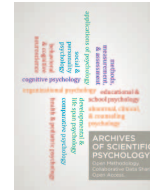
Archives of Scientific Psychology 2018, 6, 138–147
DOI: http://dx.doi.org/10.1037/arc0000042

2169-3269

American Psychological Association

Archives of Scientific Psychology

www.apa.org/pubs/journals/arc



SPECIAL SECTION: EDUCATIONAL PSYCHOLOGY

In Inductive Category Learning, People Simultaneously Block and Space Their Studying Using a Strategy of Being Thorough and Fair

Nate Kornell
Williams College

Kalif E. Vaughn
Northern Kentucky University

A B S T R A C T

Learning efficiency increases when study trials are spaced apart rather than blocked together. Interleaving is a kind of spacing in which similar items are studied in a mixed order. Previous research has shown that learners typically think blocking is more effective than interleaving. We allowed learners to control how much blocking and spacing they did. In our crucial second experiment, 61 participants studied five photographs each of 10 species of penguin. They arranged the spacing between repetitions of a given species by deciding, on each trial, which species they wanted to study next. During repetition learning (where each species was represented by a single photograph presented repeatedly), participants did more spacing and less blocking than they did during inductive category learning (where every photograph of a species was different). Rather than choosing blocking or spacing, participants chose blocking and spacing: They chose the same species on consecutive trials frequently, and when they did not their choices frequently maximized spacing. The high frequency of blocked and spaced responses was offset because participants rarely chose intermediate amounts of spacing. The authors hypothesize that participants used a "thorough and fair" strategy: Instead of consciously considering blocking or spacing when they decided how to study, participants blocked because they were trying to be thorough when they studied a given species, and spaced because they were trying to be fair to all of the species by not returning to one species until they had visited all of the others.

S C I E N T I F I C   A B S T R A C T

Studying something and then restudying it right away (blocking) is less effective than studying it and then restudying it again after a delay (spacing). Learners would do well to recognize the advantage of spacing, but they often rate it as being less effective than blocking. The authors looked at what happens when people are put in control of whether they block or space their own learning. Participants studied pictures of 5 penguins from each of 10 different species. They decided which species to study next on every trial. The results showed that participants did a lot of blocking, replicating previous research.

This article was published November 8, 2018.

Nate Kornell, Department of Psychology, Williams College; Kalif E. Vaughn, Department of Psychology, Northern Kentucky University.

This article is part of the special section "Educational Psychology." The guest editor for this section is Melody Wiseheart.

Thank you to Robert A. Bjork for his guidance and support. This research was supported by Scholar Award 220020371 from the James S. McDonnell Foundation.

The authors have made available for use by others the data that underlie the analyses presented in this paper (see Kornell & Vaughn, 2018), thus allowing replication and potential extensions of this work by qualified researchers. Next users are obligated to involve the data originators in their publication plans, if the originators so desire.

Correspondence concerning this article should be addressed to Nate Kornell, Department of Psychology, Williams College, 25 Stetson Court, Williamstown, MA 01267. E-mail: nkornell@gmail.com

When they were not blocking, though, they had a tendency to space more than would be expected by chance. The authors conclude that instead of consciously considering blocking or spacing when deciding how to study, participants blocked because they were trying to be thorough when they studied a given species, and they spaced because they were trying to be fair to all of the species by not returning to one until they had visited all of the others.

Keywords: spacing, interleaving, metacognition, choice, learning
Data repository: http://doi.org/10.3886/ICPSR36999.v1

Hundreds of experiments have demonstrated that study trials produce more learning when they are spaced apart rather than being blocked together (e.g., Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1988; Glenberg, 1979). The benefits of spacing are robust across materials (Hatala, Brooks, & Norman, 2003; Kornell & Bjork, 2008; Taylor & Rohrer, 2010), time scales (Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008), age groups (Balota, Duchek, Sergent-Marshall, & Roediger, 2006; Vlach, Sandhofer, & Kornell, 2008), and even species (Carew, Pinsker, & Kandel, 1972).

In this research we investigated a kind of spacing called *interleaving*. Spacing can involve studying information A, B, and C, doing something unrelated, and then studying A, B, and C again. In interleaving, by contrast, similar items are mixed together. Thus, for example, if one is to study A, B, C, D, and E, interleaving might mean studying them in this order: CBDAEDACEB. Like spacing, interleaving enhances learning, although not necessarily for the same reasons as spacing (Birnbaum, Kornell, Bjork, & Bjork, 2013; Carvalho & Goldstone, 2015; Kang & Pashler, 2012; Rohrer, Dedrick, & Burgess, 2014). In this article, we use the terms *spacing* when referring to events or effects that can be attributed to spacing, *interleaving* specifically when referring to mixing items together, and *blocking* to refer to studying the same item on consecutive trials.

Spacing and interleaving are attractive because they increase learning, are relatively easy to do, and do not require more study time than blocking (Dempster, 1988). Unfortunately, people tend to believe that blocking is more effective than spacing (Baddeley & Longman, 1978; Kornell, 2009; Kornell & Bjork, 2008; Simon & Bjork, 2001; Zechmeister & Shaughnessy, 1980). One potential explanation for this preference stems from the increased fluency that blocking affords, because higher fluency is associated with higher memorability ratings (e.g., Benjamin, Bjork, & Schwartz, 1998; Rhodes & Castel, 2008). If two presentations are spaced or interleaved, remembering the first presentation will be relatively difficult at the time of the second presentation. If the presentations are blocked, doing so will be relatively easy. People naturally assume that easy retrieval is a sign that information is well learned. Thus, they believe that blocking is more effective than interleaving. On the contrary, retrieval difficulty (i.e., low accessibility) often enhances learning (Bjork & Bjork, 1992; Pyc & Rawson, 2009). Spacing is a "desirable difficulty" (Bjork, 1994), that is, a manipulation that creates challenges during initial learning but enhances long-term learning.

Study Choices

If people's study decisions are guided by their metacognitive judgments, they should prefer blocking to spacing or interleaving. But study choices are not always guided by metacognitive judgments (see Kornell & Son, 2009).

Do people spontaneously choose to space their studying? A number of studies have investigated this question using paradigms in which participants could study the same word-pair multiple times. The evidence is somewhat mixed. Son's (2004) participants studied a word pair (e.g., hirsute—hairy) for one second, made a judgment of learning, and then decided whether to restudy now (i.e., block) or later (i.e., space). In general, participants blocked the word pairs they had rated as difficult but spaced the easy ones. Other studies found the opposite pattern: People blocked easy items and spaced difficult ones (Benjamin & Bird, 2006; Pyc & Dunlosky, 2010; Toppino & Cohen, 2010; Toppino, Cohen, Davis, & Moors, 2009). It appears that these disparate findings occurred because participants only prefer to block difficult pairs when, as in Son's (2004) study, they are given very little time to study and do not have to choose blocking for half of the items and spacing for the other half.

