How Should Exemplars Be Sequenced in Inductive Learning? Empirical Evidence Versus Learners' Opinions

Article *in* Journal of Experimental Psychology Applied · August 2017
DOI: 10.1037/xap0000139

5 authors, including:

Veronica X Yan
University of Texas at Austin

Elizabeth Bjotk
University of California, Los Angeles

Robert A. Bjork
University of California, Los Angeles

JOURNAL OF EXPERIMENTAL PSYCHOLOGY: APPLIED IN PRESS
Article DOI: http://dx.doi.org/10.1037/xap0000139

How Should Exemplars Be Sequenced in Inductive Learning? Empirical Evidence Versus Learners' Opinions

Veronica X. Yan[1], Nicholas C. Soderstrom[2], Gayan S. Seneviratna[3], Elizabeth Ligon Bjork[3], and Robert A. Bjork[3]

[1]Department of Educational Psychology, University of Texas, Austin

[2]Department of Psychology, Dickinson College

[3]Department of Psychology, University of California, Los Angeles

Correspondence should be addressed to Veronica Yan, email: veronicayan@austin.utexas.edu

Acknowledgements

This research was conducted at the University of California, Los Angeles, and supported by Grant No. 29192G from the James S. McDonnell Foundation. We thank Robyn Barrios, Kelly Ching, Catherine Huang, and Kenese Liang for their help in collecting data, and the members of CogFog for their insightful feedback. Aspects of this research were presented at the 55th annual meeting of the Psychonomic Society, Long Beach, CA.

Abstract

The sequencing of exemplars during study can have a large effect on category or concept induction. Counter to learners' intuitions, interleaving exemplars from different categories is often more effective for learning the different underlying categories than is blocking all the exemplars by category (e.g., Kornell & Bjork, 2008). Prior research suggests that blocking and interleaving each support different aspects of induction: Interleaving appears to enhance between-category discrimination, whereas blocking appears to promote the learning of within-category commonalities. In Experiments 1 and 2, participants studied paintings by 12 artists and were asked to induce the different artists' painting styles. We explored whether hybrid schedules can leverage the benefits of both types of schedules, comparing blocked, interleaved, and three hybrid schedules—blocked-to-interleaved, interleaved-to-blocked, and mini-blocks. The mini-blocks and blocked-to-interleaved schedules were as effective, statistically, but not better than pure interleaving. The blocked schedule led to the worst performance. In Experiments 3 and 4, we explored participants' *a priori* beliefs by having them self-schedule hypothetical future category-learning tasks. Although participants demonstrated some metacognitive sophistication with respect to the relative benefits of blocked and interleaved study, pure interleaving was the least popular schedule, despite its being one of the most, effective schedules for learning.

Keywords: Induction, interleaving, category learning, metacognition, self-regulated learning

Public Significance Statement

The present studies follows up on a counterintuitive finding that has triggered great interest—that studying exemplars from different concepts intermixed (interleaving) leads to better learning than studying exemplars one concept at a time (blocking)—and examines whether there are benefits of hybrid schedules that combine elements of both interleaving and blocking. Although hybrid schedules did not produce significantly better inductive learning than did a purely interleaved schedule, they could be just as effective and offer a valuable practical benefit over that of interleaved study in the context of real-world learning: Learners who might otherwise employ blocking in their self-guided study activities might be far more willing to employ more effective hybrid schedules.

The ability to induce categories and concepts from exposures to multiple exemplars and to apply this learning to categorizing new exemplars is fundamental to our capacity to understand, perceive, and interact appropriately with objects and events in our environments. For example, children may form categories regarding which objects are edible versus inedible; geologists learn to classify rocks as igneous, metamorphic, or sedimentary; musicians learn how to classify music into different genres; medical students learn to distinguish between different types of skin rashes; and we all form (for better or worse) stereotypes about the people around us. The primary concern of the present research is how to optimize this critical type of learning. More specifically, it focuses on how to sequence the presentation of exemplars from different to-be-learned categories in order to optimize the later ability of learners to classify never-before-seen exemplars from those categories.

One study schedule that is very common in educational settings and consistent with both learners' and teachers' intuitions for how best to learn new categories, is blocking the presentation of exemplars by concept or category. In mathematics learning, for example, it seems compelling that problems of different types and the procedures for solving those problems should be discussed in a blocked way—that is, a given type of problem should be discussed, illustrated, and practiced before moving on to a different type of problem requiring different procedure and formula to solve. To organize one's teaching in this *one-thing-at-a-time fashion* seems highly reasonable and, in fact, is ubiquitous in education. From the standpoint of learners as well, such blocking apparently seems the logical way to go about learning different categories and concepts. Learners, for example, express a strong preference for studying exemplars blocked by category over interleaving exemplars from different categories when learning to categorize bird families (Tauber, Dunlosky, Rawson, Wahlheim, & Jacoby, 2013) or artists' paintings (Kornell & Bjork, 2008; Yan, Bjork, & Bjork, 2016), and when learning to play various piano melodies (Abushanab & Bishara, 2013).

So compelling is the intuition that blocking enhances category learning that the eminent educational psychologist, Ernst Rothkopf, was quoted as saying, “Spacing is the friend of recall, but the enemy of induction” (1977, personal communication, as reported in Kornell & Bjork, 2008). That statement captured not only what Ernst Rothkopf believed, but also the consensus among researchers that blocked study should facilitate learners noticing the commonalities that tie a category together. The statement was also consistent with a large literature supporting the importance of comparison processes for the learning, transfer, and understanding of deeper relational structures (Gentner & Namy, 1999; Gentner, Loewenstein, Thompson, & Forbus, 2009; Gick & Holyoak, 1983; Oakes & Ribar, 2005; Rittle-Johnson & Star, 2011).

The literature on the role of comparison processes in category learning has typically focused on making comparisons between exemplars within the same category. Benefits, however, can also be found from making comparisons (or contrasts) across different to-be-learned categories. Despite the intuitively compelling argument that spacing the study of exemplars should be the “enemy of induction,” when Kornell and Bjork (2008) compared the relative efficacy of sequentially blocking the exemplars of to-be-learned perceptual categories versus interleaving such exemplars, they obtained the opposite pattern of results. In their basic paradigm, Kornell and Bjork had participants study six paintings by each of a number of different artists either blocked by artist (i.e., all the paintings by a given artist were studied before moving onto the next artist’s paintings) or interleaved (i.e., paintings from different artists were mixed together during study). Unexpectedly, they found that interleaved study, not blocked study, led to better classification of which artist among the set of studied artists was responsible for new paintings. This basic finding has been replicated many times and with a variety of materials (e.g., Birnbaum, Kornell, Bjork, & Bjork, 2013; Kang & Pashler, 2012; Kornell, Castel, Eich, & Bjork, 2010; Rohrer, 2012; Rozenshtein, Pearson, Yan, Liu, & Toy, 2016; Sana, Yan, & Kim, 2017; Wahlheim, Dunlosky, & Jacoby, 2011).

Why Might Interleaving the Exemplars of Different Categories Enhance Learning?

At present, the most dominant explanation for why interleaving the presentation of exemplars from different categories, rather than blocking them, has led to better inductive learning in these studies rests on the following assumptions: (a) that interleaving enhances discrimination or contrast processing across the exemplars of different categories, whereas blocking supports the processing of commonalities across the exemplars within a category, and (b) encoding the differences between exemplars of different categories is more important in terms of classifying new exemplars than is encoding the commonalities across exemplars within a category, at least for the type of stimuli used in these studies. The argument is that for materials such as artists' paintings, bird and butterfly species, and so forth, learning to discriminate between the exemplars of different categories is more critical for later correct classification than is learning to recognize the features that are shared by the exemplars within a category. Consistent with this reasoning, Birnbaum et al. (2013) showed that when the spaces between interleaved presentations of exemplars of different categories (in this case, species of butterflies and birds) were filled with unrelated trivia questions, learners lost the ability to engage in contrast processing and inductive learning was impaired.

Similarly, Kang and Pashler (2012) found a benefit of contrast processing even in a case where spacing between exemplars was eliminated. In their paradigm, participants studied paintings by three different to-be-learned artists. These paintings were presented either one-by-one (sequential) or three-at-a-time (simultaneous). When one painting by each of the three different artists was presented simultaneously (highlighting between-category differences), participants' subsequent classification of new paintings was just as accurate as when the artists' paintings were presented sequentially and in an interleaved manner. On the other hand, when three paintings by the same artist were presented simultaneously (highlighting within-category commonalities), participants' subsequent classification performance was significantly worse than in either of the two aforementioned presentation conditions, and not different to that obtained when the paintings were studied sequentially, blocked by artist.

Other studies using more cognitive and less perceptual materials have also shown similar benefits of emphasizing the making of contrasts between different to-be-learned concepts. For instance, Day, Goldstone, and Hills (2010) showed that middle school students who compared examples from two confusable types of problems (negative and positive feedback loops) were better able to distinguish between them on a later test than were students who compared two examples of the same problem type. Similarly, Taylor and Rohrer (2010) demonstrated that students who intermixed practice of four different types of math problems (calculating faces, edges, corners, and angles of prisms) scored significantly higher on a later test than did students who practiced the problems blocked by type, and they also made significantly fewer errors in discriminating between the formulae.

Although seemingly implied by the research discussed so far, it is not the case that interleaved study schedules are always better than blocked study schedules for producing inductive learning. Carvalho and Goldstone (2014) have argued that the relative efficacy of interleaving versus blocking schedules depends on the structure of the categories to be induced. According to their attentional-bias hypothesis, when categories are highly similar and, therefore, discrimination between categories is relatively more important, interleaving should draw attention to these between-category differences and produce more effective inductive learning than blocking. On the other hand, when the within-category similarity is low, making it relatively difficult to spot the commonalities within the exemplars of a given category, blocking should produce more effective learning than interleaving. To test these assumptions, they developed two sets of blob-shaped stimuli, one in which the categories were of high within- and between-category similarity and one in which the categories were of low within- and between-category similarity. They found that (a) when both within- and between-category similarity was high, learning the classification rule (a particular notch in each shape defined each category) benefited from interleaving and (b) when both within- and between-category similarity was low, learning benefited from blocking. Similarly, Zulkiply and Burt (2013) found that hard-to-discriminate categories (category exemplars contained five category-irrelevant objects) benefited from interleaving and easy-to-discriminate categories (category exemplars contained only one irrelevant object) benefited from blocking.

Might a Hybrid Schedule Be Better than either Blocking or Interleaving?

To date, the literature seems to suggest that we should use interleaved study schedules when it seems important to promote between-category discriminations for effective category learning and use blocked study schedules when it seems important to promote within category comparisons for effective category learning. But, is this approach really optimal for promoting inductive learning in general? That is, do we need to choose between promoting within-category comparisons versus between-category discriminations? If both processes play important, although different, roles in inductive learning, perhaps there are ways to combine these two types of schedules that would be better than using just one or the other. One goal of the present research was to explore this possibility by assessing the effectiveness of several varieties of hybrid schedules compared to pure blocked and pure interleaved schedules. Theoretical reasons aside, there is also an important practical reason for exploring this question: Namely, given the importance of category and concept learning in education, it is important that we understand how best to structure the teaching of new concepts to students. While the existing research literature has largely compared purely interleaved and purely blocked sequences, the occurrence of hybrid sequences is likely to be common in the real-world. For example, although concepts are often initially taught in a blocked manner (e.g., several examples of the same concept consecutively), when exam time draws near, practice problems (and the exams themselves) are often interleaved.

