

Metacognitive Control and Spaced Practice: Clarifying What People Do and Why

Thomas C. Toppino
Villanova University

Michael S. Cohen
University of California, Los Angeles

What do learners do when they control whether to engage in massed or spaced practice? According to theories by Son (2004) and by Metcalfe and Kornell (2005), the tendency for learners to choose spaced practice over massed practice should decline as item difficulty becomes greater. Support originally was obtained when pairs containing unfamiliar words were presented briefly for study, but subsequent research has suggested that, under these conditions, learners had difficulty initially encoding the members of the to-be-learned pairs. In Experiments 1 and 2, we failed to support the previously mentioned prediction in conditions in which the difficulty of learning was not correlated with the difficulty of initially encoding the pair members. Learners' relative preference for spaced practice increased, rather than decreased, with greater item difficulty, consistent with either a discrepancy-reduction-like account or an agenda-based-regulation account. In Experiment 3, we independently varied item difficulty and the point value that items were worth on the final test. Learners' relative preference for spaced practice was greater for high- than for low-value items but was unaffected by item difficulty. These results are more consistent with an agenda-based-regulation account than with a discrepancy-reduction account. More generally, learners' choices appear to be strategic and to reflect theory-based decisions, suggesting some level of appreciation for the relative benefits of massed versus spaced practice.

Keywords: metacognition, spacing effect, metacognitive control, distributed practice

Knowledge of one's own knowledge and cognitive processes (i.e., metacognition) plays an important role in learning and memory (Dunlosky & Bjork, 2008). People are able to monitor the state of their knowledge, and strategic decisions are often guided by metacognitive considerations. One manifestation of the ability to monitor is the fact that judgments of learning (JOLs) that are made following a period of study often correlate reasonably well with later memory performance. Metacognitive control is apparent when people select their learning activities on the basis of their current knowledge and/or on their understanding of the effectiveness of cognitive processes.

Allocation of Study Time in Self-Regulated Learning

The relationship between metacognition and learning has been studied extensively in the context of how people allocate study

time to items they are trying to learn. There is general agreement that learners monitor their level of knowledge and adjust study time accordingly. However, there has been some disagreement about just how this process works. According to the *discrepancy reduction model* (Dunlosky & Hertzog, 1997), learners set a desired level of learning (norm of study) as a goal and operate to reduce the discrepancy between a given item's level of learning and the target level. Thus, after an initial study period, a greater amount of study time should be allocated to items that are perceived to be relatively difficult and poorly learned (i.e., further from the norm of study) than to items that are perceived to be better learned. An alternative theoretical analysis is provided by the *region of proximal learning model* (Metcalfe, 2002; Metcalfe & Kornell, 2005). Study time presumably should be allocated preferentially to the easier-to-learn items among those that are not yet fully learned. These items are said to be in a learner's region of proximal learning. Especially when the total amount of time for studying is limited, allocating the available time to nearly learned items in the region of proximal learning would be more efficient and effective than allocating it to the most difficult, poorly learned items on which little progress could be expected.

Most recently, Ariel, Dunlosky, and Bailey (2009) argued for a more general theory of self-regulated learning, which they called the *agenda-based regulation model*. In this view, study regulation is assumed to be goal-oriented. Learners presumably construct an agenda that they use to guide the allocation of study time in a way that will efficiently maximize the attainment of task goals. How items are prioritized for study by a learner's agenda will depend on a variety of factors, including among others, item difficulty, constraints on learning time, and the reward structure of the task. Thus, the model can predict that study time will be preferentially

This article was published Online First September 6, 2010.

Thomas C. Toppino, Department of Psychology, Villanova University; Michael S. Cohen, Department of Psychology, University of California, Los Angeles.

The authors thank the members of the Memory and Cognition Lab at Villanova for their assistance with data collection and analysis and for their helpful comments and suggestions while this research was being planned and conducted. Special thanks go to Amy Moors and Ryan Corser for their help with Experiments 1 and 2; Kathleen Dempsey for her help with Experiment 2; and Matt Litke, Damien Germino, and Nick Schaefer for their help with Experiment 3. We also thank the Bjork and Castel labs at the University of California, Los Angeles, for their helpful suggestions relating to this work.

Correspondence concerning this article should be addressed to Thomas C. Toppino, Department of Psychology, Villanova University, 800 Lancaster Avenue, Villanova, PA 19085. E-mail: thomas.toppino@villanova.edu

allocated to more difficult items (e.g., when study time is unlimited and total mastery is a goal), to easier unlearned items (e.g., when study time is insufficient for mastery but the goal is to learn as many items as possible), or to valuable items regardless of their difficulty (e.g., when knowing certain items is more desirable than knowing others).

Each of the theories just reviewed has received impressive empirical support from studies in which learners regulated the amount of study time that was devoted to each item (e.g., Ariel et al., 2009; Dunlosky & Hertzog, 1997; Dunlosky & Thiede, 1998; Metcalfe & Kornell, 2003, 2005; Son & Metcalfe, 2000). A more recent concern has been how learners distribute their practice with a given item over time when the total amount of study time is held constant. This issue complements research on the allocation of study time and has great relevance for how students actually study.

Metacognition and the Distribution of Practice

The fact that distributed or spaced practice leads to better retention than does massed practice may be one of the most studied and replicated phenomena in the history of learning and memory (for reviews, see Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Dempster, 1996; and Hintzman, 1974). Yet despite the robustness of the spacing advantage, learners do not always seem aware of this effect. In motor learning, participants predict better test performance following massed practice than following spaced practice (Simon & Bjork, 2001) and may even express dissatisfaction with spaced practice (e.g., Baddeley & Longman, 1978). In the context of list learning, JOLs that are made immediately following practice may not differ for massed and spaced items or actually may favor the former (e.g., Zechmeister & Shaughnessy, 1980). Some awareness of the benefits of spacing is suggested, however, by the fact that JOLs are greater for spaced than for massed items when a delay is imposed between the final practice opportunity and the metacognitive judgment (e.g., Dunlosky & Nelson, 1994). Importantly, actual test performance is consistently better for spaced than for massed items, regardless of learners' metacognitive predictions.

When learners are allowed to control how their practice will be scheduled, conflicting results also have been obtained. In the first study to use this approach, Son (2004) presented learners with word pairs (GRE vocabulary words and their common synonyms) that were to be learned for a subsequent cued recall test. After a brief (1-s) initial study opportunity with each pair, learners made a JOL and decided whether they wanted to study the item again now (massed practice), later (spaced practice), or not at all (done). If they chose to study an item again now, it was re-presented immediately for 3 s. If they chose to study an item again later, it was re-presented for 3 s after all items had been presented at least once. Results indicated that, as item difficulty (indexed by JOLs) increased, the proportion of restudied items increased and, correspondingly, the proportion of "done" items declined. Among the restudied items, spacing was preferred for the easiest items, but as item difficulty increased, the relative tendency to choose spaced practice decreased whereas the relative tendency to choose massed practice increased.

Son (2004) interpreted her results as support for a *metacognitive hypothesis* according to which the choice of massed or spaced practice depends on the degree to which an item is perceived to be

learned. Among items that are not fully learned after the first study opportunity (and are therefore selected for restudying), those that are judged to be better learned (easier items as reflected by JOLs) should be selected for spaced practice because little would be gained from further immediate study. However, items that are more poorly learned after their initial presentation (harder items with low JOLs) are more likely to be given massed practice because, in this case, continued study would be profitable.

Son's (2004) results also can be interpreted in terms of a related hypothesis that was formulated in the context of the region of proximal learning model (e.g., Metcalfe, 2002, 2009; Metcalfe & Kornell, 2005). The key assumption is that the choice of whether to continue studying depends on metacognitive judgments of the rate of learning (jROLs), reflecting the perceived rate at which learning is taking place. When items are hard but not impossible, learners should perceive themselves to be making progress (high jROLs) at the end of the study period for an item, and they should opt to continue studying by choosing massed practice. When items are easy, jROLs may be low at the end of a study period, indicating either that learners believe the item to be fully learned or that they believe they have learned about as much as they can on this particular presentation. In the former case, they should choose *done*, and in the latter case they should select *study later*, for spaced practice.

