Journal of Experimental Psychology:
Learning, Memory, and Cognition
2014, Vol. 40, No. 1, 306–311

© 2013 American Psychological Association
0278-7393/14/$12.00 DOI: 10.1037/a0033866

RESEARCH REPORT

The Spacing Effect and Metacognitive Control

Neil W. Mulligan
University of North Carolina at Chapel Hill

Daniel J. Peterson
Knox College

Research suggests that spaced learning, compared with massed learning, results in superior long-term retention (the spacing effect). Son (2010) identified a potentially important moderator of the spacing effect: metacognitive control. Specifically, when participants chose massed restudy but were instead forced to space the restudy, the spacing effect disappeared in adults (or was reduced in children). This suggests spacing is less effective (or possibly ineffective) if implemented against the wishes of the learner. A closer examination of this paradigm, however, reveals that item-selection issues might alternatively explain the disappearance of the spacing effect. In the current experiments, we replicated the original design demonstrating that an item-selection confound is operating. Furthermore, relative to a more appropriate baseline, the spacing effect was significant and of the same size whether participants' restudy choices were honored or violated. In this paradigm, metacognitive control does not appear to moderate the spacing effect.

Keywords: the spacing effect, metacognitive control, metamemory, recall, memory

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

The effects of repetition have been a primary focus of memory research since the first formal experiments of Ebbinghaus (1885/1964). One of the enduring findings in this literature is that the spacing of repetitions produces a powerful impact on later memory: stimuli that are repeated in immediate succession (massed repetition) are less well remembered than stimuli repeated after some delay (spaced repetition; Bjork, 1979; Greene, 2008). Ongoing research continues to assess the optimum intertrial intervals and moderating influences of retention interval (e.g., as it relates to intertrial interval), but the basic phenomenon of the spacing effect is highly robust, having been replicated many times in numerous domains with various types of materials and memory tests (Cepeda, Vul, Rohrer, Wixted, & Pashler, 2008; Delaney, Verkoeijen, & Spirgel, 2010). The spacing effect is important for a theoretical understanding of human memory and has great applied relevance, for instance to educational practice (e.g., Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Son & Simon, 2012).

In this light, the recent results of Son (2010) are potentially quite important. Son found that when adult participants were allowed to choose a massing or spacing study strategy, the normal spacing effect was obtained, but when massed or spaced repetition was presented contrary to the choice of the participant, the spacing effect was eliminated (with children as participants, the spacing effect was merely reduced but not eliminated when restudy choices were not honored). This is potentially a critical limitation of the spacing effect, prompting Son to conclude that the benefits of spacing may not be "universal, particularly when the strategy is not chosen by the learner" (p. 255) and that metacognitive control may be a crucial factor in obtaining these benefits.

In Son's (2010, Experiment 1) study, adult participants were presented with a study list of synonym pairs (e.g., *hirsute—hairy*). The participants were asked to study the word pairs in preparation for a later test in which they would try to recall the second word given the first word. Furthermore, participants were told that they could choose to restudy each pair after its initial presentation if they wished, and they could also choose the manner of restudy either seeing the pair again immediately (a massed choice) or later (a spaced choice). However, the participants were warned that on a few trials their choice would not be honored. Specifically, each study trial began with an initial presentation of the word pair followed by a judgment of learning (JOL), in which the participant rated the likelihood of later recall. Next, the participant decided whether he or she had finished with that pair, would like to study it again at that time (massed choice), or would like to study it again later (spaced choice). The restudy choices were honored on two thirds of the trials: "Now" choices caused the word pair to be presented again immediately, and "Later" choices caused the pair to be presented again much later in the study phase. On one third of the trials, the restudy choice was not honored. In these cases, a Now choice caused the word pair to be presented later (forced spacing), and a Later choice caused the word pair to be presented again immediately (forced massing). Done choices were always honored; these pairs were not presented again.

The critical results were those of the twice-presented items. When choices were honored, the spaced items produced greater cued recall than did the massed items, the traditional spacing effect. However, when the choices were not honored, the spaced items (those for which massing was requested but spacing was

This article was published Online First July 29, 2013.