Aside from the Actual Effectiveness of Different Schedules, What Do Learners Prefer?

The research of Kornell and Bjork (2008) was pivotal not only for demonstrating that superior category induction could be obtained with interleaved study as opposed to blocked study, but also for showing that even after performing better on a final classification test for artists that had been studied in an interleaved fashion, participants consistently and overwhelmingly reported that studying paintings in a blocked manner had been better for their learning. These metacognitive findings—that participants are, for the most part, unaware of the scheduling that better enhances their own learning—have been replicated in further research (e.g., Kornell, Castel, Eich, & Bjork, 2010; Zulkiply, McLean, Burt, & Bath, 2012) and, strikingly, it appears that learners often hold onto their beliefs that blocking is the more effective study strategy even when presented with information about their own better learning with interleaved study (Yan et al., 2016).

That learners hold onto their beliefs that blocking is more effective even when faced with their own better performance after interleaved study could be considered just an intriguing finding. But these metacognitive findings also have critical practical implications because such beliefs can affect learners' future study choices. Tauber, Dunlosky, Rawson, Wahlheim, and Jacoby (2013) have examined the effects of *a priori* theories that participants might bring with them into the lab about how best to master categorical learning. Across four different experiments, Tauber et al. (2013) gave participants the task of learning to classify different birds in terms of the family (e.g., a type of Jay, Wren, etc.) to which they belonged. Exemplars from the different families were presented one at a time, and the participants were allowed to select, on a trial-by-trial basis, from which family they wanted to see the next exemplar. When they decided they had studied enough exemplars, they could choose to terminate study and take the classification test.

In Tauber et al.'s (2013) analysis of how participants chose to expose different exemplars for study during the learning phase, blocking was defined as studying two or more exemplars from the same bird family consecutively and interleaving was strictly defined as switching between different bird families when selecting the next exemplar to study. Additionally, switches from one block of exemplars within the same family to another block of exemplars from a different family were not counted as interleaving. If a given participant selected more than half of the exemplars in a blocked manner, then that participant was designated as a "blocker." Across the four studies conducted, between 78-100% of participants were classified as blockers.

In sum, the relevant prior studies paint a picture in which the metacognitions of learners regarding how best to optimize their categorical or inductive learning is out of sync with the actual way they learn best. In the Kornell and Bjork (2008) studies, participants' metacognitions about how they learned best was diametrically opposite to how they actually learned best, and in the Tauber et al. (2013) studies, participants, given freedom to construct their own study schedules, tended largely to select less beneficial schedules (i.e., blocking) over more beneficial schedules (i.e., interleaving).

But perhaps these judgments and behaviors were somewhat forced on the participants by the paradigms or coding strategies used in these studies. To illustrate, consider Experiment 4 by Tauber et al. (2013). In this study, unlike in their first three studies, there was no initial "familiarization phase" in which participants were exposed to every exemplar in a random order (i.e., interleaved) to familiarize them with the type of material to be learned. In their Experiment 4, 48 unique exemplars were available for study—six exemplars for each of eight bird families. On average, participants studied a total of 107 exemplars, and the average run (i.e., the average number of exemplars from the same bird family that participants chose to study successively) was 3.7. What this pattern of results reflects is that participants were likely studying multiple, spaced out "blocks" of each bird family—or in other words, they were using some combination of blocked and interleaved study (as opposed to pure blocked study), and it remains an empirical question as to whether participants were in fact choosing to study the exemplars in an effective way. Similarly, Carvalho, Braithwaite, de Leeuw, Motz, and Goldstone (2016) allowed undergraduate students to choose their own sequence of practice questions on measures of central tendency (mean, median, mode) in an online homework tutorial. Students were asked to complete a minimum of five practice questions for each of the three concepts. On average, students practiced 17 problems, and the average concept block length was 4. In other words, students overwhelmingly chose to block, but just as in Tauber et al. (2013), the summary statistic does not fully capture the approach that students chose.

Additionally, it may be that participants are ignorant of the possibility of employing more varied schedules. If the possibility of using study schedules employing various combinations of blocking and interleaving were made evident to them, perhaps their decisions about how to structure their study might reflect a more sophisticated metacognitive understanding of how to optimize their inductive learning.

Goals of the Present Research

In a series of four experiments, we examined whether participants' judgments of the effectiveness of different presentation schedules might be more in sync with reality were they to be made aware of possible study schedules different from just pure interleaving or pure blocking. In Experiments 1 and 2, we assessed both the actual effectiveness of several hybrid schedules (specifically, we looked for whether a hybrid schedule could outperform a pure interleaving schedule) and learners' perceived effectiveness of those schedules. Next, in Experiments 3 and 4, we examined how participants would choose to structure their own study schedules when we varied the number of exemplars in each category available for viewing.

Experiments 1 and 2

Our goals for Experiments 1 and 2 were twofold: (a) to compare the actual effectiveness of different types of hybrid schedules—that is, ones that combine elements of both blocking and interleaving—against the effectiveness of pure blocking and interleaving schedules; and (b) to assess participant's metacognitive judgments as to the relative effectiveness of such schedules. Although hybrid schedules can be constructed in a number of ways, we focused in Experiments 1 and 2 on three specific types of schedules—in large part, for the practical reason that these were the types of self-constructed schedules we had found participants tended to create (see Experiment 4). We refer to these three hybrid schedules as: (a) mini-blocks, in which a small number of the exemplars from a given category are blocked in small sets, with presentation of the sets themselves spaced apart; (b) blocked-to-interleaved, in which presentation of the exemplars of a given category are first presented in blocks, but then transition into being presented fully interleaved with the exemplars from other categories, and (c) interleaved-to-blocked, which starts off by presenting one exemplar from each category and ends up blocking exemplars by category. We considered the comparative effectiveness of the blocked-to-interleaved and interleaved-to-blocked hybrid schedules to be of particular interest because they both contain the same number of between-category contrasts and within-category comparisons, differing solely in whether blocking or interleaving comes first in the study schedule. A visual representation of the five different schedules is found in Figure 1.

In Experiment 1, participants studied six exemplars per category. With only six exemplars per artist, however, "blocking" in the hybrid schedules was limited to runs of only three consecutively presented exemplars from the same category, which may not have been sufficient for the benefits of blocking to emerge. We attempted to assess this possibility in Experiment 2 by presenting participants with 12 exemplars per artist, rather than just six, so as to increase their opportunity to learn (i.e., to engage in appropriate within-category comparisons and between-category contrasts) throughout all blocked and interleaved components of the hybrid schedules.

[Insert Figure 1 about here]

Method

Participants and design. One hundred and fifty-six undergraduate students (126 females, one undisclosed) and 174 undergraduate students (123 females; one undisclosed; mean age = 19.9 years, *SD* = 1.40 years, age range = 18-26) from UCLA participated for partial course credit in Experiments 1 and 2, respectively. Participants studied six (Exp. 1) or 12 (Exp. 2) paintings by each of twelve different artists via one of five schedules (illustrated in Figure 1): blocked (Exp. 1: *n* = 35; Exp. 2: *n* = 32), interleaved (Exp. 1: *n* = 31; Exp. 2: *n* = 38), mini-blocks (Exp. 1: *n* = 31; Exp. 2: *n* = 34), blocked-to-interleaved (Exp. 1: *n* = 29; Exp 2: *n* = 35), or interleaved-to-blocked (Exp. 1: *n* = 3;0 Exp 2: *n* = 35).

Materials. The experiment was programmed using Collector, an online, open-source, PHP-based platform (downloadable from www.github.com/gikeymarcia/Collector/). We used a total of 14 paintings from each of 12 artists: Georges Braque, Henri-Edmond Cross, Toni Grote, Judy Hawkins, Maryanne Jacobsen, Philip Juras, Richard Lindenberg, Lori McNamara, Marilyn Mylrea, Julie Ford Oliver, Georges Seurat, George Wexler[1]. In Experiment 1, for each participant, six paintings were randomly selected for the study phase and four were randomly selected for the test phase. In Experiment 2, 12 paintings were randomly selected for the study phase and the remaining two were used in the test phase. The paintings were resized to be as close to 500 x 400 pixels as possible.

[Insert Table 1 about here]

Study schedules. Table 1 provides a comparison of the composition of the five study schedules in Experiments 1 (left column) and in Experiment 2 (right column). In the pure blocked condition, all the paintings (6 in Exp. 1, 12 in Exp. 2) by a given artist were presented consecutively. In the pure interleaved condition, the paintings by the different artists were intermixed: In Experiment 1, the paintings were presented in six cycles and in Experiment 2, the paintings were presented in twelve cycles, where each cycle contains one painting by each artist.

[1] Seven of these artists were taken from the set used by Kornell and Bjork (2008). We could not use the exact same set of artists as Kornell and Bjork, however, as we could not find enough paintings by some of those artists to meet the requirements of Experiment 2. The new artists were found using www.dailypainters.com.

In the mini-blocks condition, the six paintings by a given artist were presented in two cycles (Exp. 1) or four cycles (Exp. 2). In each cycle, participants studied three paintings by each artist successively (e.g., AAABBBCCC…where each letter represents an artist).

In Experiment 1, in the blocked-to-interleaved condition, paintings by the artists were presented in three cycles: In the first cycle, three paintings by each artist were presented consecutively, then two paintings by each artist were presented consecutively, and finally, participants studied the one remaining painting by each artist in a randomized order (e.g., AAABBBCCC … AABBCC … ABC …). In the interleaved-to-blocked condition, the order of the three cycles was reversed (e.g., ABC … AABBCC … AAABBBCCC …). In Experiment 2, the blocked-to-interleaved and interleaved-to-blocked schedules were similar to their Experiment 1 counterparts, but an extra cycle of six consecutive paintings by each artist were presented – at the beginning in the blocked-to-interleaved condition, and at the end in the interleaved-to-blocked condition.

For all conditions, the order of the artists was randomized for each cycle and for each participant, so that the same artists would not always be juxtaposed next to each other (e.g., artist A would not always be presented next to artist B).

Procedure. Participants were informed that they would be presented with the paintings of 12 different artists (6 paintings per artist for a total of 72 paintings in Exp. 1; 12 paintings per artist for a total of 144 paintings in Exp. 2) and that their task was to learn the artists' painting styles. They were also told that, on the final test, they would be shown new paintings by these artists and asked to identify, from a list of names, the artist responsible for each painting. The participants were then randomly assigned to one of the five study schedules. In all conditions, the specific order of presentation of a given artist's paintings, and their assignment to study and test phases, were randomized for each participant. During study, participants were shown each painting for three seconds at a time. The paintings were centered on the computer screen with the name of the artist (surname only) printed below each image.