Benjamin and Bird (2006) continued the investigation of metacognitive control over the distribution of practice using a different set of procedures. In their first two experiments, Benjamin and Bird presented pairs of common words and varied their difficulty experimentally. All words were presented twice, and the duration of all presentations was 5 s. After the initial presentation, learners chose whether to study the item again sooner (massed practice) or later (spaced practice). If they chose *sooner*, the item was re-presented after one intervening pair had been presented. If they chose *later*, the item was repeated after all other items had been presented at least once. Finally, learners' choices were constrained such that they were required to select massed and spaced practice equally often. Under these conditions, Benjamin and Bird obtained results that were the opposite of those reported by Son (2004). Learners preferred spacing for the more difficult items and massing for the easier items. The authors suggested that their findings were generally consistent with the discrepancy reduction model (Dunlosky & Hertzog, 1998) because learners used the more effective study technique (spacing) for the more difficult items that needed it most.

The disparity between the findings reported by Son (2004) and by Benjamin and Bird (2006) was addressed in a recent study by Toppino, Cohen, Davis, and Moors (2009). Their first experiment was a methodological replication of Son's (2004) study except that item difficulty was manipulated experimentally, JOLs were omitted, and the duration of the initial presentation of items was varied (1 s as in Son's study vs. 5 s as in Benjamin and Bird's research). The results provided an empirical reconciliation of the discrepant findings reported by Son and by Benjamin and Bird by replicating both previous studies. As item difficulty increased, learners became more likely to restudy items, but the proportion of restudied items that were selected for spaced practice declined with increasing item difficulty when the initial presentation duration was 1 s, replicating Son's study, and increased when it was 5 s, replicating Benjamin and Bird's study.

Toppino et al. (2009) noted that the GRE vocabulary words became longer and rarer with increasing item difficulty. They hypothesized that, as a consequence, initially encoding both members of harder pairs would be more problematic and more likely to fail when a brief initial presentation duration is used. If learners also choose to see a pair right away (massed practice) when they do not fully encode both words initially, the difficulty of establishing an initial representation for both members of harder pairs might contribute to the tendency for spacing to become relatively less preferred (and massing relatively more preferred) as item difficulty increases.

Toppino et al. (2009) tested their hypothesis in a second experiment. They replicated the 1-s condition of their first experiment except that, immediately after learners chose massed practice, spaced practice, or no further practice for a pair, they were asked to report (“yes”/“no”) whether they had seen both members of the pair. The overall results replicated those of the previous experiment and of Son (2004) by showing that the relative tendency to choose spacing declined as item difficulty became greater. However, the proportion of items that learners reported fully “seeing” also declined as a function of increasing item difficulty, and, when both members of a pair were not fully encoded, learners asked that it be presented again right away (massing) with a high probability. When pairs that were not fully encoded were omitted from the data analysis, the tendency for spaced practice to become less preferred with increasing item difficulty was largely eliminated.

Theoretical Issues and the Present Research

The results reviewed thus far have left several theoretical issues unresolved. First, the status of the metacognitive and jROL hypotheses is unclear. Although neither hypothesis can account for the way learners chose to distribute their practice when a long, 5-s presentation duration was used, one can question whether the long initial presentation duration constitutes an appropriate condition for testing their predictions. According to the jROL hypothesis, whether learners will mass difficult items depends on whether they perceive that they are still making sufficient progress on these items at the end of the initial study period. If they are still making progress, they will choose to continue studying by opting for massed practice. However, if the study duration is long enough that learners perceive themselves to have learned about as much as they can on that presentation regardless of item difficulty, the prediction that spacing should be chosen relatively less often for more difficult items would no longer hold. This does not explain why the relative preference for spaced practice compared with massed practice is actually greater for hard items than for easy items with a long presentation duration, but the primary point in the present context is that a shorter presentation duration may be optimal for testing the prediction in question. (It is unclear whether an analogous claim can be made on the basis of Son’s, 2004, metacognitive hypothesis.)

Even when brief presentation durations have been used, however, questions about the metacognitive and jROL hypotheses remain. In some experiments, these hypotheses successfully predicted that learners would choose spacing relatively less as item difficulty increases under these circumstances (Son, 2004; Toppino et al., 2009, Experiment 1). However, Toppino et al.’s (2009) Experiment 2 revealed that this finding was a result of learners’

having greater difficulty encoding the individual members of harder pairs. This may reflect an early stage of processing involved in learning pairs, but other processes, such as establishing an associative connection between pair members, are also thought to contribute to the difficulty of learning. Learning difficulty frequently can be varied independently of initial encoding difficulty or under conditions that minimize the contribution of this early stage of processing. Thus, existing support for the metacognitive and jROL hypotheses seems to be confined to a limited set of circumstances. It seems that the generality of these hypotheses, and perhaps their viability, will depend importantly on whether results consistent with their predictions can be obtained when learning difficulty is not greatly affected by the initial encoding of the items contained in the to-be-learned pairs.

Experiments 1 and 2 of this article addressed this issue by attempting to extend the boundaries within which the predictions of the metacognitive and jROL hypotheses hold true. We assessed the predictions under conditions in which presentation durations were short enough to provide a valid test of the hypotheses but in which learning difficulty could not be reduced to the difficulty of initially encoding the items comprising the pairs.

A second unresolved issue is why learners’ relative preference for spaced practice increases as a function of item difficulty when the initial presentation duration is relatively long (Benjamin & Bird, 2006; Toppino et al., 2009). As suggested by Toppino et al. (2009), part of the story may be that a longer presentation duration allows sufficient time to encode the members of all pairs, regardless of item difficulty. However, that alone does not explain the greater preference for spacing with the harder items. Benjamin and Bird (2006) as well as Toppino et al. considered this finding to be consistent with a discrepancy-reduction-like explanation in the sense that learners preferentially allocated the more effective learning technique (spacing) to the harder items. However, these results can also be interpreted within the agenda-based regulation framework (Ariel et al., 2009) to the extent that learners may adopt a strategy in which difficult items are given a higher priority than easier items. The two hypotheses are similar in that they both imply that learners appreciate the relative effectiveness of massed versus spaced practice and allocate the more potent form of practice strategically to facilitate the learning of the more difficult or higher priority items, respectively.

It should be noted that these two theories were developed in the context of study time allocation and were applied post hoc to the metacognitive control of the distribution of practice. In this context, there is no clear independent evidence for either of these accounts or for the underlying assumptions that learners realize the relative effectiveness of massed versus spaced practice and that they prefer spacing in a strategic attempt to facilitate learning. We address these issues preliminarily in Experiment 2 and more systematically in Experiment 3. In the latter study, we also empirically contrast predictions derived from the discrepancy-reduction and agenda-based-regulation perspectives.

Experiments 1A and 1B

In Experiment 1A, we assessed learners’ metacognitive decisions about how to distribute practice under conditions in which the duration of the initial study opportunity was limited (1 s) and in which we avoided a systematic effect due to the difficulty of

initially encoding the individual members of each pair. Our procedures were exactly the same as those used in the 1-s condition of Toppino et al.'s (2009) first experiment, except that both members of the pairs were common words and the difficulty of learning the pairs was varied primarily by altering the normative relationship between the cues and targets. Unlike with the pairs involving GRE vocabulary words that were used in previous research by Toppino et al. and by Son (2004), in our experiment important characteristics of the individual words comprising the pairs, such as word length and frequency, did not vary systematically as a function of how difficult the pairs were to learn. Therefore, although learners might have difficulty identifying the words comprising some pairs due to the brief presentation duration, this tendency was expected to remain relatively constant across levels of learning difficulty.

