WILLIAMS COLLEGE LIBRARIES

COPYRIGHT ASSIGNMENT AND INSTRUCTIONS FOR A STUDENT THESIS

Your unpublished thesis, submitted for a degree at Williams College and administered by the Williams College Libraries, will be made available for research use. You may, through this form, provide instructions regarding copyright, access, dissemination and reproduction of your thesis. The College has the right in all cases to maintain and preserve theses both in hardcopy and electronic format, and to make such copies as the Libraries require for their research and archival functions.

___ The faculty advisor/s to the student writing the thesis claims joint authorship in this work.

___ I/we have included in this thesis copyrighted material for which I/we have not received permission from the copyright holder/s.

> If you do not secure copyright permissions by the time your thesis is submitted, you will still be allowed to submit. However, if the necessary copyright permissions are not received, e-posting of your thesis may be affected. Copyrighted material may include images (tables, drawings, photographs, figures, maps, graphs, etc.), sound files, video material, data sets, and large portions of text.

I. COPYRIGHT

An author by law owns the copyright to his/her work, whether or not a copyright symbol and date are placed on the piece. Please choose one of the options below with respect to the copyright in your thesis.

___ I/we choose not to retain the copyright to the thesis, and hereby assign the copyright to Williams College.

> Selecting this option will assign copyright to the College. If the author/s wishes later to publish the work, he/she/they will need to obtain permission to do so from the Libraries, which will be granted except in unusual circumstances. The Libraries will be free in this case to also grant permission to another researcher to publish some or all of the thesis. If you have chosen this option, you do not need to complete the next section and can proceed to the signature line.

X I/we choose to retain the copyright to the thesis for a period of 100 years, or until my/our death/s, whichever is the earlier, at which time the copyright shall be assigned to Williams College without need of further action by me/us or by my/our heirs, successors, or representatives of my/our estate/s.

> Selecting this option allows the author/s the flexibility of retaining his/her/their copyright for a period of years or for life.

II. ACCESS AND COPYING

If you have chosen in section I, above, to retain the copyright in your thesis for some period of time, please choose one of the following options with respect to access to, and copying of, the thesis.

X I/we grant permission to Williams College to provide access to (and therefore copying of) the thesis in electronic format via the Internet or other means of electronic transmission, in addition to permitting access to and copying of the thesis in hardcopy format.

> Selecting this option allows the Libraries to transmit the thesis in electronic format via the Internet. This option will therefore permit worldwide access to the thesis and, because the Libraries cannot control the uses of an electronic version once it has been transmitted, this option also permits copying of the electronic version.

___ I/we grant permission to Williams College to maintain and provide access to the thesis in hardcopy format. In addition, I/we grant permission to Williams College to provide access to (and therefore copying of) the thesis in electronic format via the Internet or other means of electronic transmission after a period of _____ years.

> Selecting this option allows the Libraries to transmit the thesis in electronic format via the Internet after a period of years. Once the restriction period has ended, this option permits worldwide access to the thesis, and copying of the electronic and hardcopy versions.

___ I/we grant permission to Williams College to maintain, provide access to, and provide copies of the thesis in hardcopy format only, for as long as I/we retain copyright.

> Selecting this option allows access to your work only from the hardcopy you submit for as long as you retain copyright in the work. Such access pertains to the entirety of your work, including any media that it incorporates. Selecting this option allows the Libraries to provide copies of the thesis to researchers in hardcopy form only, not in electronic format.

___ I/we grant permission to Williams College to maintain and to provide access to the thesis in hardcopy format only, for as long as I/we retain copyright.

> Selecting this option allows access to your work only from the hardcopy you submit for as long as you retain copyright in the work. Such access pertains to the entirety of your work, including any media that it incorporates. This option does NOT permit the Libraries to provide copies of the thesis to researchers.

Signature Redacted

Signed (student author) ______________________

Signature Redacted

Signed (faculty advisor) ______________________

Signed (2d advisor, if applicable) ______________________

Thesis title What should I study?: The short-term and long-term effectiveness of studying what you don't know

Date 5/18/2012

Library Use

Signature Redacted

Accepted By: ______________________ Date: 5/18/2012

rev. March 2010

WHAT SHOULD I STUDY?: THE SHORT-TERM AND LONG-TERM EFFECTIVENESS OF STUDYING WHAT YOU DON'T KNOW

by

KATRINA FLANAGAN

Nate Kornell, Advisor

A thesis submitted in partial fulfillment
of the requirements for the
Degree of Bachelor of Arts with Honors
in Psychology

WILLIAMS COLLEGE

Williamstown, MA

May 21, 2012

Acknowledgements

To my incredibly dedicated thesis advisor, Professor Nate Kornell:

You have been more than generous with your time, ideas, encouragement, and candy. Whenever I needed help you were there, with an average email response time of two and a half minutes. And even when you kicked me out of your office, I always knew I could come back soon. You were the best advisor I could have asked for. Thanks for everything, dude.