Son and Kornell (2009) investigated spacing decisions in a different way. Instead of examining the effect of item difficulty on spacing, they asked a more basic question: Do people prefer blocking over spacing, or vice versa? After studying a list of word pairs, participants were shown all of the cue words along with a set of "slots" in which the cues could be placed. The participants were told they would be allowed to study the pairs in the order of the slots, from top to bottom. This paradigm gave participants more freedom to arrange their study schedule than previous studies, and it allowed them to create a complete study schedule before studying. They chose to space significantly more than would be expected by chance.

Interleaving During Inductive Category Learning

The studies described so far examined verbal learning. Like most spacing research, they involved *repetition learning*: the same information was presented multiple times and then tested. Not all learning involves exact repetition, however. In the present experiments we compared repetition learning to *inductive learning*, that is, learning a concept or category via exposure to multiple different exemplars.

Inductive learning appears to benefit from interleaving. For example, Kornell and Bjork (2008) asked participants to study paintings by various artists. Presentations of an artist's paintings were either interleaved with paintings by other artists (i.e., spaced), or presented consecutively (i.e., blocked). More learning occurred in the interleaved condition, but most participants judged that blocking had been more effective than interleaving. Similar results have been obtained in subsequent studies (Birnbaum et al., 2013; Carvalho & Goldstone, 2014; Kang & Pashler, 2012; Kornell, Castel, Eich, & Bjork, 2010; McCabe, 2011; Noh, Yan, Vendetti, Castel, & Bjork, 2014; Rohrer et al., 2014; Yan, Bjork, & Bjork, 2016; Zulkiply & Burt, 2013).

In the present research, we examined study choices in both inductive learning and repetition learning. Kornell et al. (2010) replicated Kornell and Bjork's (2008) findings with respect to inductive learning, but they also examined a repetition learning condition. In this condition, instead of each artist being represented by multiple paintings, a single painting was shown for each artist, repeatedly during study and then again on the test. The benefit of interleaving was the same size for induction and repetition, but the participants' judgments were different. They judged blocking to be more effective than interleaving for inductive learning, replicating the previous finding. But they correctly judged interleaving to be more effective than blocking following repetition learning. Blocking was equally inefficient in both conditions. Kornell et al.'s study suggests that people might choose to space practice during repetition learning but block practice during inductive learning.

Study Choices During Inductive Category Learning

A recent study examined whether learners prefer to interleave or block their practice when learning novel categories. Tauber, Dunlosky, Rawson, Wahlheim, and Jacoby (2013) had participants learn various bird exemplars (e.g., Song Sparrow) from different bird families (e.g., Sparrow). During practice, participants controlled the order of exposure to these various bird families. The layout of choices differed across experiments. In Experiment 1, participants were shown buttons arranged in a 12 (number of bird families) × 6 (number of exemplars) array. They could freely select which exemplar from which family to study next. They could choose pure blocking (e.g., study six exemplars from the Chickadee family all in a row), pure interleaving (e.g., study six different species in six consecutive trials), or some combination of both. Regardless of the selection format, Tauber et al. (2013) always found that participants preferred to block their study. The percentage of exemplars that were blocked was always higher than 70% across experiments. Furthermore, 78% or more of the participants were considered "blockers" (i.e., they did more blocked studying than interleaved studying during learning). Overall, these results suggest that participants tend to block their practice when learning from bird categories.

The Present Experiments

Experiment 1 was a conceptual replication of Kornell and Bjork's (2008) first experiment. Instead of artists' paintings, participants learned to identify various species of penguins. In addition to its value as a replication, Experiment 1 served two purposes. First it tested the assumption that interleaving would be effective with the learning materials used in these studies. Second, it provided a measure of item difficulty that could be used in the analyses of Experiment 2.

Experiment 2 examined participants' decisions about how much they wanted to interleave versus block their studying. Participants were not asked to decide when they wanted to study an item in the future, because research suggests that students do not typically plan out future study sessions or future study trials (Kornell & Bjork, 2007).

Instead, their decision process seems to focus on what they want to study next. (When students do space their studying, therefore, it is not necessarily on purpose; spacing may be a byproduct of a decision process that does not consider the amount of spacing.) Thus, the procedure in Experiment 2 was similar to that used in Experiments 3 and 4 by Tauber et al. (2013): On each trial, participants were asked which species they wanted to study next. There were two groups of participants: In the inductive condition, each picture of a given species was different; in the repetition condition, each picture of a given species was the same.

Our study represents a conceptual replication of Tauber et al.'s (2013) research. It also extends that research in two important ways. First, we compared inductive learning to repetition learning. Second, we used novel approaches to data analysis that challenge existing conclusions about spacing.

Experiment 1

In Experiment 1, participants were asked to learn to identify 10 species of penguins. To do so, they were shown five different pictures of each kind of penguin. Half of the species were presented blocked and half were presented interleaved. After a multiple-choice test on novel photographs, participants were asked which condition they thought had been more effective, blocking or interleaving.

Method

Participants and design. The participants were 45 University of California, Los Angeles, undergraduates (37 female, eight male, mean age = 19.9 years) who were recruited from a psychology department subject pool and participated in exchange for credit in their psychology classes. The only independent variable was interleaving versus blocking, which was manipulated within-participants. This study was approved by the University of California, Los Angeles Office for Protection of Research Subjects.

The minimum sample size was determined based on the results of Kornell and Bjork's (2008, Experiment 1a). They found an effect size of .99 in a within-participant design. Assuming this effect size, we calculated that we needed a minimum of 16 participants to achieve 95% power.

Materials. The materials were 70 photographs of penguins (see Figure 1). There were 10 penguin categories, with seven photographs per category; 50 photographs were used during study, and 20 were used during the final test. Although each species had a label, it was not possible to use the actual names for some of the penguins because they were too descriptive (e.g., Chinstrap, Yellow-Eyed). Thus, the penguins were labeled Adelie, Conall, Gentoo, Janakima, Kohar, Lachesis, Odovacar, Peredur, Reinhard, and Shivali. (The corresponding actual names were Adelie, Chinstrap, Gentoo, Emperor, Galapagos, Snares Island, King, Fairy, African, and Yellow-Eyed).