In view of the brief presentation duration used in Experiment 1A, the metacognitive and jROL hypotheses (Metcalf, 2009; Metcalf & Kornell, 2005; Son, 2004) predict that learners' relative preference for spacing will decline as the difficulty of learning the items increases. However, if this prediction applies only when the difficulty of learning is positively correlated with the difficulty of initially encoding the individual members of the pairs, a different pattern of results may be obtained when this correlation is avoided by varying learning difficulty among pairs of common words. For example, it is possible that the relative preference for spaced practice might increase with greater item difficulty, an outcome that would be compatible with both the discrepancy-reduction and agenda-based-regulation hypotheses. A nonsignificant trend consistent with such an outcome was previously reported by Cohen (2007), who also used pairs of common words (see also Pyc & Dunlosky, 2010).

Experiment 1B was conceived as a control experiment and was run concurrently with Experiment 1A. It differed only in the fact the learning materials consisted of GRE vocabulary words and their synonyms for which learning difficulty has been found to be correlated with the difficulty of initially encoding the members of each pair (Toppino et al., 2009). We expected to replicate previous findings (Son, 2004; Toppino et al., 2009) showing that these learning materials lead to a decline in the relative preference for spaced practice as item difficulty increases, consistent with the predictions of the metacognitive and jROL hypotheses. If the results of Experiment 1A are not also consistent with the predictions of the latter hypotheses, a comparison with Experiment 1B would allow the results to be attributed unambiguously to the learning materials that were used and not to random factors such as differences in the participant population.

Method

Participants. A total of 42 introductory psychology students participated in Experiments 1A and 1B. Twenty-one were assigned randomly in blocks of two to each of the two experiments.

Materials. The materials in Experiment 1A consisted of 48 pairs (16 at each of three levels of difficulty) composed of common nouns between four and eight letters in length. The pairs were selected primarily on the basis of association norms (Nelson, McEvoy, & Schreiber, 2004), with imagery ratings (Friendly, Franklin, Hoffman, & Rubin, 1982; Paivio, Yuille, & Madigan, 1968) as a secondary factor. The normative forward associative strength between members of easy pairs ranged from .075 to .125.

For medium pairs, it ranged between .01 and .03. For hard pairs, there was no normative relationship between pair members, nor was any semantic relationship intuitively apparent. Imagery ratings were above 5.60 for the members of easy pairs and below 3.5 for hard pairs. The relative difficulty of the pairs was confirmed in pilot work on the basis of JOLs after one study opportunity and on the basis of subsequent cued recall.

The pairs used in Experiment 1B were composed of GRE vocabulary words and their synonyms and were identical to the materials used by Toppino et al. (2009). Briefly, we used 48 pairs that were selected from those originally used by Son (2004). We used her JOL data as a basis for assigning 16 pairs to each of three ranges of judged difficulty (easy, medium, and hard).

Procedure. The procedure of both experiments was exactly the same as that used in Experiment 1 of Toppino et al. (2009). Participants received detailed information about the procedure and were instructed to choose the strategy that they thought would be most beneficial for remembering the target items on the later cued recall test. During list presentation, pairs were presented successively in an order that was randomized independently for each participant. For the initial presentation of each pair, the cue and target appeared together on the computer screen for 1 s. Immediately after the pair disappeared, participants were prompted with a choice of three options: *study now* (massed practice), *study later* (spaced practice), and *done* (no further practice). If *done* was chosen, the pair was not presented a second time. If *study now* was selected, the pair was re-presented immediately for 3 s. If *study later* was chosen, the pair was added to a queue and was re-presented for 3 s after all items had been presented at least once. Thus, the second presentation of spaced items occurred at the end of the list, where they appeared in a new random order.

After a 5-min distractor task involving solving simple arithmetic problems, participants received a cued recall test. The cue words were presented alone in a new random order, and participants were given 10 s to type the target word into the computer.

Results

Study choice. The results for both Experiment 1A and Experiment 1B are shown in Figure 1 in terms of the mean proportion of pairs for which each choice (massed practice, spaced practice, or done) was chosen at each level of item difficulty (easy, medium, and hard). Although we reported the proportion of *Done* responses produced by participants, we primarily were concerned with the proportion of items that learners chose to mass and the proportion they chose to space. Together, they comprise the proportion of items that learners chose to restudy. The proportion of *Done* responses is simply the complement of the proportion of restudied items (one minus the proportion of restudied items) and, therefore, is completely redundant. For this reason, the proportion of *done* responses was not included in any of the analyses in our experiments. The significance level was set at $p = .05$ for all analyses.

The data for Experiment 1A were analyzed using a 3×2 repeated measures analysis of variance (ANOVA) involving item difficulty (easy vs. medium vs. hard) and spacing choice (massed vs. spaced). The analysis revealed a significant main effect of item difficulty, $F(2, 20) = 62.67$, $MSE = .01$, $\eta_p^2 = .76$, indicating that participants chose to restudy items more often as item difficulty increased. Equivalently, the participants chose the *Done* option

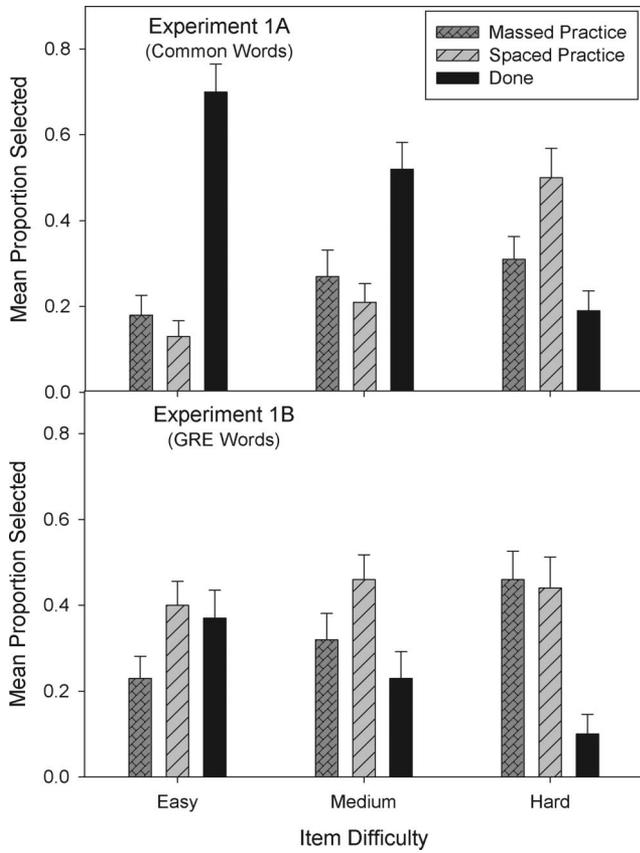


Figure 1. Mean proportion of items for which massed practice, spaced practice, and done were chosen at each level of item difficulty in Experiments 1A and 1B. Error bars represent the standard errors of the means.

less often as item difficulty became greater. This finding confirms that our a priori manipulation of item difficulty was successful. The most important finding was a reliable Difficulty \times Spacing interaction, $F(2, 40) = 5.96$, $MSE = .04$, $\eta_p^2 = .23$. For the easiest items, massed practice was preferred over spaced practice, but as item difficulty became greater, the choice of spaced practice increased more rapidly than did the choice of massed practice. That is, increasing item difficulty was associated with a greater relative preference for spaced practice.¹

An identical ANOVA was conducted on the data for Experiment 1B. There again was a significant effect of item difficulty, $F(2, 20) = 22.52$, $MSE = .01$, $\eta_p^2 = .53$, indicating that participants chose to restudy items more often as item difficulty increased. The Difficulty \times Spacing interaction was also significant, $F(2, 40) = 3.14$, $MSE = .03$, $\eta_p^2 = .14$. However, the form of the interaction was exactly the opposite of the comparable interaction in Experiment 1A. That is, spaced practice was preferred over massed practice for the easiest items in the present experiment, but as item difficulty became greater, the proportion of items selected for massed practice increased to a greater extent than did the proportion selected for spaced practice. Thus, the relative preference for spaced practice over massed practice declined with increasing item difficulty.