After all the study phase paintings had been presented, participants engaged in a game of Tetris for 45 seconds before beginning the final test. On the final test, participants were presented with new paintings sequentially (4 paintings per artist, one presented in each of 4 randomized cycles of 12 paintings in Experiment 1; 2 paintings per artist, one presented in each of 2 randomized cycles of 12 paintings) and asked to select the artist they believed responsible for each painting from a list of names. In Experiment 1, they were provided with the correct answer after each selection (i.e., presented with the painting again but now with the correct name written beneath it) and in Experiment 2, no feedback was given on the test. The test was self-paced.

Once they had completed the classification test, participants read descriptions of the five different schedules (as shown in the supplementary online materials, S1) and were reminded of the schedule in which they had studied the paintings. They were then asked to indicate which of the five schedules they believed would lead to the best learning of artist styles. Finally, participants were asked what schedule they would select, if forced to choose between a pure blocked schedule and a pure interleaved schedule.

Results

[Insert Figure 2 about here]

Experiment 1 classification performance. Participants' performance on the classification test as a function of study schedule across the four test cycles is presented in Figure 2. As shown, and as we expected, given our use of feedback during testing (which in the present situation amounted to presenting all participants with an interleaved learning condition across the test cycles), performance appears to improve from the first to the fourth test cycle. Indeed, a 4 (test cycle) x 5 (schedule) mixed ANOVA performed on the data revealed a significant main effect of test cycle, $F(3,453) = 13.16$, $MSE = .02$, $p < .001$, $\eta_p^2 = .08$. Additionally, a marginal main effect of schedule was observed, $F(1,151) = 2.12$, $MSE = .11$, $p = .08$, $\eta_p^2 = .05$, but no interaction, $F < 1$. Overall, average correct classification performance increased across test cycles from a mean performance of .44 (*SD* = .21) on Test Cycle 1 to a mean performance of .51 (*SD* = .21) on Test Cycle 4. Post-hoc pairwise comparisons revealed that the interleaved (*M* = .50; *SD* = .16) and blocked-to-interleaved (*M* = .50, *SD* = .16) conditions were both significantly better than the blocked condition (*M* = .39, *SD* = .17), $t(64) = 2.54$, $p = .01$, $d = .62$ and $t(62) = 2.46$, $p = .02$, $d = .62$. None of the other comparisons were significant, *p*s > .10.

Given that the test procedure we employed amounted to providing all participants with an interleaved learning schedule across the four test cycles, the performance obtained in Test Cycle 1 provides a relatively cleaner comparison of the relative effectiveness of the different study schedules. We, therefore, also performed a one-way between-subjects ANOVA using only the data from Test Cycle 1, which revealed a significant effect of schedule, $F(4,151) = 2.54$, $p < .05$, $\eta_p^2 = .06$. Additionally, post-hoc comparisons of performance obtained in the different study schedules revealed that interleaved study (*M* = .52, *SD* = .20) led to significantly better performance than did blocked study (*M* = .37, *SD* = .21) or interleaved-to-blocked study (*M* = .41, *SD* = .17); $t(64) = 2.96$, $p < .01$, $d = .74$, and $t(59) = 2.29$, $p < .05$, $d = .60$, respectively. Performance obtained with blocked-to-interleaved study (*M* = .47, *SD* = .21) was revealed to be marginally better than that obtained with pure blocked study, $t(62) = 1.83$, $p = .07$, $d = .46$. There was a small-to-medium effect of the mini-blocks condition (*M* = .45, *SD* = .23) being better than the blocked study condition, but this effect was not significant, $t(64) = 1.48$, $p = .14$, $d = .37$. No other differences between conditions were revealed, *p*s > .20. In other words, none of the hybrid schedules outperformed the pure interleaved study, although the blocked-to-interleaved schedule and the mini-blocks schedule were not worse.

[Insert Figure 3 about here]

Experiment 2 classification performance. Correct classification performance for each condition (averaged across the two no-feedback test cycles) is presented in Figure 3. A one-way between-subjects ANOVA confirmed a significant main effect of schedule, $F(1,4) = 5.13$, $MSE = .03$, $p < .001$, $\eta_p^2 = .11$, and thus post-hoc comparisons were subsequently performed. First, we replicated the standard benefit of interleaved ($M = .58$, $SD = .19$) over blocked study ($M = .42$, $SD = .16$). Although we might have predicted better performance, the level of performance obtained with the mini-blocks ($M = .55$, $SD = .18$) and the blocked-to-interleaved schedules ($M = .55$, $SD = .17$) did not differ significantly from that obtained in the pure interleaved schedule, and all three led to significantly better classification performance than did the blocked schedule, $t(68) = 3.99$, $p < .001$, $d = .97$; $t(64) = 3.35$, $p < .001$, $d = .84$; and $t(65) = 3.49$, $p < .01$, $d = .87$, respectively. Performance obtained with the interleaved-to-blocked schedule ($M = .48$, $SD = .19$), however, was significantly worse than that obtained with the pure interleaved schedule, $t(71) = 2.41$, $p < .05$, $d = .57$, and not significantly better than that seen in the blocked condition, $t(65) = 1.38$, $p > .05$, $d = .34$. Additionally, performance obtained with the interleaved-to-blocked schedule was marginally worse than that obtained with both the mini-blocks schedule, $t(67) = 1.78$, $p = .08$, $d = .43$, and the blocked-to-interleaved schedule, $t(68) = 1.86$, $p = .07$, $d = .45$, with performance in this latter condition falling somewhere between that obtained with the blocked and interleaved schedules.

Metacognitive judgments. In Experiments 1 and 2, the schedules most likely to be selected as "most effective" were the mini-blocks (36% and 24%), blocked (24% and 32%), and blocked-to-interleaved (21% and 33%) schedules. Interleaved-to-blocked received 16% and 7% of the votes, and interleaved study was only judged to be the most effective by 3% and 3.5% of participants, in Experiments 1 and 2 respectively.

When forced to choose between a pure blocked versus a pure interleaved schedule, an overwhelming majority of participants (86% and 82%, in experiments 1 and 2, respectively) chose the blocked schedule. Greater details about participants' metacognitive judgments, by experienced condition, are presented in the Supplementary Online Materials (S2).

Discussion

Two key findings emerged from Experiments 1 and 2. First, certain hybrid schedules—namely mini-blocks and blocked-to-interleaved schedules—can be as effective for inductive learning as a pure interleaved schedule. This finding was also very consistent when we examined the effects of the schedules for each artist. For more details about the by-artist schedule effects, see Supplementary Online Materials, S3. Our findings are also consistent with that from research on motor-skills learning. Specifically, in a study by Proteau, Bladin, Alain, and Dorion (1994) comparing a mini-blocks practice schedule to pure blocked and pure interleaved schedules for the learning of three motor sequences, the following pattern of results was observed. While pure blocking led to faster acquisition, pure interleaving led to better learning (i.e., performance on a delayed retention test). The mini-blocks schedule combined the best of both worlds, producing faster acquisition than the interleaved schedule and better learning than the blocked schedule. Participants receiving the mini-blocks schedule for the learning of motor sequences, however, did not outperform those receiving the pure interleaving schedule, as was the case in the present research for the learning of artistic painting styles.

The second main finding from Experiments 1 and 2 is that unlike pure interleaving, which (consistent with prior results) was fairly uniformly undervalued by participants as an effective learning schedule, the mini-blocks and blocked-to-interleaved schedules were more accurately judged as ones that would be effective for learning. Indeed, it would appear that prior studies have overestimated the reliance of learners on blocked study. Akin to the findings of such studies, when given only the two extremes (blocked versus interleaved), 84% of the present participants across the two studies chose blocking as the schedule that would be more effective for learning. When provided with a set of hybrid schedules as well, however, the percentage of participants who indicated that they would choose a pure blocking schedule for study dropped to 28%; and instead, the majority (60%) chose one of the three empirically most effective schedules (interleaved, mini-blocks, or blocked-to-interleaved). Although 60% could reflect "random guessing" (i.e., three out of five possible schedules), we argue that this is unlikely given that the "baseline"—shown both in the present studies as well as in prior studies—is an overwhelming preference for blocked study and an overwhelming aversion to interleaved study. Instead, we would contend that this 60% majority reflects a level of metacognitive understanding of effective learning schedules that has not been revealed in previous studies comparing only pure blocking and pure interleaving study schedules.

Experiment 3

To better capture and understand the *a priori* theories that individuals might have concerning how to optimize induction learning, we presented the participants in Experiments 3 and 4 with hypothetical learning scenarios and asked them to self-schedule their own learning.

Our intent was to explore how individuals—when drawing purely on their pre-experimental learning experiences and metacognitive beliefs about how best to learn—would elect to construct their study schedules for the learning of categories. To this end, we asked participants to construct their own study schedules for learning the painting styles of different artists, but without being shown any specific exemplars of the different artists' painting styles.

More specifically, in Experiment 3, participants were asked to arrange flashcards representing exemplars of different artists so as to best learn the particular painting style of each artist. The use of physical flashcards opened up a secondary variable of interest: We explored whether the way in which the flashcards were arranged at the start of the learning assignment would affect participants decisions on how to construct their study schedules. To do so, we manipulated the original arrangement of flashcards for each participant to be one that we thought would promote blocking or one that we thought would promote interleaving. In the block-promoting condition, the original arrangement of the flashcards was such that participants who wanted to arrange the cards for their individual study session with the least amount of effort would end up blocking their study; in the block-discouraging condition, the original arrangement of the flashcards was such that participants who wanted to arrange the cards for their individual study session with the least amount of effort would end up interleaving their study. This manipulation, therefore, allowed us to tease apart the participants' beliefs about learning from a potential confounding variable—namely, the motivation to leave the study as soon as possible.

Method

Participants and design. Twenty-nine undergraduates from the University of California, Los Angeles (UCLA) participated for partial course credit. Six stacks of flashcards, each containing six cards, were set out in front of the participants in one of two original arrangements corresponding to the block-promoting condition and the block-discouraging condition. In the block-promoting condition (n = 15), six stacks of flashcards, with each stack contained six cards representing six different exemplars of the same artist's paintings, were set in front of the participant; whereas, in the block-discouraging condition (n = 14), six stacks of flashcards, in which the six cards in each stack represented a single exemplar from each of the six different artists, were set in front of the participant. Participants were alternately assigned into the block-promoting and block-discouraging conditions. Although originally aiming to recruit a total of 30 participants, we terminated the study at the end of the academic quarter.

Materials. Thirty-six 3" x 4" index cards were used, with each card representing one painting by a given artist from the total set of six artists whose painting styles were to be learned. No paintings, however, were actually shown. Instead, on the front of each card was the name of one of the six artists, written in sufficiently large print to fill up the surface of the card, with each artist's name written on a total of six cards each and in a different ink color, so as to highlight the different names. The names of the six artists whose painting styles were to be learned were: Hawkins, Lewis, Foster, Cross, Oliver, and Juras. These artists were randomly chosen from the subset of artists that had been used in prior interleaving studies (with the constraint that none shared the same initial letter) and were unfamiliar to all the participants.