To my second reader, Professor Safa Zaki:

Thank you so much for the time and care you put into reading my thesis. Your comments have contributed both an extreme attention to detail and a fresh perspective to the writing process, and this paper would be much the worse for wear without them.

To my family:

Thank you for always encouraging me to see academic challenges not as obstacles but as opportunities. The relative serenity and composure (i.e., the relative infrequency of severe anxiety attacks) with which I was able to write this thesis would have been impossible without your love and support.

To my friends:

Thank you to each and every one of you for also writing theses. The ability to share the joys and pains of the thesis-writing process with people who are having a similar experience has been more valuable than I could have imagined. The hard work you devoted to your theses and the moral support you so willingly gave to me and each other were the positive influences I needed to write my own thesis.

Abstract

The present research explored learners' restudy choices. I start by identifying a study strategy that has been established both as one that students actually employ and as one that is in fact optimal for learning: restudy unknown items. I then move to an evaluation of the effectiveness of this strategy as compared to that of an alternative strategy (restudy known items) for both short- and long-term learning. I present a series of experiments designed to test the hypothesis that the strategy of restudying unknown items is good for short-term retention but bad for long-term retention. The results generally supported the hypothesis that restudying known items can be simultaneously a good learning strategy for short-term retention and not particularly beneficial for long-term retention. This challenges the assumption that if a study strategy is optimal for short-term learning, it will be optimal for long-term learning as well. Students and teachers should be aware of this and other errors in their beliefs about learning and memory, and they should use this awareness to improve the effectiveness of their studying and teaching.

What should I study?: The short-term and long-term effectiveness of studying what you don't know

Knowing how and what to study is an important part of effective learning. The resources a person can devote to studying are limited, and learners must make decisions about which subjects to devote their time to, how to distribute the time they spend studying each subject, and of what their studying will actually consist. The present research explored learners' restudy choices. I start by identifying a study strategy that has been established both as one that students actually employ and as one that is in fact optimal for learning. I then move to an evaluation of the effectiveness of this strategy as compared to that of an alternative strategy for both short- and long-term learning.

Envision a college student studying for the final exam in his biology class. He sits at a table in the library with an open textbook and notes spread out before him. As he starts to review the key terms, concepts, and theories he learned earlier in the semester, he encounters some items that he remembers from before and some that he does not. Naturally, he devotes what little time he has to studying the material that he could not recall, testing himself occasionally to monitor his progress and dropping items from study as they become familiar.

This anecdote illustrates one strategy that students use while studying: drop the items that they already know. This allows more time for students to study the items that they do not know, and thus to increase their chances of recalling those items. The reasoning behind this learning strategy is simple. If the student knows an item now, he believes this knowledge is likely to be available in the future (Koriat, Bjork, Sheffer, &

Bar, 2004). But if he does not know an item now, he views future recall as improbable. To achieve the goal of maximizing future recall, then, only studying items that one does not know appears to be the most useful strategy.

The widespread practice and effectiveness of making study choices based on metacognitive monitoring has been extensively researched. Nelson, Dunlosky, Graf, and Narens (1994) showed that people chose to restudy word pairs that they judged to be poorly learned. Subjects who were allowed to use this strategy performed better on a cued-recall test than both those who were assigned the objectively most difficult pairs to restudy and those who were assigned the most well learned pairs to restudy. Thus, people can and do use their metacognitions to allocate their restudy of items in the manner described above, and this allocation is effective (see also Son & Metcalfe, 2000; Son, 2004).

Thiede and Dunlosky (1999) propose a theory based on discrepancy reduction to explain why learners choose to study like this. During study, learners compare their current state of learning to their desired state of learning. The desired state of learning for any given item is knowledge or recall of that item, and if there is a discrepancy between that and the current state of learning (e.g., the item is not known/recalled), the learner will choose to restudy that item until the discrepancy has been reduced. If there is no discrepancy between the current and desired states of learning, the item will not be chosen for restudy. In addition, the items featuring the greatest gaps between current and desired learning will be chosen most frequently for restudy.