Cued recall. Performance on the cued recall test was of minor interest in the present experiment. As in previous work (Son, 2004;

Toppino et al., 2009), recall was thoroughly confounded by complex Participant \times Item interactions such that no statistical analysis seemed entirely appropriate and unequivocal interpretation seemed impossible. Nevertheless, recall generally improved with spaced practice and declined with greater item difficulty. The proportions of correct recall are presented in Table 1.

Discussion

The crucial finding is that the results of Experiment 1A involving pairs of common words failed to support the predictions of the metacognitive and the jROL hypotheses. Both hypotheses predict that the tendency to choose spaced practice over massed practice should be greatest for easy items and should decline with increasing item difficulty. Although item difficulty clearly affected how learners chose to distribute their practice, the effect was exactly the opposite of what was predicted. Learners were somewhat more likely to choose massed practice than spaced practice for easy items. As item difficulty became greater, however, the tendency to choose spaced practice increased more than did the tendency to choose massed practice, such that spaced practice was clearly preferred for the hardest pairs.

When learning materials consisted of GRE vocabulary words and their synonyms in Experiment 1B, the tendency to choose spaced practice more than massed practice declined with increasing item difficulty, replicating the previous findings by Son (2004) and by Toppino et al. (2009). The fact that Experiments 1A and 1B yielded opposite patterns of results can be attributed to the nature of the learning materials. Toppino et al. provided evidence that, as the difficulty of the GRE-related pairs increased, learners had greater difficulty encoding the words comprising each pair. We chose pairs of common words for Experiment 1A because, in contrast to the GRE-related pairs, the difficulty of encoding the individual words in these pairs should not vary as a function of the difficulty of learning the pairs. Under these circumstances, the tendency to choose spaced practice rather than massed practice increased with greater item difficulty, a pattern of results that is consistent with the expectation of the discrepancy-reduction-like hypothesis originally offered by Benjamin and Bird (2006) and with the agenda-based-regulation viewpoint, assuming that learners give more difficult items higher priority. Similar findings were observed by Toppino et al. with GRE-related pairs and a long, 5-s initial presentation duration. In both cases, the experimental procedures may have ensured that the probability of encoding both members of the pairs would be similar across levels of learning difficulty, either by providing extra processing time or by using materials in which learning difficulty and the difficulty of encoding the items in each pair are not correlated.

¹ To ensure that the critical interaction was not an artifact of the relatively small number of easy items that participants chose to restudy, we reanalyzed the data using only the two harder levels of item difficulty. The same pattern of results was obtained. Hard items were restudied more than medium-difficulty items, $F(1, 20) = 58.01$, $MSE = .01$, $\eta_p^2 = .74$, and there was a reliable Difficulty \times Spacing interaction, $F(1, 20) = 7.88$, $MSE = .04$, $\eta_p^2 = .28$.

Table 1
Proportion of Correct Recalls in Experiments 1, 2, and 3

Experiment	Easy			Medium			Hard		
	Massed	Spaced	Done	Massed	Spaced	Done	Massed	Spaced	Done
Experiment 1									
1A	.80	.74	.77	.22	.39	.49	.02	.09	.05
1B	.30	.54	.33	.15	.22	.23	.06	.17	.09
Total	.51	.59	.61	.18	.27	.41	.04	.13	.06
Experiment 2									
2.5-s duration	.36	.43	.48	.15	.18	.26	.03	.10	.11
5-s duration	.43	.47	.48	.22	.23	.28	.15	.12	.11
Total	.39	.44	.47	.18	.20	.27	.07	.11	.11
Experiment 3									
Value = 1	.38	.46	.31	.28	.23	.13	.10	.19	.11
Value = 5	.28	.56	.38	.12	.24	.23	.08	.15	.15
Total	.35	.51	.34	.22	.23	.17	.09	.17	.13

Experiment 2

The results of Experiment 1A expanded the generality of the finding that the relative preference for spaced practice over massed practice increases as the difficulty of learning the items becomes greater. In doing so, they provided no support for the predictions of the metacognitive and jROL hypotheses (Metcalf, 2009; Metcalfe & Kornell, 2005; Son, 2004). The primary purpose of the present experiment was to provide a further test of the predictions of these hypotheses.

We returned to the exclusive use of pairs consisting of GRE vocabulary words and their common synonyms. We used two initial presentation durations (2.5 s vs. 5 s), such that both durations would be long enough to allow both members of nearly all pairs to be initially encoded, and one would be short enough to preclude learning on a trial from being completed in the time allotted. The 2.5-s duration was selected because pilot work suggested that it was the shortest duration at which we could be reasonably confident that participants could initially encode both members of GRE synonym pairs, regardless of item difficulty. The 5-s duration was used by Toppino et al. (2009), who found that learners' relative preference for choosing spaced practice over massed practice increased as item difficulty became greater. We expected to replicate those findings. The condition was included to provide a point of comparison with the 2.5-s condition.

When the time required for establishing an encoding of the items comprising a pair is taken into account for the most difficult items, a 2.5-s presentation duration allows for little additional study time. At least some pairs should benefit from further immediate study. To the extent that the metacognitive and jROL hypotheses are applicable beyond the initial encoding stage, they should predict that the relative tendency to choose spaced practice over massed practice will decline as item difficulty becomes greater in the 2.5-s condition.

A second purpose of the present study was to collect preliminary data regarding learners' intent and their perception of their own strategies. To this end, participants were given a questionnaire following the study phase of the experiment but before the distractor activity and the recall test.

A final purpose of the present experiment was to minimize a confounding that could raise questions about all previous work on

how learners choose to distribute their practice. All experiments that have addressed this issue have used a procedure in which the second occurrence of spaced repetitions has been re-presented after all other pairs were presented at least once. Participants were aware of this fact in all cases. Thus, choosing spaced practice was confounded with opting for recency or a shorter retention interval. This seems unlikely to be a significant problem in view of the fact that people's metacognitions seem to be surprisingly insensitive to the effects of retention interval (Koriat, Bjork, Sheffer, & Bar, 2004), but we felt that the confounding should be minimized to be certain. We accomplished this with a procedure in which items chosen for spaced practice were repeated after four to 12 other presentations rather than after all other items had been presented. Thus, choosing to see an item later would still produce the benefits of spaced practice, but any advantage associated with reducing the retention interval would be greatly diminished, if not completely eliminated. If the results replicate previous findings, we can be more confident that learners' metacognitive choices reflect an appreciation of the relative benefits of spacing versus massing repetitions and not primarily the perceived advantage of a shorter retention interval.

Method

Participants. Sixty-nine introductory psychology students participated in the experiment. The data of six participants were excluded due to computer malfunction or for failure to follow instructions. The remaining participants were assigned randomly in blocks of two to two presentation conditions, resulting in the 2.5-s condition and the 5-s condition having 32 and 31 participants, respectively.

Materials and procedure. The materials were GRE vocabulary words and their synonyms that were identical to those used in Experiment 1B. The procedure was also identical to that of the previous experiments with the following exceptions:

First, depending on a participant's presentation-duration group, restudied items were presented for either 2.5 s or 5 s on their initial presentation. The shorter presentation duration was selected on the basis of pilot work in which we varied the initial presentation duration of pairs while using the procedure introduced by Toppino et al. (2009, Experiment 2), in which participants indicated

whether they were able to see both members of each pair. On the basis of these self-reports, a duration of 2.5 s appeared to be the shortest duration for which learners could reliably encode the members of nearly all pairs at all difficulty levels.

