Journal of Applied Research in Memory and Cognition

ISSN: 2211-3681

2024, Vol. 13, No. 3, 329–341
https://doi.org/10.1037/mac0000193

EMPIRICAL ARTICLE

General Knowledge and Detailed Memory Benefit From Different Training Sequences

Sharon M. Noh[1], Robert A. Bjork[2], and Alison R. Preston[3, 4, 5]
[1] The University of Texas at Austin, United States
[2] Department of Psychology, University of California, Los Angeles, United States
[3] Department of Psychology, The University of Texas at Austin, United States
[4] Department of Neuroscience, The University of Texas at Austin, United States
[5] Center for Learning and Memory, The University of Texas at Austin, United States

Real-world decisions require understanding generalities (e.g., sorting an album collection by genre) and the ability to remember specific events (where one acquired a particular album). Discriminating between broad categories versus individual events requires contrasting features at different levels of specificity and therefore have different representational demands. Here, we used a within-participant design to test the hypothesis that different training protocols (blocked or interleaved order) would have dissociable impacts on the representation of generalities and specifics. On each trial, participants viewed a painting from one of 12 artists along with information about its unique location. Category generalization and source memory were tested immediately and after 1 week. Interleaving enhanced generalization, while blocking improved incidental learning of episodic details. Furthermore, category knowledge remained stable over time, whereas episodic details declined. These results indicate that interleaving and blocking optimize discrimination at different levels of specificity, with differential impacts on inferring generalities and remembering specific events.

General Audience Summary

In real-world learning experiences, individuals can learn specific facts, for example, which museum collection holds a specific painting by Monet, while simultaneously extracting generalities, such as the general style of Monet's brush strokes and color palette. Both types of learning are important in classroom settings. However, they may benefit from different types of instruction. Knowing where Monet's specific paintings are located requires forming memories that discriminate between each painting and its location from all other paintings by Monet. Learning how Monet's painting style differs from Manet's requires discriminating between paintings from each artist. Here, we tested whether different learning sequences supported discriminative contrast of features required for remembering specifics and generalities. We found that interleaving paintings by different artists improved individuals' ability to learn each artist's general style. In contrast, blocking paintings by individual artists (i.e., presenting all of one artist's paintings in sequence before moving on to the next artist) improved incidental memory for the specific locations of each painting by that artist. These findings suggest that instructors have an opportunity to emphasize one type of learning over another, specific details and general knowledge, by manipulating how they present information to students in classroom settings.

Keywords: concept learning, episodic memory, inductive reasoning, mnemonic discrimination, retention

Supplemental materials: https://doi.org/10.1037/mac0000193.supp

Sharon M. Noh https://orcid.org/0000-0003-0002-7759
Sharon M. Noh is now at the Department of Cognitive Sciences, University of California, Irvine, United States.

The authors have no conflicts of interest to declare with respect to their authorship or the publication of this article. This research was supported in part by the National Institute of Mental Health of the National Institutes of Health (Grant R01MH100121) to Alison R. Preston, the National Science Foundation through the National Science Foundation Graduate Research Fellowship Program to Sharon M. Noh, and the National Institute of Neurological Disorders and Stroke of the National Institutes of Health (Grant F31NS105353) to Sharon M. Noh.

Sharon M. Noh played a lead role in conceptualization, data curation, formal analysis, investigation, methodology, project administration, and writing–original draft, a supporting role in funding acquisition, and an equal role in validation, visualization, and writing–review and editing. Robert A. Bjork played a supporting role in conceptualization, investigation, methodology, and supervision and an equal role in writing–review and editing. Alison R. Preston played a lead role in funding acquisition, resources, and supervision and an equal role in conceptualization, investigation, methodology, validation, visualization, and writing–review and editing.

Correspondence concerning this article should be addressed to Sharon M. Noh, Department of Cognitive Sciences, University of California, Irvine, 2201 Social and Behavioral Sciences Gateway, Irvine, CA 92697-5100, United States, or Alison R. Preston, Department of Neuroscience, The University of Texas at Austin, 100 East 24th Street, Austin, TX 78712-0805, United States. Email: nohsm@uci.edu or apreston@utexas.edu

In real-world contexts, we encounter situations requiring the recall of specific event details and others that may involve drawing upon general knowledge. Acquisition of episodic details and general category knowledge has typically been studied separately, but in natural learning contexts, it is often possible to extract both general knowledge and specific details from the same learning event. For instance, an art student may encounter the *Doni Tondo* for the first time and use her general knowledge of Renaissance artists to infer that the work was painted by Michelangelo while simultaneously encoding a detailed memory linking the *Doni Tondo* with her visit to the Uffizi Gallery in Florence. An open question is whether the acquisition of general versus episodic knowledge benefits from similar learning conditions.

Learning general category knowledge can be optimized by manipulating the order or *sequence* of category exemplars during the study (Brunmair & Richter, 2019). Interleaving exemplars from different categories improves learners' ability to classify new exemplars of those categories (Brunmair & Richter, 2019), a finding that is counterintuitive to learners' preference for blocking exemplars by category (Kornell & Bjork, 2008). Despite participants' intuitions that blocking might facilitate encoding of category-defining features, experimental findings support a *discriminative-contrast hypothesis*—that extracting information about the differences between exemplars of different categories, rather than the similarities among members of the same category, may be more critical for category learning (for a review and meta-analysis, see Brunmair & Richter, 2019).

Discrimination is also a central tenet in episodic memory theory (Bakker et al., 2008; Criss & Koop, 2015; Hulbert & Norman, 2015; Lohnas et al., 2018). Studies suggest that forming discriminable representations is fundamental to overcoming memory interference, allowing for the successful retrieval of episodic details that differentiate similar episodes (Chanales et al., 2021; Favila et al., 2016; McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997). While placing similar episodes close together in time might increase the potential for interference when remembering the specific details that differentiate them (Noh et al., 2023), blocking similar episodes could provide an opportunity for learners to directly compare and contrast their distinct individual features. Indeed, learners' attention is often biased toward the features that differ across successive stimulus presentations (Barron et al., 2016; Horner & Henson, 2008; Reed Hunt & Worthen, 2006; Summerfield et al., 2008; Yassa & Stark, 2011). Thus, just as category generalization may benefit from interleaving exemplars from different categories, memory for episodic details may improve when similar events are juxtaposed during encoding (i.e., blocking same-category exemplars together).

Here, we propose that discriminative contrast is a universal learning principle that benefits all types of knowledge, whether general or specific. We predict learners naturally compare and contrast adjacent trials to reduce competition or interference, thus enhancing discrimination between episodes learned in close temporal proximity. Importantly, our framework suggests that the optimal sequence for learning generalities and specifics will differ. To improve individuals' broad category knowledge, a learning sequence should juxtapose examples from different categories to allow contrast of category features that differ. In contrast, to improve memory precision of examples within a category, one should study exemplars from the same category in close temporal proximity to highlight what makes each exemplar unique. In other words, sequencing could be used to promote local contrast of stimulus features to be optimal for the desired type of learning.

To test our hypothesis that general category knowledge and detailed memory may benefit from different learning sequences—interleaved and blocked, respectively, we adapted a paradigm from Kornell and Bjork (2008), in which participants learned to categorize paintings from unfamiliar artists (categories) studied in an interleaved or blocked fashion within subjects. For each participant, six artists' paintings were presented in a blocked sequence (all paintings from one artist were presented), whereas the remaining six artist paintings were interleaved with each other (one painting from each of the six different artists was presented). Each individual painting was also paired with a unique location (real-world landmarks).

By incorporating multiple stimuli on each study trial (artist name, painting, location), our task mimics a naturalistic learning scenario in which learners might encode various event features that can be beneficial for enhancing general category knowledge and/or episodic details. We tested knowledge acquisition immediately after study and after a 1-week delay using a category generalization test which involved categorizing novel (unstudied) paintings by the studied artists. We also included two memory tests targeting different levels of specificity. The detailed memory test asked participants to remember the specific location of each painting, which required precise source memory of individual study trials and differentiation between studied paintings by the same artist. The general recognition test, on the other hand, only required participants to remember the artist name associated with a location, without needing to recall any specific paintings. We hypothesized that interleaving would enhance category generalization by facilitating between-category discrimination, whereas blocking would improve detailed memory by highlighting the distinctiveness of episodes within the same category.

