Journal of Experimental Psychology:
Learning, Memory, and Cognition
2013, Vol. 39, No. 6, 1682–1696

0278-7393/13/$12.00 DOI: 10.1037/a0032425

Do Students Think That Difficult or Valuable Materials Should Be Restudied Sooner Rather Than Later?

Michael S. Cohen and Veronica X. Yan
University of California, Los Angeles

Vered Halamish
University of Haifa

Robert A. Bjork
University of California, Los Angeles

Despite the clear long-term benefits of spaced practice, students and teachers often choose massed practice. Whether learners actually fail to appreciate the benefits of spacing is, however, open to question. Early studies (e.g., Zechmeister & Shaughnessy, 1980) found that participants' judgments of learning were higher after massed than after spaced repetitions, but more recent studies have found that participants, when allowed to choose between restudying right away and restudying later, tend to choose later, apparently reflecting an appreciation for the benefits of spacing. In these recent studies, however, choosing to restudy later also meant restudying closer to the final test, leaving open the question of what was driving participants' choices. In addition, the choice confronting participants has typically been between getting a spaced and truly massed repetition, whereas in real-world learning contexts the choice is often between a short, but not immediate, spacing interval and a longer one. In our research, we controlled final retention interval and asked participants to choose between restudying word pairs after either a relatively short (but not truly massed) interval or a longer interval. We found that participants had a clear preference for restudying higher priority (more difficult or more valuable) items sooner rather than later, even when doing so was not the most effective option. Thus, previous findings showing a preference for spaced repetition do not extend to a context in which the shorter spacing interval is substantially longer than true massing, and they may merely reflect a preference to restudy closer to the test.

Keywords: memory, metacognition, metacognitive control, spacing effects, lag effects

The finding that spaced repetition of study items leads to better long-term recall than massed repetition is one of the most well-established findings in cognitive psychology. This phenomenon, known as the *spacing effect*, was initially reported by Ebbinghaus (1885/1964) and has been shown across many different types of study materials and learning conditions (see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006, for a review). When a relatively short, but nonzero, spacing interval is compared to a longer spacing interval, the longer interval is often still better, which is commonly called the *lag effect*. Lag effects, though, are dependent on retention interval: The optimal spacing interval is generally longer when the final retention interval is longer, but a longer spacing interval can be suboptimal when it is close to or greater than the final retention interval (Cepeda et al., 2006). Still, although the effects of lag on recall appear to be nonmonotonic, the meta-analysis conducted by Cepeda et al. (2006) observed lag effects under a variety of conditions across studies.

This article was published Online First April 8, 2013.

Michael S. Cohen and Veronica X. Yan, Department of Psychology, University of California, Los Angeles; Vered Halamish, Department of Psychology, University of Haifa, Haifa, Israel; Robert A. Bjork, Department of Psychology, University of California, Los Angeles.

McDonnell Foundation Grant 29192G supported this research. Portions of this research were presented at the 51st Annual Meeting of the Psychonomic Society, Seattle, Washington. We thank Thomas Toppino, Barbara Knowlton, and members of Cogfog for helpful suggestions related to this work. We also thank Shruti Ullas, Narek Vardanian, Zach Merrill, and Brooke Hollyfield for help with data collection.

Correspondence concerning this article should be addressed to Michael S. Cohen, Department of Psychology, University of California, Los Angeles, CA 90095. E-mail: mcohen1@ucla.edu

Metacognitive Monitoring of Spacing and Lag Effects

Despite the strong evidence for spacing and lag effects, the evidence is less clear as to whether learners are able to appreciate the benefits of spaced practice. In fact, they often appear not to do so even when they clearly should, at least when making judgments after the fact. For instance, Zechmeister and Shaughnessy (1980) found that when learners were asked to make item-by-item metacognitive judgments after studying words twice, they did not appreciate that spaced repetitions were more effective than massed repetitions, and in fact tended to predict better recall for massed repetitions. Logan, Castel, Haber, and Viehman (2012) used a similar paradigm, but with multiple study–test cycles, and found that people tended to predict slightly higher recall after a spaced repetition than after a massed repetition. Still, the predicted spacing effect was consistently much smaller than the actual benefit of spacing on recall, even after people had experienced the benefits of spaced practice firsthand.

Kornell (2009) provided students with a more realistic verbal learning context, in which vocabulary flashcards were studied with either a relatively short or relatively long spacing interval between repetitions of the same card. He found that longer spacing intervals during study yield better recall on the final test, but when learners were asked to make predictions after each study session, they were unaware of this benefit. After a single study session, people tended to predict better recall for the items restudied at shorter spacing intervals, and after subsequent study sessions, they showed no effect of spacing interval on predictions. It was only after all items were restudied together, in a final session before the test, that participants tended to show some appreciation for the benefits of longer lags, but this effect did not reach significance. Pyc and Rawson (2012) extended Kornell's findings: When making item-by-item judgments after a series of retrieval practice opportunities, participants tended to predict better recall on a later test for items practiced at a short lag than for items practiced at a long lag, even though long-lag items were actually recalled better. Pyc and Rawson also found that people did not predict a difference between lag conditions on an aggregate level, even when judgments were taken after the final test—during which, presumably, they could have noticed the benefits of spacing.

Similarly, in a nonverbal learning context, Simon and Bjork (2001) found that people predict better performance on a later test after learning keystroke sequences in a blocked manner, rather than after learning those sequences in an interleaved manner. In actuality, although performance during training was better for blocked training, performance on the final test was much better after interleaved training. A similar pattern was observed by Kornell and Bjork (2008), who examined the inductive learning of categories as a function of the spacing of successive exemplars from a given category. Even after a final test on which participants' ability to identify to what category new exemplars belong was greater when the studied exemplars had been spaced, most participants thought blocked presentations of exemplars from a given category were more effective for learning than was spacing those presentations. One exception to the general pattern is that, unlike global judgments, category-level judgments taken after induction learning but before the test predict better recall for spaced categories than for massed categories (Wahlheim, Dunlosky, & Jacoby, 2011). In addition, Kornell, Castel, Eich, and Bjork (2010) found that when participants were asked to learn what artist had painted a specific painting, and each painting was presented six times, either in a row or interleaved with other paintings, spaced repetitions were most likely to be judged as best after the test. It is not entirely clear why people can appreciate the benefits of spaced practice after learning in these contexts but not in others, but the point remains that people are often insensitive to the benefits of spacing.

Metacognitive Control of Spacing

Still, in laboratory paradigms that give learners a choice between massed practice and spaced practice, the pattern of choices has tended to go in the opposite direction, presumably suggesting that people do have an appreciation for the benefits of spaced practice. Benjamin and Bird (2006) gave subjects the opportunity to decide how to restudy, with the constraint that half of the studied word pairs were to be repeated after one intervening item (i.e., almost massed), whereas the other half were assigned to a spaced repetition, which would occur at the end of the list. Under these conditions, people tended to assign more difficult items to the longer spacing interval.

Benjamin and Bird interpreted their results in the context of the discrepancy reduction model (Dunlosky & Hertzog, 1998), which suggests that the amount of study time learners allocate to an item correlates positively with the difference between what they currently know and their goal ("norm of study"). Thus, learners should allocate more study time to more difficult or less well learned items. Under many, though not all, study conditions, this prediction is indeed confirmed (Son & Metcalfe, 2000). Benjamin and Bird extended the discrepancy reduction model to apply to the context of strategy use. The inference is that when people are given control of how to schedule their practice rather than how long to study each item, they will apply what they think is the more effective spacing strategy to the more difficult items. Thus, the fact that people chose to space the more difficult items in this paradigm implies a belief that spacing, rather than massing, is more effective.

It should be noted that Son (2004) tested subjects in a somewhat similar paradigm and found the opposite pattern of results: Learners tended to choose massed practice for the more difficult items and spaced practice for the easier items. However, Toppino, Cohen, Davis, and Moors (2009) found that Son's pattern of results was attributable to the very difficult items and fast presentation rate used in her study (see also Pyc & Dunlosky, 2010). Toppino et al.'s results suggested that people chose to restudy difficult items immediately under those conditions because they were barely able to perceive the words in the given amount of time. Thus, the pattern of results in Son's study appears to rely on a different mechanism from the discrepancy-reduction-based explanation above, a mechanism that presumably only applies when available study time is less than the time it would take to complete initial processing on the words.

Toppino and Cohen (2010) provided further evidence that under typical study conditions, people prefer later restudy to immediate restudy. For instance, when the stimuli were common words, people preferred to space the more difficult items even when a fast presentation rate was used. Toppino and Cohen also examined people's preferences more directly by assigning point values (1 or 5 points) to different study items. People more frequently chose to restudy the more valuable items later than to restudy the less valuable items later. The logical connection from these results to true preferences is more parsimonious than the discrepancy-reduction explanation described above. It seems clear that if people selectively apply the strategy that they believe is more effective to a subset of items, and the items differ in value, they would choose to apply the more effective strategy to the more valuable items. Thus, the fact that participants were more likely to choose later restudy for the more valuable items strongly suggested that people truly believe that later restudy is a better strategy than immediate restudy.

Previous researchers have assumed that a preference for later restudy over immediate restudy likely means that people know that spaced restudy will be better than massed restudy. However, there are two notable problems with this interpretation. For one, if this interpretation is accurate, it is not entirely clear why people often fail to appreciate the benefits of spacing after study (e.g., Zechmeister & Shaughnessy, 1980). An intriguing explanation that has been suggested for this contradiction (Son & Kornell, 2009; Toppino & Cohen, 2010) is that people may know that spacing is better on a general level—that is, as an implicit theory (Koriat, 1997). However, after experiencing the high retrieval fluency of a massed repetition, as compared to the lower retrieval fluency of a spaced repetition, they would still predict better recall for the massed items, relying more heavily on their subjective experience than on implicit theories or beliefs about memory.