Kornell and Metcalfe (2006) have supported this strategy as one that people generally follow, with some exceptions. For some items, there may be a discrepancy between the current state of learning and the desired state of learning, yet subjects choose not to devote time to restudying those items. According to Metcalfe (2002), this is because some items are so difficult or poorly learned that they are out of the subject's region of proximal learning (RPL). This research confirms the idea from discrepancy reduction theory that learners drop known items from study. It also suggests that out of the items people do not know, however, the ones they choose to restudy are those they are closest to knowing (those that are in their region of proximal learning). The RPL framework claims not only that this is how people choose to study, but also that this is how people *should* choose to study, as it is the most effective way to learn. Indeed, participants in these experiments did better when their study choices (which were consistent with the study choices predicted by RPL) were honored than when they were not, providing support for the effectiveness of RPL.

A study by Kornell and Bjork (2008) demonstrated that the study-unknown-items approach to learning, though popular, is not always the best one. Participants in their experiments studied word pairs, and they showed a similar tendency to drop flashcards when they thought they knew the pairs. The authors showed that test performance actually suffered slightly when this strategy was adopted as opposed to a strategy of not dropping any flashcards, a result that directly opposes the idea that people make optimal decisions when choosing which items to restudy (Nelson et al., 1994; Thiede & Dunlosky, 1999; Metcalfe, 2002; Kornell & Metcalfe, 2006). In explaining these results, the authors indicated the likelihood that people often underestimate the value of additional study, and that this can lead to ineffective decision making and, ultimately, less learning.

In another study by Karpicke and Roediger (2007), a strategy of dropping recalled word pairs from study (in the form of tests with feedback) did not benefit retention on a delayed test when compared to a strategy of not dropping any pairs from study. In fact, the authors claimed that repeated retrieval of the recalled word pairs from memory was crucial to long-term retention, as final test performance was significantly worse for subjects who were allowed to drop the recalled pairs from study.

I intended to investigate further the benefits, and possibly costs, of this strategy of dropping known information and studying unknown information. An alternative is for learners to study known information and drop unknown information. It appears evident from the literature that few people take the latter approach to learning, but it may have hidden benefits, particularly for long-term learning.

Short-term benefits and long-term costs?

Let us return to our struggling college student. Based on this review of the literature, it appears he is taking the right approach to studying for his biology exam. When he goes to take it the next morning, he will probably remember not only the things he knew at the onset of his study session, but also the more difficult material that he spent the majority of his time studying later. He will probably do very well on the exam. But what if there is another exam a few months from now called the MCAT that requires knowledge of some of the material learned in that biology class? Would his study strategy result in optimal performance on this test as well? The natural assumption is that whatever strategy produces the best recall in the short term should be best in the long term as well, but this may not always be the case.

It has been theorized that learned items have a certain retrieval strength in memory (see Bjork & Bjork, 1992). This strength is important in determining whether an item will be correctly recalled in response to a given cue. If the retrieval strength of the item is higher than the threshold for recall it will be successfully recalled, but if it is lower it will not. If you are trying to learn a set of items, and your strategy is to drop items that you know (i.e., items above threshold), you will move some below-threshold items above threshold (by studying them), and end up with a group of items above threshold but only moderately well-learned. In contrast, if you use the alternative strategy of dropping unknown items and focusing on the ones you know (i.e., items above threshold), you will end up with some items that are very well-learned and well above threshold and others that languish below threshold. In the short term, the first strategy will be best for recall, because a larger number of items are above threshold. But forgetting will occur as time passes, and the memory strength of all items will decline. This will cause many of the moderately well-learned items to fall below threshold, but the items that were far above threshold in the second scenario will remain above threshold, assuming the rate of forgetting is the same for all items. As a result, even if studying unknown items seems beneficial in the short-term, studying known items may produce better recall after a delay (see Figure 1).