Neil W. Mulligan, Department of Psychology, University of North Carolina at Chapel Hill; Daniel J. Peterson, Department of Psychology, Knox College.

Correspondence concerning this article should be addressed to Neil W. Mulligan, Department of Psychology, 341B Davie Hall, University of North Carolina, Chapel Hill, NC 27599-3270. E-mail: nmulligan@unc.edu

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

implemented against the participant's wishes) produced no greater recall than did the massed items (those items for which spacing was requested but were actually massed). These results imply that spacing enhances memory when it is consistent with the learner's self-chosen strategies but has less (and perhaps no) benefit when implemented contrary to the learner's desires. This result is potentially a very important limiting condition of the spacing effect both theoretically and practically. Theoretically, it may indicate that spacing does not generally enhance memory and that metacognitive control may be a critical moderating factor. Practically, it may indicate that spacing is not always a useful study strategy and that counseling spacing against the wishes of the learner may not be useful.

However, before settling on this interpretation, an important item selection concern needs to be addressed. In this paradigm, the study items assigned to the various massed and spaced conditions are not under experimental control but are rather assigned by the participants' choices. This is intentional, of course, and necessary as the study is designed to examine metacognitive control, but it nevertheless gives rise to the potential of an item selection problem. In particular, items that are chosen for spaced practice may be easier to recall than items that are chosen for massed practice. Indeed, in this study, participants perceived these items to be easier—participants gave higher JOLs to items chosen for spacing than to items chosen for massed practice (e.g., Son, 2004, 2010). Given that participants' JOLs are at least somewhat related to recall (as evidenced by above-zero gamma correlations), then the items that are perceived to be easier are in fact better remembered. This implies that the items chosen for spacing are an easier set of items, on average, than those chosen for massing. If so, then the size of the spacing effect when the participants' choices were honored may be artifactually increased in size because the items in the honored-spaced condition are from an easier set of items and those in the honor-massed condition are from a harder set prior to any repetition. Likewise, the size of the spacing effect in the dishonored condition may be artifactually reduced for the complementary reason. Items in the forced-spacing condition were those chosen for massing, and the forced-massed items had been chosen for spacing. Consequently, the items in the forced-spaced condition may have been a more difficult set of items to begin with, potentially counteracting any positive effect that may have been due to spacing. To the extent that the items differed in average difficulty, the original findings of Son (2010) are difficult to interpret and may not indicate an eliminated, or even reduced, spacing effect in the dishonored-choice condition.

How should this be addressed? In order to interpret the effects of spacing in this paradigm, one needs baseline conditions against which to evaluate repetition effects. Such baseline conditions serve two purposes. First, they may be used to evaluate if items chosen for spacing or massing actually differed in difficulty (i.e., recallability) prior to repetition. Second, the baseline conditions would provide a comparison for evaluating any effect of spacing in the honor and dishonor conditions relative to expected recall performance for items placed into the different categories prior to any repetition. In the present study, we followed the basic paradigm of Son (2010, Experiment 1) in which participants are presented with synonym pairs to study for a later cued-recall test. After the initial presentation, participants provide a JOL and then are given the choice to study the word pairs again, either Now or Later, or to be Done with the pair. For those pairs chosen for repetition, the Now or Later selection was honored for the majority of trials and dishonored for a minority of trials. In addition to these conditions, we included a baseline condition consisting of a set of study pairs that were not presented a second time regardless of the participant's decision. The recall levels of these items allowed us to assess whether items chosen for massed or spaced repetition differed in difficulty prior to repetition. This baseline performance also allowed us to compare the effects or massed or spaced repetition relative to the recall level expected prior to repetition, a more appropriate comparison than treating the two sets of items as if they were drawn from equivalent populations.

Experiment 1

Method

Participants. Forty-seven undergraduates at the University of North Carolina at Chapel Hill participated in exchange for course credit.