The backgrounds in the photos were a mix of ice, snow, liquid water, rocks, mud, and vegetation. The background was not exclusively snow and ice in any category; snow or ice were prominently displayed in five of the categories and minimal or completely absent in the other five. The pictures were obtained from three web pages in 2007. When we entered the URLs for these pages in 2017, they all showed a message that the page could not be found.

Procedure. The study was conducted online. After reading instructions describing the procedure, participants studied five photographs from each of 10 penguin categories. The photographs were presented for 3 s each. The category name (e.g., Adelie) was printed beneath each photograph.

Half of the categories were presented blocked, meaning that all 5 photographs of a given species were presented consecutively. The other categories were presented interleaved, meaning that the species were mixed together. The presentations occurred in blocks of five items. Each block either consisted of five photographs of one category (blocked, or B) or one photograph from each of the spaced categories (interleaved, or I). The order of blocks was IBIBIBIBIB. Categories were randomly assigned to one of these positions on a by-participant basis. Within a given category, the exemplars were always shown in the same order (e.g., the first encounter with a Kohar showed the same picture for all participants).

After the study phase ended, participants were asked to count backward by threes from 547 for 15 s, and then the final test began. None of the photographs presented during the final test had been presented during study. During a test trial, a photograph was presented with 10 buttons below it. Each button was labeled with a category name (e.g., Adelie) and the participant was asked to select the species in the photograph. After making a correct response, the participants were told that they had been correct; after making an error, they were told what the correct answer was. Test trials were self-paced and feedback was presented for 2 s.

Figure 1. Sample images from Experiments 1 and 2. These images are identical to the ones shown in the actual experiments. In the experiments they were all shown at a height of 400 pixels and their width varied. The actual names for the Lachesis, Janakima, Odovacar, and Shiavli penguin are Snares, Emperor, King, and Yellow-Eyed, respectively. Top row photo credits: Thomas Mattern, Dave Houston, and Hugh Pearson. Bottom row photo credits: Josh Landis, Sean Mack, and Dave Houston.

There were 20 test trials, which were split into two blocks. All 10 categories were tested, in random order, in the first block. Then all 10 categories were tested again, using new photographs, in a new random order. The results of the second test block will not be presented for two reasons. First, there was no second test block in Experiment 2 (in Experiment 2, in the repetition condition, a second test block would not have made sense because there was only one photo per category and it was tested in the first block). Second, feedback was given during the first test. This made the first test into a spaced learning opportunity for all items, blurring the distinction between the spaced condition and the massed condition (the latter of which became a massed-then-spaced condition).

After finishing the final test, the participants were asked a final question. The terms *massed* and *spaced* were defined as follows: "MASSED—A given penguin's photos are shown consecutively (e.g., 5 consecutive paintings are all of the same kind of penguin). SPACED—A given penguin's photos are shown intermittently, with other photos in-between (e.g., 5 consecutive photos might all be of different kinds of penguins)." (The word *paintings* was used instead of *photo* due to experimenter error.) Participants were asked "Which do you think helped you learn more, massed or spaced?" (Although it would have been technically correct to refer to the conditions as blocked and interleaved instead of massed and spaced, our participants did not know this and, moreover, *massed* and *spaced* seem like more intuitive terms).

Results

The proportion of responses that were correct was significantly higher in the interleaved condition (M = .69, SD = .26) than the blocked condition (M = .50, SD = .26), $t(44) = 4.23$, $p < .0001$, $d = .64$. Thirty of 38 participants (79%) did better in the interleaved condition than the blocked condition (seven did equally well in the two conditions). Nevertheless, 40 of the 45 participants (89%) believed that they had done better in the blocked condition—despite making the rating after taking the final test. In short, interleaving was more effective than blocking for inductive learning, but blocking was rated as more effective. These findings replicate those of Kornell and Bjork (2008).

Experiment 2

In Experiment 2, participants studied the same materials and took the same type of test, as in Experiment 1. There were two important differences between the experiments. First, the order in which the categories were studied, and thus the degree of spacing, was under the participant's control. Second, there was a repetition learning condition in which each species was represented by a single photograph that was used during the study phase and again on the test. There was also an inductive learning condition, in which no photograph was ever repeated, as in Experiment 1. Based on the results of Tauber et al. (2013), we predicted that participants would choose to block study rather than interleave. We also predicted, based on the results of Kornell et al. (2010), that participants would be more inclined to space their studying during repetition learning than during inductive learning.

Method

Participants and design. The participants were 61 University of California, Los Angeles, undergraduates who were recruited from a psychology department subject pool and participated in exchange for credit in their psychology classes. (Because of a researcher error, participant demographics were not collected.) Thirty-six and 25 participants were assigned to the inductive learning and repetition learning conditions, respectively.

The sample size was determined based on the rule of thumb that between participant experiments should have roughly 30 participants or more per condition.

The repetition condition did not quite reach this threshold because random assignment was truly random and so the participants did not split into conditions evenly. We did not feel it was necessary to increase the sample size further because we expect repetition versus induction to have large effects on choice behavior and we expected participants to have a strong preference for blocking over interleaving.

Materials. The materials were 60 of the 70 penguin photographs used in Experiment 1 (see Figure 1). The reason only 60 of the photographs were used is that it was not feasible to test each category twice in the repetition learning condition, so in both conditions each species was tested only once. Of the 60 photographs used in the induction condition, only the 10 that were used on the test in the induction condition were used in the repetition condition. Thus, the same items were tested in both conditions.

Procedure. The experiment took place online. After reading instructions, participants began the study phase. They studied five photographs from each of 10 penguin species for 3 s each, like in Experiment 1. In the repetition condition, all five photographs of a given species were exactly the same during the study phase. Each trial began with a screen containing 10 buttons arranged in a single vertical column. Each button was labeled with a category name (e.g., Gentoo). The order of the species was ordered randomly on a participant-by-participant basis. Next to each button a number was displayed that represented the number of times the button could be pressed. At the outset of study, all of the numbers read 5. If the participant pressed the Gentoo button, a photograph of a Gentoo penguin was displayed for 3 s with the word *Gentoo* under it, and then the screen containing 10 buttons reappeared. Because the Gentoo button had been pressed one time and could only be pressed four more times, the number next to it now read 4. After all of the buttons had been pressed five times, a link appeared that allowed the participant to advance to the test.

During the final test, 10 photographs, one from each species, were presented one at a time in random order. The participants were shown the names of all 10 types of penguin and asked to select, and type in, the name corresponding to the photograph. Responding was self-paced. No feedback was provided during the final test. In the repetition condition, the photographs were the same ones that had been presented during study; in the induction condition, they were novel photographs.