Experiment 1: Determining Optimal Learning Schedules for Different Levels of Knowledge Specificity

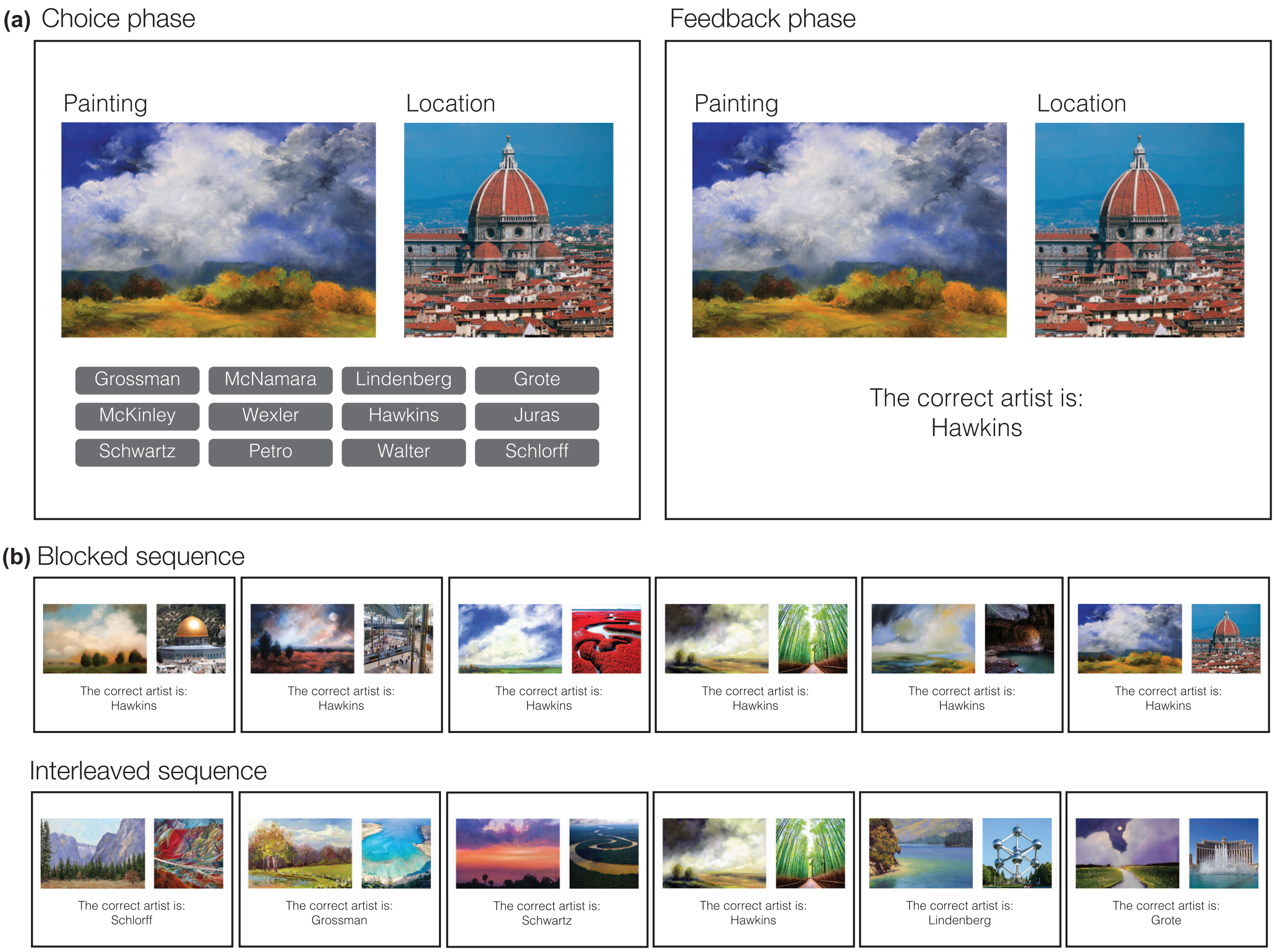
Method

Experiment 1 used a within-participants design to manipulate the presentation sequence—blocked or interleaved—in which participants studied category exemplars to examine its effects on both immediate and delayed tests of knowledge specificity. Participants completed two experimental sessions that were 1 week ± 1 day apart. The first session consisted of the learning phase (Figure 1) as well as immediate tests of general and detailed knowledge (Figure 2). The second session assessed both generalization and detailed memory after a delay, using a set of test items unique from those used for the immediate memory tests.

Participants

Participants were recruited from the University of Texas at Austin's Psychology undergraduate subject pool. Participants received course credit for their participation in the study. Informed consent was obtained for each participant prior to any data collection in accordance with the University of Texas Institutional Review Board. There were 51 participants in Experiment 1 (age = 18–31 years, $M \pm SD = 19.96 \pm 2.96$, 16 male). The target sample size was determined from a power analysis conducted on the results from Kornell and Bjork (2008): The reported effect size for the interleaving

Figure 1
Example Training Phase Trials and Learning Sequence Schematic



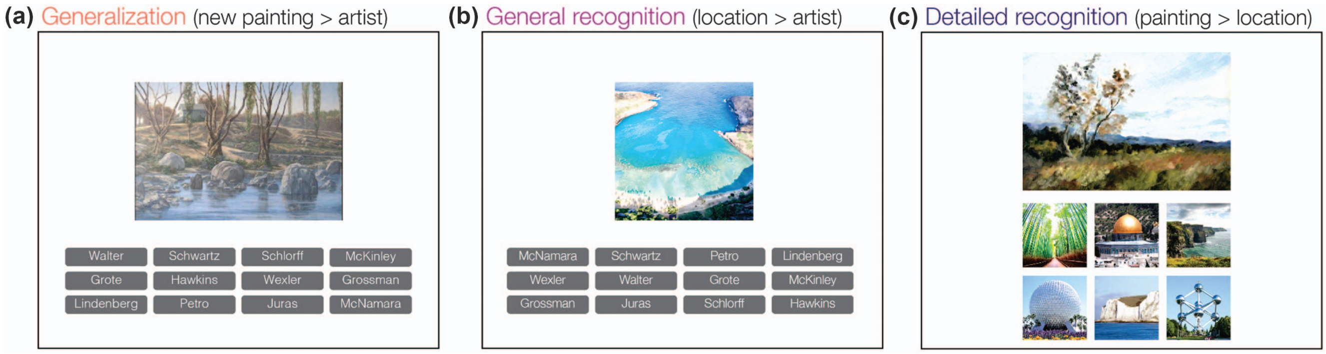
Note. (a) During training, participants view a painting paired with a location, along with 12 artist name choices. Participants learn to identify the artist of each painting through trial and error. (b) Artist paintings were either shown in a blocked or interleaved sequence using a within-subjects design. In the blocked sequence, all six paintings of the same artist were presented in a random order. In the interleaved sequence, one painting from each of the six interleaved artists was presented in a random order. For each participant, six artists' paintings were studied in a blocked sequence, whereas the other six artists' paintings were studied in an interleaved sequence. The procedure was counterbalanced between subjects such that some participants started with a blocked sequence (BIIBBIIBBIIB) and others started with an interleaved sequence (IBBIIBBIIBBI).

benefit on induction performance was $\eta_p^2 = .39$. With a significance criterion of $\alpha = .05$ and power $= .95$, our power analysis (using G*Power) revealed that the minimum sample size needed with this effect size is $N = 8$ to see an interleaving effect for generalization performance in our task using a within-subjects design. However, in the present study, we further wanted to assess an interaction of learning schedule on general category knowledge and detailed memory across a 1-week retention interval; thus, we aimed for a minimum of double the sample size required based on the power analysis (minimum sample size of 16 for each analysis). After the initial piloting of our task design, we determined that approximately 50 participants would be necessary to adequately assess retention of knowledge across our multiple memory tests, especially given the difficulty of the "Detailed Recognition" source memory task that resulted in only 19 usable data points for one particular analysis (see the Analyses subsection of the Results section in Experiment 1 for analysis details and exclusions). Of the 51 participants, data from one participant were excluded for technical issues with the software that arose in Session 2, leaving a final sample size of 50 for the target analysis (age = 18–31 years, $M \pm SD = 20 \pm 2.98$, 16 male).

Materials

The images used for the experiments were 18 landscape paintings by each of the 12 artists (Grossman, Grote, Hawkins, Juras, Lindenberg, McKinley, McNamara, Petro, Schlorff, Schwartz, Walter, and Wexler). Six unique paintings by each artist were used for the learning and memory test phases, and six unique paintings were used for each generalization test (immediate and delayed). Paintings were cropped and resized to 500 × 333 pixels. In addition,

Figure 2
Example Trials From Each Test Phase



Note. In Session 1, participants completed a training phase, followed by three different tests of knowledge specificity. Participants returned 1 week later (Session 2) to complete three delayed tests of knowledge specificity. (a) In the generalization test, participants were shown a novel (unstudied) painting and had to select the correct artist from 12 artist name options. (b) In the general recognition test, participants were shown a previously studied landmark and had to identify the artist whose work was at that location. (c) In the detailed recognition test, participants had to select the location of a previously studied painting.

we used images of 72 famous landmarks (36 natural and 36 manmade images cropped and resized to 333 × 333 pixels), each of which was paired with a unique painting during the learning phase. The landmark images also served as memory probes and choices during the source memory tests. After completing the experiment, participants were asked about their preexperimental familiarity with each artist. No participants reported any prior knowledge of the artists or their paintings.

Procedure

Session 1 Learning. During each learning trial, participants saw one painting next to an image of a unique location at the top of the screen (Figure 1a). Participants were told that their goal was to learn each artist's painting style and that in a later phase, they would be shown a new painting and would have to identify the artist based on what they learned about each artist's style. To facilitate learning the artists' style and ensure participants' sustained attention during the study, participants were asked to select the name of the artist to which each painting could be attributed from all 12 possibilities (presented as options at the bottom of the screen) on each trial. Participants had up to 10 s to make their selection. After making their choice, the correct artist's name was presented as feedback (2 s) before moving on to the next painting–landmark pair. Across learning, the screen position associated with each artist's name was shuffled trial to trial to avoid spatial response biases during learning. During the learning phase, participants were exposed to six unique paintings for each of the 12 artists (36 different paintings in total).