The 36 cards were separated into six stacks of six cards each, which were arranged on the desk in front of the participant into two rows, with each row containing three stacks of cards. In the block-promoting condition, each pile contained six cards with the same name. It would therefore be easy for participants to study the "paintings" in a blocked manner by simply picking up each stack individually in succession. In the block-discouraging condition, each of the six cards in a stack displayed a different artist's name. Furthermore, the order of the names in each stack was randomly determined, with the constraint that none of the top cards showed the same name. Thus, for participants to study the artists' paintings in a blocked manner, they would have to dig through each of the stacks to find the appropriate name.

Procedure. The stacks of cards appropriate for a given participant's assigned condition were first laid out in front of the participant, who was then given an instruction sheet to read that described the task. Participants were told to pretend they were art students trying to learn the styles of six different artists by studying six different exemplars of each artist's paintings. The instructions also listed the six names and informed the participants that each index card represented one painting by one of the six artists. Participants were instructed that they were to create one single stack of these cards that would represent the order in which they would study the paintings represented by the cards.

After the participants read the instruction sheet, the experimenter verbally repeated the instructions to ensure they were understood. Participants were allowed to take as long as they needed to create the order in which they would study the cards. Once finished, they gave their stack of cards to the experimenter, who confirmed with the participant that the top index card was intended to be the first to-be-studied card and then recorded the order each participant had created.

Finally, participants filled out a questionnaire asking them to describe the strategy they had used to schedule their study of the artists' paintings. They were also provided with a list of possible study orders that they could have chosen and asked both to indicate which order their own schedule most resembled and which order they thought would lead to the best learning of the different artists' painting styles. All parts of the study were self-paced.

Results and Discussion

[Insert Table 2 about here]

The results of Experiment 3—how participants scheduled their study (as categorized by the experimenters into the possible study orders provided in the post-test questionnaire) and which type of schedule participants thought would be most effective—are presented in Table 2, organized from the most blocked to the most interleaved schedules.

Self-constructed study orders. Across both card-arrangement conditions, participants showed an overwhelming tendency to schedule their study in a blocked order. Twenty-four of the 29 participants—all 14 participants from the block-discouraging condition and 10 from the block-promoting condition—constructed a pure blocked schedule, choosing to study all six paintings by a given artist before moving on to the next artist.[2] In other words, even when the stacks of cards were arranged so as to make blocked scheduling more difficult, participants went out of their way to create a less effective study schedule.

Of the remaining five participants, one chose to study three paintings by a given artist at a time; one chose to study two paintings by a given artist at a time; one chose to study one painting by each artist first and then to alternate their study in pairs (i.e., ABCDEF-ABABABABAB-CDCDCDCDCD-DFDFDFDFDF, where each unique letter corresponds to one artist); and only two chose to study the artists all mixed up (i.e., a "pure" interleaved schedule).

Post-test questionnaire. When participants were asked to select (from the list provided in the post-test questionnaire) which order they thought would be most effective for learning, the most popular choice (*n* =11, or 42%) was blocked-to-interleaved, followed closely by pure blocking (*n* = 9, or 29%). Only one participant selected pure interleaving.

---

[2] A chi-square test of independence of arrangement and schedule was not conducted given that so few participants chose to do something other than purely block their study and thus the assumptions of the chi-square analysis (that there should be at least 5 responses expected in each cell) were not met.

Consistent with the results of Tauber et al. (2013), whether we looked at how participants schedule their own learning or how they chose from a list of possible study orders, the participants demonstrated an overwhelming preference for blocked study. That a large proportion of participants judged that the blocked-to-interleaved schedule would be the most effective schedule, however, suggests that learners may see some benefits to interleaving, even though most participants did not spontaneously include any form of interleaving in their own schedules.

Experiment 4

The self-constructed study orders of participants in Experiment 3 demonstrated an overwhelmingly preference to study all six exemplars by a given artist successively—in other words, to block their study. Their post-test questionnaire responses, however, indicated an appreciation that pure blocking might not be the optimal method of study. One possibility for the discrepancy between participants' self-constructed study orders and what they later judged would be the most effective study order is that six exemplars is simply not enough for learners to feel comfortable in using a schedule that includes some interleaving. It may be that participants would engage in more metacognitively sophisticated scheduling if they were given more exemplars from each of the to-be-learned categories with which to work. In other words, with only six exemplars to study, blocking might be prioritized, but with more exemplars, participants may begin to mix learning up. We explored this possibility in Experiment 4 by giving participants 20 exemplars of each artist's painting style to use in constructing their study schedules. Our primary questions were whether learners would now introduce at least some degree of interleaving into their study schedules and, if so, at what point would such interleaving first appear. Under these conditions, we did not expect that learners would remain so firm in their apparent preference for blocked study as to block 20 exemplars from the same artist successively.

Method

Participants. Fifty UCLA undergraduates participated for partial course credit. The data from two participants who did not follow instructions (i.e., did not write down the appropriate number of paintings per artist) were discarded.

Materials and procedure. Participants were first given a single page of text to read containing the instructions for the task and displaying a grid of 10 rows with 10 cells per row, with this grid to be used in constructing their study schedules. Instructions followed the same scheme as in Experiment 3: Participants were told that they needed to learn the painting styles of different artists by studying exemplars of each artist's paintings in preparation for a final exam on which they would be shown new paintings and have to identify which of the studied artists was responsible for each new painting. In contrast to Experiment 3, participants were told that they were to learn the painting styles of five artists and they had 20 exemplars of each artist's paintings available to them for study (as opposed to learning the styles of six artists, each with 6 exemplars available for study). Another difference from Experiment 3 was that, instead of organizing a stack of flashcards, participants had to write one of the letters A through E (where each letter represented one of the five artists) and one number between 1 and 20 (where each number represented one of the 20 paintings by a given artist) in each cell of the 10 x 10 grid that had been given to them at the start of the instruction phase. Participants were instructed to work from left to right, top to bottom, and gaps were printed between each row to clarify this order.

Once they had finished constructing their proposed study schedules on the grid, participants were handed two more pages, one containing a post-test questionnaire (see

Supplementary Online Materials, S4) and the other containing demographic questions. On the questionnaire, they were asked (a) to describe the rationale behind the study schedule they had constructed; (b) to select from the sample schedules described on the questionnaire the one that best matched the schedule they had constructed; (c) to select from the sample schedules listed on the questionnaire the one that they believed would be most effective for learning the different artistic styles; (d) to choose between either a pure blocked or pure interleaved study schedule with respect to which would be more effective for learning, and (e) to explain why they would block and why they would interleave their study of the paintings by each artist. The latter two questions were added to the study after data collection was begun, and thus, were not asked of the first nine participants. Participants were allowed to work at their own rate and generally took between five to ten minutes to complete the required tasks.

Results and Discussion

[Insert Table 3 about here]

Self-constructed study orders. The results of Experiment 4—how participants scheduled their own study (as categorized by the experimenter into the possible study orders provided in the post-test questionnaire) and their later judgments of which type of study schedule from the list provided on the post-test questionnaire would be the most effective—are presented in Table 3, organized from the most blocked to the most interleaved schedules.

Of the results displayed in Table 3, the most striking is the fact that even with 20 exemplars per artist, 14 (29%) of the participants still chose to study all 20 exemplars from the same artist consecutively and an additional two participants chose to study either 18 or 19 paintings by one artist consecutively ("familiarize, then block"). As is evident from Table 3 and from Figure 4—which illustrates, as another measure of blocking tendency, the maximum run lengths appearing in the participants' study schedules (i.e., the longest consecutive block of paintings by the same artist)—participants had a tendency to block their study of paintings by the same artist. Very few participants chose a pure interleaving strategy (where two paintings by the same artist are never studied consecutively); very few even chose to switch up artists every two or three paintings.

[Insert Figure 4 about here]

It is interesting, however, that when asked to select from the list of possible study schedules presented in the post-test questionnaire the schedule the participants thought would be most effective for learning artists' styles, 52% (25 out of 48) of participants chose the hybrid blocked-to-interleaved schedule. Why this discrepancy between participants' self-constructed schedule and their post-test questionnaire selections? Although the present data cannot differentiate among them, at least two possible explanations come to mind. First, perhaps the idea of creating hybrid schedules did not occur to participants at the time of schedule construction, but after the hybrid scheduling strategies were described to them, they were able to recognize that these were schedules that could be more effective than pure blocking. In fact, anecdotally, one participant asked for the study instructions to be clarified multiple times because it simply did not occur to this individual that there could be any other scheduling strategy than purely blocking one's study. Another possibility, however, is that participants did consider hybrid options during schedule construction, but the perceived complexity of creating such hybrid schedules dissuaded them from writing those types of schedules down on the response sheet.

Despite an overall tendency for participants to block study of paintings by the same artist, the 34 participants, who did not simply block all 20 paintings of an artist, created schedules that were strategic to some degree. For instance, some chose to see one painting by each artist at the beginning of their study to familiarize themselves with the range of painting styles. Others chose to interleave paintings by different artists in the middle or at the end in order to test themselves or to examine the differences between the artists. Examples of the rationales that participants gave for how they chose to schedule their study are shown in the Supplementary Online Materials (S5).

Blocking versus interleaving. When forced to make a choice between studying all the paintings in a blocked manner or in a fully interleaved manner, 32 (82%) of the 39 participants responded that it would be better to study the paintings blocked by artist, highlighting again that individuals appear to have an overwhelming a priori belief that blocking is more useful for learning concepts than is interleaving.

On the other hand, answers to the open-response questions, "If/When you put paintings by the same artist next to each other, why did you do so?" and "If/When you mixed up the paintings by different artists, why did you do so?" indicated that the participants had at least some appreciation of the possible benefits of interleaving as well as how interleaving and blocking benefits might differ. We coded participants' responses about the benefits of blocking and the benefits of interleaving into different categories. For blocking, we coded responses into the following categories: to see similarities, blocking is better organized, no reason, and other. For interleaving, we coded responses into the following categories: to see differences, to test oneself, no reason, and other. A majority (30, or 77%) of the participants responded that they blocked in order to see the similarities within the different paintings of a given artist, and 20 (51%) responded that they interleaved in order to learn the differences between the different artists' styles. These types of responses are consistent with the hypotheses that cognitive researchers have advanced as being the relative benefits of blocking and interleaving (e.g., Carvalho & Goldstone, 2014; Kang & Pashler, 2012; Kornell & Bjork, 2008).

Seven (18%) of the participants also reported that interleaving paintings by different artists was useful as a way of self-testing whether the different styles had been learned. Very few responses fell into the other categories (8% stated that blocking was more organized; 8% and 18% did not give a reason or explicitly stated that there was no reason for blocking and interleaving, respectively; and only 5% of responses fell into the "other" category, for both blocking and interleaving).