Materials. The critical materials were 80 synonym pairs taken from practice lists for the Graduate Record Examination (e.g., *abnormal—aberrant*, *lethargic—torpid*). In a given study list, 20 pairs were assigned to the baseline condition. These pairs were not displayed a second time regardless of the choice of the participant. Four versions of the study list were created such that each pair was in the baseline condition in one of the lists. The other 60 critical pairs in each list were in the honor and dishonor conditions of the experiments (as described later). The baseline and other critical pairs were randomly intermixed. An additional 12 filler pairs were placed at the beginning of each study list and were not tested on the recall test.

Procedure. The general procedure was modeled on that of Son (2010, Experiment 1). At the outset of the experiment, the participants were told that they would be presented with a series of synonym pairs to study. For each pair, they were asked to rate their confidence that they would remember the second word when given only the first word on a later memory test, using a scale from 0 to 100 where 100 indicated complete confidence and 0 indicated no confidence at all. Furthermore, the participants were told that they would have another chance to study the pair again if they choose. After the confidence rating, participants were told that they could choose to study the pair again Now, study it again Later, or be Done with the word pair. The participants were informed that when they requested to study the pair again at that time, the pair would be presented again immediately—most of the time. They were warned, however, that on a small number of trials, the pair would be presented again later instead. When they chose to study the pair again later, the list would move on to the next pair, and the current word pair would be saved and presented again later in the list—again, most of the time. On a small number of trials, the word pair would be presented again immediately rather than later. The participants were also informed that if they choose Done, the word pair would not appear again. Finally, the participants were told that "although there will be a few trials during which you will get the opposite of what you selected, please be as sincere as possible in your restudy choices." These instructions were modeled closely (and the last selection was taken verbatim) from the instructions in Son (2010, p. 257).

Each trial of the study phase proceeded as follows. First, the study pair was presented on the computer screen for 1 s, preceded and followed by a blank screen for 100 ms. Next, the participant was prompted for the JOL rating, which was entered by typing. The participant was then prompted for the restudy decision, typing *N* for study Now, *L* for study Later and *D* for Done. For the first 12 (noncritical) trials, the choice was always honored. For Now choices, the word pair was immediately re-presented for 3 s. For Later choices, the study pair was saved and presented again (for 3 s) after all other pairs had received their initial presentation. For Done choices, the study pair was not shown again. These noncritical pairs were not tested later. For the critical pairs not in the baseline condition, the Now and Later study choices were honored on two thirds of the trials and proceeded as described previously. For one third of these trials, the participant got the opposite of the request: following a Now choice, the pair was presented later rather than immediately. and following a Later choice, the pair was presented again immediately but not later. Done choices were always honored. Finally, for critical pairs in the baseline condition, the pairs were never presented again regardless of the participant's choice. It should be noted that the participants did not know until after the restudy choice whether the choice would be honored or not.

After the study phase, participants were given a 5-min distractor task of math problems followed by the memory test. In the test, participants were presented with the first word from each of the critical pairs and asked to recall the synonym presented with it during the study phase. The cue word was presented in the middle of the computer screen. Participants recalled the synonym aloud and continued on to the next trial by hitting the space bar (the test was self-paced).

Results

Participants were unconstrained in their restudy choices (as in Son, 2010), raising the possibility that zero items might be placed in one of the critical conditions (the honored or dishonored Now or Later conditions, or the baseline Now or Later conditions). Seven participants were replaced because they had zero study items in one of these conditions, leaving an effective sample of 40.[1]

JOL results. A preliminary analysis indicated that JOLs did not differ across study items that were subsequently honored, dishonored or in the baseline condition ($F < 1$), as one would expect given that these items were indistinguishable at the time the JOLs were rendered. The JOLs did significantly differ as a function of the restudy choice, $F(2, 70) = 97.59$, mean square error (MSE) = 142, $p < .001$. Planned contrasts indicated that all three conditions significantly differed, with Now choices corresponding to the lowest JOLs ($M = 32$, $SD = 18.5$), Later choices in the middle ($M = 46$, $SD = 13.9$), and Done choices corresponding to the highest JOLs ($M = 71$, $SD = 20.7$). The relationship between JOLs and subsequent recall performance was assessed with the gamma correlation, although as noted by Son (2010), this measure is muddied in the present design by the additional study presentation for most of the items. Regardless, the gamma correlations were substantial ($M = 0.49$) and comparable to those reported in Experiment 1 of Son (2001; $M = 0.43$).