Although this explanation seems plausible, recent work suggests that it is probably not correct. When people are told that a particular item or set of items will be either spaced or massed, they do not predict better recall for the to-be-spaced items; if anything, they seem to predict better recall for to-be-massed items (Litke & Toppino, 2011; Toppino, Litke, Cohen, & Halamish, 2012). If the preference for spaced practice in previous studies were being driven by an implicit theory that spaced practice is better, we would expect people to predict better recall for to-be-spaced items. Thus, it appears that some other factor is probably responsible for people's tendency to prefer later restudy when they are given a choice.

There is, however, another alternative interpretation that must be ruled out before interpreting the previous findings as reflecting a true preference for spaced practice. That is, when people chose later restudy over immediate restudy, they were choosing to restudy the word pair at the end of the list, which would be closer to the test. It is clear that the basic effect of spacing interval on memory does not depend on this confounding between spacing and retention interval, as many previous studies have demonstrated spacing and lag effects while controlling for retention interval (e.g., Cepeda et al., 2009; Toppino, Hara, & Hackman, 2002). However, the usual ways of controlling for retention interval—for example, using a between-subjects design, or varying the timing of the initial presentation so that the timing of the second presentation is equated—cannot be implemented when allowing participants to choose how they would like to restudy each item. Thus, in previous studies of spacing choice, it may be that people preferred later restudy because they wanted to restudy harder or more valuable items closer to the test, and with fewer intervening items, rather than because they appreciated that a greater spacing between study opportunities would be better. Such a preference would be consistent with the well-known phenomenon that recall is better after a shorter retention interval (Ebbinghaus, 1885/1964) and/or with fewer intervening items (McGeoch, 1932), but it would not allow us to draw conclusions about whether people actually appreciate the benefits of spaced practice.

Toppino and Cohen (2010) partially addressed this issue in their Experiment 2, as items chosen for later restudy were presented after four to 12 intervening items, rather than at the end of the list. Under these conditions, people still showed a stronger preference for later restudy on harder items. Similarly, Son and Kornell (2009) found that when learners were allowed to freely position items in a study list, the spacing interval for items studied twice was greater than would be expected by chance. They noted that this could merely be due to divergence between what people think is a random distribution and what is actually random, but it could also reflect a real appreciation that spacing restudy opportunities is beneficial. Additionally, Susser and McCabe (2013) found that college students report being more likely to space study sessions across multiple days in real-world learning situations when an upcoming exam is more difficult or more important, or when they have fewer competing academic priorities. Thus, there is some reason to believe that the choice of later restudy in previous laboratory studies does reflect an appreciation for the benefits of spaced practice, but the question remains to be fully resolved.

The Current Approach

The present studies represent a different approach toward this question. In the present experiments, the restudy opportunity is always equally far from the test regardless of the spacing interval, which we accomplished by either varying the position of the test along with the restudy opportunity or varying the position of the initial study opportunity. We should note that the designs used in the present studies did not allow us to compare true massed practice with spaced practice. Instead, we compare a relatively short spacing interval to a relatively long spacing interval. This change does make it somewhat harder to compare the present work to previous studies, but it also has an important strength: The choice between a short lag and a long lag is something that learners are more likely to face in real-life learning situations than a choice between an immediate and a later repetition. Even if people realize that a truly massed repetition is bad, they may still favor shorter lags over longer lags in real-life learning situations. The procedures used herein allow us to directly address people's preferences between short and long lags, as well as whether they appreciate the effects of increased spacing when spacing and retention interval are not confounded.

Experiment 1

Experiment 1 examined people's choices when they were allowed to assign word pairs—on a pair-by-pair basis—to either a restudy-sooner condition or a restudy-later condition. Choices were made immediately after a pair was presented. The choice that a participant made affected both when the item would be reshown and when it would be tested. The interval between the end of the restudy period and the beginning of the test was 5 min for both the sooner and later restudy period. Items chosen for sooner restudy were tested before items for later restudy were re-presented. This design meant that the long spacing interval would necessarily be longer than the retention interval (see Figure 1).

As noted above, Cepeda et al. (2006) found that a longer spacing interval is not necessarily better when the spacing interval is longer than the retention interval. Therefore, we included an initial list in which items are randomly assigned to study conditions. This initial list allows us to accurately assess how the spacing and retention intervals used in this experiment affect recall. Including an initial list also has the advantage of allowing participants to have a very clear idea of the procedures before they make their choices, since they will have already experienced the full procedure before they ever make choices. Because the procedure is somewhat complicated, this design allows us to have greater confidence in the reliability of people's choices on the second list. At the same time, it is a possibility that choices will be influenced in unpredictable ways by the experience that people have with the initial list. We address this potential concern in Experiment 6.

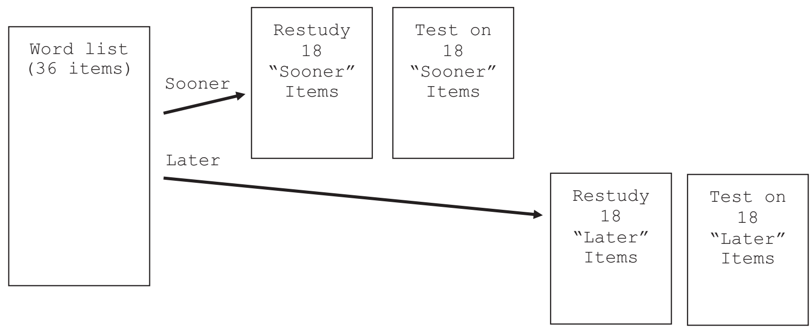

Figure 1. Diagram of basic procedure for each of the two lists in Experiments 1–3.

Method

Participants. Forty students from the University of California, Los Angeles, participated for course credit. Our participant pool for this experiment, and for all other experiments in this article (with the exception of Experiment 6, as described below), included a mixture of introductory psychology students, introductory linguistics students, and advanced psychology students.

Materials. Stimuli were 72 pairs of nouns, composed in part from a pool used previously in Experiment 1A of Toppino and Cohen (2010). Difficulty was varied in terms of normative associative strength (Nelson, McEvoy, & Schreiber, 2004), with imagery ratings (Coltheart, 1981; Friendly, Franklin, Hoffman, & Rubin, 1982; Paivio, Yuille, & Madigan, 1968) varied as a secondary factor. Normative forward associative strength ranged from .075 to .125 for easy pairs (e.g., *doctor–lawyer*) and .01 to .03 for medium pairs (e.g., *freedom–justice*). Difficult pairs (e.g., *position–irony*) had no normative association, nor any readily apparent association after pairing. Imagery ratings were above 5.40 for words in easy pairs, between 3.40 and 5.60 for words in medium pairs, and below 4.5 for words in difficult pairs, on a scale from 1 to 7.

Procedure. Two word lists were constructed, with 36 word pairs in each list, 12 at each of the three difficulty levels. List order was counterbalanced across participants.

On List 1, first word pairs were presented for 2.5 s each. Easy, medium, and hard items were printed in blue, purple, and red font, respectively, so participants would be aware of normative item difficulty. After each word pair was presented, the computer indicated whether that word would be shown again “sooner” or “later.” Participants then pressed the space bar to go on to the next pair. Half of the items were assigned to the sooner list, and half were assigned to the later list, with assignment of items to study conditions counterbalanced. Items assigned to be restudied sooner were re-presented at the end of the list, after all items had been shown once. Then, after a 5-min puzzle distractor task, participants took a cued-recall test on only the sooner items, in which they were shown the cue word from each pair and asked to type the target word via keyboard. The test was self-paced; participants could advance to the next item by pressing the space bar, with up to 10 s allowed for each item. After the sooner test, and a 3-min distractor task of arithmetic problems, participants then restudied the word pairs assigned to the later list. This was again followed by a 5-min distractor and a cued-recall test on the “later” items in the same format as that for the “sooner” items. Figure 1 shows the basic design; this figure was shown to participants during the initial instruction phase.

Instructions for List 2 were presented immediately after the final test of List 1 was complete. List 2 followed the same basic structure as List 1, but with one important difference. After each item was shown, participants, rather than being told how the item would be restudied, were allowed to choose whether to restudy that item sooner or later. Choices were restricted such that each list could have no more than 18 items (i.e., 50% of the list). This restriction was put in place out of a concern that participants might assign a small number of high-priority items to one list, in order to have those items tested as part of a smaller set. In that case, it would be difficult to interpret whether choices reflect a preference for a particular spacing interval or a preference for a smaller set of items on the test.

Participants were made aware of the restriction on their choices, and the number of items that had already been placed in each list was indicated on-screen as the choice was being made. After 18 items had been assigned to one list, participants were required to choose the other list for remaining items. Because these items were not freely assigned to a list, they were not included in any analyses of the List 2 choice data.[1] They were, however, included in analyses of the List 2 recall data, as we assumed that study conditions would be comparable regardless of whether an item was freely assigned to a given condition.

Results

List 1 recall. We first examine the recall data from List 1 (see Table 1). In this and all other recall analyses, we corrected for spelling and other typographical errors before scoring recall responses. A 2 × 3 (Spacing × Difficulty) repeated-measures analysis of variance (ANOVA) revealed a significant main effect of difficulty, $F(2, 78) = 227.154$, $MSE = 349.80$, $p < .001$, with more difficult items recalled more poorly. The effect of spacing was marginal, $F(1, 39) = 2.834$, $MSE = 297.69$, $p = .10$, with items restudied later tending to be recalled better than items

[1] Note that the number of items that could theoretically be excluded on this basis ranged from one to 18 items per subject. The actual number of items excluded in each of the three experiments that used item-by-item choices are as follows: Experiment 1: $M = 3.25$, $SD = 3.16$; Experiment 2: $M = 3.07$, $SD = 2.54$; Experiment 3: $M = 3.64$, $SD = 4.36$.