Taken together with the results of Experiment 3 and the metacognitive results of Experiments 1 and 2, the present findings demonstrate that although participants hold a general belief that blocking exemplars by category is more effective for learning—a finding consistent with that of previous studies (e.g., Tauber et al., 2013)—they also have some measure of metacognitive sophistication in that they tend to understand that interleaving highlights differences between artists and generally prefer to have some degree of interleaving as well as blocking. Any interleaving of paintings by different artists, however, is most likely to be introduced after first studying paintings blocked by artist: Across both Experiments 3 and 4, only 2 out of 77 participants judged the interleaved-to-blocked schedule as being the most effective, whereas 36 participants—almost half—judged the blocked-to-interleaved schedule as being the most effective. Similarly, with the exception of a handful (n = 14) of participants, almost all (n = 65) participants began their study with a block of at least four painting by the same artist.

General Discussion

The Effects of Hybrid Schedules

Experiments 1 and 2 produced converging results. While certain combinations of blocked and interleaved study schedules were better for learning than was a pure blocked schedule, and matched that of pure interleaving, no combination led to better overall inductive learning than did pure interleaving. Consistently—across experiments and across specific artists (see Supplementary Online Materials, S3)—we also found that an interleaved-to-blocked study schedule was not as effective for learning as was a pure interleaving study schedule or as either of the other two hybrid study schedules (blocked-to-interleaved and mini-blocks). This result for the interleaved-to-blocked study sequence is interesting from the perspective that blocking enhances comparisons and interleaving enhances contrasts, as the blocked-to-interleaved and interleaved-to-blocked schedules should afford the same opportunities for making comparisons and contrasts. One potential explanation for this pattern could be that it is more important to notice commonalities before noticing differences, and the point at which a learner transitions from focusing on commonalities to differences may be important (see for example, progressive alignment theory, Gentner & Markman, 1994) in which learning easier relational alignments first can scaffold learning of more distant or structural alignments). Carpenter (2014) made a similar prediction, suggesting that for difficult concepts, building familiarity with a given concept first via blocking may be important before transitioning to interleaving. This explanation, however, does not account for why the pure-interleaving schedule (where attention is never drawn to within-category similarities and familiarity with a given concept is not first built up) is just as good as the blocked-to-interleaved schedule.

These results might also be interpreted in light of other explanations of the interleaving benefit, namely, the spacing effect. There are two ways in which the spacing inherent to non-blocked schedules has been proposed to benefit inductive learning: the study-phase retrieval hypothesis (which has typically been applied to explaining spacing benefits for memory, e.g., Appleton-Knapp, Bjork, & Wickens, 2005; Benjamin & Tullis, 2010; Thios & D'Agostino, 1976) and the forgetting-as-abstraction hypothesis (Vlach, 2014; Vlach, Sandhofer, & Kornell, 2008).

In the present studies (as is also the case in many other instances of category learning), category induction is only one part of the whole learning process. Learners have to induce each artist's style (perhaps through noticing commonalities and differences within and between categories), but they also have to learn the different names of the artists, and then learn to associate each of the induced styles with the corresponding artist's name. In other words, there are memory components to the task. Interleaving study of the paintings by different artists not only allows learners to spot the differences among the artists, it also inherently introduces spacing between the presentations of paintings by the same artist. Such spacing may be beneficial if the spaced presentations act as "reminders" and encourage retrieval of prior paintings by the same artist, and it may also enhance the learning of the association between the name of artists and their induced painting styles. This associative learning is not a trivial component of category learning, particularly when there are many categories to be learned. Although Kang and Pashler (2012) concluded that spacing is advantageous in category learning only when it promotes discriminative contrast, their materials consisted of only three different artists, meaning that the necessary associative learning was very easy in their study. Using 16 categories of butterflies, Birnbaum et al. (2013) showed that when the number of discriminative contrasts was held constant, presentation of a greater number of intervening items between exemplars from the same category led to greater inductive learning.

In the present experiments, the blocked-to-interleaved study schedule is analogous to an expanding study schedule: Presentations of a given artist's paintings are increasingly spaced out, with no spacing between the first three paintings and an average interval of 11 intervening paintings (by different artists) by the end of the learning task. In contrast, the interleaved-to-blocked schedule is analogous to a contracting study schedule in which the presentations of a given artist's paintings become increasingly closer to each other as the learning task proceeds. This difference in the two types of schedules is important because research has shown that expanding schedules are particularly beneficial for learning when forgetting occurs rapidly between the first and second presentations of the material to be learned (Landauer & Bjork, 1978; Storm, Bjork, & Storm, 2010). Additionally, assuming that it is impossible to associate an artistic style to the corresponding artist's name before that style has been induced, spacing should be relatively more effective for such associative learning when it occurs toward the end of study (i.e., after a style has been induced) rather than toward the beginning. It is important to note, however, that despite these apparent advantages of the blocked-to-interleaved schedule, it was still no better for learning than was the pure interleaved schedule. Potentially, the pure interleaved schedule simply represents a greater total number of retrieval (or, reminder) events. If we consider that blocking leads to no reminding (because there has been no forgetting), then the hybrid schedules contain only 2-4 reminder events, whereas the pure interleaved schedule contains six or 12 reminder events, in Experiments 1 and 2, respectively. This greater number of reminders in the pure-interleaving condition may have offset any benefit that might otherwise be conferred from the hybrid schedules.

It remains unclear how the presently observed pattern of results might be affected by and generalize to other stimuli, such as cognitive concepts (e.g., mathematics concepts) rather than perceptual concepts, or to stimulus sets with different category structures. In the present studies, we did not design for rigorous control of within-category and between-category similarities. As we have argued above, however, learning categories from exemplars almost certainly involves multiple processes: the search for similarities among the exemplars of the same category (Goldstone, 1996); the search for differences across different categories; the need to associate the induced categories with the correct category labels (particularly when there is a large number of to-be-learned categories); and perhaps even retrieval or reloading of previously learned formulae or rules (e.g., as in the case of math learning or other rule-based categories). While the first process—that is, noticing commonalities (e.g., by spotting common features or discovering some rule to which all members of a category conform) within a category—is likely to benefit from blocking (Carvalho & Goldstone, 2014), the other processes may benefit more from interleaving and spacing. Another type of category structure—whether category learning relies on an explicit, rule-based system or an implicit, information-integration system—has also been shown to be an important factor in determining optimal study schedules (Noh, Yan, Bjork & Maddox, 2016), but again may only play a role in the initial category formation (versus maintenance in long-term memory).

Thus, to tease apart the multiple processes underlying category learning and develop a more complete understanding of how we learn from exemplars, future studies should explore the effectiveness of different study schedules while varying the nature of the to-be-learned stimuli. Additionally, it would be important to examine whether the patterns found in the present Experiments 1 and 2 are maintained over longer retention intervals.

Examining Learners' Beliefs about How Best to Study

Across the present four experiments, participants were asked to judge which study schedule, among a number of choices provided to them, would be the most effective for inductive learning. An analysis of their choices suggests that participants' judgments regarding effective study schedules for inductive learning are not as inaccurate as previous research would seem to suggest. In our post-test questionnaires, when participants were forced to choose between only the two extremes of pure blocking versus pure interleaving study schedules, 84% (across Experiments 1, 2 and 4) chose blocking over interleaving as being the more effective way to learn. When provided with a range of options that lay between these two extremes, however, the majority (60%) of participants in Experiments 1 and 2 chose one of the top three (and equally effective) schedules: interleaved, mini-blocks, or blocked-to-interleaved. In fact, when presented with various hybrid options, the proportion of participants who reported believing that pure blocking was the most effective dropped to 28% in Experiments 1 and 2, 29% in Experiment 3, and 17% in Experiment 4.

Furthermore, beyond post-test questionnaire responses, an examination of participants' self-constructed study schedules in Experiment 4 showed that quite a few participants shied away from pure blocking when a large number of exemplars were available for study. To illustrate, in Experiment 4, in which participants had 20 exemplars of each artist's style available for study, only a third constructed a pure blocked schedule of study (as compared to 83% of the participants in Experiment 3, who had only 6 exemplars of each artist available for study). Moreover, when asked directly why they would block or interleave their study, participants were generally able to express that blocking would help them to spot the commonalities within a given to-be-learned category and a sizeable proportion of the participants were, additionally, able to indicate that interleaving would help them distinguish between the different artists' styles. That these participants were responding to open-ended questions (that is, not just picking from possible explanations suggested by the experimenters) makes their ability to express such considerations even more impressive. Taken together, then, the results of the present studies provide a more complete and nuanced picture of what learners believe about their own learning than has been suggested by prior studies.

Another result of the present research adding to a fuller picture of what learners understand about effective learning strategies is our finding that participants preferred blocked-to-interleaved study schedules over interleaved-to-blocked study schedules. Given that the findings of Experiments 1 and 2 showed that both a blocked-to-interleaved and a mini-blocks schedule were just as effective for inductive learning as was a pure interleaved study schedule, the participants who judged either of these two hybrid schedules as being the most effective for learning could arguably be considered as being metacognitively accurate. The reasons, however, for participants preferring blocked-to-interleaved over interleaved-to-blocked study schedules may not necessarily align with the reasons why the former is empirically more effective than the latter. Although we did not directly question participants about their choice of blocked-to-interleaved schedules over interleaved-to-blocked schedules, the impression emerging from their open-ended responses about their self-constructed schedules is that many participants used interleaving at the end to test themselves—thus implying a belief that interleaving should only be introduced into one's study schedule after learning (i.e., after inducing the different to-be learned categories). This reasoning would be consistent with the findings of Kornell and Son (2009) indicating that people choose to test themselves at the end of study (instead of restudying) based on a desire to diagnose their level of learning, rather than from an appreciation of the testing effect per se (see Roediger & Karpicke, 2006).

Experiments 3 and 4 were designed as an examination of the *a priori* beliefs that learners have developed over their learning experiences. A natural question that follows from these studies, however, is whether participants would actually have learned better using their self-constructed schedules. In the present studies in which participants self-constructed study schedules, no actual study or test of the artists' painting styles took place. In Experiment 3, where there were six exemplars per artist to sequence, the possibility that participants would have learned better by using their overwhelmingly-blocked schedules seems unlikely given the findings from the present Experiment 1 and prior studies (e.g., Kornell & Bjork, 2008; Yan et al., 2016) showing a robust interleaving benefit. In Experiment 4, however, where participants were more likely to use some form of hybrid schedule, how an actual pattern of results might appear is less clear.

Concluding Comments

The present studies reveal new findings on two fronts. First, they reveal that—despite the theoretical reasons for why combining blocking (which presumably facilitates noticing of within-category commonalities) and interleaving (which presumably facilitates between-category discrimination) within the same study schedule should lead to optimal inductive learning—no hybrid schedule of study produced better category induction than did a pure interleaving schedule. This pattern of results suggests the possibility that these types of rich, perceptual-based categories may be learned in some more implicit way, rather than as a result of noticing particular features, and our conjecture is that hybrid schedules might indeed lead to the most effective learning of categories that are more feature-based or rule-based.