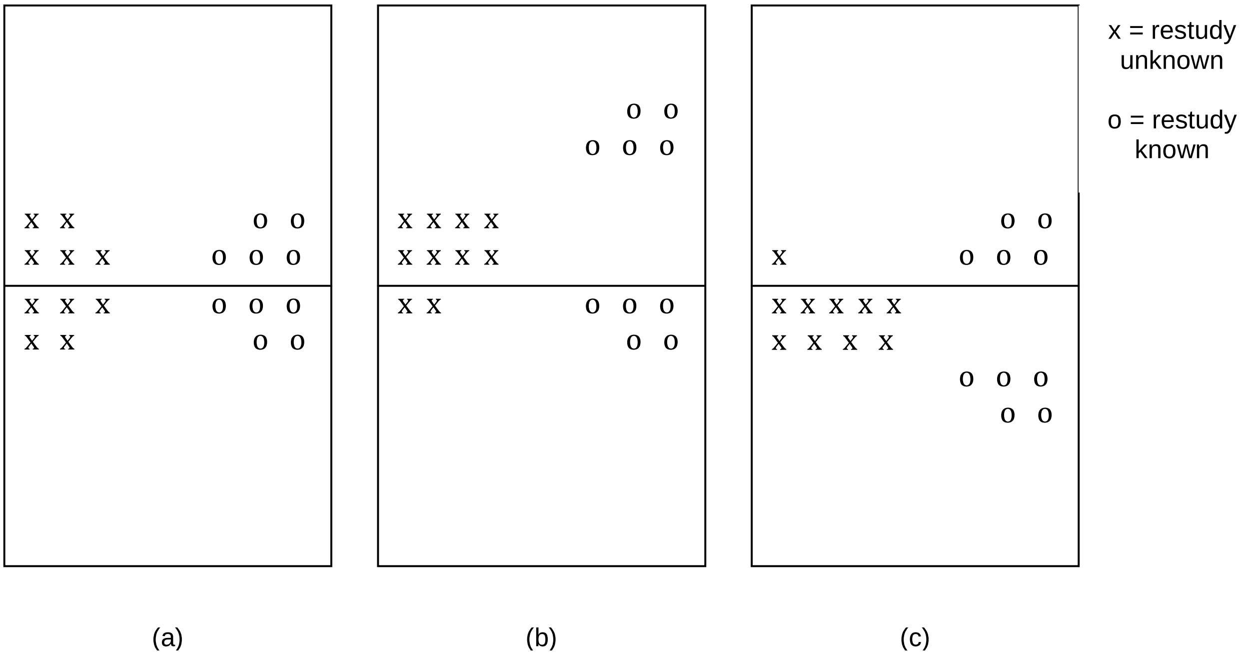

Figure 1. **Hypothesized model of item distribution for both the restudy unknown strategy (x's) and the restudy known strategy (o's). The line in the center of each panel is the recall threshold, and retrieval strength is indicated by an item's position on the y-axis. Shown in panel (a) is the distribution of items we would expect after the initial study phase. Panel (b) shows the distribution after some items have been restudied (this is also the distribution we would expect for an immediate test, when there has been little time for forgetting to occur). Panel (c) shows the distribution of items at a significant delay.**

This idea is supported by research comparing short- and long-term outcomes of studying in a different context. The finding that one strategy can be good for immediate tests but bad for delayed tests (and vice versa) has been established. For example, Roediger and Karpicke (2006a) found that participants who studied a set of items and then were tested on those items did worse on an immediate test, but better on a delayed test, than those who studied the items twice without being tested on them. Similarly, Wheeler, Ewers and Buonanno (2003) found that if items were studied then repeatedly tested they were more likely to be recalled on a delayed test than items that were repeatedly studied and not tested. These authors concluded that tests are instrumental in preventing, or at least slowing the rate of, forgetting.

There also appears to be theoretical and empirical support for the claim that there are indeed circumstances under which studying known items in particular might be better in the long term. Kornell, Bjork, and Garcia (2011) propose an explanation for the beneficial effects of tests without feedback as a better tool for long-term retention than simply studying. They explain apparently slower rates of forgetting for items that have been tested by showing that the distribution of items by their memory strength becomes bifurcated when tested but not when studied. The bifurcation occurs because items that can be successfully retrieved when tested are considerably strengthened in memory and move way above the threshold for recall, but the items that cannot be retrieved remain weak and far below threshold. Studying increases item strength by a lesser amount than tests (see Roediger & Karpicke, 2006b), but the strengthening happens for all items that are restudied. When forgetting occurs across all items, more of the studied items fall below threshold than the tested items, because the memory strengths of the tested items above threshold are greater than those of the studied items above threshold.

I predicted a similar bifurcation of items when the “study known” strategy is used. By restudying known items, you move those items even farther above threshold, and you allow the other items to stay below threshold. Thus, what happens to this group of items is similar to what happens to the group of items that is tested without feedback: the items that were initially known end up far above threshold and the items that were initially unknown remain below threshold.

This interpretation of slower forgetting provides theoretical support for my idea that the ideal strategy for long-term retention (when resources are limited) might be to study in such a way that causes a bifurcation of the item distribution. Halamish and Bjork (2011) provide additional support for this idea by showing the benefits of testing (and thus a bifurcated distribution of items) not only after a long delay, but also when the final test is particularly difficult. It appears, then, that the strategy of choosing to study the items you know and drop the items you do not might be effective in some situations. While I anticipated that people in my experiments would continue to choose to study in the manner that RPL and discrepancy reduction predict, I expected my results to contradict a separate claim of these theories: that the study decisions people make are optimal.