Each painting was paired with a unique landmark during learning. Participants were instructed that the landmarks corresponded to the nearest recognizable location where the painting could be found (e.g., a painting shown next to a picture of the Washington Monument indicated that the painting was located near that landmark). Participants were told to use the landmarks as an aid for learning but were not explicitly instructed that they would be tested on the associations between paintings and landmarks. Assignment of landmarks to individual paintings was determined randomly for each participant with a broad constraint that of the six paintings for each artist, half were paired with man-made landmarks, and the other half were paired with natural landmarks.

To assess the effects of different learning sequences on the acquisition of general (category) knowledge and detailed (source location) memory, we manipulated study sequence within subjects by presenting the 12 artist categories using one of two possible sequence types: blocked and interleaved (Figure 1b). For each participant, six of the artists (determined randomly for each participant) were assigned to each sequence type (six blocked artists, six interleaved artists). For blocked sequences, all six paintings from one artist were presented sequentially (Figure 1b, "Blocked sequence"). For interleaved sequences, one painting from each of six artists was presented sequentially (Figure 1b, "Interleaved sequence"). In both types of sequence, the same set of six paintings per artist was shuffled and repeated a total of five times, to allow for better encoding of episodic details for each trial. The six blocked sequences (each with six paintings from a unique artist) and six interleaved sequences (each with one unique painting from the same six artists) were presented to each participant in a counterbalanced order: BIIBBIIBBIIB or IBBIIBBIIBBI.

Generalization Test. After the training phase, participants were tested on the general category-level knowledge they acquired during training. During the generalization test (Figure 2, "Generalization"), participants were shown 72 new (unstudied) paintings (six new paintings per artist) and asked to identify the correct artist based on what they learned about each artist's painting style during the learning phase. Participants were shown a painting and had up to 15 s to select among the 12 artist name options. After participants made their selection (or after 15 s), they were shown the correct artist's name for 2 s before moving on to the next painting. The 72 test trials were pseudorandomly presented as six blocks of 12 paintings, with each block consisting of one novel painting from each artist presented in a random order.

Detailed Recognition Test. After the generalization test, participants were tested on their memory for specific details presented during the training phase. In the detailed recognition test (Figure 2, "Detailed Recognition"), participants were asked to select the location that was paired with specific paintings that were studied in the training phase. Participants were shown a painting and asked to identify the location of the painting (i.e., the landmark that was paired with that painting during training). Participants were tested on half of the paintings they studied during the training phase (the other half was reserved for the retention test a week later) without any feedback. For each of the 36 tested paintings, participants had 15 s to choose among six previously seen landmarks. The paintings for this phase were randomly selected (three studied paintings per artist) and presented in a pseudorandom order across three blocks, with each block consisting of one painting from each artist presented in a random order. The response choices for each painting were landmarks that were paired with paintings by the same artist during training, which made the detailed recognition test highly specific and particularly difficult, as learners had to remember the specific painting's location and could not use general category-level information to eliminate any of the landmark options provided.

General Recognition Test. After the detailed recognition test, participants then completed a more general test of memory. The general recognition test differed from the detailed recognition test in that participants were asked to identify the artist associated with a given location from the training phase, without having to recall the specific details of the painting that was associated with that location. During the general recognition test (Figure 2, "General Recognition"), participants were shown a previously studied landmark and asked to remember the name of the artist whose painting was located at that landmark. Participants were tested on the 36 landmarks that were paired with the paintings tested in the previous phase. The 36 landmarks used were presented in a pseudorandom order, without feedback, across three blocks, with each block consisting of one painting from each artist presented in a random order. Participants were shown a picture of a landmark and had up to 15 s to select from 12 artist name options to make their choice. While this general recognition test still probes memory, it is less dependent on remembering the specific details of any individual painting. Thus, the general recognition test serves as an easier and more general test of memory relative to the detailed recognition test.

Tests of Knowledge Retention. Participants returned 1 week later and completed delayed tests of generalization, detailed recognition, and general recognition. The delayed tests were identical to the immediate tests in their structure (e.g., number of trials, timing, shuffling). During the generalization test, participants were tested on 72 new paintings (six new paintings per artist) and asked to identify the artist based on what they learned a week prior. For the detailed and general recognition tests, participants were tested on the remaining half of the paintings that were studied during the training phase in Session 1.

Analyses

A 2 (Schedule: Blocked vs. Interleaved) × 2 (Session: Immediate vs. Delayed) within-subjects analysis of variance (ANOVA) was conducted for each test (generalization, detailed recognition, and general recognition) to assess how training with different schedules impacted generalized knowledge and memory specificity across sessions. Participants who showed chance levels of overall learning performance (regardless of schedule) on any immediate test were excluded from the analysis of that particular test, as we are unable to adequately assess retention of knowledge on Session 2 in cases of no initial learning. For each test phase, we used a binomial probability distribution to determine chance levels of performance using the number of trials in that test phase and cumulative probability of success from guessing on every trial. Importantly, chance levels of performance on each test were assessed using average performance, collapsed across the blocked and interleaved conditions. Because our sequence manipulation was within subjects, we were thus able to average condition data within participants when considering exclusions. Therefore, exclusions were determined based on overall performance, were blind to condition-level effects (i.e., blocked vs. interleaved), and thus did not bias our results. After applying these exclusion criteria to our 50 participants, we were left with 44 data points in the generalization test (accuracy ≥ 0.153), 19 in the detailed recognition test (accuracy ≥ 0.277), and 37 in the general recognition test (accuracy ≥ 0.166). These numbers satisfied the minimum sample size requirements as determined by our power analysis (see the Participants subsection of the Method section in Experiment 1). Materials and analysis code for this study are available on GitHub (https://github.com/prestonlab/Paintings_JARMAC).

Results

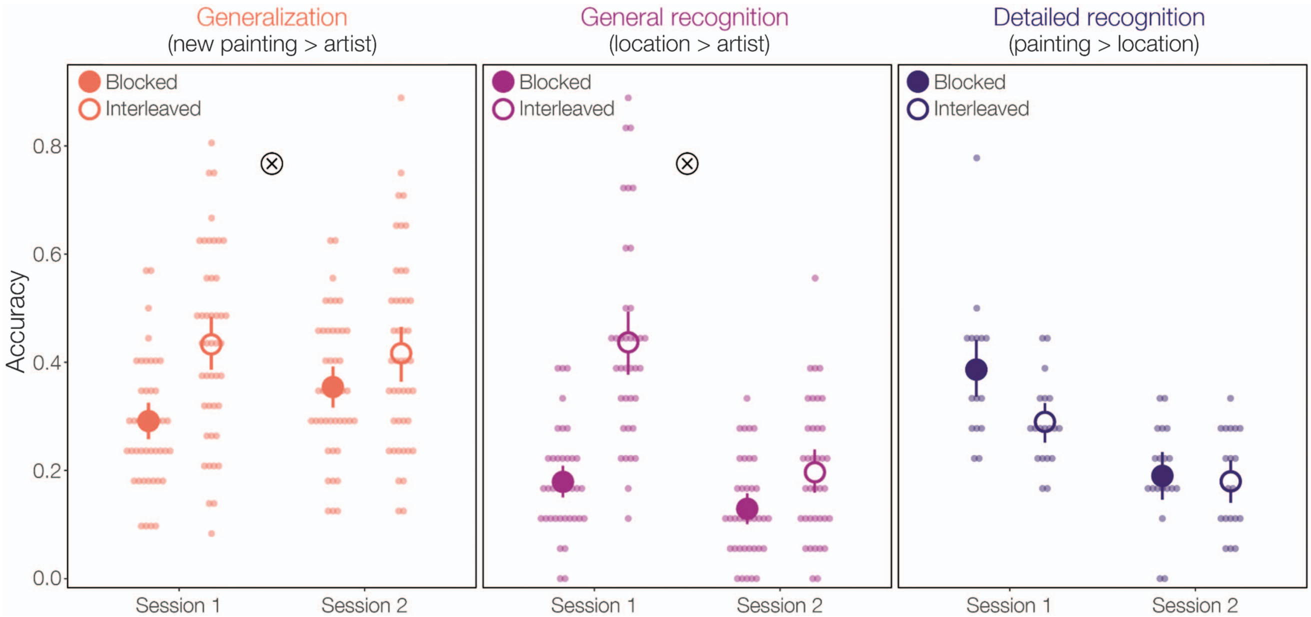
Generalization Test

The average generalization performance for each condition is presented in Figure 3 ("Generalization"). A 2 (Schedule: Blocked, Interleaved) × 2 (Session: Immediate, Delayed) within-subjects ANOVA revealed a main effect of schedule, $F(1, 43) = 22.96$, $MSE = .022$, $p < .001$, $\eta_p^2 = .35$, such that accuracy was higher for identifying novel paintings of interleaved artists ($M = .43$, $SD = .18$) relative to blocked artists ($M = .32$, $SD = .12$). There was also a marginally significant effect of session, $F(1, 43) = 3.90$, $MSE = .008$, $p = .055$, $\eta_p^2 = .08$, such that generalization accuracy showed improvements in Session 2 ($M = .39$, $SD = .15$) relative to Session 1 ($M = .36$, $SD = .15$). Notably, there was a Schedule × Session interaction, $F(1, 43) = 19.74$, $MSE = .004$, $p < .001$, $\eta_p^2 = .32$. Interleaved training led to superior generalization accuracy ($M = .43$, $SD = .18$) relative to blocked training on an immediate test ($M = .28$, $SD = .12$), $t(43) = 6.28$, $p < .001$. However, the interleaving benefit was attenuated after a delay, as the blocked condition showed significant improvements in performance on Session 2 ($M = .35$, $SD = .13$) relative to Session 1 ($M = .28$, $SD = .11$), $t(43) = 3.21$, $p = .003$, whereas the interleaved condition showed no significant difference in performance from Session 1 ($M = .43$, $SD = .17$) to Session 2 ($M = .42$, $SD = .18$), $t(43) = 1.03$, $p = .307$. The overall improvement in the blocked group may be due to the fact that feedback was provided during the generalization test, thereby giving the blocked group an opportunity to gain additional training at an interleaved study schedule and improve over the course of the generalization test. Similarly, Kornell and Bjork (2008) found that when participants received feedback during their test of category induction (generalization), the blocked group showed improvements across the four test blocks.