Second, the metacognitive results obtained across all of the present studies suggest that learners' beliefs about how they learn best, though seriously flawed in some respects, are not as seriously flawed as we and others have inferred from previous studies. Our participants still judged pure blocking to be a more effective study schedule than pure interleaving (which was categorically wrong with the type of material used in the present learning tasks), but they were also highly likely to judge that certain hybrid schedules—namely, the blocked-to-interleaved and the mini-blocks schedules, which were, empirically, as effective as pure interleaving—would be the most effective for learning, even when they had not themselves introduced any level of interleaving into their own self-constructed study schedules. Moreover, and perhaps most impressive, evidence for some recognition of the different types of processing promoted by blocking versus interleaving was revealed in our participants' answers to our free-response questions. The majority of participants made statements indicating some awareness that blocking helped them to notice the commonalities across different exemplars of a given category and a substantial proportion was also able to appreciate that interleaving would be beneficial for learning to distinguish among the artists' styles. Overall, then, our findings suggest that learners, although far from being truly metacognitively sophisticated about learning, nonetheless understand more about how to learn than has been revealed by prior studies.

Finally, although we did not find that hybrid schedules (mini-blocks and blocked-to-interleaved, specifically) combined the best of both blocking and interleaving in terms of promoting learning through noticing within-category commonalities and between-category differences, we did find that they combined the best of both worlds in a different way. Namely, they enhanced learning to the same extent as pure interleaving while retaining the popularity of blocking. With so much of our learning being self-regulated, it is critical that we find ways of encouraging learners to use effective learning strategies. Yan et al. (2016) showed that learners are surprisingly resistant to using pure interleaving study strategies; hybrid study strategies may provide an appealing and effective counter to such resistance.

References

Abushanab, B., & Bishara, A. J. (2013). Memory and metacognition for piano melodies: Illusory advantages of fixed-over random-order practice. *Memory & Cognition*, *41*, 928-937.

Appleton-Knapp, S. L., Bjork, R. A., & Wickens, T. D. (2005). Examining the spacing effect in advertising: Encoding variability, retrieval processes, and their interaction. *Journal of Consumer Research*, *32*, 266-276.

Benjamin, A. S., & Tullis, J. (2010). What makes distributed practice effective? *Cognitive psychology*, *61*, 228-247.

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*, 392-402.

Carpenter, S. K. (2014). Spacing and interleaving of study and practice. In V. A. Benassi, C. E. Overson, & C. M. Hakala (Eds.), *Applying the science of learning in education: Infusing psychological science into the curriculum* (pp. 131-141). American Psychological Association.

Carvalho, P. F., Braithwaite, D. W., de Leeuw, J. R., Motz, B. A., & Goldstone, R. L. (2016). An in vivo study of self-regulated study sequencing in introductory psychology courses. *PloS one*, *11*, e0152115.

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,* 481-495.

Day, S., Goldstone, R. L., & Hills, T. (2010). The effects of similarity and individual differences on comparison and transfer. *In Proceedings of the Thirty-Second Annual Conference of the Cognitive Science Society (pp. 465-470).*

Gentner, D., Loewenstein, J., Thompson, L., & Forbus, K. D. (2009). Reviving inert knowledge: Analogical abstraction supports relational retrieval of past events. *Cognitive Science*, *33*, 1343-1382.

Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science*, *5*, 152-158.

Gentner, D., & Namy, L. L. (1999). Comparison in the development of categories. *Cognitive Development*, *14*, 487-513.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, *15*, 1-38.

Goldstone, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, *24*, 608-628.

Kang, S. H. K., & Pashler, H. (2012). Learning painting styles: Spacing is advantageous when it promotes discriminative contrast. *Applied Cognitive Psychology, 26,* 97-103.

Kornell, N. & Bjork, R. A. (2008). Learning Concepts and Categories: Is Spacing the "Enemy of Induction"? . *Psychological Science, 19*, 585-592.

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*, 498-503.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. *Practical aspects of memory*, *1*, 625-632.

Kornell, N., & Son, L. K. (2009). Learners’ choices and beliefs about self-testing. *Memory*, *17*, 493-501.

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.

Oakes, L. M., & Ribar, R. J. (2005). A comparison of infants' categorization in paired and successive presentation familiarization tasks. *Infancy*, *7*, 85-98.

Proteau, L., Blandin, Y., Alain, C., & Dorion, A. (1994). The effects of the amount and variability of practice on the learning of a multi-segmented motor task. *Acta Psychologica*, *85*, 61-74.

Rittle-Johnson, B., & Star, J. R. (2011). The Power of Comparison in Learning and Instruction: Learning Outcomes Supported by Different Types of Comparisons. *Psychology of Learning and Motivation-Advances in Research and Theory*, *55*, 199.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*, 181-210.

Rohrer, D. (2012). Interleaving helps students distinguish among similar concepts. *Educational Psychology Review, 24*, 355-367.

Rozenshtein, A., Pearson, G. D. N., Yan, S. X., Liu, A. Z., & Toy, D. (in press) Effect of massed versus interleaved teaching method on performance of students in radiology. *Journal of American College of Radiology*. *13,* 979-984

Sana, F., Yan, V. X., & Kim, J. A. (2017). Study sequence matters for the inductive learning of cognitive concepts. *Journal of Educational Psychology. 109,* 84-98.

Storm, B. C., Bjork, R. A., & Storm, J. C. (2010). Optimizing retrieval as a learning event: When and why expanding retrieval practice enhances long-term retention. *Memory & Cognition*, *38*, 244-253.

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,* 356-363.

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

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*, 529-536.

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*, 163-168.

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children’s memory and category induction. *Cognition*, *109*, 163-167.

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*, 750-763.

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,* 918-933.

Zulkiply, N., & Burt, J. S. (2013). The exemplar interleaving effect in inductive learning: Moderation by the difficulty of category discriminations. *Memory & Cognition*, *41*, 16-27.

Zulkiply, N., McLean, J., Burt, J. S., & Bath, D. (2012). Spacing and induction: Application to exemplars presented as auditory and visual text. *Learning and Instruction*, *22*, 215-221. doi:10.1016/j.learninstruc.2011.11.002.

Table 1

Comparison of the Five Schedule Conditions Across Experiments 1 and 2

Schedule	Experiment 1	Experiment 2
Blocked	6 consecutively	12 consecutively
Mini-blocks	3-3	3-3-3-3
Blocked-to-interleaved	3-2-1	6-3-2-1
Interleaved-to-blocked	1-2-3	1-2-3-6
Interleaved	1 (x6)	1 (x12)

Note. Numbers represent the number of paintings by a given artist studied consecutively before moving on to a new artist.

Table 2

Number of Participants whose Self-constructed Study Schedule Matched each Type of Study Schedule Listed in the Post-test Questionnaire (as Categorized by the Experimenters) and Number Judging that Type of Schedule to be Most Effective for Learning the Artist Styles in Experiment 3.

Types of Study Schedules (Listed in Post-test Questionnaire)	Categorization of Self-constructed Study Orders	Judged Most Effective Order
Pure blocked	24	9
Familiarize with one painting by each artist, then block study	1	4
Study two or three from each artist at a time	2	2
Blocked-to-interleaved	0	11
Interleaved-to-blocked	0	0
Pure interleaved	2	1

Table 3

Number of Participants whose Self-constructed Study Schedule Matched each Type of Study Schedule Listed in the Post-test Questionnaire (as Categorized by the Experimenter) and Number Judging that Type of Schedule to be Most Effective for Learning the Artist Styles in Experiment 4.

Types of Study Schedules (Listed in Post-test Questionnaire)	Categorization of Self-constructed Study Orders	Judged Most Effective Order
Blocked	14	8
Familiarize, then block	2	8
Ten at a time	8	4
Five at a time	10	0
Blocked to interleaved	9	25
Interleaved to blocked	0	2
Pairs or triples	2	1
Interleaved	3	0

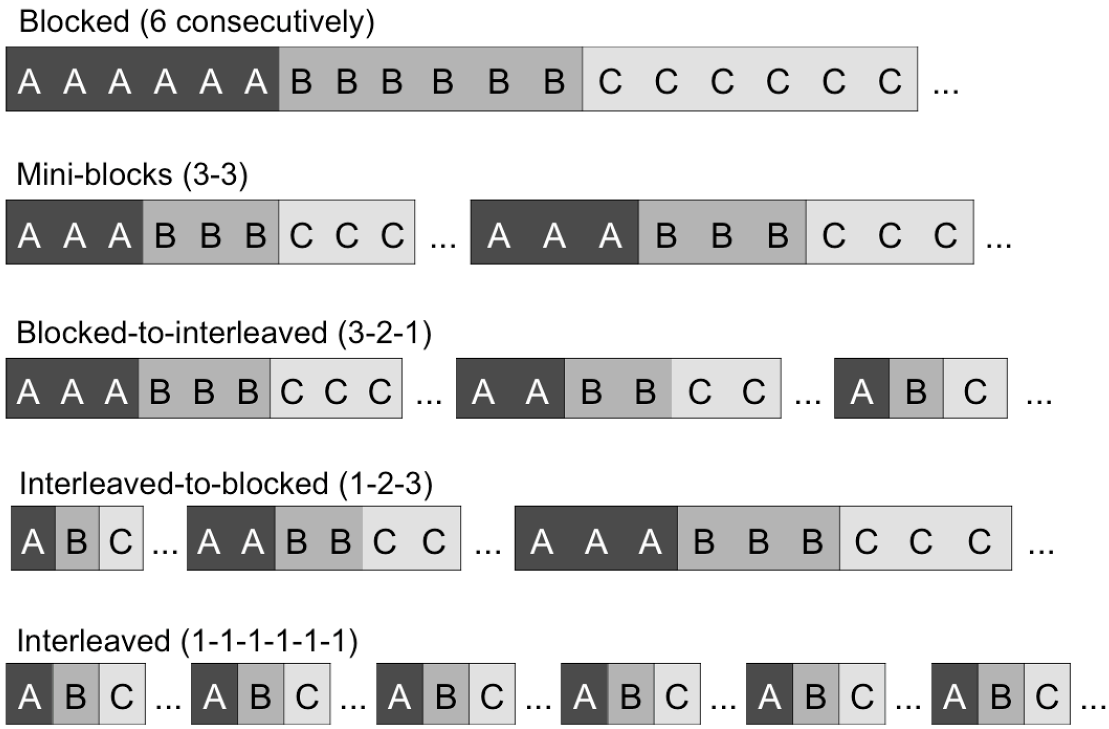


Figure 1. A visual representation of the five schedules in Experiment 1. Each letter represents a different image by a given artist; the specific order of the artists, however, was block randomized, so paintings by artist B were not always juxtaposed against paintings by artists A and C, and so on.