Recall results. The most important results are those of the recall test. For purposes of displaying and analyzing the data, the Now and Later choices that were honored are referred to as the *honored-massed* and *honored-spaced* conditions, respectively. The Now choices that were not honored are referred to as the *forced-spaced condition* (because the participant requested massed repetition but received spaced repetition). Likewise the Later choices not honored are referred to as the *forced-massed condition*. Recall performance in these conditions is presented in Figure 1a. We first analyze these unadjusted recall scores before considering recall in the baseline conditions.

Proportion recall was analyzed with a 2 × 2 analysis of variance (ANOVA) using spacing (massed vs. spaced) and honor condition (honored vs. forced) as within-subject factors. The analysis revealed a main effect of spacing, $F(1, 39) = 5.52$, $MSE = 0.039$, $p = .024$, and a significant interaction between spacing and honor condition, $F(1, 39) = 4.35$, $MSE = 0.035$, $p = .044$. Analyses of the simple effects indicated a significant effect of spacing in the honored condition, $t(39) = 3.81$, $p < .001$, but not in the forced condition, $|t| < 1$. That is, there was an apparent spacing effect in the honored condition but not in the forced condition, replicating the results of Son (2010). Less critically, recall for Done items ($M = 0.28$, $SD = 0.19$) was numerically greater than for any other condition, significantly so for all (ts > 2.8, ps $< .008$) except the honored-space condition ($p = .37$).

The unadjusted recall scores do not take into account possible pre-repetition differences between the items chosen for massing or spacing. Recall in the baseline condition is presented in Figure 1b and shows that study pairs chosen for spaced study are, in fact, easier to recall than items chosen for massed study even before any repetition: recall is significantly greater for the Later baseline items than for the Now baseline items, $t(39) = 2.19$, $p = .035$. Given that difference, the unadjusted recall scores are not interpretable. Adjusted recall scores were computed for each participant by computing the increase in recall due to the (massed or spaced) repetition relative to the appropriate baseline condition. The honored-massed and forced-spaced conditions (both composed of Now choices) were assessed relative to the Now baseline, and the honored-spaced and forced-massed conditions (both composed of Later choices) were assessed relative to the Later baseline. The adjusted recall scores (Figure 1c) were analyzed with a 2 (spacing) × 2 (honor condition) ANOVA, revealing only an effect of spacing, $F(1, 39) = 5.28$, $MSE = 0.040$, $p = .027$ (other Fs $<$ 1). That is, spacing enhanced memory more than massing relative to the appropriate baseline condition, and by the same amount, whether the metacognitive choice was honored or not.[2]

[1] Of the remaining participants, four had zero study items in the Done condition, although this is unimportant as the critical analyses focus on the items chosen for restudy.

[2] The adjusted recall scores embody the assumption of an additive effect of repetition, which is an oversimplification in the case of stimulus repetition; more widely spaced repetitions produce superadditive effects (Benjamin & Tullis, 2010). Given that the primary goal is to determine if the spacing effect is equivalent across the honor conditions, rather than directly estimate the "true" spacing effect within each honor condition, this simplification is not of great concern. It may be that the superadditivity in the spaced conditions entail some underestimate of the true spacing effect, but this underestimate should be equal across honor conditions, leaving the comparison of the spacing effect across the honored and forced conditions uncompromised. This issue is revisited, and an alternative analysis of the adjusted recall, yielding consistent results, is reported after Experiment 2.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