To minimize the influence of feedback and more accurately examine retention of category knowledge, we conducted another 2 (Schedule: Blocked, Interleaved) × 2 (Session: Immediate, Delayed) within-subjects ANOVA on the generalization test but restricted our analysis to the last 12 test trials from Session 1 and the first 12 test trials of Session 2. This restricted analysis revealed two main effects: Interleaved training ($M = .42$, $SD = .26$) was still superior to blocked

Figure 3

Results of the Generalization (Left), General Recognition (Middle), and Detailed Recognition (Right) Tests in Experiment 1



Note. Large circles indicate group means, and small circles indicate individual participant performance. Error bars indicate 95% confidence intervals of the mean.

training ($M = .31$, $SD = .19$), $F(1, 43) = 11.16$, $MSE = 0.041$, $p = .002$, $\eta_p^2 = .21$, but the main effect of session revealed that performance does not in fact improve across a delay. We instead observe a significant decrease in generalization performance when comparing performance at the end of Session 1 ($M = .40$, $SD = .23$) relative to the beginning of Session 2 ($M = .32$, $SD = .21$), $F(1, 43) = 10.39$, $MSE = 0.041$, $p = .002$, $\eta_p^2 = .20$.

General Recognition Test

The average performance for each condition on the general recognition test is presented in Figure 3 ("General Recognition"). A 2 (Schedule: Blocked, Interleaved) × 2 (Session: Immediate, Delayed) within-subjects ANOVA revealed a main effect of schedule, $F(1, 36) = 49.16$, $MSE = .021$, $p < .001$, $\eta_p^2 = .58$; accuracy was higher when participants had to remember artists associated with a given landmark for interleaved artists ($M = .32$, $SD = .16$) relative to blocked artists ($M = .15$, $SD = .09$). There was also a main effect of session, $F(1, 36) = 54.54$, $MSE = .014$, $p < .001$, $\eta_p^2 = .60$; performance was higher in Session 1 ($M = .31$, $SD = .14$) relative to Session 2 ($M = .16$, $SD = .11$). There was also a Schedule × Session interaction, $F(1, 36) = 22.32$, $MSE = .014$, $p < .001$, $\eta_p^2 = .38$. This interaction revealed that there was a large performance benefit of interleaved training ($M = .44$, $SD = .19$) relative to blocked training ($M = .18$, $SD = .10$) on Session 1, $t(36) = 7.12$, $p < .001$, $d = 1.17$, but the interleaving benefit decreased significantly in Session 2, $t(36) = 3.02$, $p = .005$, $d = .50$. The attenuation of the interleaving benefit was due to the fact that the interleaved condition suffered greater performance declines from Session 1 ($M = .44$, $SD = .19$) to Session 2 ($M = .20$, $SD = .13$), $t(36) = 6.91$, $p < .001$, $d = 1.14$, whereas the magnitude of performance declines from Session 1 ($M = .18$, $SD = .10$) to Session 2 ($M = .13$, $SD = .09$) was much smaller in the blocked group, $t(36) = 2.70$, $p = .011$, $d = .44$, likely due to the fact that initial performance was already much closer to chance levels in the blocked condition relative to the interleaved condition.

Detailed Recognition Test

The average performance for each condition of the detailed recognition test is presented in Figure 3 ("Detailed Recognition"). A 2 (Schedule: Blocked, Interleaved) × 2 (Session: Immediate, Delayed) within-subjects ANOVA revealed a main effect of schedule; blocked training led to better performance ($M = .29$, $SD = .10$) relative to interleaved training ($M = .23$, $SD = .08$) on this test of item-level specificity, $F(1, 18) = 8.34$, $MSE = 0.014$, $p = .01$, $\eta_p^2 = .32$. There was also a main effect of session; participants performed better on Session 1 ($M = .34$, $SD = .10$) relative to Session 2 ($M = .19$, $SD = .09$), $F(1, 18) = 55.87$, $MSE = 0.014$, $p < .001$, $\eta_p^2 = .76$. Performance in Session 2 was no greater than chance for either the blocked ($M = .19$, $SD = .09$) or interleaved ($M = .18$, $SD = .09$) condition, suggesting that no detailed information was retained after a 1-week delay. The interaction between schedule and session was not significant, $F(1, 18) = 2.69$, $MSE = 0.014$, $p = .118$, $\eta_p^2 = .13$.

Experiment 2: Reexamination of Sequencing Effects in the Absence of Feedback

In Experiment 1, we saw an improvement in generalization performance from Session 1 to Session 2 (Figure 3, Generalization). This improvement is attributable to the fact that the generalization test in Experiment 1 provided feedback which led to improved induction performance in the blocked condition across test blocks (consistent with Kornell & Bjork, 2008). In the original Kornell and Bjork (2008) study, feedback provided during the generalization test phase served as an additional learning opportunity, particularly in the blocked condition, which showed significant performance improvements over the course of the test phase. Additional learning during the generalization test could be especially problematic in our task due to the fact that our experiment design added two subsequent tests of memory, as well as a retention component, whose results may have been contaminated by the additional learning opportunity provided by feedback during the generalization test. To eliminate this potential confound, we removed feedback during the generalization test and otherwise repeated our experiment in Experiment 2.

Method

Participants

Participants were recruited from the University of Texas at Austin's Psychology undergraduate subject pool. Participants received course credit for their participation in the study. There were 51 participants in Experiment 2 (age = 18–22, $M \pm SD = 18.94 \pm 1.10$, 10 male). The sample size was selected to match that of Experiment 1. We aimed for approximately 50 participants to adequately assess the retention of knowledge for different learning schedules across multiple memory tests. This resulted in 21 usable data points for the detailed recognition test analysis (see the Analyses subsection of the Results section in Experiment 2 for analysis details and exclusion criteria).

Materials and Procedure

The materials and procedures in this experiment were nearly identical to those in Experiment 1. Experiment 2 differed from Experiment 1 solely in that no feedback was provided during either session of the generalization test.

Analyses

A 2 (Schedule: Blocked, Interleaved) × 2 (Session: Immediate, Delayed) within-subjects ANOVA was conducted for each test (generalization, detailed recognition, and general recognition) to assess the effects of training with the different schedules across sessions on generalized knowledge and specificity. Participants who showed chance levels of overall learning performance (regardless of schedule) on any immediate test were excluded from the analysis of that particular test, as we are unable to adequately assess retention of knowledge on Session 2 in cases of no initial learning. For each test phase, we used a binomial probability distribution to determine chance levels of performance using the number of trials in that test phase and cumulative probability of success from guessing on every trial. Again, these exclusions were applied to average performance collapsed across the blocked and interleaved conditions and thus performed in an unbiased manner. After applying these exclusion criteria to our 51 participants, we were left with 44 data points in the generalization test (accuracy ≥ 0.153), 21 in the detailed recognition test (accuracy ≥ 0.277), and 40 in the general recognition test (accuracy ≥ 0.166). These numbers satisfied the minimum requirements as determined by our power analysis (see the Participants subsection of the Method section in Experiment 2).

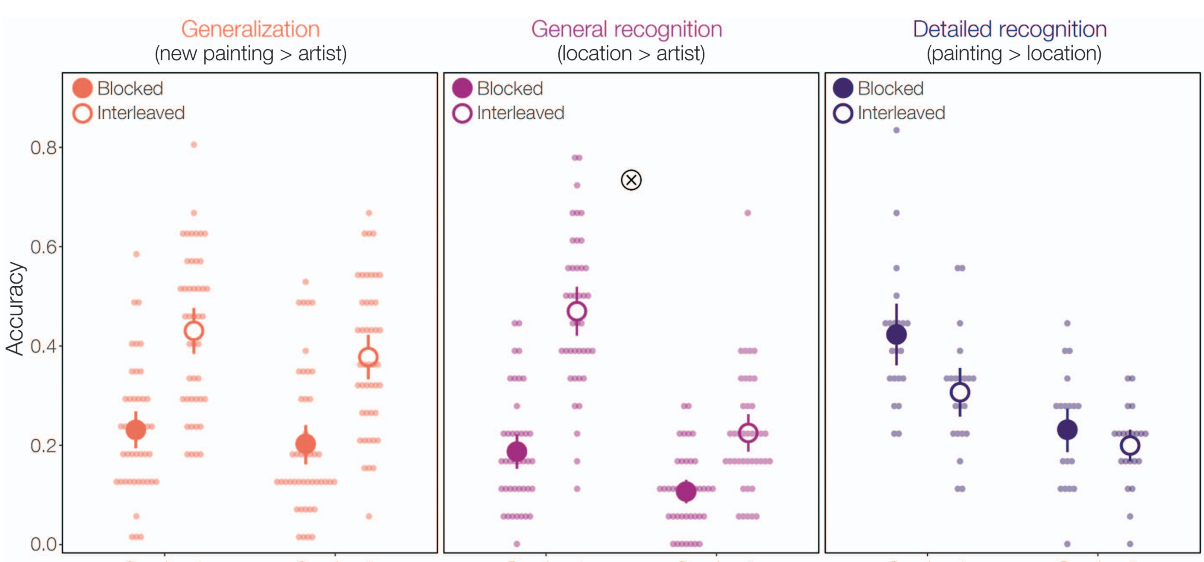
Results

Generalization Test

The average generalization performance for each condition is presented in Figure 4 (left, "Generalization"). A 2 (Schedule:

Figure 4

Results of the Generalization (Left), General Recognition (Middle), and Detailed Recognition (Right) Tests in Experiment 2



Note. Large circles indicate group means, and small circles indicate individual participant performance. Error bars indicate 95% confidence intervals of the mean.