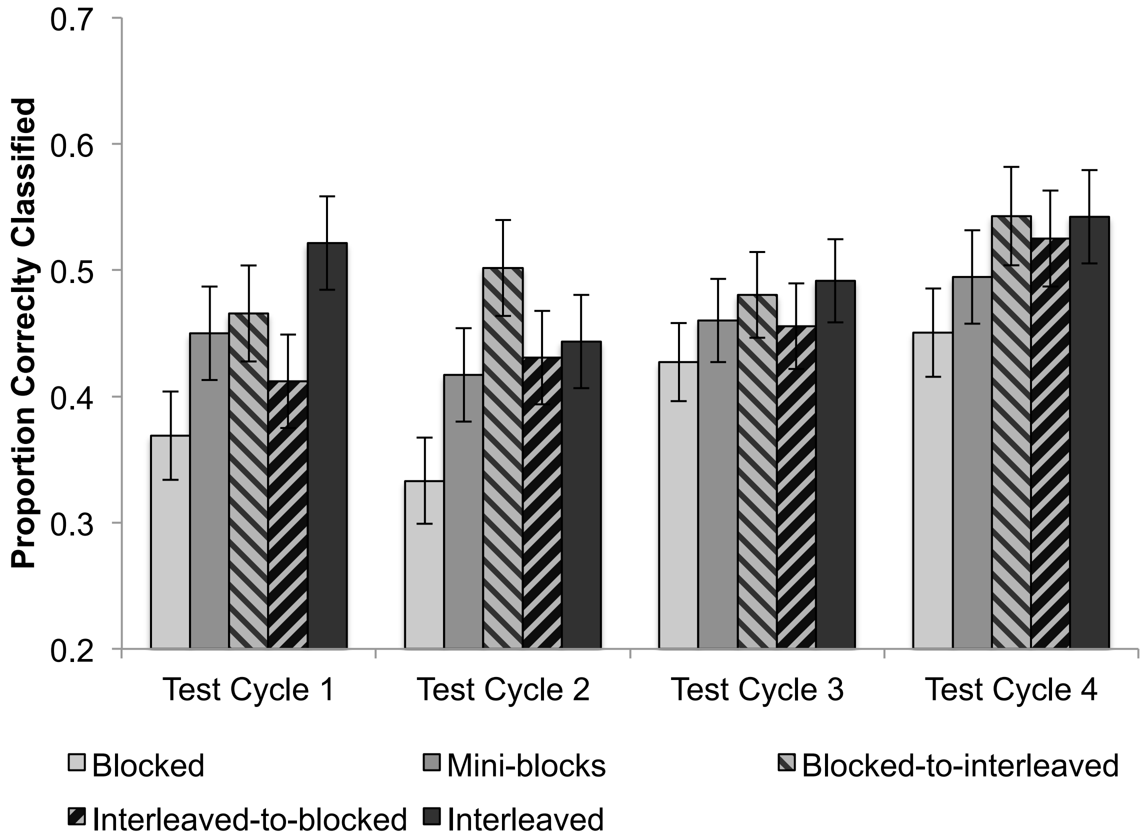
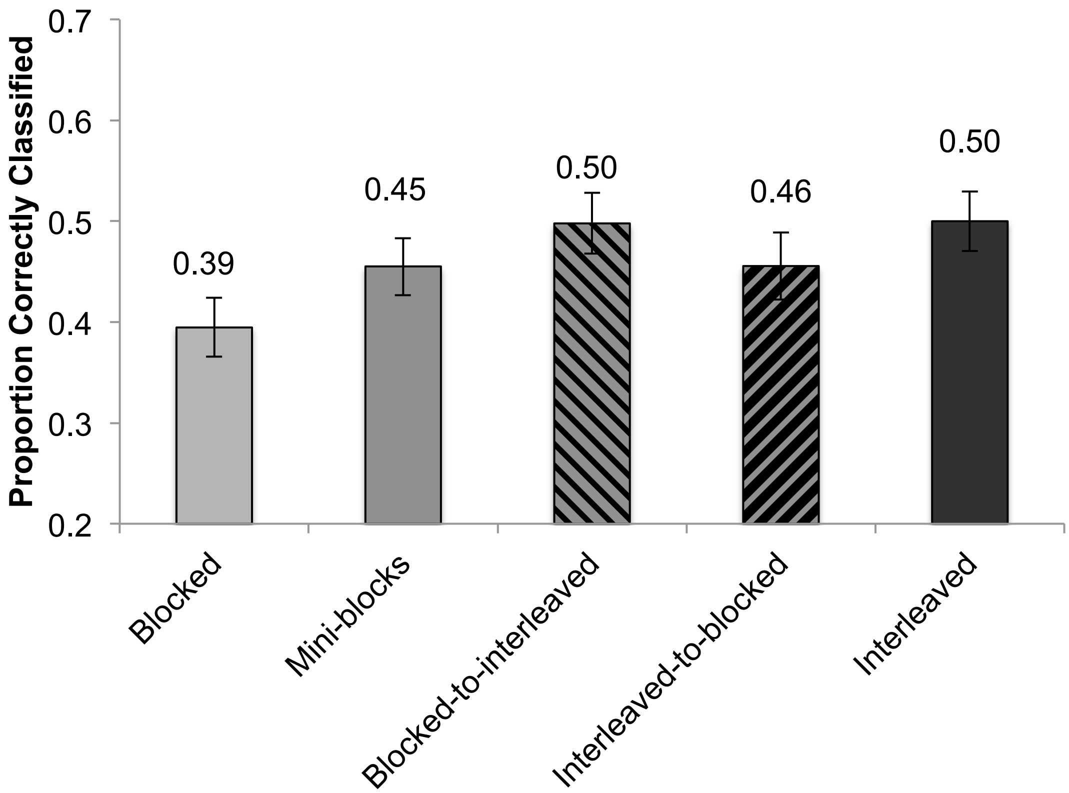


Figure 2. Proportion correct on the classification test for each type of study schedule across test cycles in Experiment 1. Error bars represent standard error of the mean.


0.7
0.6
0.5
0.4
0.3
0.2
Proportion Correctly Classified
0.39
0.45
0.50
0.46
0.50
Blocked
Mini-blocks
Blocked-to-interleaved
Interleaved-to-blocked
Interleaved

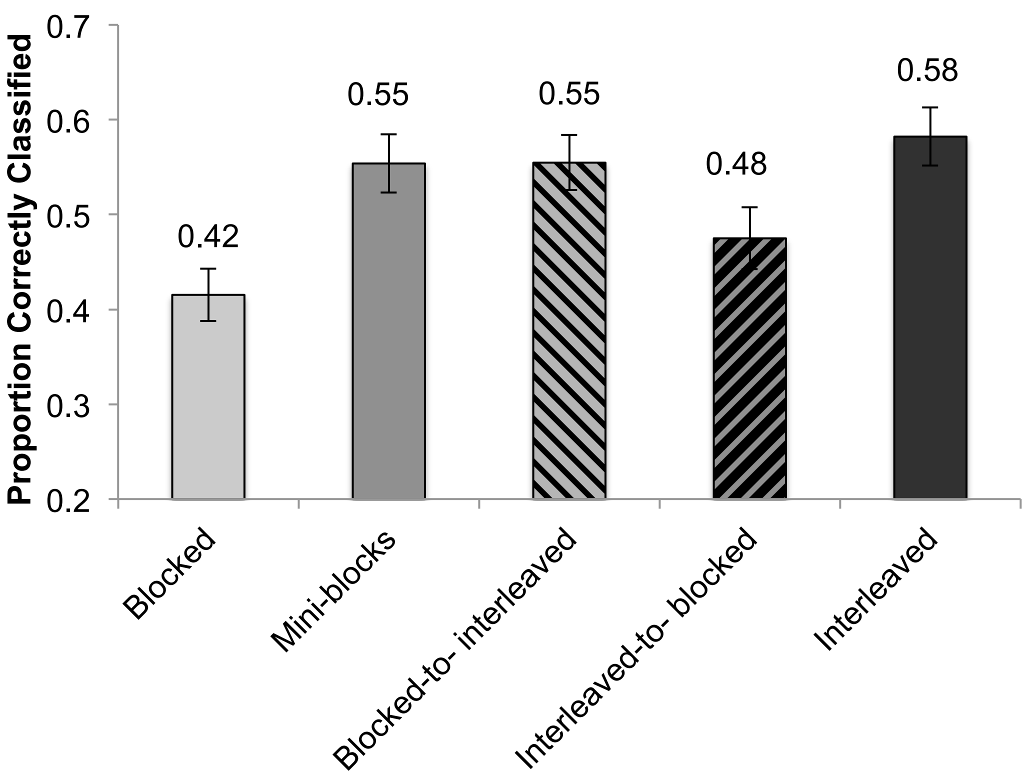

0.7
0.6
0.5
0.4
0.3
0.2
Proportion Correctly Classified
0.42
0.55
0.55
0.48
0.58
Blocked
Mini-blocks
Blocked-to- interleaved
Interleaved-to- blocked
Interleaved

Figure 3. Proportion correct on the classification test as a function of study schedules in Experiments 1 (top) and 2 (bottom). Error bars represent standard error of the mean.

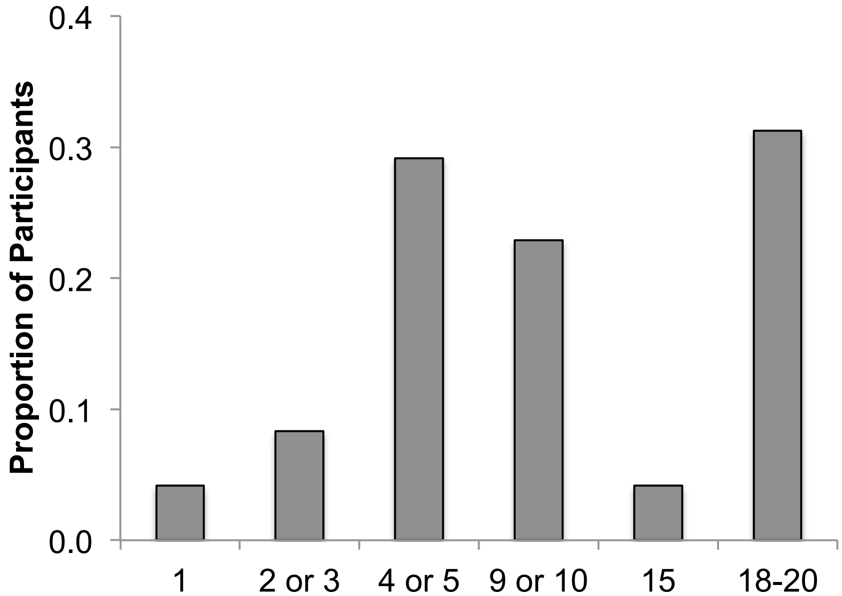


Figure 4. Maximum run lengths (i.e., maximum number of paintings by the same artist occurring consecutively) in the study schedules constructed by the participants in Experiment 4. The height of the bars conveys the proportion of participants for whom the value on the abscissa was the maximum run length in their self-constructed study schedule.

Supplemental Online Materials

S1: Post-test Description of the Five Schedules in Experiment 1

In this study, there were in fact five different types of study schedules. You just experienced **[PARTICIPANTS' SCHEDULE]** schedule of studying.