Which of these study strategies are best for learning? The studies described suggest that it really depends on how and when learning is going to be evaluated. If there is a threshold amount of memory strength for item recall, studying to push as many items as possible past that threshold is good for performance in the short run, but not in the long run. This may only be true for learning that is measurable, however, as there is likely some benefit to keeping items relatively near threshold. The “drop known” strategy could result in several items below threshold after a delay, but it might be relatively easy to relearn those items. In this paper, I will stick to the benefits and costs of each strategy in the context of measurable learning only.

In the present research, I tested the hypothesis that the “drop known items” strategy, common among learners, while beneficial for short-term retention, is a relatively poor approach when the goal is long-term retention. I hypothesized that subjects who followed this strategy would outperform on an immediate test those who studied items they knew, but that performance on a delayed test would indicate the superiority of the latter approach.

The results of a recent study done by Tullis and Benjamin (in press) provide support for this hypothesis. The authors were investigating the effectiveness of learners' study choices as a function of age. Participants of all ages chose the study strategy previously discussed of dropping the items they knew already and restudying the ones they did not know. Younger (college age) learners did better on the final test when their study choices were honored than when they were dishonored. But older learners (60-84 years old) actually did better on the test when their study choices were dishonored rather than honored. The strategy of dropping the items they already knew did not work for them; in fact, the opposite promoted the most learning.

The parallel to my research is this. Older learners forget items at a faster rate than younger learners. The strategy they used when choosing which items to restudy resulted in a bunch of moderately well-learned items that were just above the recall threshold, but by the time they took the test ten minutes later, many of those items had fallen below threshold because of the accelerated rate of forgetting that occurs among older learners. The older learners, then, are like the learners in my study who take the test two weeks later, after much forgetting has occurred. For younger learners, the items remain above threshold long enough for the "drop known" strategy to work well for them; they parallel the subjects in my study who take the final test immediately after studying. Thus, the results reported by Tullis and Benjamin (in press) are encouraging for my own research as well.

In the rest of this paper, I present a series of experiments designed to test the hypothesis that the strategy of restudying unknown items is good for short-term retention but bad for long-term retention. In the first experiment, easy and difficult word pairs were used as a proxy for known and unknown items, and the effectiveness of the "restudy unknown" strategy was compared to that of the "restudy known" strategy for both immediate and delayed tests.

Experiment 1

Method

Participants

Participants were recruited from Amazon's Mechanical Turk, a website on which people can sign up to do miscellaneous tasks for pay. Participants were offered $1.00 for completing Part One of the experiment, which took about 20 minutes, and another $1.00 for completing Part Two, which took about five minutes. Of the 53 participants included in the analysis, 35 were women, and ages ranged from 19 to 65. All participants lived in the United States and were fluent in English.

One hundred participants were recruited for this experiment. People were excluded from analysis if they did not complete the experiment, if they had major technical issues during the experiment, or if they indicated in a post-experimental questionnaire that they were familiar with the experimental materials prior to participating. Part Two of the experiment was required of both immediate and delayed participants. Although only the immediate test was included in the analysis for the immediate participants, they were asked to do Part Two so that any self-selection problems would be present in both delay conditions. Of the 53 participants included in the final analysis, 31 were in the delayed condition and 22 were in the immediate condition.

Materials

The stimuli were 60 word pairs, half of which were easy (e.g., FIRM:GRIP) and half of which were difficult (e.g., MEEK:CLOTHES). The easy pairs were taken from Nelson, McEvoy, and Schreiber (1998) and had forward association strengths of .10. This means that when people were shown the first word in each pair (e.g., FIRM), 10% of those people identified the second word in the pair (e.g., GRIP) as the first thing that came to mind. The difficult pairs were formed by randomly matching up cues and targets from word pairs that had forward association strengths of .11, a process designed to create non-associates.

Design

The experiment had a 2 x 2 x 2 mixed design, with item difficulty and restudy manipulated within subjects and test delay manipulated between subjects. Participants took the final test at either a five minute or one week delay.

Procedure

The experiment was conducted online. There were three stages: initial study, restudy, and final test (see Figure 2). Participants in the immediate test condition were also asked to take the delayed test one week later (in an effort to avoid the influence of self-selection on the results as discussed earlier), but their performance on that test was not included in the analysis.

During the initial study phase, participants studied each word pair once. The cue and target of each pair were shown on the screen until the participant decided to continue on to the next pair. After studying all of the pairs, participants proceeded to the restudy phase.

Half of the pairs were randomly selected for restudy. During the restudy phase, participants restudied this selection of pairs twice. The restudy phase consisted of test trials with feedback for each item. Participants were presented with the first word in each pair and asked to fill in the second word. Participants were given as much time as needed to enter an answer. Once they did, the correct answer was displayed on the screen for two seconds, and then the next trial began.