Blocked, Interleaved) × 2 (Session: Immediate, Delayed) within-subjects ANOVA revealed two main effects. There was a main effect of schedule such that accuracy was higher for identifying novel paintings of interleaved artists (M = .40, SD = .15) relative to blocked artists (M = .32, SD = .13), $F(1, 43) = 48.80$, MSE = .005, $p < .001$, $\eta_p^2 = .53$. There was also a main effect of the session such that generalization accuracy was higher in Session 1 (M = .33, SD = .14) than in Session 2 (M = .29, SD = .15), $F(1, 43) = 17.45$, MSE = .005, $p < .001$, $\eta_p^2 = .29$. There was no Schedule × Session interaction, $F(1, 43) = 1.81$, MSE = .005, $p = .185$, $\eta_p^2 = .04$.

General Recognition Test

The average performance for each condition on the general recognition test is presented in Figure 4 (middle, "General Recognition"). A 2 (Schedule: Blocked, Interleaved) × 2 (Session: Immediate, Delayed) within-subjects ANOVA revealed a main effect of schedule, $F(1, 39) = 122.49$, MSE = .013, $p < .001$, $\eta_p^2 =$ .76. Accuracy was higher for the interleaved condition (M = .34, SD = .14) relative to the blocked condition (M = .15, SD = .10). There was also a main effect of session, $F(1, 39) = 131.93$, MSE = .008, $p < .001$, $\eta_p^2 = .77$; performance was higher in Session 1 (M = .33, SD = .14) than Session 2 (M = .16, SD = .10). There was also a Schedule × Session interaction, $F(1, 39) = 17.38$, MSE = .016, $p < .001$, $\eta_p^2 = .31$. This interaction revealed that there was a performance benefit of interleaved training initially (M = .47, SD = .16) relative to blocked training (M = .19, SD = .12) during Session 1, $t(39) = 9.10$, $p < .001$, $d = 1.44$. However, this interleaving benefit was attenuated in Session 2, $t(39) = 8.77$, $p < .001$, $d = .83$, as the interleaved condition suffered greater performance declines from Session 1 to Session 2, $M_{\text{diff}} = .25$, SD = .18, $t(39) = 8.77$, $p < .001$, $d = 1.39$, relative to the blocked condition, $M_{\text{diff}} = .08$, SD = .13, $t(39) = 3.97$, $p < .001$, $d = .63$.

Detailed Recognition Test

Average performance for each condition of the detailed recognition test is presented in Figure 4 (right, "Detailed Recognition"). A 2 (Schedule: Blocked, Interleaved) × 2 (Session: Immediate, Delayed) within-subjects ANOVA again revealed a main effect of schedule. Blocked training led to better performance (M = .32, SD = .13) relative to interleaved training (M = .25, SD = .10) on the detailed recognition test, $F(1, 20) = 9.51$, MSE = 0.018, $p = .006$, $\eta_p^2 = .32$. There was also a main effect of session; participants performed better on Session 1 (M = .36, SD = .13) than Session 2 (M = .21, SD = .10), $F(1, 20) = 46.51$, MSE = 0.018, $p < .001$, $\eta_p^2 = .70$. Performance in Session 2 was no greater than chance for either the blocked (M = .23, SD = .11) or interleaved (M = .20, SD = .08) condition, suggesting that no detailed information was retained after a 1-week delay. The interaction between schedule and session was not significant, $F(1, 20) = 1.75$, MSE = 0.018, $p = .201$, $\eta_p^2 = .08$.

Combined Memory Test Results Across Experiments

Experiment 1 differed from Experiment 2 in that feedback was provided during the generalization test. The general recognition and detailed recognition tests, however, were identical in design across Experiments 1 and 2 and had the same pattern of results. Since the generalization test came before the two tests of memory, we wanted to address possible concerns about how feedback during the generalization test may have provided participants with additional learning opportunities (particularly in the blocked condition) and contaminated performance on the memory tests that came after it. To determine if the feedback provided during Experiment 1 influenced the outcomes of the subsequent tests that followed, we repeated our ANOVA for the general recognition and detailed recognition tests by combining both data sets and added "Experiment" as a between-subjects factor to determine if there were significant differences between results across the two experiments. For each analysis (general recognition test, detailed recognition test), we conducted a 2 (Session: Immediate, Delayed) × 2 (Schedule: Blocked, Interleaved) × 2 (Experiment: 1, 2) mixed-effects ANOVA. Schedule (blocked, interleaved) and session (immediate, delayed) were defined as within-subjects factors, whereas experiment (Experiment 1, Experiment 2) was defined as a between-subject factor.

For the general recognition test, we found no significant difference in performance across the two experiments, $F(1, 75) = .313$, $MSE =$ .022, $p = .577$, $\eta_p^2 = .004$, nor were there any significant interactions that included the experiment factor. The combined analysis across Experiments 1 and 2 reinforced the Schedule × Session interactions that were reported previously (in Experiments 1 and 2) for the general recognition test, $F(1, 75) = 39.33$, MSE = .015, $p < .001$, $\eta_p^2 = .34$.

For the detailed recognition test, we again found no significant difference in performance between the two experiments, $F(1, 38) =$ 1.96, MSE = .012, $p = .169$, $\eta_p^2 = .049$, nor were there any significant interactions that included the Experiment (between-subjects) factor. The combined analysis across Experiments 1 and 2 reinforced the significant main effect of schedule (blocked > interleaved), $F(1, 38) =$ 17.32, MSE = .016, $p < .001$, $\eta_p^2 = .31$, and significant main effect of session (immediate > delayed), $F(1, 38) = 100.36$, MSE = .016, $p <$ .001, $\eta_p^2 = .73$. However, the combined analysis revealed a significant Schedule × Session interaction that was not present in the individual reporting of results in Experiment 1 or 2, $F(1, 38) = 4.28$, $MSE =$ .016, $p < .045$, $\eta_p^2 = .10$. The interaction was driven by the fact that there was initially a statistically significant benefit of blocked training (M = .40, SD = .13) relative to interleaved training (M = .30, SD = .10) on the detailed recognition test in Session 1, $t(39) = 3.70$, $p =$ .001, $d = .58$. In contrast, the blocked and interleaved conditions did not differ significantly in performance on Session 2, $M_{\text{diff}} = .02$, $SD =$ .13, $t(39) = 1.15$, $p = .256$, $d = .182$, as both conditions showed performance levels that were no greater than chance.

Interaction Between Sequence and Knowledge Type on Immediate Test

Both Experiments 1 and 2 showed different patterns of results with respect to training sequences at various levels of knowledge specificity. Thus, we conducted a 3 (Test Type: Generalization, General Recognition, Detailed Recognition) × 2 (Session: Immediate, Delayed) × 2 (Schedule: Blocked, Interleaved) within-subjects ANOVA to assess any potential interactions that included the test type variable. Participants were only included for this analysis if they passed the exclusion criteria for all three tests (above-chance performance on immediate test for all three tests), which left 15 participants in Experiment 1 and 18 participants in Experiment 2.

Scores within each test type (generalization, general recognition, and detailed recognition) were z-scored across sessions (Sessions 1 and 2) to allow for more direct comparisons in performance across the three tests. This analysis revealed a significant three-way interaction for Test Type × Session × Schedule in both Experiment 1, $F(2, 28) = 3.62$, $MSE = 2.09$, $p = .040$, $\eta_p^2 = .21$, and Experiment 2, $F(2, 34) = 5.07$, $MSE = 2.34$, $p = .012$, $\eta_p^2 = .23$. The three-way interaction is further analyzed below.

In Experiment 1, the two-way Test Type × Schedule interaction was significant, $F(1, 14) = 14.84$, $MSE = 8.13$, $p < .001$, $\eta_p^2 = .52$, consistent with our pattern of results showing different tests benefit from different training schedules (Figure 3). Post hoc tests restricted to two test types (generalization vs. general recognition, general recognition vs. detailed recognition, and generalization vs. detailed recognition) revealed that the Test Type × Schedule interaction was significant for generalization versus detailed recognition, $F(1, 14) = 12.72$, $MSE = 7.31$, $p = .003$, $\eta_p^2 = .48$, and for general recognition versus detailed recognition, $F(1, 14) = 39.27$, $MSE = 15.55$, $p < .001$, $\eta_p^2 = .74$. This interaction is driven by the fact that detailed recognition benefitted from a blocked training schedule, whereas general recognition and generalization benefitted from an interleaved schedule. The Test Type × Schedule interaction was not significant for generalization versus general recognition, as both tests showed an interleaving benefit, $F(1, 14) = 2.28$, $MSE = 1.53$, $p = .153$, $\eta_p^2 = .14$.