Second, when pairs were selected for spaced practice, their second occurrences were intermingled with the presentation of new items rather than being saved for re-presentation at the end of the list, as in previous experiments. This was achieved in the following way: Unbeknown to the participants, lists were structured in terms of successive groups of four pairs each, with the groups presented in a set order. Two different structures were created by means of independent random assignments of pairs to groups and of groups to serial order. Half of the participants in each presentation-duration condition received a list based on each list structure. Within each group of pairs, the order in which they were presented was determined randomly for each individual participant. When an item was selected for spaced practice, it was placed in a queue to be presented, in a randomized order, at the end of the following group of words. Thus, the order of groups was A B A' C B' D C', etc., where primed letters (e.g., A') designate the repetition of items selected for spaced practice. With this arrangement, the lag separating the two presentations of a spaced-practice pair was generally between four and 12 intervening presentations.² Participants were informed of this in the explanation of the *study later* option in the instructions.

Third, following the study phase of the experiment, but before the distractor task, participants were asked to fill out a questionnaire, answering questions about their strategies and the reasons for their choices.

Results

Study choice. Figure 2 depicts the mean proportion of pairs for which each choice (massed, spaced, or done) was selected for items at each level of difficulty (easy, medium, and hard). Preliminary analyses indicated that homogeneity of variance could be better approximated by means of an arcsine transformation that is often recommended for proportional data (e.g., Edwards, 1968; Guilford, 1965; Snedecor & Cochran, 1989). These data were submitted to a 2 (presentation duration) \times 3 (item difficulty) \times 2 (spacing choice) ANOVA with repeated measures on the last two factors. The results yielded a significant effect of presentation duration, indicating that learners whose initial study duration was 2.5 s chose to restudy pairs more often (and, consequently, chose the *done* option less often) than did participants who initially studied the pairs for 5 s, $F(1, 61) = 4.19$, $MSE = .81$, $\eta_p^2 = .06$. There also was a significant effect of item difficulty, $F(2, 122) = 63.70$, $MSE = .07$, $\eta_p^2 = .51$, indicating that learners chose to restudy items more often as item difficulty increased, and an effect of spacing, $F(1, 61) = 21.54$, $MSE = 2.73$, $\eta_p^2 = .26$, indicating that learners generally chose spaced practice more often than massed practice. Most importantly, there was a reliable Difficulty \times Spacing interaction, $F(2, 122) = 8.01$, $MSE = .15$, $\eta_p^2 = .12$, indicating that increasing item difficulty led to a greater increase in the choice of spaced practice than in the choice of massed practice. No other effects were reliable ($F_s < 1.28$).

Questionnaire. The questionnaire given after the study portion of the experiment was designed to gain preliminary insight into what learners thought they were doing when they made their

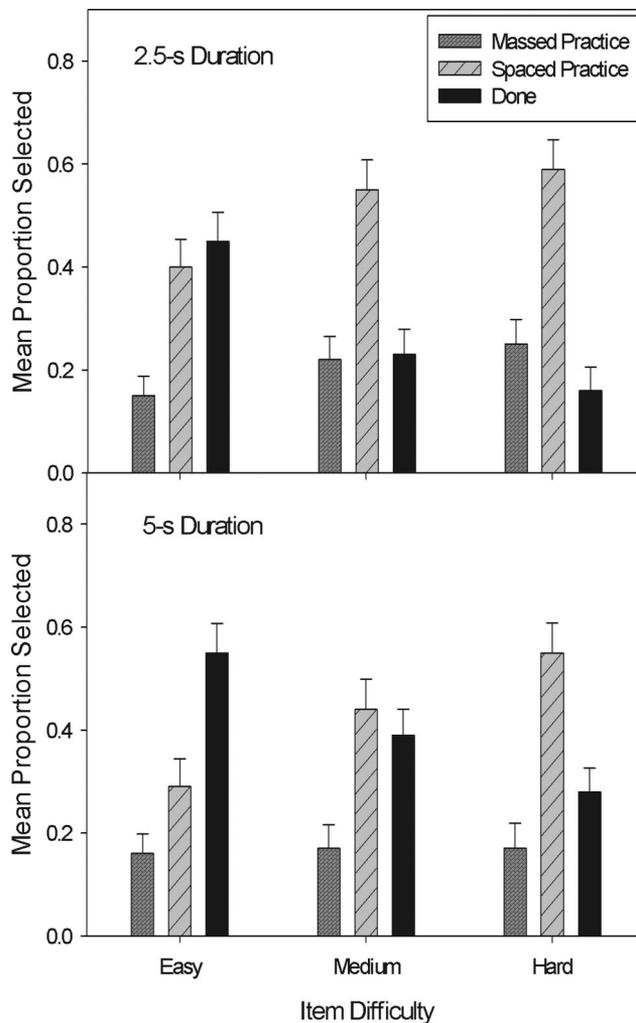


Figure 2. Mean proportion of items for which massed practice, spaced practice, and done were chosen at each level of item difficulty in the 2.5-s- and 5-s-duration conditions of Experiment 2. The error bars represent the standard errors of the means.

choices. Of the 57 participants who made use of the *done* option, 93% said they used it for items that they knew or expected to remember. The questionnaire asked about the strategies that learners had used in making their selections, and 62 participants gave interpretable responses. The most striking finding was that 71% of these participants indicated that they either chose spaced practice for most pairs or used spaced practice on harder pairs and massed practice for easier pairs, whereas 29% said they used mostly massed practice or allocated massed practice to harder items and spaced practice to easier items. However, these data should be viewed with considerable caution because at least 31% of the

² Depending on a participant's choices, a lag as long as 15 was mathematically possible. In practice, however, such lags occurred infrequently. Also, lags could be shorter than four intervening presentations for pairs in the last group of four items, but this could not be anticipated by participants and, thus, would not be expected to affect their choice of study strategies.

participants reported strategies that seemed to be at odds with the choices they actually made. For example, of 15 learners (9 in the 5-s duration condition and 6 in the 2.5-s condition) who reported using a strategy of massing harder items and spacing easier items, only 3 (all in the 2.5-s condition) actually behaved in that fashion. Finally, regardless of the strategy that they claimed to use, the vast majority of learners (84%) endorsed the idea that they were more likely to use the practice schedule they considered to be more generally effective for the harder items as opposed to the easier items.

Cued recall. Performance on the cued recall test again was of secondary interest due to inherent confounds that made interpretation futile. Nevertheless, recall tended to be better for spaced practice than for massed practice, and it declined with increasing item difficulty (see Table 1).

Discussion

Learners chose to restudy items less often (or, equivalently, chose *Done* more often) following a 5-s initial presentation than after a 2.5-s presentation. This implies that learners believed they made significant progress in learning during the extra time afforded by a 5-s presentation so that further study seemed less necessary. It also implies that they did not perceive themselves to have learned all they could about a pair in 2.5 s. Thus, it appears that the 2.5-s duration was short enough to provide an appropriate test of the predictions of the metacognitive and jROL hypotheses. Nevertheless, as in Experiment 1A, the results did not conform to expectations based on these hypotheses. That is, contrary to the predictions of the metacognitive and jROL hypotheses, learners' preference for spaced practice increased, rather than decreased, in both presentation-duration conditions as item difficulty became greater.

The increasing preference for spaced practice appeared to be somewhat less pronounced in the 2.5-s duration condition, which probably reflects the fact that a small number of learners in this condition appeared to have adopted the strategy predicted by the metacognitive and jROL hypotheses. These participants identified themselves as using the predicted strategy, and their choice behavior was consistent in that they exhibited a tendency to choose spaced practice less and massed practice more with increasing item difficulty. Nevertheless, it is clear that the majority of learners did not exhibit this strategy. Overall, our results are more consistent with the generalization that, when both items comprising a pair can be initially encoded regardless of item difficulty, most learners prefer to choose spaced practice rather than massed practice for the more difficult items. This pattern of behavior, in turn, is consistent with either a discrepancy-reduction-related account (e.g., Benjamin & Bird, 2006) or an agenda-based-regulation account (Ariel et al., 2009).