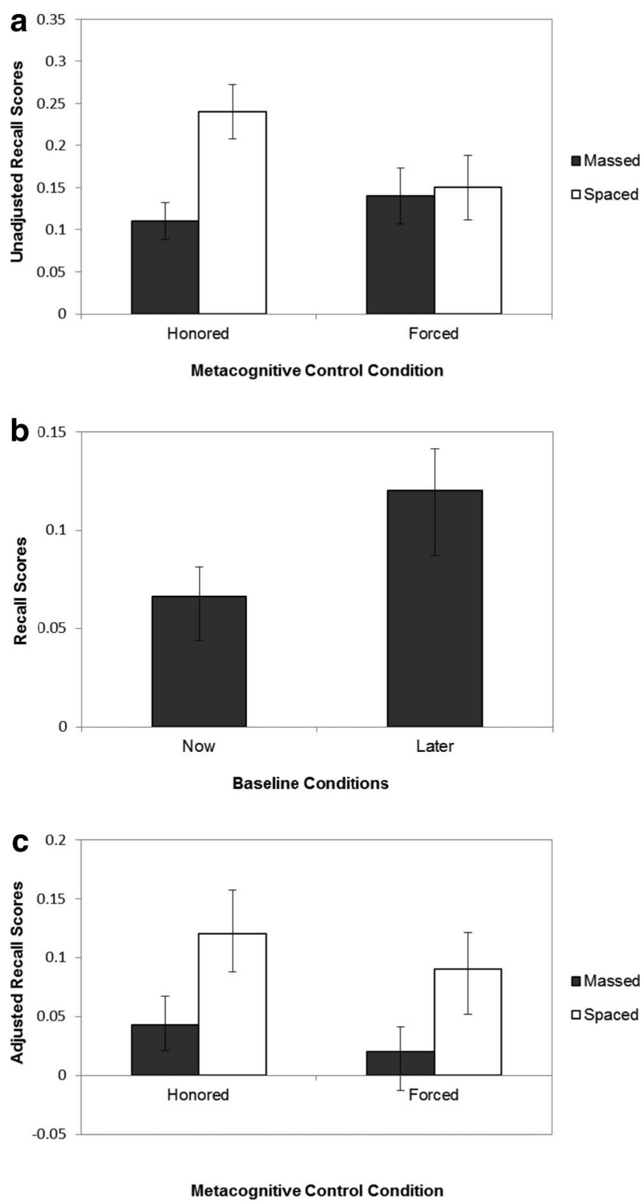

Figure 1. Results of Experiment 1. Mean unadjusted proportion recalled (± *SE*) as a function of metacognitive control and spacing (Panel a). Mean proportion recalled (± *SE*) in the baseline conditions (Panel b). Mean adjusted proportion recalled (± *SE*) as a function of metacognitive control and spacing (Panel c). Honored versus forced = choice honored or forced choice; Now versus later = restudy time immediate or delayed.

Discussion

Before considering the critical results, it is important to first note that Experiment 1 replicated the results of Son (2010). First, JOLs were lowest for Now items, intermediate for Later items, and highest for Done items (see also Son, 2004). Second, the relationship between JOLs and recall (assessed with the gamma correlation) was quite comparable for the present experiment and for Son (2010, Experiment 1). Third, the unadjusted recall scores revealed an apparent spacing effect in the honored condition but not in the dishonored (forced) condition. The replication of these results is reassuring and implies that our modified version of this experimental paradigm is tapping the same memory and metamemory processes as the original. This in turn provides additional confidence that the present concerns apply to the original experiments.

The addition of the baseline items allowed us to assess whether items chosen for immediate restudy (Now items) are equivalent to items chosen for later restudy (Later items). They are not. As suggested by the JOL ratings, the latter items are perceived to be easier and, in fact, are easier to recall on the final test. Consequently, analyses of the unadjusted recall scores is misleading, as the spaced and massed items are not equivalent prior to the repetition. When recall is adjusted relative to the appropriate baseline condition (i.e., adjusted relative to the level of expected recall prior to repetition), then both the honored and the forced conditions exhibit a spacing effect that is not measurably different in size.

Experiment 2

In Experiment 2, we replicated Experiment 1 without the Done option. That is, all study pairs were restudied, and participants chose whether to restudy Now or Later. Experiment 2 was conducted for several reasons. First, the results of Experiment 1 required replication. Second, elimination of the Done option placed more items in the Now and Later categories, increasing the power of the study and causing fewer participants to be eliminated from the analysis. Third, although Experiment 1 was capable of discerning differences between conditions (e.g., between baseline Now and Later items, between adjusted spaced and massed items, and so on), recall performance was somewhat low in some of the critical conditions. Elimination of the Done option placed more of these higher recall items into the more important Now and Later categories, increasing recall levels in these conditions.