The five schedules were:

1. **Blocked:** Study one artist at a time (see six paintings by an artist in a row).
2. **Mini-blocks:** Study three paintings by each artist at a time.
3. **Interleaved:** Study all the paintings and artists all mixed up together.
4. **Decreasing blocks**: Start with three paintings by an artist in a row, followed by two in a row, and then all 12 mixed up together.
5. **Increasing blocks:** Start with all 12 artists mixed up, then see two paintings by the same artist in a row, and then see three by the same artist in a row

S2: Experiment 1 and 2 Metacognitive Judgments

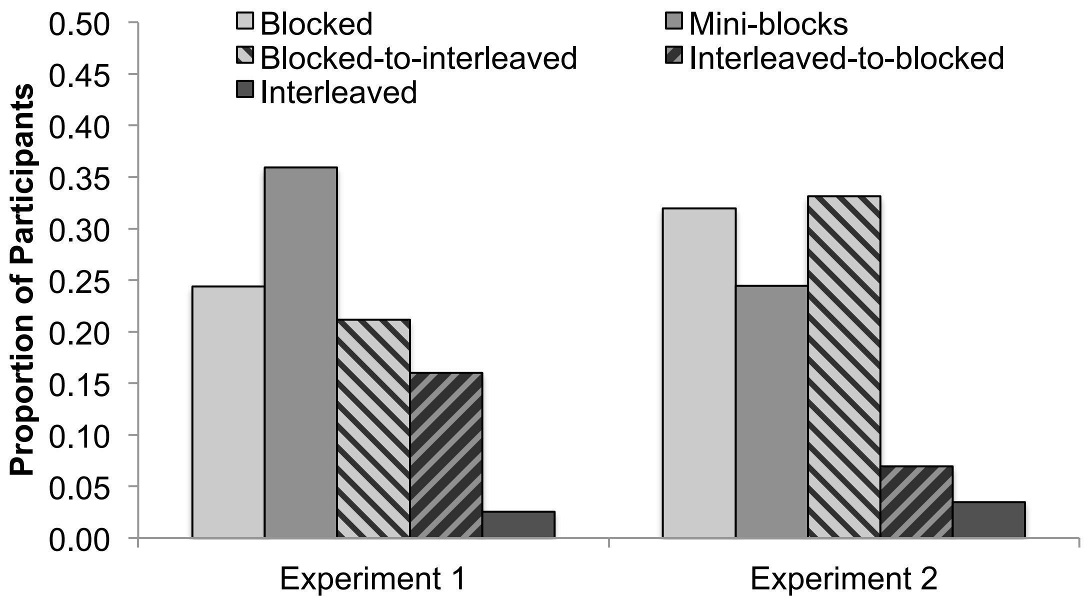


Figure S1. Post-experiment metacognitive judgments regarding which type of schedule would be the most effective for learning in Experiments 1 and 2.

Participants' beliefs about which study schedule would be the most effective for learning the artists' different painting styles for Experiments 1 and 2 are shown in the right half of Figure S1, collapsed across the assigned study schedules. We conducted a chi-square test of independence to test whether efficacy judgments (eliminating the interleaved and interleaved-to-blocked responses owing to insufficient numbers) were affected by the study schedule experienced. This analysis failed to find evidence that judgment and experienced schedule were independent in both Experiments 1 and 2, $\chi^2(8) = 13.17$, $p = .11$ and $\chi^2(8) = 12.25$, $p = .14$.

Participants' forced choice between blocked and interleaved schedules were not dependent on experienced schedule in Experiment 2, $\chi^2(4) = 3.78$, $p > .20$, but they were dependent on experienced schedule in Experiment 1, $\chi^2(4) = 14.25$, $p = .01$. Not a single person who had experienced pure blocking chose pure interleaving. However, 27% of those who had been in the mini-blocks condition and 26% of those who had been in the interleaved condition thought that interleaving would be more effective than blocking. Only 14% and 7% of participants from the blocked-to-interleaved and interleaved-to-blocked conditions, respectively, indicated that pure interleaving would be more effective than pure blocking.

S3: Effect of Schedules on the Induction of Specific Artistic Styles

Table S1

Average classification performance of each artist under each study schedule, combined across in Experiments 1 and 2

Artist	Interleaved	Blocked-to-interleaved	Mini-blocks	Interleaved-to-blocked	Blocked	Overall
Jacobsen	0.27	0.22	0.29	0.25	0.22	0.25
Lindenberg	0.32	0.29	0.32	0.19	0.25	0.28
Grote	0.34	0.40	0.34	0.24	0.22	0.31
Hawkins	0.36	0.36	0.36	0.21	0.27	0.32
Oliver	0.62	0.48	0.51	0.48	0.38	0.50
Cross	0.60	0.57	0.66	0.52	0.42	0.55
Juras	0.63	0.55	0.64	0.54	0.42	0.56
McNamara	0.63	0.58	0.56	0.56	0.47	0.56
Wexler	0.72	0.72	0.58	0.48	0.45	0.59
Mylrea	0.72	0.72	0.56	0.62	0.47	0.62
Seurat	0.70	0.65	0.69	0.64	0.63	0.66
Braque	0.88	0.81	0.76	0.74	0.56	0.75
Mean	*0.57*	*0.53*	*0.52*	*0.46*	*0.40*	

There are reasons to expect that the learning of different artistic styles would not always benefit from interleaving. Artists whose styles are more easily discriminated from those of other artists might, for example, benefit less from having exemplars of their paintings juxtaposed with those of the other artists. Combining the data from the first test cycle from Experiment 1 and the two test cycles from Experiment 2, the mean classification accuracy, sorted from most difficult to least difficult artist (defined by overall accuracy of classification) is displayed in Table S1. For the set of stimuli used in the present Experiments 1 and 2, however, we found what appeared to be consistent effects of schedules across the different artists. With the exception of the paintings by Grote in Experiment 1, pure blocking led to poorer performance than did pure interleaving for the induction of all other artists' painting styles (including Grote in Experiment 2). In fact, there is a remarkable similarity in the pattern of results across the artists.

Table S2.

Mean Rank Sum of each Study Schedule, when Examined by Artist, for Experiments 1 and 2 Separately and Combined (where rank of 1 = worst strategy, 5 = best strategy)

Schedule	Experiment 1	Experiment 2	Combined
Blocked	1.79	1.21	1.21
Interleaved-to-blocked	2.58	2.21	2.08
Blocked-to-interleaved	3.21	3.54	3.50
Mini-blocks	3.29	3.83	3.79
Interleaved	4.13	4.21	4.42

To analyze these results, we ranked each type of schedule from worst (1) to best (5) for the learning of each artist's style, and the mean rank sum for each of the five schedules is shown in Table S2, both separately and combined for the two experiments. A Friedman test confirmed that there were reliable differences between the five schedules, Friedman $\chi^2(4) = 34.88$, $p < .001$, and Wilcoxon rank sum tests revealed that all pairwise comparisons were significant, *p*s < .05, except for the difference between the mini-blocks and pure interleaved schedules, $p = .09$ (which was only marginally significant), and between the blocked-to-interleaved and mini-blocks schedules, $p = 1.00$.

These results show that, compellingly, the rank ordering of the effectiveness of the five tested schedules was remarkably consistent, even when examining the schedules by artist category, rather than just overall performance, suggesting that our findings are not unduly influenced by some specific-item effect.

S4: Scheduling study of artists — Post-test questionnaire for Experiment 4

Please describe any rationale or strategy behind the way you scheduled your study of the artists:

Which of the following orders does your schedule most resemble (select one):

A. Study all paintings by one artist consecutively (e.g. AAA…BBB…CCCC…)
B. Study all the artists mixed up together (e.g., ABCDE, BCDEA, CAEDB…)
C. Familiarize self with one painting by each of the 5 artists first, then study one artist at a time (e.g. ABCDE, Ax19, Bx19….)
D. Study two or three from each artist consecutively (e.g., AABBCC… or, AAABBB…)
E. Study five from each artist consecutively
F. Study 10 from each artist consecutively
G. Start by studying one at time and then mix them up (e.g. starting with blocks of AAAAA, BBBBB, etc., and then end with ABCDE, BCDEA, CAEDB….)
H. Start by mixing them up and then one at a time (e.g. starting with ABCDE, BCDEA, CAEDB, etc. and then ending with blocks of AAAAA, BBBBB, etc.)

Which of the above listed orders (A through H) do you think would lead to the best learning of artists' styles?

____

[please turn page over to continue]

Given the choice between options A (studying one artist at a time) and B (mixing the artists up), which would you prefer?

If/When you put paintings by the same artist next to each other, why did you do so?

If/When you mixed up the paintings by different artists, why did you do so?

What year in college are you? Freshman Sophomore Junior Senior+

What is your GPA? ____________________ What is your age? _______________

What is your gender? M / F

S5: Examples of Study Rationales in Experiment 4

Below is a sample of participants' responses to "Describe any rationale or strategy behind the way you scheduled your study of the artists," sorted by how they self-scheduled.

Pure blocking:
"I'd study each 20 paintings by one artist at a time. Mixing them, I think would confuse me."
"It is better to master an artist's method of painting by looking at different art works. Then, study another artist's paintings and analyze what his or her style is."
"You can't really determine the style of a painter or artist until you have looked at several pieces of their work. And so, by looking at all of the work of one artist at a time, you're better able to pick up on the pattern."

Blocks of ten:
"I like organizing material into big groups. I normally will study one group of related materials at a time and then move onto the next group."
"I would try to recognize 10 different paintings by for each artist, and by then I'd hope to get a little idea of their style, so when I do the next 10, it would further reinforce the little ideas I already had."

Interleave-Block (x2): *"I think first, I have to know the difference between 5 artists, so a pic by each artist will be shown first. Then to know more about each artist, I'll be shown 9 of their paintings to know more about their style. To make sure I got the differences between their paintings, one painting from each artist is shown again.*

Block of 15, then five: *"I'll examine each artist individually first by looking at 15 paintings done by each artist. I think that will definitely help me grasp their style and technique. Then I will cycle back through each artist's paintings, examining the last five to test my knowledge and make comparisons between different artistic styles."*

Blocks of five:
"I did it so you get exposure to each artist, can then test yourself a couple of times before finally reviewing. Also, you can compare them to other styles."
"Every five I chose to change an artist since I think it might be better to view the same artist's paintings again after a little break to test if I can recognize them immediately and refresh my memory."
"Studying the artist's paintings in a group to get an idea of the style; Mixing it up with the other artists to have a chance for review later on."

Blocked-to-interleaved:

“Well, I would spend some time on each artist to figure out their style and then I would mix it up to test if I could see patterns to prepare.”
“I wanted to mix the artists up as during the test, it is unlikely they would be arranged side to side. I also wanted to leave one artist at the end to heavily study so I could distinguish that specific artist from the rest, allowing me to note differences between Artist E and the rest, as well as differences between Artists A-D.”
“I first thought that going through 10 paintings of each artist would help get a feel for each one’s style. Then I did consecutive pairs, thinking it would highlight differences between artists .I did it backwards to try and mix it up a little, but not too varied at first. Then I mixed it up, first in order (ABCDE), then at the end, random.”

Pairs or triples:
“I feel like if you familiarize yourself with several paintings of one artist, then several of another, you’re more likely to recognize distinct patterns in one artist in comparison to the other artists.”
“First I wanted to get a sense of the artist’s style by just studying “bigger” groups of one artists. Then I would start comparing them to each other, by alternating the artist’s paintings.”

Interleaved:
“I believe it’s more effective to study that way because I will be able to distinguish between the paintings better instead of learning them in order. The teacher will most likely mix up the paintings on the test, so why not study that way?”