What has been lacking is independent evidence that learners' tendency to use spaced practice for more difficult items represents a strategy of allocating the more effective study technique to these items. Some tentative support for this hypothesis, however, is provided by our questionnaire data, in which the majority of learners claimed to adopt a strategy in which they used spaced practice for the hardest items and also claimed that the form of practice they used was the one that they considered to be generally more effective. We return to this issue in Experiment 3.

Finally, the present experiment used procedures that minimized and made less apparent the confounding between spacing repetitions and retention interval that had been characteristic of previous studies. Nevertheless, learners exhibited a strong preference to choose spaced repetitions, especially for the more difficult items. There was nothing in the results to suggest that minimizing the extent to which spaced practice was associated with a shorter retention interval affected learners' choice behavior, consistent with Koriat et al.'s (2004) conclusion that learners' metacognitive judgments can be relatively insensitive to the effect of retention interval.

Experiment 3

The results of the prior experiments in conjunction with those reported previously by Toppino et al. (2009) strongly indicate that, when the difficulty of initially encoding both members of the pairs is not an issue, learners exhibit an increasing tendency to choose spaced practice over massed practice as items become more difficult to learn. This finding can be construed to be consistent with the discrepancy reduction model as suggested by Benjamin and Bird (2006) or with an extension of the agenda-based-regulation model (Ariel et al., 2009).

The key ideas underlying discrepancy reduction are that learners strive to reduce the discrepancy between their perceived level of current learning and a desired or normative level of learning and that, in attempting to do so, they preferentially allocate study resources to more difficult items. As applied to the distribution of practice, this suggests that item difficulty (or perceived difficulty) should be the primary factor determining whether items are massed or spaced, with learners choosing the more effective strategy (spacing) for items that are more difficult or less well learned (Benjamin & Bird, 2006; Toppino et al., 2009).

The agenda-based-regulation model is more flexible. Self-regulated learning is viewed as a goal-oriented process in which learners construct agendas that they use to guide the allocation of study resources in pursuit of their goals. Under conditions in which difficult items are considered to be a priority, the agenda-based-regulation model would make the same prediction as would discrepancy reduction, namely that the more effective study technique (spaced practice) would be used for the more difficult items in order to maximize the total number of items recalled. However, the goal-driven nature of the agenda-based-regulation model suggests that a different pattern of results could be obtained under conditions involving different goals and agendas.

In Experiment 3, we independently varied item difficulty and learners' goals, with the latter accomplished by assigning point values for remembering different sets of items (e.g., Castel, 2007). Learners were told that their objective was to get the highest score possible on the final cued recall test. Half of the items at each difficulty level were assigned a value of 5 points, with the remaining items assigned a value of 1 point. Using a similar manipulation in the context of the self-regulation of study time, Ariel et al. (2009) found that learners allocated study time to the more valuable items rather than to the more difficult items. The analogous prediction in the context of the distribution of practice is that learners' tendency to choose spaced practice over massed practice will be greater for 5-point items than for 1-point items. Whether item difficulty will also affect learners' choices in this situation is

an open question, but it is important that an effect of item difficulty is not required by this hypothesis. As in the previous experiment, the participants in Experiment 3 also completed a questionnaire in which we probed their beliefs about the relative effectiveness of massed and spaced practice and the circumstances under which they chose each form of practice.

Method

Participants and procedure. Thirty introductory psychology students participated in an experiment that was procedurally identical to Experiment 1B with the following exceptions: The pairs of GRE words and their synonyms were presented for 5 s on their initial presentation. After the study phase of the experiment but before the distractor task, participants completed a questionnaire about the strategies that they used. Half of the items at each level of item difficulty were randomly assigned a value of 5 points, and the other half were assigned a value of 1 point. This randomization was conducted independently for each participant. All participants were instructed that, on the final cued recall test, they would receive points for the recall of each item, with the number of points depending on the value that had been assigned to the pair. Their task was to get as high a point total as possible on the final test.

Results

Study choice. Figure 3 presents the mean proportion of pairs for which each choice (massed, spaced, or done) was selected for items of each value (1 or 5 points) at each level of difficulty (easy, medium, and hard). The data were submitted to a 2 (value) \times 3 (item difficulty) \times 2 (spacing) ANOVA with repeated measures on all factors. The results indicated that 5-point items were restudied more than were 1-point items, $F(1, 29) = 7.34$, $MSE = .02$, $\eta_p^2 = .20$. In other words, *done* was chosen less often for the more valuable items. In addition, restudying was chosen more frequently as item difficulty increased, $F(2, 58) = 7.92$, $MSE = .01$, $\eta_p^2 = .22$, and spaced practice was chosen more often than massed practice overall, $F(1, 29) = 25.48$, $MSE = .53$, $\eta_p^2 = .47$. There was also a Value \times Difficulty interaction, $F(2, 58) = 4.97$, $MSE = .01$, $\eta_p^2 = .15$. The proportion of restudied items increased as a function of item difficulty when items were worth 5 points, $F(2, 58) = 10.87$, $MSE = .01$, $\eta_p^2 = .27$, but not when they were worth 1 point, $F(2, 58) = 1.55$, $MSE = .01$, $\eta_p^2 = .05$. Finally and most importantly, there was a Value \times Spacing interaction, $F(1, 29) = 6.50$, $MSE = .21$, $\eta_p^2 = .18$. Spacing was preferred both for 5-point items, $F(1, 29) = 38.37$, $MSE = .30$, $\eta_p^2 = .57$, and for 1-point items, $F(1, 29) = 7.26$, $MSE = .43$, $\eta_p^2 = .20$, but the preference for spaced practice over massed practice was markedly greater for higher valued items. No other effects were reliable ($F_s < 2.16$).

Questionnaire. In this questionnaire, we asked learners whether they thought the most effective strategy for learning was *study now* (massed), *study later* (spaced), or a mix of studying now for some items and studying later for others. Of the 30 participants, 7% endorsed massed practice, 53% endorsed spaced practice, and 40% endorsed a mixed strategy. When we probed the mixed-strategy participants about the details of their strategies, two reported accurately that they had used massed practice for the hardest, high-value items. The remaining 10 reported that they preferred spacing for more difficult items, higher value items, or

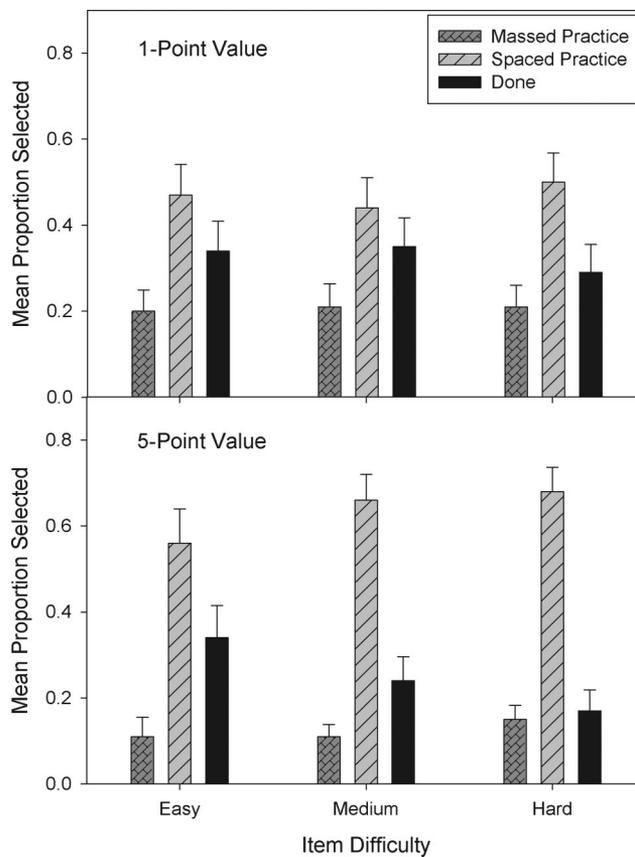


Figure 3. Mean proportion of items for which massed practice, spaced practice, and done were chosen at each level of item difficulty in the 1-point and 5-point value conditions of Experiment 3. The error bars represent the standard errors of the means.

some combination. Although there was some discrepancy between the details of what they said and what they seemed to do, they all accurately indicated that they had preferred spaced practice for difficult, high-valued items. To summarize, the questionnaire responses suggest that most learners believe that spaced practice is a more effective learning technique than is massed practice and that learners are generally aware of using a strategy in which they choose spaced practice for the higher valued and more difficult items.