After completing the test, participants were asked the following question: "Which do you think is a more effective way of studying in this experiment?" They were provided with two response options. One represented blocked study: "Study one kind of penguin repeatedly, then move on to the next kind." The other represented spaced study: "Study one kind of penguin just once or twice at a time, then move on to other types, and come back to revisit previously studied penguins again later."

Results

We begin the results section with an analysis of blocking rate, a metric that was reported previously in similar studies (Tauber et al., 2013). We then turn to an additional set of novel analyses. We also wanted to estimate how similar participants' responses were to what would be expected from random responding (i.e., responding that was not guided by any strategy). We estimated random responding by running a Monte Carlo simulation in which we simulated 10,000,000 sessions of random responding and analyzed the pattern of results.

Blocking rate. To examine the blocking rate during practice, we classified each trial as either a blocked or interleaved following the process used by Tauber et al. (2013). Trial N was classified as blocked if the same species was studied on the prior trial ($N - 1$) or the next trial ($N + 1$); otherwise it was categorized as interleaved. For example, in the study sequence of gentoo, gentoo, lachesis, lachesis would be labeled as two sets of blocked trials (i.e., four blocked trials) and would not count the switch from gentoo to lachesis as an interleaving trial. If the study sequence had continued as gentoo, gentoo, lachesis, lachesis, janakima, kohar, kohar, then the sequence would be labeled as six blocked trials and one interleaved trial.

Another way to understand the computation of blocking rate is by examining Figure 2, which displays two participants' response patterns. These participants were selected because their responses were near average on a number of variables, including blocking rate, other analyses to be described shortly such as total spacing, and distribution of blocking across the course of the session. As an example of how blocking rate was computed, the bottom row of Figure 2A represents five study trials on a species, two that were interleaved and three that were blocked. In the bottom row of Figure 2B, there are four blocked trials and one interleaved trial.

After each response had been classified (except the very first, which was not analyzed), a blocking rate was computed by dividing the number of blocked trials by the total number of trials. Participants were then categorized as blockers if their blocking rate was greater than 50%.

Our simulation showed that random responding would produce an average of 7.67 blocked responses out of a possible 49 responses that could be blocked, a blocking rate of 15.6%. The actual blocking rate was 47.2% (SD = 31.6%) in the inductive learning condition and 31.9% (SD = 33.3%) in the repetition learning condition. Both scores were significantly higher than 15.6%: $t(35) = 6.00$, $p < .0001$, d = 1.25, and $t(24) = 2.45$, $p = .02$, $d = .73$, respectively.

The blocking rate was higher in the induction condition than the repetition condition. This difference, which was predicted based on the results of Kornell et al. (2010), was significant based on a one-tailed test, $t(59) = 1.81$, $p = .04$, $d = .47$. Tauber et al. (2013) analyzed the average length of a run of blocked trials (e.g., studying the same category three times would be a run of length three). In their first experiment, for example, their participants studied an average of 6.0 exemplars from the same category before moving to a new category. In our data, the average run length was 2.03 trials (SD = 1.29) in the induction learning condition and 1.66 trials (SD = 1.46) in the repetition learning condition. Tauber et al. provided six exemplars for study, which means that on average their participants viewed every exemplar before moving on to the next category (their participants could study a category more than six times, but if they did they would have to see exemplars they had seen before). Thus, their data show that their participants typically studied all of the exemplars in a category before moving on to the next category. Our data are quite different, as participants rarely studied every exemplar in a category in a single run. In the induction condition, 32 of 36 participants never studied a category five times in a row, and of those who did, two did so only once, one did so eight times, and one did so 10 times. In the repetition condition, 23 of 25 participants never studied all of the exemplars in a category in a row, with one doing so once and one doing so 10 times. Our simulation showed that by chance, a category would be studied five times in a row less than once in 1,000 sessions.

Like Tauber et al. (2013), we classified participants as blockers if their blocking rate was above 50%. In the inductive learning condition, half of the participants were categorized as blockers (18/36 = 50%). In the repetition learning condition, six of 25 = 24% were blockers. We also classified participants using a cutoff of 15.6%, which is the blocking rate that would be expected by chance in our paradigm. Using this cutoff, 26 of 36 = 72% were blockers in the induction learning condition and 13 of 25 = 52% were blockers in the repetition learning condition.

Three main conclusions can be drawn from these findings. First, our participants chose to block at a rate much higher than would be

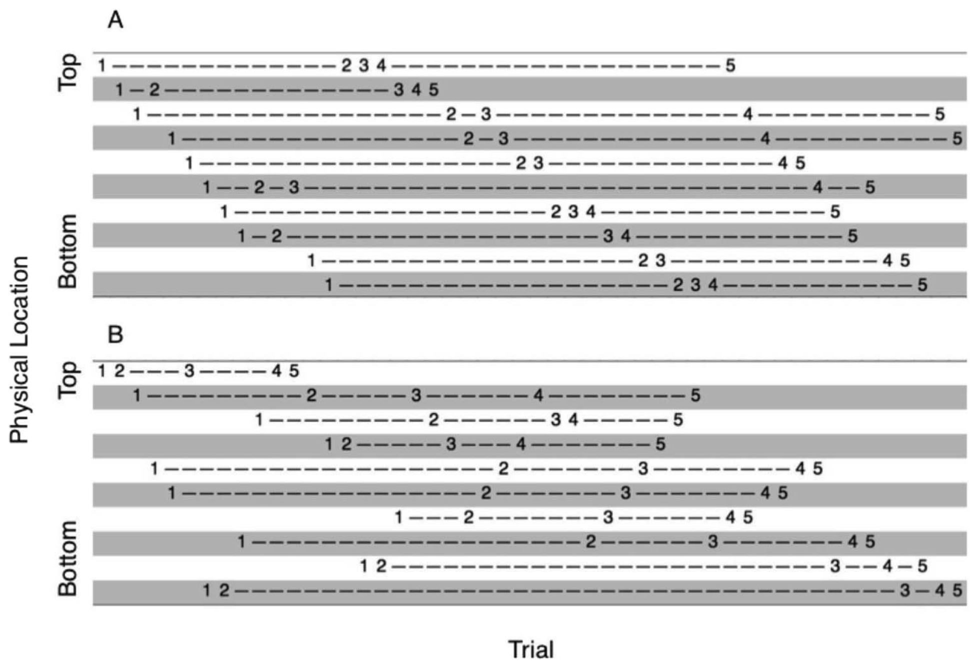


Figure 2. The responses of two participants, one in the induction condition (A) and one in the repetition condition (B). The vertical axis represents the vertical positions of the buttons the participants pressed to select a species for study. The horizontal axis represents the 50 trials in the experiment. The presence of a number within the figure signifies that that row's button was selected on that trial; the value of the number represents the number of times that button had been pressed. Dashes are included between the first and last response on a button. Thus, for example, the first five responses in Panel A were to buttons 1, 2, 3, 2, and then 4.