Table 1
Percentage of Items Recalled in List 1 of Experiments 1–3

Experiment	Easy pairs		Medium pairs		Hard pairs	
	Studied sooner	Studied later	Studied sooner	Studied later	Studied sooner	Studied later
1	76.7	82.5	50.8	56.7	17.1	16.7
2	81.1	85.6	58.7	67.8	29.2	23.5
1 and 2[a]	78.9	84.1	54.8	62.2	23.1	20.1
3	85.3	84.1	66.7	67.1		

[a] Combined data are shown for Experiments 1 and 2 because the procedure was identical for these two experiments up to the end of List 1.

restudied sooner. A closer examination of the means in Table 1 suggested that restudy later was in fact only beneficial for easy and medium items, whereas restudy sooner tended to be better for difficult items. However, the Spacing × Difficulty interaction was not significant ($F < 1$).

List 2 choices. The effect of item difficulty on the percentage of items chosen to be restudied sooner in List 2 was analyzed with a one-way ANOVA. There was a significant main effect of difficulty, $F(2, 78) = 3.881$, $MSE = 1273.26$, $p = .025$, in that harder items were more likely to be chosen for restudy sooner than were easier items (see Figure 2). In addition, we used t tests to examine whether the proportion of items chosen for restudy sooner differed from 50% at any difficulty level. The proportion was significantly greater than 50% for difficult items, $t(39) = 2.574$, $p = .014$, but

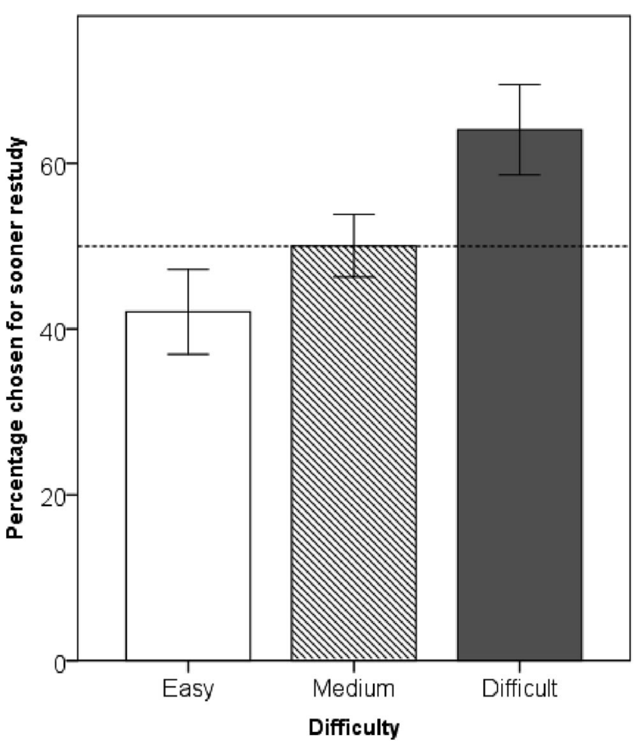

Figure 2. Percentage of items chosen for sooner restudy in Experiment 1. The dashed line represents 50% sooner restudy, which is the overall target for study choices. Error bars represent ±1 standard error.

was not significantly different from 50% for easy items, $t(39) = -1.546$, or for medium items, $t(39) < 1$.

List 2 recall. Because the recall data from List 2 are confounded by item selection effects, it would be difficult to draw any strong conclusions from these data. In addition, 15 of the 40 participants did not have any items in at least one combination of spacing and difficulty, so they could not be included in any statistical analysis. Thus, because statistical analysis did not seem appropriate for these problematic data, we simply report recall data for all participants in Table 2 and describe the data qualitatively. The expected effects of difficulty are apparent, with easier items recalled better than harder items. There is also no apparent benefit from restudying later; if anything, there is a slight trend such that restudying sooner is slightly better than restudying later.

Discussion

These results suggest that participants tend to prefer to restudy sooner, and thus prefer a shorter spacing interval. On the basis of a discrepancy-reduction explanation of people's study choices, people will apply what they believe is the more effective strategy to the items that will benefit the most from it, particularly when that strategy is a limited resource, as is the case here. All else being equal, it would follow that people should apply the more effective strategy to the more difficult items. By this logic, the preference for restudying difficult items reflects an incorrect belief that restudy sooner is better than restudy later.

However, the List 1 recall data complicate this interpretation. Specifically, restudy later tended to be better for easy and medium items, but restudy sooner tended to be better for later items. The interaction was not significant in Experiment 1, and thus should be interpreted with caution (although, as is discussed further below, a significant Spacing × Difficulty interaction was in fact obtained across Experiments 1 and 2, in which an identical procedure was used on the first list). If participants were somehow aware of the apparent interaction between spacing and difficulty, their pattern of choices would in fact be an adaptive response. It seems somewhat unlikely that this is the case because, for one, poststudy self-reports were generally more consistent with a discrepancy-reduction explanation. In addition, although an interaction between spacing and difficulty under certain conditions does follow from a

Table 2
Percentage of Items Recalled in List 2 of Experiments 1–3

Experiment	Value	Easy pairs		Medium pairs		Hard pairs	
		Studied sooner	Studied later	Studied sooner	Studied later	Studied sooner	Studied later
1		81.5	79.2	48.9	47.7	13.3	8.3
2	1	90.4	75.5	63.6	61.1	22.1	16.3
	5	87.2	78.0	62.3	62.8	30.2	19.8
3	1	83.3	80.2	66.6	60.9		
	5	90.5	74.9	66.7	69.7		

Note. Recall data were averaged across participants, but the mean score for each cell was based on a different number of items for different participants. This table includes data from all participants, including those who did not contribute to some cells of the design due to their pattern of choices. We do, however, exclude five individuals who did not take the later recall test in Experiment 3 because of an experimenter error.

study-phase retrieval theory of the spacing effect (e.g., Thios & D'Agostino, 1976), as is discussed further in the General Discussion, this prediction is counterintuitive. It seems unlikely that participants would be aware of the precise ways in which spacing interval, retention interval, and difficulty interact to determine the magnitude or presence of the spacing effect. Experiments 2 and 3 were conducted in part to rule out this interpretation.

Experiment 2

Although it is possible to interpret people's choices as a function of item difficulty by using a discrepancy-reduction interpretation, this interpretation requires assumptions that may or may not be correct. A more direct way to determine which strategy people think is most effective is to assign different point values to different items, such that some items are high value and some are low value. Ariel, Dunlosky, and Bailey (2009) showed that people will preferentially choose to restudy more valuable items, and Toppino and Cohen (2010) showed that people are also sensitive to point values when choosing study strategies. Thus, it is reasonable to assume that if we vary the value of items in the present paradigm, participants will preferentially choose to apply the study strategy that they believe is most effective to the more valuable items.

Method

Participants. Forty-four students from the University of California, Los Angeles, participated for course credit.

Materials. Experiment 2 used the same word pairs as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, with one exception. Word pairs in List 2 were randomly associated with a point value of either 1 point or 5 points, with half of the items at each difficulty level assigned to each value condition and assignment of item to value counterbalanced. Participants were instructed that they should try to get the highest possible score across both tests. Point values were shown on-screen above the word pair on the initial presentation. Note that items still varied in terms of difficulty as in Experiment 1, and the color coding from List 2 of Experiment 1 was maintained in List 2 of Experiments 2 and 3.

Results

List 1 recall. We first examined the proportion of items recalled on List 1 (see Table 1). A 2 × 3 (Spacing × Difficulty) repeated-measures ANOVA showed a main effect of difficulty, $F(2, 86) = 221.415$, $MSE = 332.32$, $p < .001$, with easier items recalled better than harder items. There was not a significant main effect of spacing, $F(1, 43) = 1.207$, $MSE = 384.34$, $p = .278$, but there was a Spacing × Difficulty interaction, $F(2, 86) = 4.858$, $MSE = 259.25$, $p = .01$. Probing the interaction showed that items with a longer spacing interval tended to be recalled better for both easy and medium items; this trend was not reliable for easy items, $t(43) = -1.354$, $p = .183$, but it was significant for medium-difficulty items, $t(43) = -2.028$, $p = .049$. For hard items, the trend was in the opposite direction, with items given a shorter spacing interval tending to be recalled better, and this trend was marginally significant, $t(43) = 1.827$, $p = .075$.

Because the procedure up to the end of List 1 was identical for Experiments 1 and 2, we also report statistics for the combined List 1 recall data from the two experiments in order to increase power in our statistical analyses. A 2 × 3 × 2 (Spacing × Difficulty × Experiment) mixed ANOVA, with repeated measures on the first two factors, revealed a significant main effect of difficulty, $F(2, 164) = 448.754$, $MSE = 340.636$, $p < .001$, with easier items being recalled better. The main effect of spacing was marginally significant, $F(1, 83) = 3.753$, $MSE = 343.13$, $p = .056$, as a longer spacing interval tended to lead to better recall. This effect was qualified, however, by a significant Spacing × Difficulty interaction, $F(2, 166) = 4.912$, $MSE = 260.94$, $p = .008$. Probing the interaction showed a significant benefit for longer spacing on both easy items, $t(83) = -2.285$, $p = .025$, and medium items, $t(83) = -2.339$, $p = .022$. For difficult items, there was a trend toward a benefit for sooner restudy, but this trend was not significant, $t(83) = 1.423$, $p = .159$. Finally, there was a significant main effect of experiment, $F(1, 82) = 4.265$, $MSE = 1688.81$, $p = .042$, with Experiment 2 showing better recall performance overall. There was not, however, a significant interaction between experiment and difficulty ($F = 1.346$, $MSE = 340.636$, $p = .26$), experiment and spacing ($F < 1$), or a three-way interaction ($F < 1$).