Method

Participants. Forty-six undergraduates at the University of North Carolina at Chapel Hill participated.

Materials and procedure. The materials were identical to Experiment 1. The procedures were also identical with one change. During the study phase, the Done option was eliminated. Participants were told that all study pairs would be restudied and that they would choose to restudy the pair now or later. Otherwise, the instructions and presentation of study pairs was identical to Experiment 1.

Results and Discussion

Two participants were replaced because they had zero study items in one of the conditions, leaving an effective sample of 44.

JOL results. A preliminary analysis indicated that JOLs did not differ across study items that were subsequently honored, dishonored, or in the baseline condition ($F < 1$). JOLs were significantly greater for items chosen for restudy Later ($M = 52$, $SD = 19.7$) compared with those chosen for restudy Now ($M = 36$, $SD = 18.8$), $t(43) = 5.79$, $p < .001$. The relationship between JOLs and subsequent recall performance was assessed with the gamma correlation ($M = 0.53$) and is comparable to that found in Experiment 1.

Recall results. The unadjusted recall scores (Figure 2a) were analyzed with a 2 × 2 ANOVA using spacing (massed vs. spaced) and honor condition (honored vs. forced) as within-subject factors. The analysis revealed a main effect of spacing, $F(1, 43) = 8.52$, $MSE = 0.028$, $p = .006$, and a significant interaction, $F(1, 43) = 5.14$, $MSE = 0.032$, $p = .028$. Analyses of the simple effects indicates a significant effect of spacing in the honored condition, $t(43) = 4.99$, $p < .001$, but not in the forced condition, $|t| < 1$.

Recall in the baseline condition (Figure 2b) again shows that study pairs chosen for spacing are easier to recall than items chosen for massing even without any repetition: recall was significantly greater for the Later baseline items than for the Now baseline items, $t(43) = 2.49$, $p = .017$. Adjusted recall scores were computed as in Experiment 1 (Figure 2c) and analyzed with a 2 (spacing) × 2 (honor condition) ANOVA. The analysis revealed only an effect of spacing, $F(1, 43) = 8.48$, $MSE = 0.051$, $p = .006$ (other Fs < 1). That is, spacing enhanced memory more than massing relative to the appropriate baseline condition, and by the same amount, whether the metacognitive choice was honored or not.