The two-way Test Type × Schedule interaction in Experiment 2 was also found to be significant, $F(1, 17) = 20.84$, $MSE = 15.08$, $p < .001$, $\eta_p^2 = .55$, consistent with our results demonstrating the differential effects of training schedules on general and detailed knowledge (Figure 4). Post hoc tests restricted to two test types (generalization vs. general recognition, general recognition vs. detailed recognition, and generalization vs. detailed recognition) revealed the same pattern of results as in Experiment 1: The Test Type × Schedule interaction was significant for generalization versus detailed recognition, $F(1, 17) = 16.36$, $MSE = 18.22$, $p = .001$, $\eta_p^2 = .49$, and for general recognition versus detailed recognition, $F(1, 17) = 47.40$, $MSE = 26.29$, $p < .001$, $\eta_p^2 = .74$, but not for generalization versus general recognition, $F(1, 17) = 1.47$, $MSE = .74$, $p = .242$, $\eta_p^2 = .08$. Overall, the significant Test Type × Schedule interactions found across both Experiments 1 and 2 reveal that different levels of knowledge benefit from different training schedules.

Interaction Between Sequence and Knowledge Type Across a 1-Week Delay

The two-way Test Type × Session interaction was also significant in both Experiment 1, $F(1, 14) = 19.75$, $MSE = 8.27$, $p < .001$, $\eta_p^2 = .59$, and Experiment 2, $F(1, 17) = 10.08$, $MSE = 3.60$, $p < .001$, $\eta_p^2 = .37$, in Session 2. Post hoc tests restricted to two test types (generalization vs. general recognition, general recognition vs. detailed recognition, and generalization vs. detailed recognition) revealed that the Test Type × Session interaction was significant for generalization versus general recognition in both Experiment 1, $F(1, 14) = 21.59$, $MSE = 6.29$, $p < .001$, $\eta_p^2 = .61$, and Experiment 2, $F(1, 17) = 11.26$, $MSE = 4.05$, $p = .004$, $\eta_p^2 = .40$. The differences in retention for generalization versus general recognition are being driven by the fact that general recognition suffers much greater performance declines over time relative to generalization performance, which is relatively stable.

The Test Type × Session interaction was also observed for generalization versus detailed recognition in both Experiment 1, $F(1, 14) = 40.00$, $MSE = 16.21$, $p < .001$, $\eta_p^2 = .74$, and Experiment 2, $F(1, 17) = 16.72$, $MSE = 6.47$, $p = .001$, $\eta_p^2 = .50$. This interaction is driven by the fact that detailed recognition suffers large performance declines after a 1-week delay, whereas generalization performance is more stable. The Test Type × Schedule interaction was not significant for general recognition versus detailed recognition, as both tests show performance declines after 1 week in Experiment 1, $F(1, 14) = 4.12$, $MSE = 2.30$, $p = .062$, $\eta_p^2 = .23$, and Experiment 2, $F(1, 17) = .86$, $MSE = .28$, $p = .365$, $\eta_p^2 = .05$. Collectively, the Test Type × Session interactions found across Experiments 1 and 2 suggest that while memory processes (general recognition and detailed recognition) show performance decrements after a delay, general category knowledge (category induction) is much more stable.

Replicated Interaction Between Learning Sequences and Knowledge Specificity

To address possible concerns related to the small sample sizes ($N = 15$ in Experiment 1 and $N = 18$ in Experiment 2) used in the previous analyses of implied interactions between learning sequence and our three measures of knowledge specificity, we analyzed an additional data set that was collected as part of another study as a replication analysis for the Test Type × Session interaction. The data set was collected from a single session and used six (instead of 12) artist categories. In this study, the assignment of participants to the learning schedule was a between-participant factor, so 52 participants learned about the artists in a blocked order, and 48 participants were assigned to the interleaved condition. Both participant groups completed the same induction and source memory tests, as described in Experiments 1 and 2. Detailed methods can be found in the Supplemental Material. Critically, no performance-based exclusion criteria were applied to the data, as this version of the experiment did not have an appropriate way to exclude based on overall performance regardless of learning condition (due to the learning sequence manipulation being between participants) nor did it have a retention component to assess.

We conducted a 3 (Test Type: Generalization, General Recognition, Detailed Recognition) × 2 (Schedule) mixed ANOVA. This analysis replicated the Test Type × Schedule interaction observed in Experiments 1 and 2 of the present study, $F(2, 196) = 18.075$, $MSE = 0.502$, $p < .001$, $\eta_p^2 = .16$. Post hoc tests restricted to two test types (generalization vs. general recognition, general recognition vs. detailed recognition, and generalization vs. detailed recognition) revealed that the Test Type × Schedule interaction was significant for generalization versus detailed recognition, $F(1, 98) = 37.79$, $MSE = .896$, $p < .001$, $\eta_p^2 = .27$, and for general recognition versus detailed recognition, $F(1, 98) = 18.72$, $MSE = .575$, $p < .001$, $\eta_p^2 = .16$. This interaction is driven by the fact that detailed recognition benefitted from a blocked training schedule, whereas general recognition and generalization benefitted from an interleaved schedule. The Test Type × Schedule interaction was not significant for generalization versus general recognition, as both tests showed an interleaving benefit, $F(1, 98) = 1.23$, $MSE = 0.035$, $p = .272$, $\eta_p^2 = .012$. Thus, we were able to replicate the same pattern of results that we found in Experiments 1 and 2 with respect to the effects of sequencing on different levels of knowledge using an independent data set even when the learning sequence was manipulated between subjects.

General Discussion

Here, we demonstrate that, when learning natural categories, the acquisition of category-defining features and episodic details benefits from different sequences. Our study design assesses the acquisition and retention of both category knowledge and recognition memory acquired from the same learning experience and quantifies the interaction between sequence and knowledge specificity. While category generalization and general recognition memory benefitted from interleaving, highly detailed memory about individual training episodes (i.e., painting locations) benefitted from blocking. Additionally, we found retention differences in knowledge specificity: While category generalization is largely retained, both general and detailed recognition memory suffer greater losses with time.

Our generalization results align with prior research demonstrating an interleaving benefit for natural categories (for a meta-analysis, see Brunmair & Richter, 2019). Our generalization results corroborate claims that interleaving facilitates the encoding of features that differentiate categories (Birnbaum et al., 2013; Carvalho & Goldstone, 2015; Kang & Pashler, 2012; Kornell & Bjork, 2008). Such discriminative contrast may be particularly important when artists have similar styles, consistent with work demonstrating interleaving benefits when there is high between-category similarity (Carvalho & Goldstone, 2014).

For instance, while differentiating a Michelangelo painting from a Picasso may be easy given their dramatically different artistic styles, discerning Michelangelo's work from those of other Renaissance painters like Raphael is more challenging. Because our task used landscape artists, most of whom used an impressionistic style, our categories likely had high perceptual overlap and between-category similarity. By highlighting discriminative category features, interleaving may reduce ambiguity between artists with similar painting styles and facilitate category generalization.

Reducing *episodic* ambiguity, however, may require contrasting features different from those used for categorical judgments. For example, an art student's ability to recognize the *Doni Tondo* as a painting by Michelangelo would not necessarily help her remember that the *Doni Tondo* is located in the Uffizi Gallery. Our data indicate that for those who successfully learned the painting–location associations, recalling individual painting locations benefits from blocking. By extension, remembering that the *Doni Tondo* is located in Florence, while the *Entombment* is located in London, may benefit from learning these facts juxtaposed (blocked) rather than separated by time (interleaved). Seeing the two Michelangelo paintings in close proximity may support the encoding of details that differentiate the two paintings, thus supporting superior recognition.

Consistent with this idea, work in human memory has shown that high event similarity triggers adaptive "repulsion" mechanisms, which bias neural codes to store more differentiated representations to reduce memory interference and improve accuracy (Chanales et al., 2017, 2021; Favila et al., 2016). One intriguing possibility is that enhanced discrimination of episodic details may come at a cost to acquiring general category knowledge. While our study was not designed to quantify trade-offs between acquiring different types of knowledge, these interactions could be examined in future studies.

Our results also introduced interesting nuances with respect to blocked versus interleaved benefits across different levels of knowledge specificity. For instance, our general and detailed recognition tests are both tests of episodic source memory, yet each benefitted from different learning schedules. General (but not detailed) recognition benefitted from interleaving, similar to the generalization test. One interpretation, based on our local discrimination framework, might explain why only detailed recognition showed blocking benefits. In our task, each trial consisted of three components: an artist name, painting, and location. In a blocked sequence, only two components (painting and location) differed between trials, as the artist was the same across the 30 consecutive trials within a single artist's block.

If learners are naturally inclined to discriminate between successive trials and/or similar episodes (consistent with local discrimination), learners would be biased to encode *differences* between successive trials (paintings and locations), rather than similarities (artist name). This bias would lead to superior performance in detailed recognition, which directly tested the painting–location association. On the other hand, in an interleaved sequence, all three components (an artist name, painting, and location) varied from trial to trial, as paintings from different artists were presented in sequence. Given we provided explicit instructions that emphasized the importance of the artist's name (e.g., the goal was to learn each artist's style), it is reasonable to assume that learners prioritized learning of the artist information. As a result, both generalization (which tested category knowledge of artists' painting style) and general recognition (which tested the mapping of artist names to locations) may have benefitted from interleaved training, which contained variable artist information across trials. Because artist information does not vary across trials in the blocked condition, it may be difficult to override the natural bias to encode trial-by-trial differences (e.g., painting–location information) despite explicit instruction to prioritize the artist information.

