), 247–258. https://doi.org/10.1002/icd.1933
Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn Locally, Think Globally: Exemplar Variability Supports Higher-Order Generalization and Word Learning. *Psychological Science*, *21*(12), 1894–1902. https://doi.org/10.1177/0956797610389189
Rau, M. A., Aleven, V., & Rummel, N. (2013). Interleaved practice in multi-dimensional learning tasks: Which dimension should we interleave? *Learning and Instruction, 23*, 98-114. https://doi.org/10.1016/j.learninstruc.2012.07.003
Rau, M. A., Aleven, V., & Rummel, N. (2010). Blocked versus interleaved practice with multiple representations in an intelligent tutoring system for fractions. In V.Aleven, J.Kay, & J. Mostow(Eds.), Intelligent tutoring systems (pp.413–422).Berlin,Germany:Springer.Retrieved from http://dx .doi.org/10.1007/978-3-642-13388-6_45
Rohrer, D. (2012). Interleaving Helps Students Distinguish among Similar Concepts. *Educational Psychology Review, 24*(3), 355-367. https://doi.org/10.1007/s10648-012-9201-3
Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science, 35*(6), 481-498. https://doi.org/10.1007
Schneider, V. I., Healy, A. F., & Bourne, L. E. (2002). What Is Learned under Difficult Conditions Is Hard to Forget: Contextual Interference Effects in Foreign Vocabulary Acquisition, Retention, and Transfer. *Journal of Memory and Language, 46*(2), 419-440. https://doi.org/10.1006/jmla.2001.2813
Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and Memory, 5*(2), 179-187. https://doi.org/10.1037/0278-7393.5.2.179
Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When Looks Are Everything: Appearance Similarity Versus Kind Information in Early Induction. *Psychological Science*, *18*(2), 179–185. https://doi.org/10.1111/j.1467-9280.2007.01869.x
Son, J. Y., Smith, L. B., & Goldstone, R. L. (2008). Simplicity and generalization: Short-cutting abstraction in children's object categorizations. *Cognition*, *108*(3), 626-638.
Spencer, J. P., Perone, S., Smith, L. B., & Samuelson, L. K. (2011). Learning words in space and time: Probing the mechanisms behind the suspicious-coincidence effect. *Psychological Science*, *22*(8), 1049-1057.
Taylor, K., & Rohrer, D. (2009). The effects of interleaved practice. *Applied Cognitive Psychology, 24*(6), 837-848. https://doi.org/10.1002/acp.1598
Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*(5), 529-536.
Twomey, K. E., Ranson, S. L., & Horst, J. S. (2014). That's More Like It: Multiple Exemplars Facilitate Word Learning. *Infant and Child Development*, *23*(2), 105–122. https://doi.org/10.1002/icd.1824
Vlach, H. A. (2014). The spacing effect in children's generalization of knowledge: Allowing children time to forget promotes their ability to learn. *Child Development Perspectives*, *8*(3), 163-168.
Vlach, H. A. (2016). How we categorize objects is related to how we remember them: The shape bias as a

memory bias. *Journal of experimental child psychology*, *152*, 12-30.
Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, *109*(1), 163-167.
Vlach, H., & Sandhofer, C. M. (2012). Fast mapping across time: Memory processes support children's retention of learned words. *Frontiers in psychology*, *3*, 17291.
Vlach, H. A., & Kalish, C. W. (2014). Temporal dynamics of categorization: forgetting as the basis of abstraction and generalization. *Frontiers in psychology*, *5*, 1021.
Vlach, H. A., Kaul, M., Hosch, A., & Lazaroff, E. (2022). Attending less and forgetting more: Dynamics of simultaneous, massed, and spaced presentations in science concept learning. *Journal of Applied Research in Memory and Cognition, 11*(3), 361-373. https://doi.org/10.1016/j.jarmac.2021.10.007
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. https://doi.org/10.3758/s13421-010-0063-y
Wang, J., Liu, Z., Xing, Q., & Seger, C. A. (2020). The benefit of interleaved presentation in category learning is independent of working memory. *Memory, 28*(2), 285-292. https://doi.org/10.1080/09658211.2019.1705490
Yan, V. X., & Sana, F. (2021). The Robustness of the Interleaving Benefit. *Journal of Applied Research in Memory and Cognition, 10*(4), 589-602. https://doi.org/10.1016/j.jarmac.2021.05.002
Yan, V. X., Bjork, E. L., & Bjork, R. A. (2016). On the difficulty of mending metacognitive illusions: A priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General,145*(7),918-933. https://doi.org/10.1037/xge0000177
Yan, V. X., Soderstrom, N. C., Seneviratna, G. S., Bjork, E. L., & Bjork, R. A. (2017). How should exemplars be sequenced in inductive learning? Empirical evidence versus learners' opinions. *Journal of Experimental Psychology: Applied, 23*(4), 403-416. https://doi.org/10.1037/xap0000139