expected by chance. Second, the blocking rate was higher in the inductive learning condition than in the repetition learning condition. Third, our results diverged from those of Tauber et al. (2013). In their studies, participants blocked 70–90% of the exemplars they studied and tended to study every exemplar from a category before moving to the next category. Our participants blocked less than 50% of their exemplars and, with the exception of three participants, almost never blocked all of a category's exemplars in a row. There are a variety of differences between our procedure and Tauber et al.'s (2013) that could explain this difference and without further research there is no way to identify which mattered, as we discuss further in the general discussion.

Our results thus far can be summarized as showing that our participants' strategy was to block their learning. It follows that they chose to interleave less often than would be expected by chance. Nonetheless, the majority of their responses were interleaved.

There are countless choices to be made when one interleaves; for example, the sequences ABABCDCDEFEF and ABCDEFABCDEF are both completely interleaved, but they differ in the amount of spacing between repetitions of an item (and in how much learning should occur; Cepeda et al., 2006). The next three analyses focus on the spacing of our participants' responses.

Total spacing. We computed total spacing as the number of trials that intervened between the first and last study trial. For example, if a category was studied five times in a row, the total spacing would be three items because that is how many items separate the first study trial from the last; if it was the first and last item studied in the session, its total spacing would be 48 items. For example, in the top row of Figure 2A, the total spacing is 35 items (because there are 32 dashes and three numbers between the first and last study trial) and in the top row of Figure 2B the total spacing is 10 items.

The total spacing between the first and last study trial is important, because research suggests that when participants learn via study trials (or tests with feedback), the total spacing between the first and last trial affects learning, but the arrangement of items in between the first and last trial is less important (Balota, Duchek, & Logan, 2007; Karpicke & Roediger, 2007, 2010; Landauer & Bjork, 1978; Storm, Bjork, & Storm, 2010).

Hypothetically, a participant could have an average total spacing anywhere from 3 to 39 items. Random responding would result in an average total spacing of 33 items. The average total spacing scores of the induction condition ($M = 30.71$, $SD = 9.73$) and repetition condition ($M = 32.34$, $SD = 8.01$) did not differ significantly from the rate expected based on random responding. Neither did they differ from each other (all p values $> .15$). In isolation, this finding would suggest that participants were indifferent to blocking versus spacing, but of course this interpretation is inconsistent with our analysis of blocking rates. This inconsistency can be explained by the next set of analyses.

Exemplar-to-exemplar spacing. We next examined the spacing between any two selections of a given category. The unit of analysis here was pairs of responses within a category, which means that we analyzed four pairs of responses for each category (response 1–2, 2–3, 3–4, and 4–5). For example, if a category was studied twice in a row (i.e., blocked), the spacing of that pair of trials would be zero; if it was studied and then all of the other categories were studied, once each, before the original category was studied again, its spacing would be nine. In the top row of Figure 2A, the four exemplar-to-exemplar spacing values are 13, 0, 0, and 19; in the top row of Figure 2B they are 0, 3, 4, and 0.

As the histogram in Figure 3 shows, two responses stood out in both the induction condition and repetition condition: Studying a category twice in a row (i.e., blocking) and studying after an interval of nine other trials. Just as important, degrees of spacing in between zero and nine, that is, pairs separated by between one and eight trials, stood out by their absence. In other words, participants chose to block consecutive trials or they chose to space, but they avoided short spacing gaps

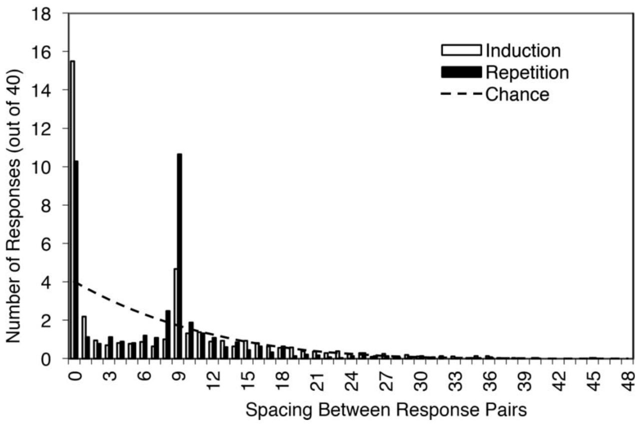


Figure 3. A histogram showing the number of response pairs that were separated by 0, 1, 2 . . . , 48 trials. The dashed line represents random responding. There were 40 total response pairs in a session (i.e., four response pairs for each species).

in between. Pairs separated by more than nine trials occurred at a rate roughly equivalent to chance.

The analyses of blocking rate, above, showed that participants blocked their study (i.e., studied the same category on consecutive trials) more than would be expected by chance. To assess whether participants spaced their nonblocked trials, we excluded blocked trials from the next analysis. Based on random responding, excluding blocked trials, any two repetitions of a category would be spaced by an average of 8.33 other trials. The actual spacing, excluding blocked trials, was significantly higher than 8.33 in the induction condition ($M = 12.54$, $SD = 5.32$), $t(34) = 4.69$, $p < .0001$, $d = .79$, and the repetition condition ($M = 11.36$, $SD = 5.10$), $t(23) = 2.91$, $p = .008$, $d = .59$. Thus, when participants did not block, they tended to space at a level greater than chance.

In short, participants chose to both block and space, while avoiding responses in between. This outcome seems to explain why total spacing did not differ from chance: On a pair-by-pair basis, the blocking rate was high but so was the spacing rate, and these two rates cancelled each other out with respect to total spacing.

There is a difference between a response pattern and a strategy. Our participants' response patterns showed preferences for both blocking and spacing, but we hypothesize that their strategy did not actually involve consideration of blocking or spacing. Their strategy seemed to be as follows: Choose a category that was a high priority because it had not been studied recently, and study it thoroughly, meaning often more than once. Then assign that category a low priority, because it had been studied recently, and choose a different high-priority category. This strategy will result in a pattern of blocking and spacing with few responses in between, which is exactly what we found.