Our results may appear to run counter to a popular sequencing theory referred to as the *sequential attention theory* of categorization (Carvalho & Goldstone, 2017). This theory posits that blocking highlights within-category similarities, whereas interleaving emphasizes between-category differences. Strict application of this theory might predict blocking would bias learners in our task to encode similarities across episodes and impair detailed recognition, as no distinct episodic details would be encoded during the study. Yet our results show that blocking showed superior detailed recognition performance (in those who demonstrated incidental encoding of episodic details), which would not occur if participants were encoding within-category similarities during learning. However, our study differs from those validating the sequential attention theory in that there were potentially competing learning objectives: While participants were instructed to learn category-level information, learners were also presented with additional episodic details that could compete with the encoding of goal-relevant features. When faced with a problem that does not have competing information, the *sequential attention theory* may hold true because learners apply strategies consistent with the learning objective and selectively attend to goal-relevant features (e.g., "these items belong to the same category so I should identify their similarities").

While our instructions emphasized learning at the category level, our task was designed to assess incidental learning of episodic details. When examining individuals who show reliable learning of detailed episodic information, we saw that blocking enhanced differentiation of highly similar exemplars (i.e., paintings from the same artist). Although the blocking benefit was observed in a subset of the original sample (in Experiment 1) who showed evidence of learning above chance, this effect was replicated across three independent data sets

(Experiment 1, Experiment 2, and Supplemental Material online methods), suggesting that the incidental learning of episodic details appears to benefit from blocking.

When interpreting our results, it is important to consider interactions between sequencing and spacing effects in memory (Bjork et al., 2013; Cepeda et al., 2008). Earlier studies referred to interleaving as "spaced" and blocking as "massed" learning (Kornell & Bjork, 2008; Kornell et al., 2010; Kang & Pashler, 2012), but recent distinctions disambiguate the terms such that "massing" and "spacing" refer to *temporal* learning manipulations, whereas "blocking" and "interleaving" refer to the sequence or arrangement of learning materials. Blocking/interleaving typically applies to category learning problems, where materials can be grouped by (blocked) or mixed across (interleaved) different categories (Brunmair & Richter, 2019; Dunlosky et al., 2013). Temporal (e.g., "spacing") and sequencing (e.g., "interleaving") manipulations often covary (i.e., interleaved category exemplars are spaced in time, whereas blocked category exemplars are massed) and likely interact to influence learning.

In this study, blocking/interleaving was applied at the *category level*, consistent with prior work using similar designs (Kornell & Bjork, 2008; Kornell et al., 2010). As such, the sequence manipulation also covaried with massed/spaced practice: Blocked paintings were massed within that artist's temporal block and never revisited, whereas interleaved artists were spaced out and revisited (albeit with different exemplars) throughout the experiment. While interactions between temporal and sequencing dynamics were outside the scope of this article, our task design is similar to those who have directly examined interactions between spacing and sequencing of categories (Birnbaum et al., 2013; Kang & Pashler, 2012). Based on those prior studies, we believe the interleaving benefit in category generalization is largely driven by discriminative contrast, rather than spacing. Nonetheless, spacing and sequencing may interact differently across our three tests of knowledge specificity. To our knowledge, the interactions between spacing and sequencing have not been systematically examined in episodic memory and thus warrant further investigation.

In our task, unique trials were equally spaced, regardless of sequence. At the exemplar level, each exemplar was spaced similarly across repetitions (roughly once every six trials) regardless of the sequence (blocking/interleaving). In fact, the exemplar-level sequence only differed in that adjacent exemplars were from the same (blocked) or different (interleaved) categories. The consistent spacing of individual exemplars highlights how same- versus different-category comparisons between adjacent episodes can dramatically change the features extracted from those individual events. By juxtaposing items from the same category, blocking may bias learners to encode subtle within-category differences between similar episodes and improve detailed recognition.

On the other hand, general recognition memory benefitted from interleaving. We provided one possible account for this difference: Blocking varied painting–location information, causing learners to ignore the invariant artist information necessary for successful general recognition performance. Another account may be that temporal spacing reduces memory interference between episodes that share overlapping information (Noh et al., 2023; Schlichting et al., 2015) by separating competing information across time.

While temporal spacing might improve general recognition by reducing memory interference, spacing may be insufficient at resolving memory interference for detailed recognition. Adding temporal spacing between exemplars has been shown to eliminate the interleaving benefit for category generalization by disrupting discriminative-contrast processes (Birnbaum et al., 2013; Kang & Pashler, 2012). Temporal spacing might similarly eliminate the blocking benefit for detailed episodic memory by making it harder to recall specific details that differentiate each painting. Thus, our findings provide a novel contribution to the memory literature by demonstrating how different sequences can be used as a learning tool to enhance memory for episodes that may otherwise be vulnerable to memory interference. The interactions between spacing and sequencing could benefit from further investigation, both in the domain of concept learning and memory. For instance, one interesting avenue of research could be to consider how blocked categories might benefit from repeated spaced practice (e.g., revisiting/repeating blocked exemplars after studying other categories).

It is also important to recognize that factors other than discriminability might influence optimal learning sequences. Studies have shown that the best sequence for learning can vary based on multiple factors such as category structure (Noh et al., 2016), training (active vs. passive) format (Carvalho & Goldstone, 2015), learning problem (Flesch et al., 2018, 2023), serial position effects (Ge et al., 2021), representational goals (Schlichting et al., 2015; Zhou et al., 2023), or alignment between task instructions and goals (Abel, Brunmair, et al., 2021; Abel, 2024; Abel, Niedling, et al., 2021; Miyatsu et al., 2020; Noh et al., 2014). While some factors such as active versus passive learning format are likely to have relatively minor influences on our results (e.g., our active learning manipulation may have exaggerated the interleaving benefit seen in generalization), others may have had a larger role in contributing to our findings. While our study was not designed to test these other variables, we acknowledge that discriminability alone may not fully explain our results.

For instance, the present study did not account for primacy and/or recency effects (Gershberg & Shimamura, 1994; Helstrup, 1979) on memory performance nor did we control for differences in between- and within-category transitions, which may differ in the potential to perform discriminative contrast across learning episodes. Specifically, it is possible that serial position effects may bias encoding for the first and last categories studied in a blocked fashion, whereas no such bias exists in the interleaved categories, which are intermixed throughout the experiment. Relatedly, another contributing factor to the interleaving benefit may be that interleaving allows for better sampling of the range of the stimulus space, whereas the first category in a blocked sequence would not provide adequate information about the range of stimuli and features that make up the problem space. Additionally, it is possible that interleaving improves category generalization not necessarily because it enhances between-category discriminability, but simply because there are more category-level transitions relative to blocking (during which there is a transition only when moving from one category block to the next), which provide more opportunities to compare and contrast at the category level. Future study designs should provide better controls and interpretations for these inherent variances that exist in many blocked/interleaved category designs to better isolate the mechanisms that drive sequencing effects.

Finally, our results also revealed differences in the retention of knowledge across different levels of specificity. We found that categorization performance is largely retained over time, whereas episodic details are forgotten (Brainerd & Reyna, 1990, 1998; Sekeres et al., 2016). The comparison between category generalization and general recognition in Experiment 2 best illustrates this dissociation between how different levels of knowledge are retained (Figure 4). On the immediate test, participants show similar performance levels in generalization and general recognition. However, after a delay, general recognition suffers sharp performance declines, whereas generalization shows minimal decrements. These findings suggest a dissociation between the mechanisms that support generalization and episodic memory. Future studies could examine the neural mechanisms associated with acquiring and retaining general versus detailed knowledge to better assess this possibility.

Our study holds significant practical implications, particularly in academic settings. In real-world learning scenarios, there are times when the learning objective is to master general concepts and others when it is necessary to memorize specific details. Although learners typically prefer structured study methods such as grouping (blocking) learning materials by concept (Bjork et al., 2013; Tauber et al., 2013), our findings suggest that learners should strategically sequence their learning to effectively enhance the type of knowledge (e.g., general vs. specific) consistent with one's learning goal (e.g., mastering concepts vs. memory). We designed our task to mimic naturalistic learning by combining multiple stimuli on each trial to create learning episodes composed of multiple features, which allowed learners to acquire episodic details and conceptual knowledge simultaneously. Our approach reveals how different types of knowledge are acquired and retained and suggests that learning can be optimized through training regimes that emphasize discrimination at varying levels of specificity.

While interleaving benefits (e.g., "distributed practice") are increasingly being examined in classroom settings (Agarwal & Agostinelli, 2020; Hartwig et al., 2022; Hopkins et al., 2016; Samani & Pan, 2021; Sana & Yan, 2022), these studies measure conceptual mastery, as the goal of most courses is conceptual understanding rather than rote memorization. In real-world contexts, however, there are many cases when one should remember specific episodic details. For example, a lawyer may need to differentiate between similar witness accounts to accurately attribute the unique details provided by each witness to prepare his case. Or perhaps you meet a close friend's extended family and want to remember each person's name for when you see them again. Our findings suggest that sequencing can be used as a learning tool to highlight discriminative features depending on one's learning goals.