List 2 choices. The choice data were analyzed as a function of value and normative difficulty (see Figure 3). A 2 × 3 (Value × Difficulty) ANOVA on the percentage of items chosen for sooner restudy revealed a main effect of value, $F(1, 43) = 12.839$, $MSE = 1462.73$, $p = .001$, with 5-point items more likely to be chosen for sooner restudy. There was no main effect of difficulty, $F(2, 86) < 1$, and the interaction also was not significant, $F(2, 86) = 1.690$, $MSE = 371.23$, $p = .191$. In addition, we performed one-sample

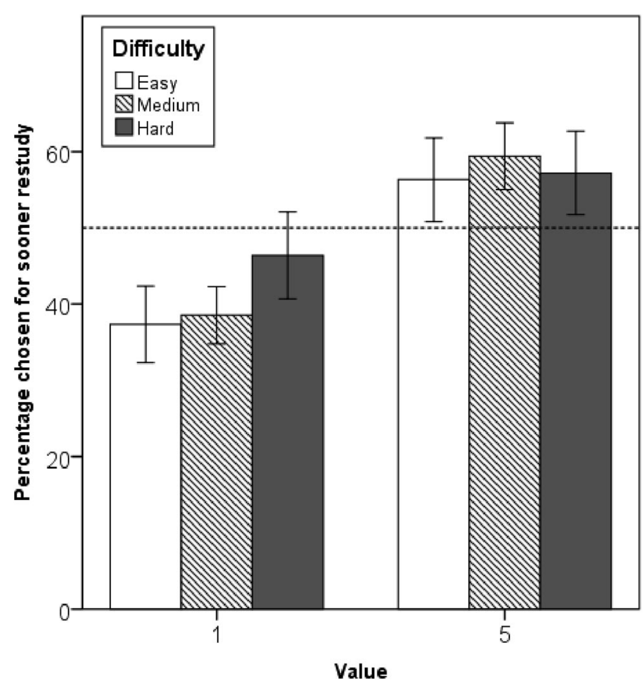

Figure 3. Percentage of items chosen for sooner restudy in Experiment 2. The dashed line represents 50% sooner restudy, which is the overall target for study choices. Error bars represent ±1 standard error.

t tests to examine whether the likelihood of choosing restudy sooner was significantly different from 50% at each value level, collapsing across difficulty levels. (Note that this analysis included somewhat different numbers of items from each difficulty level for each participant, based on the distribution of choices for that individual.) One-point items were chosen to be restudied sooner significantly less than 50% of the time, $t(43) = -3.616$, $p = .001$, whereas 5-point items were chosen to be restudied sooner significantly more than 50% of the time, $t(43) = 2.842$, $p = .007$.

List 2 recall. As in Experiment 1, it is difficult to analyze and interpret the recall data for List 2 because of item selection effects. In addition, some individuals did not select items of all types for both spacing durations, and even those that did select items of all types did not do so evenly. Thus, no statistical analysis seemed appropriate, so we again merely report raw recall data in Table 2, and describe the data qualitatively. The expected effects of difficulty are apparent, but again there is, if anything, a reverse spacing effect, with items restudied sooner tending to be recalled better. The apparent reversal of the spacing effect may be due in part to the fact that people tended to choose to restudy more valuable items sooner. However, value effects themselves appear to be somewhat inconsistent. In any case, because it is impossible to determine all of the ways that item selection effects might have affected recall in List 2, these recall results should be interpreted with caution.

Discussion

These data provide stronger evidence that people prefer sooner rather than later restudy. There is a clear preference for sooner restudy on the more valuable items and, correspondingly, a clear preference for later restudy on the less valuable items. Still, because a longer spacing was not in fact reliably better across all difficulty levels in this paradigm, even on the first list, we wanted to replicate this effect when a longer spacing would in fact be better. The following experiments aim to provide this replication.

Experiment 3

A longer spacing interval was better than a shorter spacing interval for items of easy and medium difficulty in Experiments 1 and 2, whereas this was not the case for the most difficult items. Thus, we repeated Experiment 2 using only items of easy and medium difficulty.

Method

Participants. Twenty-eight students from the University of California, Los Angeles, participated for course credit.

Materials. The items were 72 word pairs, 36 of which were easy pairs and 36 of which were medium-difficulty pairs. These were composed in part of the 24 word pairs at each difficulty level from Experiments 1 and 2. In addition, 12 new pairs were generated at each difficulty level with the same criteria as were used for the original pairs.

Procedure. The procedure was the same as in Experiment 2, except that only two levels of difficulty were used.

Results

List 1 recall. We first examine the effects of difficulty and of spacing interval on recall performance for List 1 (see Table 1). We submitted the percentage of items recalled to a 2 × 2 (Difficulty × Spacing) repeated-measures ANOVA. There was a significant effect of difficulty, $F(1, 27) = 40.530$, $MSE = 220.30$, $p < .001$, as recall was better for easy items than for medium-difficulty items. There was not a significant effect of spacing interval, however ($F < 1$), nor was there an interaction ($F < 1$).

List 2 choices. We then examine the choice data from List 2 (see Figure 4). We submitted the percentage of items chosen for restudy sooner to a 2 × 2 (Value × Difficulty) repeated-measures ANOVA. There was a significant effect of value, $F(1, 27) = 7.049$, $MSE = 945.73$, $p = .013$, as participants were more likely to choose to restudy sooner for 5-point items than for 1-point items. There was not a significant effect of difficulty, $F(1, 27) = 1.171$, $MSE = 931.23$, $p = .289$, nor was there an interaction ($F < 1$).

As in Experiment 2, we also examine whether items at each individual value level, collapsed across difficulty, were significantly above or below a 50% likelihood of restudying sooner. The percentage of 5-point items chosen to restudy sooner was significantly above 50%, $t(27) = 3.568$, $p = .001$. However, the percentage of 1-point items chosen to restudy sooner did not differ significantly from 50%, $t(27) = -0.421$, $p = .677$.

List 2 recall. Finally, we examine recall for List 2. As in previous experiments, these data are contaminated by item selection effects and other issues, so we merely report means (see Table 2) and describe the data qualitatively. We again see the expected effects of item difficulty, with easier items recalled better. There is also a tendency for items restudied sooner to be recalled better, and

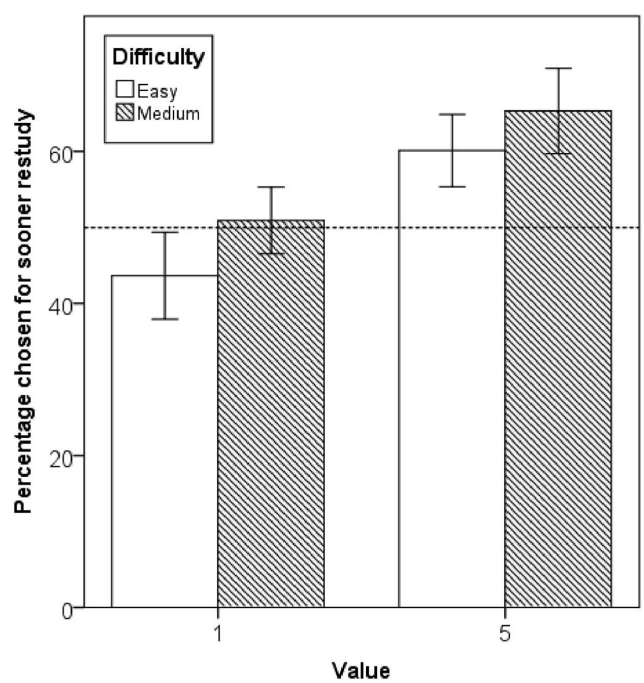

Figure 4. Percentage of items chosen for sooner restudy in Experiment 3. The dashed line represents 50% sooner restudy, which is the overall target for study choices. Error bars represent ±1 standard error.

somewhat inconsistent effects of point value. As in Experiment 2, the possible reversal of the spacing effect may result from the tendency to allocate more of the valuable items to sooner restudy, as well as other item selection effects.

Discussion

This experiment replicated our primary finding from Experiment 2, that people were more likely to choose to restudy sooner for more valuable items. In addition, it is notable that in this experiment, only the preference for 5-point items differed from 50%. People's choices for 1-point items, when they were allowed to freely choose a spacing interval, were no different from chance. (Note that this outcome is possible because we are not counting choices made after participants had already assigned 50% of the items to a single strategy, at which point they were no longer allowed to freely choose a study strategy.) This finding, in combination with the responses to our postexperiment questionnaires, gives us reason to believe that the effects are being driven by a preference to restudy valuable items sooner, not by a preference to restudy less valuable items later.

Contrary to our expectations, this experiment did not provide us with a situation in which restudying later would lead to better baseline recall than restudying sooner across item types. Instead, there was no effect of spacing in List 1. Thus, while the present experiment shows that people prefer restudying sooner when they should be indifferent to the spacing interval, it does not tell us whether they would prefer restudying sooner even when restudying later is consistently better. The experiments that follow address this question in a different way.