Category-to-category spacing. It is impossible to prove our hypothesis that our participants used the strategy we are calling thorough and fair. This hypothesis does make another testable prediction, though: The number of categories studied in between repetitions of a given category should be about nine. We examine this prediction next.

Earlier we examined the number of trials that intervened between repetitions, but this number should vary depending on how much blocking participants do. The number of other categories that intervene between repetitions of a category is not affected by the amount of blocking that occurs. For example, suppose someone studied categories A–E in the following order: ABBBCCDDDBBEEAA. The category-to-category spacing between the first and second A would be five, because five other blocks of categories (B, C, D, B, E) were studied between the first and second block of A trials. If another participant chose an order of ABCDBEA, the category-to-category spacing would be exactly the same (B, C, D, B, E), but these two sequences differ with respect to blocking rate, total spacing, and exemplar-to-exemplar spacing.

We examined category-to-category spacing by counting each block of consecutive trials on the same category as a single study event and determining the number of these study events between pairs of selections of a given category. For example, in the top row of Figure 2A, there are three study events (1, 234, and 5) and the category-to-category spacing was 13 and 10; in the top row of Figure 2B there were three study events (12, 3, and 45) and the category-to-category spacing was 3 and 3.

The benefit of this analysis is that it is the only way to examine how fair the participants' responses were (i.e., how many other categories they visited before returning to a given category) independent of how thorough their responses were (i.e., how many times they studied a given category before moving on to another category). A perfectly fair response strategy with 10 categories would always result in a category-to-category spacing of nine, regardless of how much blocking was involved.

The category-to-category spacing means were 8.35 and 8.55 in the induction and repetition conditions, respectively. The median and mode were 9 in both conditions.

Moreover, intervals of nine other categories accounted for 30% and 42% of all responses in the induction and repetition conditions, respectively, far more than any other interval (see Figure 3). In short, after leaving a category, participants tended to study every other category before returning to study again, which supports the hypothesis that they were guided by the principle of being fair in selecting categories.

Individual differences. One important result was difficult to quantify: Every participant's response pattern was unique. Given our paradigm, the number of possible response patterns was over 4^{43}, a number larger than the mass, in kilograms, of 10 trillion suns. Figure 2 displays the responses of two of our participants. The participant in the induction condition in Figure 2A visited most of the categories once, then blocked practice in the middle of the session, and then paid a last visit to most of the categories at the end of the session (i.e., crammed). The participant in the repetition condition (Figure 2B) did less blocking than the induction participants. Ironically, his or her responses also resulted in less total spacing than the induction participant's.

The responses of these typical participants fit with the thorough and fair strategy described above. Both participants tended to study a category and then either stick with it or come back to it after a meaningful delay. Figure 2 shows that participants did not necessarily come back to a category after studying nine other exemplars, but they did tend to come back after studying about nine other categories. Clearly, though, there was variability in participants' response patterns and they were neither always thorough nor always fair.

Metacognitive judgments. After completing the final test, participants were asked which they thought was more effective, blocking or spacing. Sixty-seven percent said spacing. In Experiment 1, by contrast, 88% said blocking. Consistent with Kornell et al.'s (2010) results, more participants favored spacing in the repetition learning condition (76%) than the in the inductive learning condition (61%).

Why did 88% of participants rate blocking as more effective than spacing in Experiment 1, whereas only 33% did so in Experiment 2? One possibility is that whereas both experiments defined blocking as studying a species all at once, they differed in the way spacing was defined: In Experiment 1, it was defined as never repeating a species but in Experiment 2 the definition was, "Study one kind of penguin just once or twice at a time, then move on to other types, and come back to revisit previously studied penguins again later." It is possible that this description of a mix of blocking and spacing was more appealing than pure blocking or pure spacing and, in particular, participants may have wanted to avoid pure spacing in both experiments. This possibility is consistent with the mix of blocking and spacing evident in the study choices in Experiment 2.

This finding also suggests that previous research on metacognitive judgments about spacing may have missed something important. Previous research, and Experiment 1, suggests that people think pure blocking is a better way to learn than pure interleaving. This false belief is a damning indictment of students' metacognitive abilities. Our results suggest that this might be a false choice between two bad options. Our participants seemed to think a combination of interleaving and blocking was preferable to pure blocking (which they prefer to pure interleaving). This belief is much less damning and, in fact, may be adaptive; some research has shown that allowing students benefit from being allowed to make their own decisions about studying (Carvalho, Braithwaite, de Leeuw, Motz, & Goldstone, 2016).

Final test accuracy. Proportion correct on the final test was significantly higher in the repetition learning condition (M = .93, SD = .17) than the inductive learning condition (M = .55, SD = .25), $t(59) = 6.57$, $p < .0001$, $d = 1.76$. This finding is to be expected because the tested photograph was being presented for the sixth time in the repetition learning condition but only the first time in the inductive learning condition.

We did not analyze the impact of spacing versus blocking on learning. One reason for this decision is that it is not obvious how people should be divided into groups, given that a single individual might be considered a heavy blocker and a heavy spacer at the same time. The other reason is that without randomly assigning participants to conditions, we do not think it is possible to accurately estimate the benefit of spacing and blocking.

Results summary. The initial question was, would people block or space their study? The answer appears to be both. Participants often blocked their studying by choosing the same category twice or more in a row. When they did move away from a category, though, they tended not to come back to it until they had studied a good number of other categories; that is, they avoided intermediate gaps and chose either blocking or a large amount of spacing between pairs of exemplars in a category. Inductive learners favored blocking more than repetition learners. Judgments of which was more effective, blocking or spacing, showed that the majority of participants in both groups judged a mix of blocking and spacing to be more effective than pure blocking, although this preference was especially pronounced in the induction learning condition.

General Discussion

A bee foraging for nectar will visit a flower and spend as long as necessary to extract the nectar, "blocking" its attention if the nectar is abundant. It will then move on to the next flower and put off returning to the first flower for a long time. That is, it will "space" its visits by visiting the other flowers before returning to the first one. Participants in Experiment 2 may have used the same strategy when foraging for information (see Metcalfe & Kornell, 2005; Pirolli & Card, 1999), a strategy we call being thorough and fair. They were thorough when studying a category, often studying it multiple times, which resulted in blocking. They were fair when choosing the next category they would study, often choosing the category studied least recently, which resulted in spacing. We cannot prove that this strategy guided our participants, but it would produce the pattern of data we obtained.