Cued recall. As in the previous experiments, cued recall performance was of secondary interest. The proportions of correct recall are presented in Table 1, and once again, performance tended to be better for spaced practice than for massed practice and to decline with increasing item difficulty. Interestingly, high-value items were recalled only slightly better than were low-value items (.28 and .25, respectively), but it is not possible to know what, if anything, this means in view of the fact that the recall data are completely confounded with learners' strategies and Learner \times Item interactions.

Discussion

The critical results of this experiment are that learners tended to choose spaced practice over massed practice and that this tendency

was greater for higher value (5-point) than for lower value (1-point) items. Introducing point value as a goal-determining factor in this experiment, however, resulted in a diminished effect of item difficulty. Learners restudied harder items more frequently than easier items only when the items were of high value. And in contrast with the results of previous experiments, their relative preference for spaced practice was not reliably influenced by item difficulty at all. These findings are consistent with expectations derived from extending the agenda-based-regulation model to learners' self-regulation of the distribution of practice. As we discuss later, these findings also imply that learners appreciate the relative benefits of massed and spaced practice and strategically select spaced practice in order to achieve their learning goals.

General Discussion

Son's (2004) metacognitive hypothesis and Metcalfe and Kornell's (2005; Metcalfe, 2009) jROL hypothesis predict that the relative degree to which learners choose massed or spaced practice should depend on the difficulty of learning the items, with the relative preference for spaced practice declining as learning difficulty increases. In Experiment 1B, we replicated previous findings (Son, 2004; Toppino et al., 2009) that seemed to support this prediction under conditions involving a brief 1-s initial presentation duration and pairs consisting of GRE vocabulary words and their synonyms. Toppino et al. (2009) previously showed that, under these conditions, this finding can be attributed to learners having greater difficulty initially encoding the members of the more difficult pairs. Our present results strongly indicate that, when the initial difficulty of encoding the words comprising each pair does not vary with learning difficulty, the relative preference for spaced practice increases with item difficulty, contrary to the predictions of the metacognitive and the jROL hypotheses.

In Experiment 1A, we used a brief, 1-s presentation duration but with pairs of common words. These materials were selected because the difficulty of initially encoding the words comprising the pairs should be similar at all levels of learning difficulty, with difficulty depending primarily on how hard it is to establish an associative connection between members of the pairs. We found that learners' relative preference for spaced practice over massed practice increased rather than decreased as a function of how difficult it was to learn the items.

In Experiment 2, we used the GRE-related pairs that had been employed in Experiment 1B and in previous work (Son, 2004; Toppino et al., 2009), but we neutralized the effect of initial-encoding difficulty by using longer initial-presentation durations (2.5 and 5 s). Although the shorter duration was long enough to ensure encoding of the words comprising each pair, it was not so long that learners had completed studying for the trial within this time. This was evidenced by the fact that learners chose to restudy fewer items in the 5-s condition than in the 2.5-s condition, suggesting that the former group of participants perceived themselves to have learned more during the additional study time that was available to them. Regardless of the presentation duration, however, we found that learners' relative preference for spaced practice increased as the difficulty of learning the items became greater. This outcome was consistent with the results of Experiment 1A but not with the expectations of the metacognitive and the jROL hypotheses.

Together the results of Experiments 1A, 1B, and 2 suggest restrictive boundaries on the applicability of the metacognitive and jROL hypotheses. We were unable to obtain evidence that these hypotheses applied in any conditions other than those in which the difficulty of initially encoding pair members is correlated with the difficulty of learning the pairs. The hypotheses may capture the operation of an early stage of learning, but this stage of learning may often have little influence on the effect of learning difficulty, either because it is unrelated to learning difficulty (e.g., Experiment 1A) or because its effect is neutralized by allowing enough time for encoding to take place regardless of difficulty (e.g., Experiment 2). Finally, we cannot rule out the possibility raised by Toppino et al. (2009; see also Pyc & Dunlosky, 2010) that the difficulty of initial encoding may primarily reflect perceptual processes and may have little to do with learning as it usually is conceptualized.

Thus, our first two experiments established, in conjunction with several previous studies (Benjamin & Bird, 2006; Pyc & Dunlosky, 2010; Toppino et al., 2009), that learners generally prefer spaced practice to massed practice and that the relative preference for spaced practice becomes greater as the items become harder to learn. However, a theoretical account of this pattern of findings remained unclear. Benjamin and Bird (2006) and Toppino et al. (2009) speculated that the results were consistent with a discrepancy-reduction-like account in which learners apply the more effective learning technique (spacing) to the more difficult items for which the level of learning is perceived to be further from the desired level. However, the results also were potentially consistent with an extension of the agenda-based-regulation model (Ariel et al., 2009).

In Experiment 3, we provided the first experimental assessment of these hypotheses by independently varying item difficulty and the point value that was assigned to each item. Under these circumstances, learners' relative preference for spaced practice over massed practice increased as the point value of the items became greater but did not vary as a function of item difficulty. This finding favors an agenda-based-regulation account of the findings over a discrepancy-reduction account.

The fact that learners' relative preference for spaced practice depended on the reward structure of the task suggests that the degree to which learners chose massed or spaced practice was guided primarily by goal-oriented agendas in which items assigned a high point value were given priority. The discrepancy-reduction hypothesis (as applied to the distribution of practice) might account for the greater preference for spacing with high-value items by assuming that learners adopt a higher norm of study for these items than for low-value items. However, it is not clear how the hypothesis can accommodate the lack of an effect of item difficulty on the distribution of practice. Contrary to its predictions, there was no evidence that the choice of massing or spacing an item was guided by the degree to which the item's current level of learning was judged to deviate from a desired level of learning.

When an effect of item difficulty on the distribution of practice was obtained in previous experiments, the to-be-learned items varied only with respect to their relative difficulty. Under these circumstances, learners may adopt agendas in which they give priority to more difficult items. In this case, the discrepancy-reduction account of control over the distribution of practice may be seen as a special case of the agenda-based-regulation account in

which learners' agendas are determined primarily by the relative difficulty of learning the items.

Related to the previously mentioned theoretical issues is the more general question of the extent to which learners are cognizant of the relative effectiveness of massed and spaced practice. Theories have not always been in agreement on this point. For example, according to the jROL hypothesis (Metcalf, 2009; Metcalfe & Kornell, 2005), there may be no need to assume that learners appreciate the relative benefits of massing versus spacing practice. Learners are assumed to monitor their rate of learning with respect to the current item. Depending on whether they perceive themselves to be making progress, they will decide to continue studying the item or not. The point is that the decision is based on judgments about learning during the current presentation of an item and not on expectations about the benefits of a second presentation. Massing occurs when learners decide to continue studying the current item, whereas spacing is simply what happens (by default) when one decides to stop studying the current item. In contrast, both the agenda-based-regulation model and the discrepancy-reduction hypothesis seem to entail the assumption that learners allocate the more effective learning technique, respectively, to high-priority or more difficult items. This implies that they appreciate the relative effectiveness of massed and spaced practice and use this knowledge to make theory-based decisions about how to distribute their study opportunities.