After the restudy phase, the participants in the delayed test condition were thanked for their participation and told they might be contacted for a follow-up experiment. Participants in the immediate test condition played a game called Asteroids for three minutes as a distractor task and then were given a test without feedback on all 60 word pairs. All participants were emailed one week later and asked to take the delayed test on all of the word pairs. Only participants who participated in the second session of the experiment and completed this final test were included in the analysis.

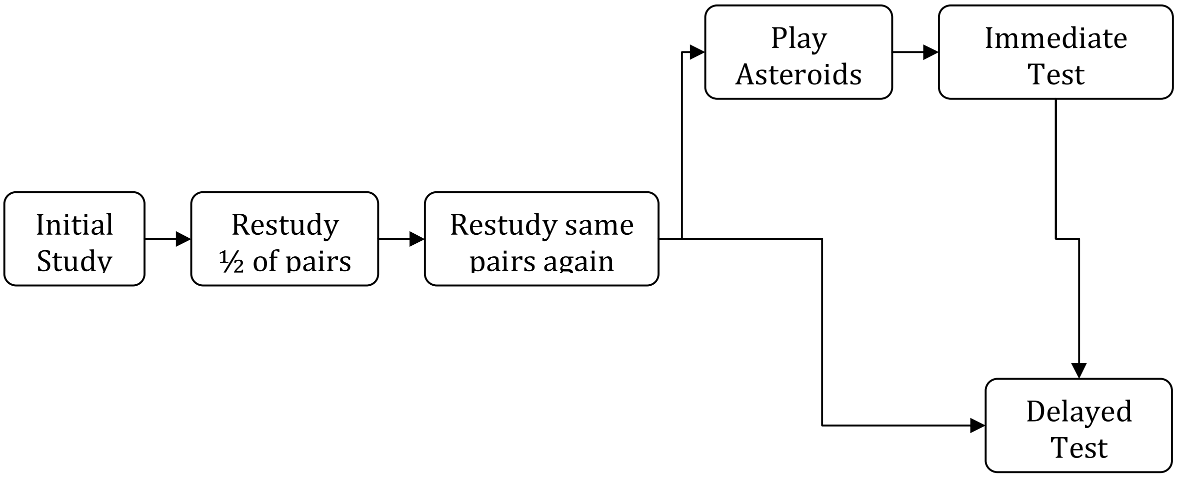

Figure 2. **Procedure for Experiment 1.**

Results and Discussion

For the analysis, a new "study strategy" variable was formed out of the four within subjects conditions: easy/restudied items, easy/not restudied items, hard/restudied items, and hard/not restudied items. One level of this new variable was "hard focus," which was designed to correspond to the strategy, described in detail in the introduction, of dropping items that you already know and restudying items that you do not. The hard focus condition was created by combining hard/restudied items with easy/not restudied items. The other level was "easy focus," which was just the opposite: a strategy of dropping items that you do not know and choosing to restudy those you do. The easy focus condition was created by combining easy/restudied items with hard/not restudied items.

To test the hypothesis that there would be an interaction between study strategy and test delay, a two-way mixed model ANOVA was conducted (see Figure 3). The interaction was not significant, $F(1,51) = .951$, $p = .334$. There was a significant main effect of study strategy, $F(1,51) = 20.450$, $p < .05$, with the easy focus condition outperforming the hard focus condition on the final test. There was also a significant main effect of test delay, $F(1,51) = 13.665$, $p < .05$, with the immediate condition unsurprisingly outperforming the delayed condition. A paired samples t-test was also conducted to compare performance in the easy focus and hard focus conditions on the delayed test. The difference between the two was significant, $t(30) = 2.986$, $p < .05$. (See Table 1 for means.)