A participant guided by the principles of thoroughness and fairness could have produced the data we observed without ever thinking about the value or blocking or spacing. Instead of asking "When do I want to study this category next?," their decision process may have focused on "What do I want to study right now?" In other words, it is possible that our participants did little or no planning ahead. Spacing and blocking may have been byproducts, rather than goals, of our participants' study strategies.

How Thorough?

One might wonder whether "thorough" is the right term for participants who tended to study categories two or three times in a row. Our participants almost never studied a category from start to finish on consecutive trials, whereas Tauber et al.'s (2013) participants chose this strategy frequently. Were our participants less than thorough? We have two answers to this question. First and foremost, the fact that they blocked their studying suggests that being thorough guided their decisions. A strategy of being fair but not thorough would mean switching categories every trial. This is not the strategy our participants used, which suggests that in addition to being fair they also tried to be thorough.

Second, being thorough does not mean persisting for as long as possible. A bee is not more thorough if it stays on a flower after the nectar is gone; similarly, a thorough studier sticks with a category until they are gaining little or no learning from it, and then moves on. In a thorough strategy, therefore, the decision to depart from a category depends on one's judgment of how much one is learning. This kind of judgment has been called a judgment of rate of learning (jROL) or a judgment of "information uptake" (Metcalfe & Kornell, 2005; Kornell & Metcalfe, 2006). jROLs are distinct from judgments of learning (JOLs); in a JOL one judges how well one will do on an upcoming test, whereas in a jROL one judges how much one's knowledge is changing as one studies (JOL is a bit of a misnomer, because what is being judged is actually future knowledge, not current learning). Research has shown that people tend to stop studying an item when continuing to study seems fruitless, even if they do not think they know the answer; that is, they stop when their jROLs become low, even if their JOL is still low (Metcalfe & Kornell, 2005). We presume that our participants switched categories after two or three repetitions because their jROLs were low, but we do not have direct evidence for this claim. Thus, our participants were thorough, but we do not claim to know how thorough they were.

Relationship to Prior Studies

One important question is why our participants chose to block less than 50% of their exemplars, whereas Tauber et al.'s (2013) participants blocked 70–90% of their exemplars. It is impossible to give a definitive answer because the studies were different in many ways.

One important factor might be the stimuli themselves. For example, our photographs were probably more interesting to look at, because they showed penguins in their natural environment, caring for their young, and interacting with other penguins, whereas the birds used by Tauber et al. were similar to each other in size and shape, and they were all pictured against a brown background perched in the same position. Study choices are influenced by interestingness (e.g., Son & Metcalfe, 2000), and it is possible that Tauber et al.'s (2013) participants would have blocked less if the pictures they were shown had been more interesting. This proposal could be tested by manipulating the interestingness of the materials in a category learning study. The materials also differed with respect to the variability within, and between, categories. Like interestingness, variability is difficult to quantify, but our photos might have had more within-category variability (i.e., variability among the pictures of a single species), because they varied in terms of the position and angle of the penguins, the number of penguins in the photo, and what the penguins were doing. Our photos probably had more between-category variability as well, given that we showed penguins that varied in size, shape, coloration pattern, texture, and more. A study that manipulated study materials could examine whether or not stimuli with less variability tend to be blocked more.

We speculate that a smaller methodological difference might also have mattered: Tauber et al. let their participants decide when to stop studying whereas our participants had to study each species exactly five times. Learners usually like to cram before a test (Kornell, 2009) but trying to do so would have meant different things for Tauber et al.'s participants than for ours. Tauber et al.'s participants could study all six exemplars from a category in a row and then cycle back, at the end of the session, to cram in some of restudy each of the categories. If our participants completely blocked their study, they would have used up all their chances to study and been unable to cram. Thus, ironically, our participants might have chosen to space their studying because they wanted to cram. (It is possible that our participants' strategy was actually, be thorough, be fair, and cram at the end, but we do not feel the evidence for cramming is strong enough for such a claim.) These speculations could be tested by manipulating whether or not participants who wanted to cram needed to hold a final study trial or two in reserve.

Although it is important to try to explain why our data differ from those of Tauber et al. (2013), as we have just done, it is no less essential to acknowledge that the gap between the two studies holds an important lesson of its own. The methodological differences between our study and the studies by Tauber et al. are relatively small. The differences among students in how they study are much larger. If our small differences produced such large differences in blocking rates, we can only imagine that blocking rates in real life must vary even more. Thus, it is important to highlight that the question "do people block or interleave their learning when studying categories" might only have one answer: it depends.

It is also important to ask whether Tauber et al.'s (2013) data disprove our hypothesis about people using the strategy of being thorough and fair. Tauber et al.'s participants were clearly thorough, given that they blocked so much. It is not possible to say whether they were fair, because Tauber et al. (2013) did not present analyses of spacing. Thus, the results presented by Tauber et al. (2013) are not inconsistent with participants being thorough and fair.

Limitations

A limitation of the present experiment, and previous research on study decisions regarding spacing, is that participants decided how to mix items within a single study session. In other words, participants made decisions about interleaving versus blocking. Spacing can also mean distributing entire study sessions across time. To date, the literature says nothing about how people would choose, for example, to distribute three hours of study time over the course of a week (e.g., "Should I study for 3 hours on Sunday or for 1 hour each on Monday, Thursday, and Sunday?").

Decisions about study sessions are probably more relevant than decisions about interleaving to the choices students face in real life. In real life, we suspect, participants do decide what they want to study right now, rather than deciding when in the future they are going to study something. The decision about what to study right now is probably largely determined by what due date or exam is most imminent (Kornell & Bjork, 2007). It would be interesting to know whether choices across sessions tend toward a combination of both spacing and blocking, and whether they are consistent with the thorough and fair strategy proposed here.

Conclusion

We began with a simple question: "Do people prefer to block or interleave their studying during inductive learning?" Previous research suggested that they preferred to block. The answer we arrived at is not so simple. Our participants blocked their study more than would be expected by chance, but they interleaved slightly more than they blocked. Moreover, they made sure their interleaved trials were more spaced apart than would be expected by chance. From another perspective, though, the answer to our initial question may be neither. We have argued our participants used a strategy that, in terms of their mental processes, was not driven by a desire to block or space their study, per se. Instead, they were thorough and fair in studying categories, in a manner reminiscent of foraging, and this strategy resulted in blocking and spacing.

References

Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics, 21,* 627–635. http://dx.doi.org/10.1080/00140137808931764

Balota, D., Duchek, J., & Logan, J. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extant literature. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger, III* (pp. 83–105). New York, NY: Psychology Press.