Our results are consistent with the latter view. On responses to questionnaires, participants in our studies indicated that they used spaced practice for harder or higher valued items and that, in doing so, they believed they were using the more effective study strategy. Stronger support is provided by the fact that learners preferentially chose spaced practice for higher value items in Experiment 3. The distinction between high- and low-valued items was entirely arbitrary and uncorrelated with any factor affecting learning or the perception of learning, thus seeming to eliminate other possible bases for learners' decisions. Thus, the data from questionnaires administered in Experiments 2 and 3 and the choice data of Experiment 3 converged on the conclusion that learners do have at least some knowledge of the relative effectiveness of massed and spaced practice and that they used this knowledge to make theory-driven decisions about how practice should be distributed.

Although the present results clarify what people do and why they do it when they are in control of the way their practice is distributed, there are at least two unresolved issues that should be addressed by future research. First, all studies (including our own) that have examined metacognitive control of the distribution of practice have offered learners only two alternatives: massed practice versus a single level of spaced practice. Although learners appear to be cognizant of the relative benefits of massed and spaced practice, we cannot tell from this design whether they primarily appreciate the futility of massed practice or whether they also appreciate that increased spacing has benefits that go beyond an opportunity to avoid massed practice. What would learners do if their options included multiple levels of spacing? Second, despite the evidence that learners understand the relative benefits of massed and spaced practice, almost all learners chose massed practice some of the time. It remains unclear why they do this rather than maximizing their learning by choosing spaced practice for all items (see Toppino et al., 2009, for a previous discussion of this issue).

Monitoring Versus Control

The research on how learners control the distribution of practice strongly suggests that learners are cognizant of the relative effectiveness of massed and spaced practice. This seems to be at odds, however, with studies on metacognitive monitoring that indicate that, immediately after studying an item for the second time, learners often fail to predict that they will have better memory for spaced items and sometimes predict the opposite (e.g., Zechmeister & Shaughnessy, 1980). These findings, however, may be less paradoxical than they seem.

Metacognitive judgments are often analyzed in terms of the cues on which they are based (e.g., Koriat, Nussinson, Bless, & Shaked, 2008). It seems likely that choosing to study an item by means of spaced practice and predicting how well one will remember an item that previously received spaced or massed practice will be based on different cues, leading to imperfectly related outcomes. Adopting the terminology of Koriat (e.g., Koriat, 1997; Koriat & Bjork, 2006), choosing to study an item by spaced practice rather than by massed practice would be a theory-based decision using extrinsic cues, whereas predicting the probability of recalling a previously studied item is more likely to reflect an experience-based decision involving mnemonic cues. That is, choosing spaced study over massed study is likely to be based on a belief that certain conditions of learning (spacing) generally lead to better memory than do other conditions of learning. However, predictions of how well a previously studied item will be remembered are often attributed to the subjective feelings of processing fluency or accessibility related to that particular item, which learners interpret as indicators of the item's memorability (e.g., Benjamin, Bjork, & Schwartz, 1998; Dunlosky & Nelson, 1992). Future memory performance may be overestimated for massed items relative to spaced items because immediate accessibility or retrieval fluency may be greater after massed than after spaced practice (Peterson, Hillner, & Saltzman, 1962).

Failing to predict that an item with a history of spaced practice will be remembered better than one with a history of massed practice constitutes an error in metacognitive judgment, but the error does not seem to stem primarily from learners' failure to apply their theory that spaced practice leads to better memory. In fact, the theory is a generalization that applies best to large groups of items. Learners may correctly intuit that a history of massing or spacing would be an unreliable index of the future memorability of any one item. They quite reasonably may prefer to rely on an indicator of how well they currently know a particular item, while discounting the importance of the item's learning history (Koriat, 1997). The true source of the metacognitive error seems to lie primarily in overestimating the reliability of accessibility or fluency as indicators of learning and their utility as predictors of future memory performance. As Benjamin et al. (1998) noted, people have limited knowledge of the factors that affect fluency and little appreciation that the effects of some factors on fluency are transient. With respect to the distribution of practice, people may fail to realize that, although massed practice leads to greater immediate retrieval fluency than does spaced practice, the effect is short-lived.

References

- 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
- Baddeley, A. D., & Longman, D. J. A. (1978). The influence of length and frequency of training session on the rate of learning to type. *Ergonomics*, *21*, 627–635. doi:10.1080/00140137808931764
- 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
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General*, *127*, 55–68. doi:10.1037/0096-3445.127.1.55
- Castel, A. D. (2007). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation: Skill and strategy in memory use* (Vol. 48, pp. 225–270). London, England: Academic Press. doi:10.1016/S0079-7421(07)48006-9
- 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
- Cohen, M. S. (2007). How do subjects use judgments of item difficulty to guide study strategies in selection of spaced or massed practice? A comparison of theories. Unpublished master's thesis, Villanova University.
- Dempster, F. (1996). Distributing and managing the conditions of encoding and practice. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 317–344). San Diego, CA: Academic Press. doi:10.1016/B978-012102570-0/50011-2
- Dunlosky, J., & Bjork, R. A. (2008). *Handbook of metamemory and memory*. New York, NY: Psychology Press.
- Dunlosky, J., & Hertzog, C. (1997). Older and younger adults use a functionally identical algorithm to select items for restudy during multitrial learning. *Journal of Gerontology: Psychological Sciences*, *52B*, 178–186.
- 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.
- Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOLs) and the delayed-JOL effect. *Memory & Cognition*, *20*, 373–380.
- Dunlosky, J., & Nelson, T. O. (1994). Does the sensitivity of judgments of learning (JOLs) to the effects of various study activities depend on when the JOLs occur? *Journal of Memory and Language*, *33*, 545–565. doi:10.1006/jmla.1994.1026
- Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect people's self-paced study and yield "labor-and-gain" effects. *Acta Psychologica*, *98*, 37–56. doi:10.1016/S0001-6918(97)00051-6
- Edwards, A. L. (1968). *Experimental design in psychological research*. New York, NY: Holt, Rinehart and Winston.
- 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.
- Guilford, J. P. (1965). *Fundamentals of statistics in psychology and education*. New York, NY: McGraw-Hill.
- Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Hillsdale, NJ: Erlbaum.
- 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. (2006). Mending metacognitive illusions: A comparison of mnemonic-based and theory-based procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 1133–1145. doi:10.1037/0278-7393.32.5.1133
- 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
- Koriat, A., Nussinson, R., Bless, H., & Shaked, N. (2008). Information-based and experience-based metacognitive judgments: Evidence from subjective confidence. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 117–135). New York, NY: Psychology Press.
- Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General*, *131*, 349–363. doi:10.1037/0096-3445.131.3.349
- Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science*, *18*, 159–163. doi:10.1111/j.1467-8721.2009.01628.x
- Metcalfe, J., & Kornell, N. (2003). The dynamics of learning and allocation of study time to a region of proximal learning. *Journal of Experimental Psychology: General*, *132*, 530–542. doi:10.1037/0096-3445.132.4.530
- Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language*, *52*, 463–477. doi:10.1016/j.jml.2004.12.001
- 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*, *35*, 402–407.
- Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *66*, 206–209.
- Peterson, L. R., Hillner, K., & Saltzman, D. (1962). Time between pairings and short-term retention. *Journal of Experimental Psychology*, *64*, 550–551. doi:10.1037/h0046079
- 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
- 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
- Snedecor, G. W., & Cochran, W. G. (1989). *Statistical methods*. Ames, IA: Iowa State University Press.
- 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., & 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
- 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
- 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 17, 2009

Revision received May 5, 2010

Accepted July 6, 2010 ■