Experiment 4

One unusual aspect to the procedures in the preceding experiments is that the long spacing interval is always longer than the retention interval. This aspect is a consequence of the fact that the first test comes before the second restudy period. In another experiment that will not be reported in full here, we tried altering the previously described procedure, such that participants viewed the later restudy list immediately after the sooner restudy list. This was followed by the sooner test, and then by the later test. This change allowed the retention interval to be longer than the spacing interval. However, there still was not an advantage for items restudied later, presumably because those items were affected by output interference from the items that had been restudied sooner. Thus, a different approach was needed.

A different way to vary the spacing interval, while still ensuring that spacing and retention interval are not confounded, is to vary the scheduling of the initial presentation. Specifically, in Experiments 4 and 5, we presented participants with items in separate sublists. All items were then re-presented in a random order after the final sublist. Thus, items shown earlier have a longer spacing interval, but the average retention interval is constant across conditions. One downside to this procedure is that participants are not able to choose to restudy sooner or later on an item-by-item basis. However, it is possible for participants to express a preference for a particular sublist before it is presented. One way to do this is to ask participants to choose which of two sublists should be made more valuable (cf. Litke & Toppino, 2011). This is the way we assessed preferences in the present experiment and in those that follow.

Method

Participants. Twenty-four students from the University of California, Los Angeles, participated for course credit.

Materials. Experiment 4 used the same set of 72 word pairs used in Experiment 3.

Procedure. As in earlier experiments, the set of word pairs was randomly divided into two lists of 36 items each, with assignment of items to list counterbalanced. In the present experiment, each list was further subdivided into two sublists of 18 items each, with nine items at each level of difficulty in each sublist, and assignment of item to sublist counterbalanced. These were described as "Part 1" and "Part 2" in the instructions. On List 1, participants were first shown the Part 1 word pairs for 2.5 s each. Then, after a 3-min distractor task of arithmetic problems, they were shown the Part 2 word pairs, again for 2.5 s each. Note that because choices were not being made on an item-by-item basis, items were not color coded based on difficulty, as we had done in previous experiments. This was immediately followed by the restudy list, in which all 36 word pairs were shown for 2.5 s each. Then, after a 5-min math puzzle distractor task, participants were given a cued recall test on each of the 36 items, with up to 10 s available for each item. Figure 5 shows a diagram of the basic procedure for each list in this experiment.

The procedure was similar for List 2, with one important exception. Participants were told at the beginning of List 2 that they would now get to choose how valuable items on each part of the list would be. They were given two options. Either all items on Part 1 would be worth 5 points each and all items on Part 2 would be worth 1 point each or Part 1 items would be worth 1 point and Part 2 items would be worth 5 points. Participants entered their choice into the computer at the end of the instruction period. A second screen asked them to confirm that their choice had been entered correctly. The program then went on to display the words

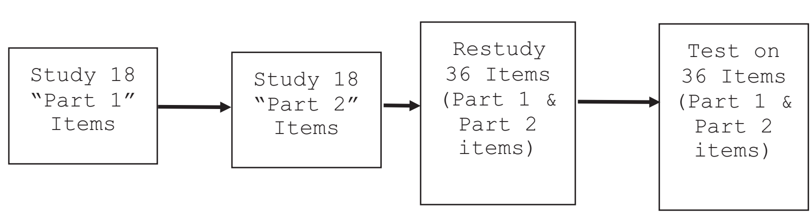

Figure 5. Diagram of basic procedure for each of the two lists in Experiment 4. Experiment 5 is similar in structure, except that each of the two lists is divided into three parts rather than two.

in the same manner as on List 1, except that values were also shown.

Results

List 1 recall. We first examine recall data from List 1 (see Table 3). The primary factors of interest are spacing and difficulty. However, in order to see how recall performance during List 1 relates to subsequent study choices in list 2, we also include choice in the analysis. Thus, we submitted the percentage of items recalled on List 1 to a 2 × 2 × 2 (Difficulty × Spacing × Choice) mixed ANOVA with repeated measures on the first two factors. We found a significant effect of spacing, $F(1, 23) = 6.318$, $MSE = 170.45$, $p = .020$, such that items with a longer spacing (i.e., items in Part 1) were remembered better than those with a shorter spacing (i.e., items in Part 2). We also found a significant effect of difficulty, $F(1, 23) = 16.195$, $MSE = 225.10$, $p = .001$, such that easier items were remembered better than medium-difficulty items. In addition, the main effect of choice is marginally significant, $F(1, 22) = 2.999$, $MSE = 1379.12$, $p = .097$. Individuals who chose to make Part 2 more valuable on List 2 tended to have worse recall on List 1 than individuals who chose to make Part 1 more valuable on List 2. None of the interactions in this analysis approached significance (Fs < 1).

List 2 choices. We then examined people's choices using a chi-square test. This analysis showed that the distribution of choices was significantly different from what would be expected if people were choosing randomly between making Part 1 and Part 2 more valuable ($\chi^2 = 8.167$, $p = .004$). Specifically, participants were significantly more likely to choose to make Part 1 more valuable (observed $n = 19$) than would be expected by chance (expected $n = 12$).

List 2 recall. We also examined the recall data for List 2 (see Table 3). We should note that although there are no item selection effects in this experiment, these data are still affected by people's choices, in that Part 1 is more valuable for some participants, whereas Part 2 is more valuable for other participants. We therefore used a 2 × 2 × 2 (Difficulty × Spacing × Choice) mixed ANOVA, with repeated measures on the first two factors, to analyze the proportion of items recalled from List 2. There was a significant effect of difficulty, $F(1, 22) = 54.259$, $MSE = 132.00$, $p < .001$, such that recall was better for easier items. There was not a significant main effect of spacing, $F(1, 22) < 1$, but there was a marginally significant interaction between spacing and choice, $F(1, 22) = 3.629$, $MSE = 345.17$, $p = .07$, as is described in more detail below. There was not a significant main effect of choice, $F(1, 22) < 1$, and none of the other interactions approached significance (all Fs < 1).

Table 3
Percentage of Items Recalled in Experiment 4

List	Preferred sublist	Easy pairs		Medium pairs	
		Part 1	Part 2	Part 1	Part 2
1	Part 1	84.2	76.0	69.0	55.6
	Part 2	62.2	60.0	53.3	44.4
2	Part 1	86.0	70.8	63.7	54.4
	Part 2	73.3	77.8	48.9	55.6

Because there was a trend toward an interaction between spacing and choice despite the relatively low power of this analysis, we examined the effect of spacing separately for the two groups using paired-samples t tests. For individuals who chose to make Part 1 more valuable, there was a significant spacing effect, $t(18) = -3.385$, $p = .003$, such that items in Part 1 (longer spacing) were remembered better than items in Part 2. For those who chose to make Part 2 more valuable, there was not a significant effect of spacing, $t(4) = 0.443$, $p = .681$. Indeed, for these individuals, the trend was in the opposite direction, suggesting that our failure to find a typical spacing effect was not simply due to lack of power. Instead, it appears that people tended to have better recall for items in the more valuable sublist, regardless of the spacing interval. As in earlier experiments, one should therefore be cautious in interpreting the recall data from List 2.

Discussion

In the present experiment, a longer spacing interval led to better baseline recall than a shorter spacing interval. Unlike in previous experiments, participants also expressed a reliable preference for the longer spacing interval. Still, in this case, one potential confounding factor is that people may have been choosing to make Part 1 items more valuable because of an appreciation of primacy effects, rather than because they appreciated that the longer spacing interval would be more effective. Castel (2008) has shown that participants do show an awareness of primacy effects when predicting later recall, at least under certain conditions, suggesting that this is a plausible expectation in the present study. Experiment 5 modifies the design to eliminate the confounding between longer spacing and primacy.

Experiment 5

In Experiment 5, we modified one major feature of the design from Experiment 4: We divided each list into three sublists instead of two. Participants were asked to make items either in Part 2 or Part 3 of List 2 more valuable. Items in the other part would then have the lowest values, whereas items in Part 1 were always given intermediate values. In this case, a preference to make Part 2 items more valuable would reflect a preference for a longer spacing interval. On the other hand, a preference to make Part 3 items more valuable would indicate a preference for a shorter spacing interval. In addition to this change, we used items reflecting three levels of difficulty, as in Experiments 1 and 2.

Method

Participants. Thirty-six students from the University of California, Los Angeles, participated for course credit.

Materials. This experiment used the same set of 72 word pairs as in Experiments 1 and 2, with 36 items assigned to each list.

Procedure. The procedure was similar to that of Experiment 4, with one major exception. That is, on List 2, Part 1 items were always worth 3 points each. Before beginning to study the items in List 2, participants were allowed either to make all Part 2 items worth 5 points and all Part 3 items worth 1 point or to make all Part 2 items worth 1 point and all Part 3 items worth 5 points.

Results

List 1 recall. We first examined the effects of difficulty and spacing on recall in List 1, as well as differences in List 1 recall that predict subsequent study choices in List 2 (see Table 4). We conducted a 3 × 3 × 2 (Difficulty × Spacing × Choice) mixed ANOVA, with repeated measures on the first two factors. There was a significant main effect of difficulty in the expected direction, $F(2, 68) = 262.569$, $MSE = 391.85$, $p < .001$. There was not a significant interaction between difficulty and choice, $F(2, 68) = 2.237$, $MSE = 391.85$, $p = .115$. There was, however, a significant main effect of spacing interval, $F(2, 68) = 9.636$, $MSE = 557.40$, $p < .001$, with longer spacing intervals yielding better recall, as well as a significant interaction between spacing interval and subsequent choice, $F(2, 68) = 3.435$, $MSE = 557.40$, $p = .038$, which is discussed further below. The main effect of choice in this experiment was marginally significant, $F(1, 34) = 2.980$, $MSE = 2531.81$, $p = .093$. Individuals who chose to make a longer spacing (Part 2) more valuable tended to have worse recall than those who made a shorter spacing (Part 3) more valuable. In addition, there was a marginally significant interaction between spacing and difficulty, $F(4, 136) = 2.006$, $MSE = 551.45$, $p = .097$. This trend is not theoretically meaningful, so it will not be discussed further. There also was no three-way interaction ($F < 1$).