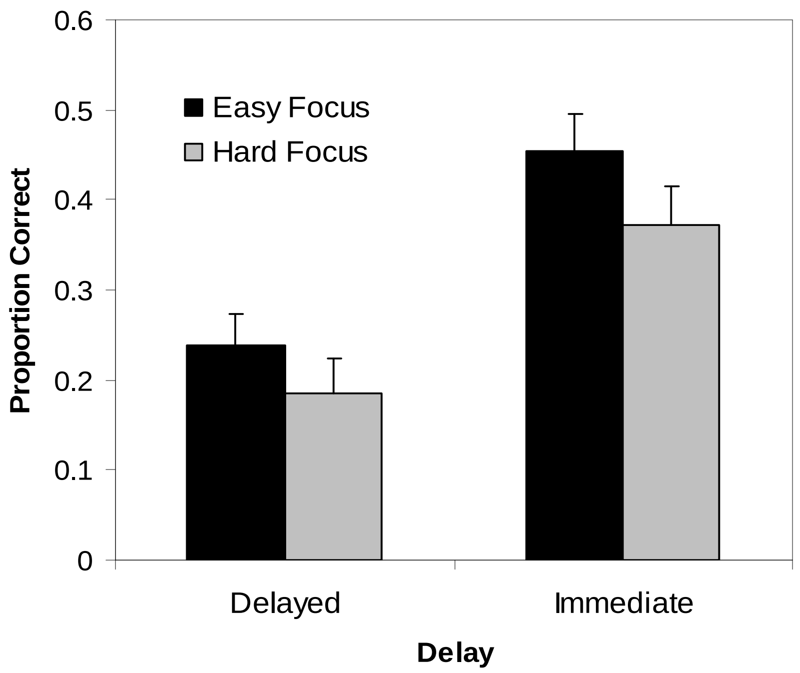

Figure 3. **Proportion of word pairs correctly recalled on the final test in Experiment 1 as a function of test delay and study strategy (easy focus or hard focus).**

These results were generally unsupportive of my hypothesis, as the effectiveness of the two different study strategies did not vary according to the delay between the study and test phases. And as Figure 3 shows, somewhat counter-intuitively, the easy focus strategy resulted in better performance on both the immediate and delayed tests. This could be due to the fact that the study strategies as I have defined them in this experiment do not match up precisely with those outlined in the introduction of this paper.

In real life, learners base their restudy decisions on whether they know each item or not. In this experiment, I used item difficulty as a proxy for this knowledge, making the assumption that easy items correspond to known items and hard items correspond to unknown items. In fact, the fairly low initial test accuracy achieved by the participants in my experiment indicates that the easy items may not have been so easy. This could account for the somewhat surprising finding that the easy focus condition outperformed the hard focus condition. The "easy" focus condition was actually made up of many items that were unknown to learners, but they were easier and thus closer to being learned than those in the hard focus condition. The items in the hard focus condition were likely so difficult that restudying them was not sufficient to push many above the recall threshold for the final test.

Experiment 2

In this experiment I moved to a more precise measure of item knowledge than item difficulty.

Method

Participants

Forty participants were recruited from Amazon's Mechanical Turk. Eighteen were excluded from analysis for not completing the experiment, technical difficulties, or self-reported familiarity with the experimental materials. Of the 22 participants included in the analysis, 18 were female, and ages ranged from 19 to 62. All participants lived in the United States and were fluent in English. Part One of the experiment took about 25 minutes, and Part Two took about five minutes. Participants were paid $1.50 for completing Part One and another $1.00 for completing Part Two. Eleven participants were randomly assigned to the immediate condition and eleven to the delayed condition.

Materials

The stimuli for this experiment were taken from the same set of associates as those in Experiment 1 (Nelson et al., 1998). For this experiment the 30 easy word pairs from Experiment 1 (which have forward association strengths of .10) were used, as well as 30 new word pairs with forward association strengths of .11 (e.g., AGONY:DEFEAT).

Design

The design for Experiment 2 is the same as that for Experiment 1 with one exception. Instead of manipulating item difficulty (easy vs. hard), I manipulated item knowledge (known vs. unknown). This was done in order to test my hypothesis (which is described in detail in the introduction of this paper) more precisely. Items were assigned to the known and unknown conditions based on each participant's performance during the initial test phase.

Procedure

The procedure in Experiment 2 was almost identical to that in Experiment 1, but a phase was added between the initial study phase and the restudy phase. After studying all 60 word pairs during the initial study phase, participants took an initial test without feedback on all of the word pairs. Performance on this initial test was used to divide items into the known and unknown conditions. That is, if a person correctly recalled an item on this initial test that item was placed into the known condition, and if he or she failed to recall the item it was placed into the unknown condition. After this initial test phase, the restudy and final test phases were the same as Experiment 1.

Results and Discussion

As in Experiment 1, a new "study strategy" variable was created to be used in the analysis. The inputs for this variable were the four conditions that were within subjects: known/restudied items, known/not restudied items, unknown/restudied items, and unknown/not restudied items. There were two levels of this new variable: "restudy known" corresponds to the strategy described in the introduction of restudying items you already know and dropping those you do not, and "restudy unknown" corresponds to the opposite strategy of restudying items you do not know and dropping those you do. Performance in the known/restudied and unknown/not restudied conditions was averaged to create the restudy known condition, and performance in the unknown/restudied and known/not restudied conditions was averaged to create the restudy unknown condition.