In summary, our results demonstrate how different levels of knowledge specificity are optimally acquired and retained over time, and how different learning conditions are beneficial for improving specific outcomes: Interleaving improves more general knowledge, such as general recognition memory and category knowledge, whereas blocking enhances discrimination between same-category exemplars. Our results highlight the importance of discrimination as a general learning principle for both learning and memory and show how local discrimination can promote differentiation at multiple levels of specificity. Furthermore, we found that while category generalization is largely preserved over time, episodic memory content is mostly lost after a delay. These results illustrate the importance of leveraging the appropriate learning strategies that enhance performance and knowledge retention for various tasks.

References

Abel, R. (2024). Some fungi are not edible more than once: The impact of motivation to avoid confusion on learners' study sequence choices. *Journal of Applied Research in Memory and Cognition*. *13*(1), 103–112. https://doi.org/10.1037/mac0000107

Abel, R., Brunmair, M., & Weissgerber, S. C. (2021). Change one category at a time: Sequence effects beyond interleaving and blocking. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(7), 1083–1105. https://doi.org/10.1037/xlm0001003

Abel, R., Niedling, L. M., & Hänze, M. (2021). Spontaneous inferential processing while reading interleaved expository texts enables learners to discover the underlying regularities. *Applied Cognitive Psychology*, *35*(1), 258–273. https://doi.org/10.1002/acp.3761

Agarwal, P. K., & Agostinelli, A. (2020). Interleaving in math: A research-based strategy to boost learning. *American Educator*, *44*(1), 24–28.

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, *319*(5870), 1640–1642. https://doi.org/10.1126/science.1152882

Barron, H. C., Garvert, M. M., & Behrens, T. E. J. (2016). Repetition suppression: A means to index neural representations using BOLD? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *371*(1705), 1–14. https://doi.org/10.1098/rstb.2015.0355

Birnbaum, M. S., Kornell, N., Bjork, E. L., & Bjork, R. A. (2013). 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

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, *10*(1), 3–47. https://doi.org/10.1016/0273-2297(90)90003-M

Brainerd, C. J., & Reyna, V. F. (1998). When things that were never experienced are easier to "remember" than things that were. *Psychological Science*, *9*(6), 484–489. https://doi.org/10.1111/1467-9280.00089

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

Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & Cognition*, *42*(3), 481–495. https://doi.org/10.3758/s13421-013-0371-0

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

Cepeda, N. J., Vul, E., Rohrer, D., Wixted, J. T., & Pashler, H. (2008). Spacing effects in learning: A temporal ridgeline of optimal retention. *Psychological Science*, *19*(11), 1095–1102. https://doi.org/10.1111/j.1467-9280.2008.02209.x

Chanales, A. J. H., Oza, A., Favila, S. E., & Kuhl, B. A. (2017). Overlap among spatial memories triggers repulsion of hippocampal representations. *Current Biology*, *27*(15), 2307–2317.e5. https://doi.org/10.1016/j.cub.2017.06.057

Chanales, A. J. H., Tremblay-McGaw, A. G., Drascher, M. L., & Kuhl, B. A. (2021). Adaptive repulsion of long-term memory representations is triggered by event similarity. *Psychological Science*, *32*(5), 705–720. https://doi.org/10.1177/0956797620972490

Criss, A. H., & Koop, G. J. (2015). Differentiation in episodic memory. In J. G. W. Raaijmakers, A. H. Criss, R. L. Goldstone, R. M. Nosofsky,

& M. Steyvers (Eds.), *Cognitive modeling in perception and memory: A festschrift for Richard M. Shiffrin* (pp. 112–125). Psychology Press.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, *14*(1), 4–58. https://doi.org/10.1177/1529100612453266

Favila, S. E., Chanales, A. J. H., & Kuhl, B. A. (2016). Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nature Communications*, *7*(1), Article 11066. https://doi.org/10.1038/ncomms11066

Flesch, T., Balaguer, J., Dekker, R., Nili, H., & Summerfield, C. (2018). Comparing continual task learning in minds and machines. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(44), E10313–E10322. https://doi.org/10.1073/pnas.1800755115

Flesch, T., Saxe, A., & Summerfield, C. (2023). Continual task learning in natural and artificial agents. *Trends in Neurosciences*, *46*(3), 199–210. https://doi.org/10.1016/j.tins.2022.12.006

Ge, Y., Li, F., Li, X., & Li, W. (2021). What is the mechanism underlying the interleaving effect in category induction: An eye-tracking and behavioral study. *Frontiers in Psychology*, *12*, Article 770885. https://doi.org/10.3389/fpsyg.2021.770885

Gershberg, F. B., & Shimamura, A. P. (1994). Serial position effects in implicit and explicit tests of memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(6), 1370–1378. https://doi.org/10.1037/0278-7393.20.6.1370

Hartwig, M. K., Rohrer, D., & Dedrick, R. F. (2022). Scheduling math practice: Students' underappreciation of spacing and interleaving. *Journal of Experimental Psychology: Applied*, *28*(1), 100–113. https://doi.org/10.1037/xap0000391

Helstrup, T. (1979). *Serial position effects in short-term memory of serial and free recall lists*. University of Bergen Norway.

Hopkins, R. F., Lyle, K. B., Hieb, J. L., & Ralston, P. A. S. (2016). Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, *28*(4), 853–873. https://doi.org/10.1007/s10648-015-9349-8

Horner, A. J., & Henson, R. N. (2008). Priming, response learning and repetition suppression. *Neuropsychologia*, *46*(7), 1979–1991. https://doi.org/10.1016/j.neuropsychologia.2008.01.018

Hulbert, J. C., & Norman, K. A. (2015). Neural differentiation tracks improved recall of competing memories following interleaved study and retrieval practice. *Cerebral Cortex*, *25*(10), 3994–4008. https://doi.org/10.1093/cercor/bhu284

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology*, *26*(1), 97–103. https://doi.org/10.1002/acp.1801

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

Lohnas, L. J., Duncan, K., Doyle, W. K., Thesen, T., Devinsky, O., & Davachi, L. (2018). Time-resolved neural reinstatement and pattern separation during memory decisions in human hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(31), E7418–E7427. https://doi.org/10.1073/pnas.1717088115

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review*, *105*(4), 724–760. https://doi.org/10.1037/0033-295X.105.4.734-760

Miyatsu, T., Nosofsky, R. M., & McDaniel, M. A. (2020). Effects of specific-level versus broad-level training for broad-level category learning in a complex natural science domain. *Journal of Experimental Psychology: Applied*, *26*(1), 40–60. https://doi.org/10.1037/xap0000240

Noh, S. M., Cooper, K. W., Stark, C. E. L., & Bornstein, A. M. (2023). *Multi-step inference across the lifespan can be improved with individualized memory interventions*. PsyArXiv. https://doi.org/10.31234/osf.io/3mhj6

Noh, S. M., Yan, V. X., Bjork, R. A., & Maddox, W. T. (2016). Optimal sequencing during category learning: Testing a dual-learning systems perspective. *Cognition*, *155*, 23–29. https://doi.org/10.1016/j.cognition.2016.06.007

Noh, S. M., Yan, V. X., Vendetti, M. S., Castel, A. D., & Bjork, R. A. (2014). Multilevel induction of categories: Venomous snakes hijack the learning of lower category levels. *Psychological Science*, *25*(8), 1592–1599. https://doi.org/10.1177/0956797614535938

Reed Hunt, R., & Worthen, J. B. (Eds.). (2006). *Distinctiveness and memory*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169669.001.0001

Samani, J., & Pan, S. C. (2021). Interleaved practice enhances memory and problem-solving ability in undergraduate physics. *NPJ Science of Learning*, *6*(1), Article 32. https://doi.org/10.1038/s41539-021-00110-x

Sana, F., & Yan, V. X. (2022). Interleaving retrieval practice promotes science learning. *Psychological Science*, *33*(5), 782–788. https://doi.org/10.1177/09567976211057507

Schlichting, M. L., Mumford, J. A., & Preston, A. R. (2015). Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. *Nature Communications*, *6*(1), Article 8151. https://doi.org/10.1038/ncomms9151

Sekeres, M. J., Bonasia, K., St-Laurent, M., Pishdadian, S., Winocur, G., Grady, C., & Moscovitch, M. (2016). Recovering and preventing loss of detailed memory: Differential rates of forgetting for detail types in episodic memory. *Learning & Memory*, *23*(2), 72–82. https://doi.org/10.1101/lm.039057.115

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM-retrieving effectively from memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166. https://doi.org/10.3758/BF03209391

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. https://doi.org/10.1038/nn.2163

Tauber, S. K., Dunlosky, J., Rawson, K. A., Wahlheim, C. N., & Jacoby, L. L. (2013). Self-regulated learning of a natural category: Do people interleave or block exemplars during study? *Psychonomic Bulletin & Review*, *20*(2), 356–363. https://doi.org/10.3758/s13423-012-0319-6

Yassa, M. A., & Stark, C. E. L. (2011). Pattern separation in the hippocampus. *Trends in Neurosciences*, *34*(10), 515–525. https://doi.org/10.1016/j.tins.2011.06.006

Zhou, Z., Singh, D., Tandoc, M. C., & Schapiro, A. C. (2023). Building integrated representations through interleaved learning. *Journal of Experimental Psychology: General*, *152*(9), 2666–2684. https://doi.org/10.1037/xge0001415

Received June 27, 2023
Revision received June 12, 2024
Accepted July 8, 2024 ■