Because our hypotheses depend on the relationship between recall for Part 2 and Part 3 items, we collapsed across difficulty levels and sublist preference, and conducted a paired-samples *t* test as a planned comparison on these two groups (see Figure 6). We found that items initially studied in Part 2 ($M = 46.5$, $SE = 3.5$) were recalled significantly better than items from Part 3 ($M = 39.8$, $SE = 3.7$), $t(35) = -2.195$, $p = .035$. We also found that items initially studied in Part 1 ($M = 53.5$, $SE = 3.2$) were recalled significantly better than items from Part 2, $t(35) = -2.223$, $p = .033$, indicating a further benefit of a longer spacing interval.

We also probed the Spacing × Choice interaction in an effort to better understand the patterns in the data, again collapsing across difficulty, using separate one-way repeated-measures ANOVAs for each choice group (see Figure 6). For individuals who subsequently chose to make Part 2 more valuable, there was a significant effect of spacing on List 1, $F(2, 32) = 10.707$, $MSE = 201.95$, $p < .001$, with better recall for longer spacing intervals. For individuals who later chose to make Part 3 more valuable, there was not a reliable effect of spacing ($F < 1$). Thus, there appear to be differences on recall during List 1 that relate to people's later choices.

List 2 choices. Next we examined the actual choice data. Here we found that the number of people who chose to make Part 2 more valuable ($n = 17$) and the number of people who chose to make Part 3 more valuable ($n = 19$) did not differ from a random distribution ($\chi^2 = 0.111$, $p = .739$). Thus, overall, people did not show a preference for either Part 2 or Part 3.

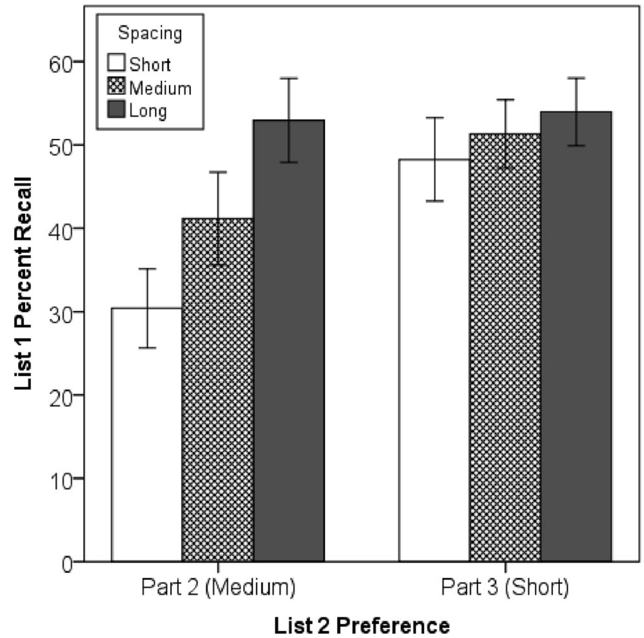

Figure 6. Percentage of items recalled on List 1 of Experiment 5 as a function of spacing, collapsed across difficulty levels. Participants are split based on the choice that they subsequently made as to which part of List 2 should be made more valuable. Error bars represent ±1 standard error.

List 2 recall. Finally, we examined the percentage of items recalled in List 2 (see Table 4) using a 3 × 3 × 2 (Difficulty × Spacing × Choice) mixed ANOVA with repeated measures on the first two factors. It should be noted that, as in Experiment 4, this analysis is confounded by the fact that people's choices affected how valuable different items would be. There was a significant main effect of difficulty in the expected direction, $F(2, 68) = 244.166$, $MSE = 445.35$, $p < .001$, with no significant interaction between difficulty and choice, $F(2, 68) = 1.743$, $MSE = 445.35$, $p = .183$. We also found a significant main effect of spacing, $F(2, 68) = 3.253$, $MSE = 505.75$, $p = .045$, as well as a significant interaction between spacing and choice, $F(2, 68) = 3.276$, $MSE = 505.75$, $p = .044$. As in Experiment 4, this interaction is likely due to the confounding of spacing and point value as a function of choice condition and is not otherwise theoretically meaningful, so it will not be examined further. There was a main effect of choice,

Table 4
Percentage of Items Recalled in Experiment 5

		Easy pairs			Medium pairs			Hard pairs		
List	Preferred sublist	Part 1	Part 2	Part 3	Part 1	Part 2	Part 3	Part 1	Part 2	Part 3
1	Part 2	77.9	75.0	52.9	58.8	36.8	35.3	22.1	11.8	2.9
	Part 3	85.5	85.5	75.0	64.5	52.6	51.3	11.8	15.8	18.4
2	Part 2	79.4	77.9	61.8	44.1	47.1	39.7	8.8	16.2	11.8
	Part 3	93.4	85.5	81.6	73.7	55.3	60.5	25.0	15.8	22.4

$F(1, 34) = 6.968$, $MSE = 2285.48$, $p = .012$, as people who chose to make the shorter spacing interval (Part 3) more valuable produced better overall recall on List 2. There was no interaction between spacing and difficulty, $F(2, 68) = 1.359$, $MSE = 452.89$, $p = .252$, nor was there a three-way interaction ($F < 1$).

Discussion

In this experiment, items with a moderately long spacing interval (Part 2) were remembered better than items with a shorter spacing interval (Part 3). However, people did not appear to show a preference between these two schedules, unlike in the previous studies. We were also somewhat surprised to see that individuals who had better recall overall, that is, better learners, tended to choose to make the items with a shorter spacing interval more valuable. This finding differs from the outcome in Experiment 4, where individuals with better recall on List 1 tended to make the sublist with a longer spacing interval more valuable.

It appears, therefore, that people's choices of the longer spacing interval in the present experiment were not being driven by expertise. However, it is also notable that the individuals who chose to make Part 2 items more valuable actually experienced a greater spacing effect than individuals who chose to make Part 3 items more valuable. This finding suggests that the experience from List 1 might have directly influenced people's choices. Those individuals who recalled more items as the spacing duration became longer might have learned about the benefits of spacing (cf. Kornell, 2009). By contrast, those who recalled all items approximately equally regardless of spacing might stick with a natural tendency to prefer a shorter spacing. Experiment 6 was designed to test this interpretation.

Experiments 6A and 6B

Experiments 6A and 6B were designed to examine whether people's preferences for assignment of values to sublists of List 2 in Experiments 4 and 5 were affected by their experience with List 1. In Experiments 6A and 6B, participants did not have that firsthand experience before making their choices. Instead, the paradigms were described in detail, and people were asked to choose which sublist they would make more valuable, following the constraints from Experiments 4 and 5 for Experiment 6A and 6B, respectively. Based on the overall recall data from Experiments 4 and 5, the longer spacing option should be superior in both experiments. However, Experiments 1–3 suggest that people will often prefer a shorter spacing interval when given a choice. If that pattern is replicated here, it would suggest that the experience with the initial list is what led people to prefer a longer spacing interval in Experiment 4 and to be indifferent to spacing interval in Experiment 5. Instead, we might expect to see a preference for a shorter spacing interval in the experiments that follow.

Method

Participants. Participants were students from the University of California, Los Angeles, who participated for course credit. This experiment was run together with another study for which we had to exclude students in advanced psychology courses; thus, the participant pool for Experiment 6 was mostly composed of introductory psychology and linguistics students. For Experiment 6A, we report data from 28 students. For Experiment 6B, we report data from 20 students. An additional nine individuals were run in Experiment 6A, and an additional four individuals were run in Experiment 6B. These individuals were excluded because post-study questionnaires suggested that they may not have fully understood the paradigm when making their choices. Still, the reported findings do not change based on whether these individuals are or are not included.

Materials and procedure. Participants were given a sheet of paper that described the relevant experiment. In Experiment 6A, the page described the procedure from Experiment 4, whereas in Experiment 6B, the page described the procedure from Experiment 5. The description included a diagram of the procedure, similar to Figure 5. A version of the diagrams in Figures 1 and 5 had been shown as part of the instructions in the preceding experiments. Note that unlike in the earlier experiments, the procedure was only described as being run once; that is, there was no initial list in which people would study words but not make study choices. In addition, to further eliminate the effects of experience-based cues, no actual word pairs were shown.

Participants indicated with a check mark on the bottom of the page how they would allocate point values. For Experiment 6A, this meant that they could choose to make Part 1 items worth 5 points each and Part 2 items worth 1 point each, or to make Part 1 items worth 1 point and Part 2 items worth 5 points. For Experiment 6B, they were told that all Part 1 items would be worth 3 points, but they could either make all Part 2 items worth 5 points and all Part 3 items worth 1 point or make all Part 2 items worth 1 point and all Part 3 items worth 5 points.

Results

We analyzed the data for each experiment using a chi-square test. In Experiment 6A, 17 participants chose to make Part 2 worth 5 points each—that is, more valuable than Part 1 items—which is more than would be expected by chance (expected $n = 10$; $\chi^2 = 9.800$, $p = .001$). In Experiment 6B, 20 participants chose to make Part 3 items worth 5 points each—that is, more valuable than Part 2 items—which is also more than would be expected by chance (expected $n = 14$; $\chi^2 = 5.143$, $p = .023$).