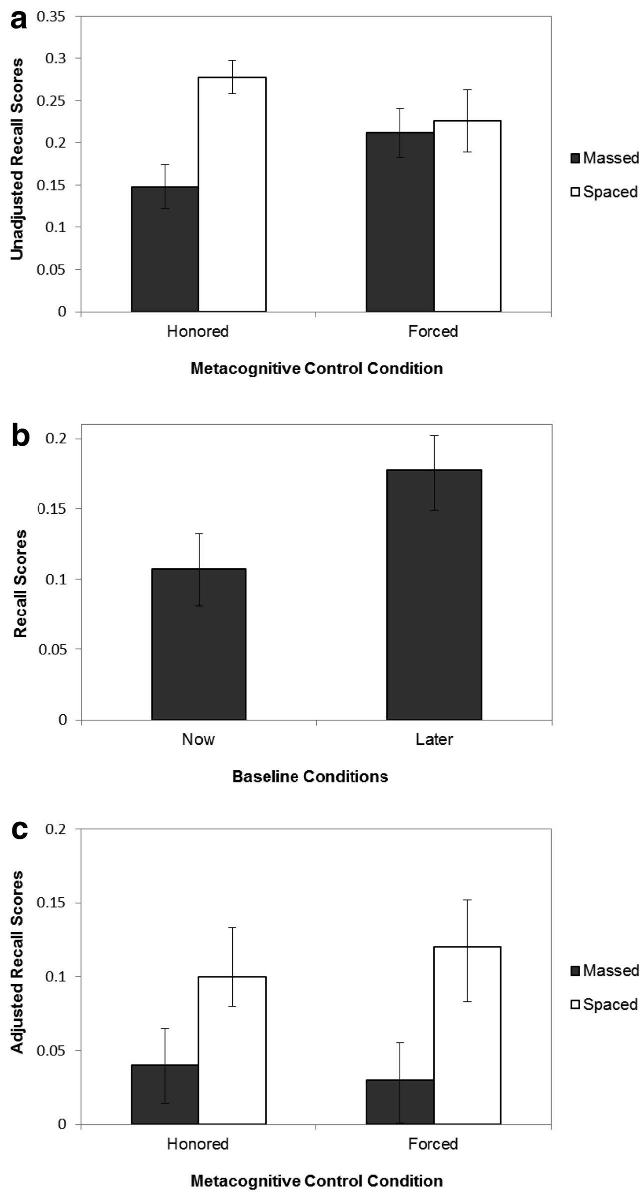

Figure 2. Results of Experiment 2. Mean unadjusted proportion recalled (± *SE*) as a function of metacognitive control and spacing (Panel a). Mean proportion recalled (± *SE*) in the baseline conditions (Panel b). Mean adjusted proportion recalled (± *SE*) as a function of metacognitive control and spacing (Panel c). Honored versus forced = choice honored or forced choice; Now versus later = restudy time immediate or delayed.

We met our goals in Experiment 2. Elimination of the Done option increased the number of items placed into the Now and Later categories leading to the exclusion of fewer participants and increasing, at least somewhat, final recall in these conditions. More important, the results of Experiment 1 were replicated. First, the JOL and unadjusted recall results replicated both Experiment 1 and the original Son (2010) study: JOLs were higher for Later than for Now choices, gamma correlations were relatively high (and comparable to Experiment 1 and to Son, 2010), and a spacing effect occurred for the unadjusted recall scores in the honored condition but not in the spaced condition. Second, like Experiment 1, the results of the baseline condition reveal the differences between items chosen for massed and spaced study, the latter being easier to recall than the former before any repetition occurs. Third, the use of the baseline measures provides a more appropriate assessment of the effects of repetition in this paradigm, showing that the spacing effect was equivalent across the honored and forced conditions.

To provide a more powerful analysis of the size of the spacing effect in the honored and forced conditions, we combined the adjusted recall scores from Experiments 1 and 2. The resulting analysis demonstrated that the size of the spacing effect on the adjusted recall scores was statistically equivalent across honor conditions (i.e., the interaction between the spacing and honor factors was not significant, $F < 1$). Furthermore, the size of the spacing effect was numerically larger in the forced condition (.08) than in the honored condition (.07), indicating no hint of a trend for the spacing effect to be smaller with forced spacing.

An alternative evaluation of adjusted recall scores is based on an item (rather than a subject) analysis. This adjusts recall levels controlling for item difficulty, a potentially important issue given that items of different average difficulty are placed in the Now and Later categories by the participants. Adjusted recall scores computed over items produced the same results as the subject analyses reported earlier. In Experiment 1, the effect of spacing was significant, $F(1, 62)$[3] $= 8.80$, $MSE = 0.039$, $p = .004$, indicating greater recall in the spaced than in the massed condition, and the interaction between the spacing and honor factors was not significant ($F < 1$). The same results were found in the item analysis of Experiment 2: spacing was significant, $F(1, 73) = 21.17$, $MSE = 0.027$, $p < .001$, and the interaction with the honor factor was not significant, $F = 1.3$, $p = .258$. Combining the item analyses of Experiments 1 and 2 into a single, more powerful analysis likewise indicated no interaction between the spacing and honor factors ($F < 1$). As with the subject analysis, the spacing effect in the adjusted recall scores was numerically larger for the forced (.10) than honored (.06) condition. Thus, whether adjusted recall scores

[3] Given that participants placed items into different conditions by their re-study choices, not all items occurred in each critical condition over subjects. Such items cannot be included in the analysis, leading to 63 (or 80) items used in the item analysis of Experiment 1 and 74 (out of 80) used in Experiment 2.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

are computed by subjects or by items, the same results are found: adjusted recall exhibits a spacing effect unmoderated by honor condition.