Balota, D. A., Duchek, J. M., Sergent-Marshall, S. D., & Roediger, H. L., III. (2006). Does expanded retrieval produce benefits over equal-interval spacing? Explorations of spacing effects in healthy aging and early stage Alzheimer's disease. *Psychology and Aging, 21,* 19–31. http://dx.doi.org/10.1037/0882-7974.21.1.19

Benjamin, A. S., & Bird, R. (2006). Metacognitive control of the spacing of study repetitions. *Journal of Memory and Language, 55,* 126–137. http://dx.doi.org/10.1016/j.jml.2006.02.003

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127,* 55–68. http://dx.doi.org/10.1037/0096-3445.127.1.55

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. http://dx.doi.org/10.3758/s13421-012-0272-7

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). Cambridge, MA: MIT Press.

Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. In A. F. Healy, S. M. Kosslyn, & R. M. Shiffrin (Eds.), *From learning processes to cognitive processes: Essays in honor of William K. Estes* (Vol. 2, pp. 35–67). Hillsdale, NJ: Erlbaum.

Carew, T. J., Pinsker, H. M., & Kandel, E. R. (1972). Long-term habituation of a defensive withdrawal reflex in aplysia. *Science, 175,* 451–454. http://dx.doi.org/10.1126/science.175.4020.451

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*(3), e0152115. http://dx.doi.org/10.1371/journal.pone.0152115

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. http://dx.doi.org/10.3758/s13421-013-0371-0

Carvalho, P. F., & Goldstone, R. L. (2015). What you learn is more than what you see: What can sequencing effects tell us about inductive category learning? *Frontiers in Psychology, 6,* 505. http://dx.doi.org/10.3389/fpsyg.2015.00505

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132,* 354–380. http://dx.doi.org/10.1037/0033-2909.132.3.354

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

Dempster, F. N. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43,* 627–634. http://dx.doi.org/10.1037/0003-066X.43.8.627

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7,* 95–112. http://dx.doi.org/10.3758/BF03197590

Hatala, R. M., Brooks, L. R., & Norman, G. R. (2003). Practice makes perfect: The critical role of mixed practice in the acquisition of ECG interpretation skills. *Advances in Health Sciences Education: Theory and Practice, 8,* 17–26. http://dx.doi.org/10.1023/A:1022687404380

Kang, S. H. K., & Pashler, H. (2012). Learning Painting Styles: Spacing is Advantageous when it Promotes Discriminative Contrast. *Applied Cognitive Psychology, 26,* 97–103. http://dx.doi.org/10.1002/acp.1801

Karpicke, J. D., & Roediger, H. L., III. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 704–719. http://dx.doi.org/10.1037/0278-7393.33.4.704

Karpicke, J. D., & Roediger, H. L., III (2010). Is expanding retrieval a superior method for learning text materials? *Memory & Cognition, 38,* 116–124. http://dx.doi.org/10.3758/mc.38.1.116

Kornell, N. (2009). Optimizing learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology, 23,* 1297–1317. http://dx.doi.org/10.1002/acp.1537

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14,* 219–224. http://dx.doi.org/10.3758/BF03194055

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19,* 585–592. http://dx.doi.org/10.1111/j.1467-9280.2008.02127.x

Kornell, N., Castel, A. D., Eich, T. S., & Bjork, R. A. (2010). Spacing as the friend of both memory and induction in young and older adults. *Psychology and Aging, 25,* 498–503. http://dx.doi.org/10.1037/a0017807

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 609–622. http://dx.doi.org/10.1037/0278-7393.32.3.609

Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory, 17,* 493–501. http://dx.doi.org/10.1080/09658210902832915

Kornell, N., & Vaughn, K. (2018). *In inductive category learning, people simultaneously block and space their studying using a strategy of being thorough and fair* (ICPSR36999.v1) [Data set]. Ann Arbor, MI: Inter-University Consortium for Political and Social Research. http://dx.doi.org/10.3886/ICPSR36999.v1

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). London, UK: Academic Press.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition, 39,* 462–476. http://dx.doi.org/10.3758/s13421-010-0035-2

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52,* 463–477. http://dx.doi.org/10.1016/j.jml.2004.12.001

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

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review, 106,* 643–675. http://dx.doi.org/10.1037/0033-295X.106.4.643

Pyc, M. A., & Dunlosky, J. (2010). Toward an understanding of students' allocation of study time: Why do they decide to mass or space their practice? *Memory & Cognition, 38,* 431–440. http://dx.doi.org/10.3758/MC.38.4.431

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 58,* 646–668.

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137,* 615–625. http://dx.doi.org/10.1037/a0013684

Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review, 21,* 1323–1330. http://dx.doi.org/10.3758/s13423-014-0588-3

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 907–912. http://dx.doi.org/10.1037/0278-7393.27.4.907

Son, L. K. (2004). Spacing one's study: Evidence for a metacognitive control strategy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 601–604. http://dx.doi.org/10.1037/0278-7393.30.3.601

Son, L. K., & Kornell, N. (2009). Simultaneous decisions at study: Time allocation, ordering, and spacing. *Metacognition and Learning, 4,* 237–248. http://dx.doi.org/10.1007/s11409-009-9049-1

Son, L. K., & Metcalfe, J. (2000). Metacognitive and control strategies in study-time allocation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 204–221. http://dx.doi.org/10.1037/0278-7393.26.1.204

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. http://dx.doi.org/10.3758/MC.38.2.244

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. http://dx.doi.org/10.3758/s13423-012-0319-6

Taylor, K., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24,* 837–848. http://dx.doi.org/10.1002/acp.1598

Toppino, T. C., & Cohen, M. S. (2010). Metacognitive control and spaced practice: Clarifying what people do and why. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1480–1491. http://dx.doi.org/10.1037/a0020949

Toppino, T. C., Cohen, M. S., Davis, M. L., & Moors, A. C. (2009). Metacognitive control over the distribution of practice: When is spacing preferred? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1352–1358. http://dx.doi.org/10.1037/a0016371

Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition, 109,* 163–167. http://dx.doi.org/10.1016/j.cognition.2008.07.013

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. http://dx.doi.org/10.1037/xge0000177

Zechmeister, E. B., & Shaughnessy, J. J. (1980). When you know that you know and when you think that you know but you don't. *Bulletin of the Psychonomic Society, 15,* 41–44. http://dx.doi.org/10.3758/BF03329756

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. http://dx.doi.org/10.3758/s13421-012-0238-9

Received March 21, 2017
Revision received January 6, 2018
Accepted January 11, 2018 ■