Discussion

In both experiments, most people chose to make the sublist that provided a shorter spacing interval more valuable than the alternative sublist that provided a longer spacing interval. This finding suggests that even under conditions in which a longer spacing interval is actually better, people will prefer a shorter spacing interval. It appears that people's preferences in Experiments 4 and 5 were affected by their experience on List 1. Specifically, learners are apparently able to ascertain that a longer spacing interval can lead to better recall under certain conditions. Still, it seems reasonable to suggest that the preference for shorter spacing intervals observed in Experiments 6A and 6B is a more accurate measure of what people would typically prefer without direct task experience.

Experiment 7

To explore what learners appreciate about the benefits of spacing in more educationally realistic situations, we presented students with a learning scenario and asked them to schedule their own learning. Participants were asked to schedule their study either for an exam with a 1-day retention interval or for an exam with a 1-week retention interval. Note that in this experiment, the final retention interval is at least as long as the spacing intervals. Thus, we assumed—on the basis of previous findings (e.g., Cepeda et al., 2006)—that evenly spacing study opportunities will lead to superior long-term recall, and also likely would not reduce short-term performance on the exam, as compared to a "cramming" strategy of concentrating study time immediately before the test. Between the two conditions, however, we might expect participants in the 1-week retention interval condition to "value" learning the information more, as they have to retain it for longer. Thus, if participants appreciate the spacing effect, we should see a greater amount of spacing in the 1-week condition than in the 1-day condition.

Methods

Participants and design. Sixty-eight students from the University of California, Los Angeles, participated for course credit. Participants were asked to schedule study with either a 1-day retention interval (n = 33) or a 1-week retention interval (n = 35).

Materials and procedure. This study was conducted in a two-page paper packet. On the first page, the scenario was described to participants, and participants were presented with the image of a monthly calendar, with one Monday marked as "Today" and another Monday marked as "Exam."

Participants were given a scenario in which they had 7 days (Monday/Today through Sunday) to prepare for an upcoming exam. They were told that due to other commitments, they could only allow themselves 12 hr toward studying for that exam. Their goal was to schedule these 12 study hr to get the highest grade possible in the exam. Note that the number of days available for study does not divide evenly into the number of study hours that participants were given. Still, if learners generally believe that maximal spacing (i.e., evenly spacing study time across days) is best, any variability caused by this constraint should cancel out across subjects. Thus, there would still be no reason to expect a systematic preference for allocating more or less study time to the days closer to the test, given a preference for maximal spacing.

In the 1-day retention interval condition, participants were allowed to study up to the night before the exam. In the 1-week retention interval condition, participants were told that the exam was in 2 weeks, but that in the week before the exam they would be away on a hiking trip and would not be able to study (these days were marked off as "hiking trip" on the calendar). In this condition, then, the last study opportunity would be 1 week before the exam.

Participants were asked to write the number of hours they would study in each day, making sure not to exceed a total of 12 hr, or to schedule outside the permitted 7 days (starting from the Monday marked "Today" to the next Sunday).

After they completed creating their schedule, participants were asked to turn to the second page and write a few sentences as to their reasoning behind their study schedule, explaining why they thought their schedule would help them attain the highest grade possible. Finally, they completed a demographics questionnaire.

Results

Figure 7 shows the average number of hours scheduled per day in each condition. As the figure shows, both lines slope upward, indicating that both groups had a tendency to place a greater number of study hours closer to the test. However, the slopes clearly also differ: Those in the 1-day retention interval condition were much more likely to leave the majority of study to the last 2 days, whereas those in the 1-week retention interval condition spread their study out relatively more evenly across the 7 days. We conducted linear regression analyses, predicting scheduled hours from the day, for each individual. There was a reliable tendency to allocate more study time closer to the test, both with a 1-day retention interval (average $\beta = 0.41$, $SE = .06$) and with a 1-week retention interval (average $\beta = .17$, $SE = .04$). One-sample t tests showed that both regression slopes are reliably different from 0 ($ps < .001$). An independent-samples t test showed that the two slopes are also reliably different from each other, $t(66) = 3.24$, $p < .001$. There was no correlation between the slope and any of the following individual measures: age, year in college, or grade point average.

We also coded participants' free responses about why they scheduled their study the way that they did into three categories: (a) those who talked about their schedule being better for long-term memory, (b) those who talked about their schedule as being better for achieving a good grade, and (c) all other reasons (e.g., motivated by impending deadlines, desiring days off). Those in the 1-week retention interval condition (45.7%) were significantly more likely than those in the 1-day retention interval condition (9.1%) to mention their schedule benefiting long-term memory,

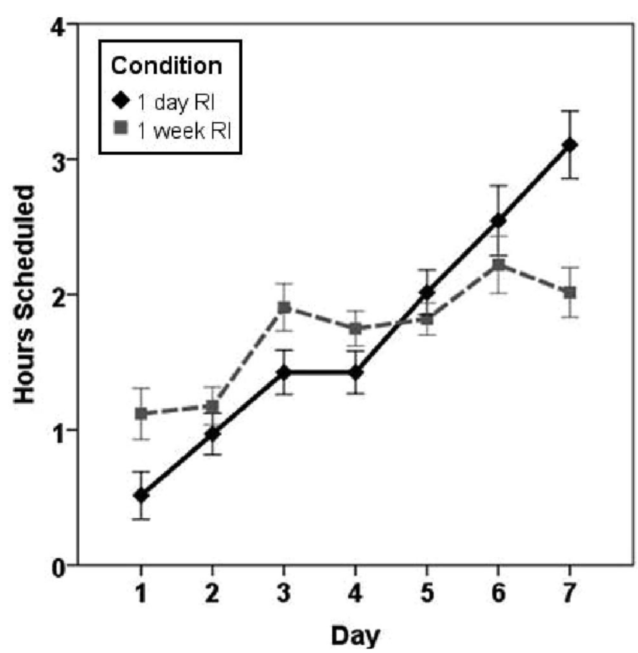

Figure 7. Mean number of hours allocated for study across each of the 7 available days in Experiment 7. Error bars represent ±1 standard error. RI = retention interval.

$\chi^2(1) = 11.32$, $p < .01$. There was no difference between the two conditions in the likelihood of mentioning either schedules benefiting grades or other reasons, nor was there a significant association between variance and likelihood of mentioning exam performance (*p*s > .05).

Discussion

Taken together, these results suggest that learners are relatively more likely to space when they know they have to retain information for longer and more likely to cram when the retention interval is shorter. People do seem to have some appreciation for the fact that spacing practice has greater benefits at a longer retention interval. This result is consistent with Susser and McCabe's (2013) finding that people tend to space out their study more when the to-be-tested material is more difficult or more important. At the same time, even with a 1-week retention interval, people still seem to prefer for more of their available study time to come later in the study period rather than earlier. Thus, even when studying for a test that will occur a full week after the end of the study period, learners do not seem to prefer maximally spacing their study, and instead still show some tendency to cram. The findings are therefore also consistent with the other experiments in the present article.

Another interesting point to note, related to the results of Experiment 7, is that past work by Koriat, Bjork, Sheffer, and Bar (2004) found that people generally were not sensitive to retention interval when predicting recall on a later test. Learners predicted the same level of recall for a test in 1 week as on an immediate test, at least when retention interval was varied between subjects. The reason why participants in the present experiment changed their study strategies based on retention interval, whereas participants in the Koriat et al. study did not change their metacognitive predictions, is not entirely clear, but it could be a promising direction for future study.

General Discussion

Our findings show that when people are given the choice between a shorter spacing interval and a longer spacing interval, they will prefer the shorter spacing interval under a number of conditions. Table 5 summarizes the design and results of our seven experiments in abbreviated form. In Experiments 1–3, we observe a preference for a shorter spacing interval when people make item-by-item decisions under a constraint that only half of the items can be assigned to each spacing interval. In these three experiments, people did gain prior experience with the paradigm before the list in which they were allowed to choose study strategies, and a longer spacing interval did not necessarily lead to better recall in that initial study list. However, recall performance on the initial list also should not have led to a preference to restudy sooner. If anything, their experience should have led to indifference between the two study strategies, not the preference for restudying sooner that we actually observe.

In Experiment 6, we extend the finding that people prefer shorter spacing intervals to a paradigm in which a longer spacing interval does lead to better recall. It is notable that in Experiments 4 and 5, after people had experience on an initial list showing them that a longer spacing interval was better, they did seem to express a preference for a longer spacing interval. Importantly, however, in Experiment 6 these same paradigms were described, but people were not given firsthand experience with them before making choices. Under those conditions, they again tended to prefer the shorter spacing interval. This allows us to conclude more strongly that people do seem to prefer shorter spacing intervals, at least as a baseline preference, even under conditions in which they clearly should prefer a longer spacing interval.

Finally, in Experiment 7, we extended our findings to a more realistic learning situation. When asked to allocate limited study time across a series of days before a hypothetical exam, participants tended to allocate more time to the days closer to the exam, rather than spacing study evenly across the study period. At the same time, there does appear to be an intuition that distributing study is good for long-term learning, whereas cramming is effective in the short term. Thus, it may be that in real life, a short-term goal of passing an upcoming exam outweighs longer term learning goals, leading people to choose a shorter spacing interval than what they know to be optimal for long-term learning.

Are People Aware of the Benefits of Spaced Practice?