Before continuing, we should note that although the present paradigm has often been used to investigate metacognition and spacing (e.g., Benjamin & Bird, 2006; Son, 2004, 2010; Toppino & Cohen, 2010; Toppino, Cohen, Davis, & Moors, 2009) it does have limitations. First, the results of the paradigm are sensitive to the presentation times of the study items. For example, in the present experiments, participants chose to mass items they perceived to be more difficult (replicating Son, 2004, 2010). Other studies have found that participants prefer to space rather than mass difficult items (e.g., Benjamin & Bird, 2006). Toppino et al. (2009) demonstrated that the duration of the initial study presentation is a critical factor in whether participants choose to space or mass difficult items (see also Toppino & Cohen, 2010). This complexity is not particularly germane to the present analysis, however, which focuses only on the paradigm as employed by Son (2004, 2010). However, it does indicate that the item selection problem uncovered in the present experiment may manifest in other forms for different variants of this paradigm (e.g., with longer initial study times, easier rather than more difficult items might be overrepresented among Now decisions).

A second limitation of the design is that the second presentation of the spaced items occurs after the massed items are presented; that is, recency and spacing are confounded (see Son, 2004; Toppino et al., 2009, for discussion). Given the goals of the present study, it was necessary to use this paradigm and thus inherit this limitation. Two issues should be noted in this regard. First, Toppino et al. (2009) used a variant of the present paradigm, controlling for recency, and found results generally consistent with the standard version of the paradigm. Second, in the present case, the critical issues have to do with the presence of the item selection problem exhibited by the baseline conditions and with the size of the spacing effect in the adjusted scores across the honor conditions. Neither of these comparisons is compromised by the potential inflation of spacing effects in this paradigm by the recency confound.

Conclusions

The spacing effect is a ubiquitous result in studies of the effect of repetition on memory. Son (2010) reported a potentially critical limiting condition of this effect, in which spacing contrary to the wishes of the adult learners produced no spacing effect (or a reduced spacing effect in the case of children), whereas spacing consistent with the learners' wishes produced a robust spacing effect. This implied that metacognitive control might be a critical moderating factor in the effectiveness of spaced repetition. The present experiments examined an important item selection confound in this paradigm: the choices of the participants categorized which items are placed in the various massed and spaced conditions, giving rise to the possibility that items chosen for spacing may be easier to begin with than items chosen for massing. The present experiments indicate that this is the case. Consequently, the unadjusted recall scores generated in this paradigm are not interpretable. Adjusting recall relative to the baseline conditions demonstrated that both the honored and dishonored conditions produced a spacing effect of comparable size. The results of this study provide evidence against the notion that metacognitive control is a moderating influence on the effects of spacing in memory. Such evidence may potentially be adduced from other sources, but not from the present paradigm.

References

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. (1979). Information-processing analysis of college teaching. *Educational Psychologist, 14,* 15–23. doi:10.1080/00461527909529203

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. doi:10.1111/j.1467-9280.2008.02209.x

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. In B. H. Ross (Ed.), *Psychology of learning and motivation* (Vol. 53, pp. 63–147). San Diego, CA: Elsevier Academic Press. doi:10.1016/S0079-7421(10)53003-2

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

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology*. New York, NY: Dover. (Original work published 1885)

Greene, R. L. (2008). Repetition and spacing effects. In H. L. Roediger (Ed.), *Learning and memory: A comprehensive reference*. Cognitive psychology of memory (pp. 65–78). Oxford, England: Elsevier. doi:10.1016/B978-012370509-9.00168-6

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. (2010). Metacognitive control and the spacing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 255–262. doi:10.1037/a0017892

Son, L. K., & Simon, D. A. (2012). Distributed learning: Data, metacognition, and educational implications. *Educational Psychology Review, 24,* 379–399. doi:10.1007/s10648-012-9206-y

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.

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. doi:10.1037/a0016371

Received March 5, 2013
Revision received June 10, 2013
Accepted June 17, 2013 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.