What does all this mean for the broader question of whether, or when, people are aware of the benefits of spaced practice? One possibility is that the previous studies that found a preference for later restudy over immediate restudy (e.g., Benjamin & Bird, 2006;

Table 5
Summary of Design and Results for All Experiments

Experiment	Type of strategy allocation	Importance manipulation	Which spacing actually better on List 1?	Preferred spacing on List 2?	Potential confounding factors
1	Item-by-item	Difficulty	Longer (marginal)	Shorter	Spacing × Difficulty interaction
2	Item by item	Value and difficulty	Neither	Shorter	None
3	Item by item	Value and difficulty	Neither	Shorter	None
4	Aggregate	Value	Longer	Longer	Longer spacing = primacy List 1 experience
5	Aggregate	Value	Longer[a]	Mixed[b]	List 1 experience
6	Aggregate	Value	Not tested (longer?)[c]	Shorter	None
7	Aggregate	Limited study time	Not tested (longer?)[d]	Shorter[e]	None

[a] Longer only reliably better for individuals who subsequently preferred longer spacing. [b] Longer after longer better in List 1, shorter after little effect of spacing in List 1. [c] Longer assumed better based on Experiments 4 and 5. [d] Longer assumed better based on prior literature. [e] Preference for shorter spacing reduced with a long retention interval.

Toppino & Cohen, 2010) found such a preference only because the spacing interval was confounded with retention interval. Thus, participants in the previous studies might have chosen later restudy for more difficult and more valuable items because they would see those items closer to the test, not because the second presentation would be further from the initial presentation. When spacing interval and retention interval are no longer confounded, it seems that people make choices consistent with a preference for a shorter spacing interval. If people do not actually appreciate the benefits of spaced practice when they are allowed to choose how to schedule study opportunities, it would be consistent with a number of previous findings from a slightly different context. As discussed above, people generally do think massed practice is better than spaced practice when they are asked after experiencing both types of practice (e.g., Kornell & Bjork, 2008; Simon & Bjork, 2001; Zechmeister & Shaughnessy, 1980).

Other explanations are possible, however. One possibility is that people do understand that immediate restudy (i.e., true massing) is not effective and that spaced restudy is better than immediate restudy. Still, this preference may be more limited than was apparent from previous studies. That is, people may only express a preference for spaced repetitions when the spacing interval for a massed repetition is actually or nearly 0.

Another possibility is that people may only be focusing on one repetition when making their choices, consistent with the stability bias reported previously by Kornell and Bjork (2009). Kornell and Bjork showed that when people are asked to predict recall after an initial study opportunity, predictions are the same regardless of the number of additional study opportunities they expect to have. In the experiments reported here, when people show a preference for a shorter spacing interval, they may merely be trying to minimize the temporal distance between the initial presentation of an item and when they are tested on it. Thus, it might be said that the present experiments only change how the spacing interval is confounded with retention interval, rather than eliminating the confounding factor. Further work will be necessary to definitively rule out these alternative explanations.

Interpreting the Interaction of Item Difficulty and Spacing

There is one other important point to note from these data. That is, the recall data from Experiments 1 and 2 show an interaction between spacing and difficulty, such that a longer spacing interval tends to be better for easier items, whereas a shorter spacing interval tends to be better for harder items. This interaction is consistent with the study-phase retrieval explanation of the spacing effect (e.g., Appleton-Knapp, Bjork, & Wickens, 2005; Thios & D'Agostino, 1976), when combined with the new theory of disuse (Bjork & Bjork, 1992). Retrieval strength is defined by Bjork and Bjork (1992) as the current ease of retrieving an item from memory and is lost over time. Storage strength is the long-term strength of the memory trace and is never lost. The key aspect of the theory is that if retrieval strength is lower at the time of restudy, then a restudy opportunity will lead to a greater increase in storage strength. This is true unless retrieval strength is so low that the initial presentation cannot be retrieved at all.

Thus, a longer spacing interval is typically better because the initial presentation has a lower retrieval strength at the time of the second presentation than would be the case with an immediate repetition. However, if the spacing interval is long enough that the initial presentation is forgotten entirely by the time the item is presented again, then a shorter spacing interval may be more beneficial. Retrieval strength would tend to decline more quickly for more difficult items than for easier items, and thus we would expect more of the difficult items to be forgotten by the time of the second presentation. This will cause a reduced, or potentially reversed, spacing effect for difficult items under conditions that produce a typical spacing effect with easier items. This type of pattern has not been reported previously in the literature, but it is exactly the pattern of results that we observe on List 1 of Experiments 1 and 2. Thus, the present article provides new support for the study-phase retrieval explanation of the spacing effect and the new-theory-of-disuse framework.

Conclusion

Whether learners understand what study strategies are most effective has been a topic of growing interest in recent years, not just among researchers (e.g., Hartwig & Dunlosky, 2012; Karpicke, 2009; Karpicke, Butler, & Roediger, 2009; Kornell & Bjork, 2007; McCabe, 2011; Susser & McCabe, 2013), but also among the broader public (e.g., Carey, 2010; Sundem, 2012). Central to that broader question is whether people understand that restudying later, rather than sooner, can enhance long-term recall; that is, are people aware of the benefits of spacing? Overall, the present findings suggest that learners do not fully appreciate the benefits of spacing, but our results also demonstrate that learners' decisions are based on multiple considerations, such as the importance of remembering a given item, which actually appears to decrease the likelihood of choosing to space, how close a restudy opportunity is to an upcoming test, and so forth. Bjork, Dunlosky, and Kornell (2013) have argued recently that learning how to learn is the ultimate survival tool, but our results, together with other recent findings, suggest that becoming good with that tool is a complex matter.

References

Appleton-Knapp, S., 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.

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General, 138,* 432–447. doi:10.1037/a0015928

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

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.

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

Carey, B. (2010, September 6). Forget what you know about good study habits. *The New York Times*. Retrieved from http://www.nytimes.com/2010/09/07/health/views/07mind.html

Castel, A. D. (2008). Metacognition and learning about primacy and recency effects in free recall: The utilization of intrinsic and extrinsic cues when making judgments of learning. *Memory & Cognition, 36,* 429–437. doi:10.3758/MC.36.2.429

Cepeda, N. J., Coburn, N., Rohrer, D., Wixted, J. T., Mozer, M. C., & Pashler, H. (2009). Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology, 56,* 236–246. doi:10.1027/1618-3169.56.4.236

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. doi:10.1037/0033-2909.132.3.354

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology: Section A: Human Experimental Psychology, 33,* 497–505. doi:10.1080/14640748108400805

Dunlosky, J., & Hertzog, C. (1998). Training programs to improve learning in later adulthood: Helping older adults educate themselves. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 249–275). Mahwah, NJ: Erlbaum.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H. A. Ruger, C. E. Bussenius, & E. R. Hilgard, Trans.). New York, NY: Dover. (Original work published 1885)

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation, 14,* 375–399. doi:10.3758/BF03203275

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review, 19,* 126–134. doi:10.3758/s13423-011-0181-y

Karpicke, J. D. (2009). Metacognitive control and strategy selection: Deciding to practice retrieval during learning. *Journal of Experimental Psychology: General, 138,* 469–486. doi:10.1037/a0017341

Karpicke, J. D., Butler, A. C., & Roediger, H. L. (2009). Metacognitive strategies in student learning: Do students practice retrieval when they study on their own? *Memory, 17,* 471–479. doi:10.1080/09658210802647009

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126,* 349–370. doi:10.1037/0096-3445.126.4.349

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133,* 643–656. doi:10.1037/0096-3445.133.4.643

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

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

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

Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: Overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General, 138,* 449–468. doi:10.1037/a0017350

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

Litke, M. D., & Toppino, T. C. (2011, November). *What's the value of massing and spacing? Learners' expectations about the effect of distributing practice*. Poster presented at the 52nd Annual Meeting of the Psychonomic Society, Seattle, WA.

Logan, J. M., Castel, A. D., Haber, S., & Viehman, E. J. (2012). Metacognition and the spacing effect: The role of repetition, feedback, and instruction on judgments of learning for massed and spaced rehearsal. *Metacognition and Learning, 7,* 175–195. doi:10.1007/s11409-012-9090-3

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

McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review, 39,* 352–370. doi:10.1037/h0069819

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida word association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36,* 402–407. doi:10.3758/BF03195588

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, 76,* 1–25. doi:10.1037/h0025327

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

Pyc, M. A., & Rawson, K. A. (2012). Are judgments of learning made after correct responses during retrieval practice sensitive to lag and criterion level effects? *Memory & Cognition, 40,* 976–988. doi:10.3758/s13421-012-0200-x

Simon, D. A., & Bjork, R. A. (2001). Metacognition in motor learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 907–912. doi: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. doi: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. doi: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. doi:10.1037/0278-7393.26.1.204

Sundem, G. (2012, January 29). Everything you thought you knew about learning is wrong [Web log message]. Retrieved from http://www.wired.com/geekdad/2012/01/everything-about-learning/

Susser, J. A., & McCabe, J. (2013). From the lab to the dorm room: Metacognitive awareness and use of spaced study. *Instructional Science, 41,* 345–363. doi:10.1007/s11251-012-9231-8

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. doi:10.1016/0022-5371(76)90047-5

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. doi:10.1037/a0020949

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

Toppino, T. C., Hara, Y., & Hackman, J. (2002). The spacing effect in the free recall of homogeneous lists: Present and accounted for. *Memory & Cognition, 30,* 601–606. doi:10.3758/BF03194961

Toppino, T. C., Litke, M. D., Cohen, M. S., & Halamish, V. (2013). *Learners' metacognitive predictions fail to predict benefits of spacing prior to restudy*. Manuscript in preparation.

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. doi:10.3758/s13421-010-0063-y

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.

Received November 14, 2012
Revision received February 19, 2013
Accepted February 20, 2013 ■