A two-way mixed model ANOVA was conducted to test the hypothesis that there would be an interaction between study strategy and test delay (see Figure 4). The interaction was significant, $F(1,21) = 31.784$, $p < .05$. There was also a significant main effect of study strategy, $F(1,21) = 65.688$, $p < .05$, with the restudy unknown condition generally outperforming the restudy known condition. The main effect of test delay was significant as well, $F(1,21) = 138.828$, $p < .05$, with participants doing better on the immediate test than on the delayed test. A paired samples t-test was also conducted to compare performance in the restudy known and restudy unknown conditions on the delayed test. The results of this test were not significant, $t(10) = -1.573$, $p > .05$. (See Table 1 for means.)

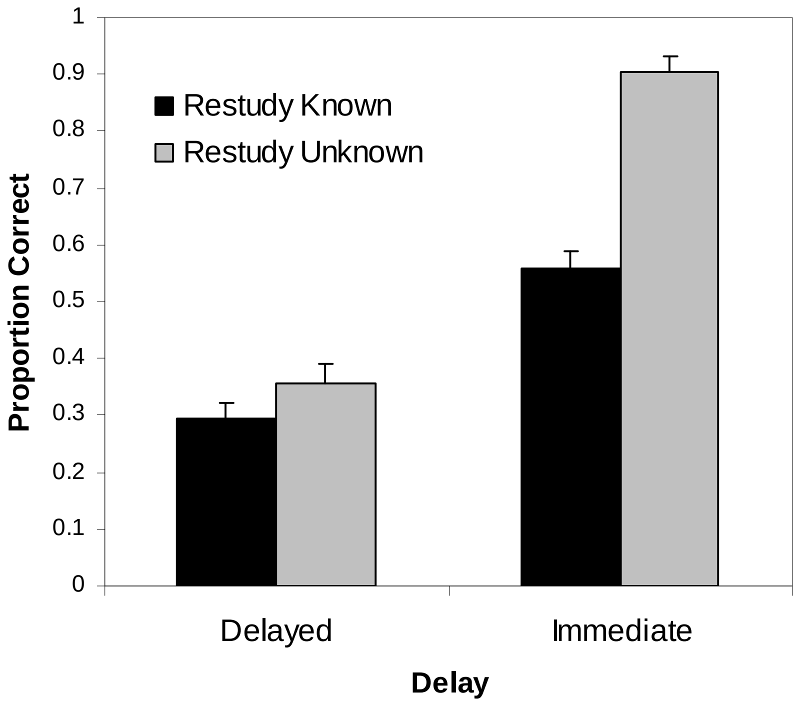

Figure 4. **Proportion of word pairs correctly recalled on the final test in Experiment 2 as a function of test delay and study strategy (restudy known or restudy unknown).**

These results supported the hypothesis that the effectiveness of each study strategy depends on when learners are tested. While the restudy unknown strategy resulted in much better performance on the immediate test than the restudy known strategy, the performance gap narrowed for the delayed test, as can be seen in Figure 4. This means that more forgetting occurred when using the restudy unknown strategy than when using the restudy known strategy. So while there were more items above the recall threshold in the restudy unknown condition to begin with, only slightly more than one third of those items remained above threshold after one week. Conversely, while there were initially fewer items above threshold in the restudy known condition, more than half remained there after a week had passed.

This is consistent with the theory that restudying items you already know increases the memory strength of those items so that they end up far above threshold and thus are not forgotten as quickly.

Experiment 3

In Experiment 3, I increased the delay from one to two weeks to see whether the restudy known strategy could actually be better for recall than restudy unknown. I expected an increase in the delay to have this effect because it would allow for more forgetting, which would have a larger effect on accuracy in the restudy unknown condition than in the restudy known condition. That is, the relative resiliency of the restudy known items that are above threshold will result in them staying above threshold for longer than the items in the restudy unknown condition.

Method

Participants

Fifty-eight participants were recruited from Amazon's Mechanical Turk. Of these, 26 were excluded from analysis for not completing the experiment, serious technical problems, or self-reported familiarity with the materials. Of the 32 included in the analysis, 22 were female, and ages ranged from 19 to 58. All participants lived in the United States, and all were fluent in English. Part One of the experiment took about 25 minutes, and Part Two took about five minutes. Participants were paid $1.00 for completing Part One and another $1.00 for completing Part Two. Fifteen of the participants were randomly assigned to the delayed condition, and the other 17 were in the immediate condition.

Materials

In this experiment, my desire was to both simulate more realistic study conditions and create a bigger gap between the known and unknown items. I thus went back to the same set of word pairs used in Experiment 1, where half the pairs were relatively easy to learn and half were relatively difficult to learn. Item difficulty was not included as an independent variable in the data analysis, however.

Design

The design was the same as that in Experiment 2.

Procedure

The procedure was the same as that in Experiment 2. The only change was in the length of time between parts one and two of the experiment. In Experiment 2, the delay before the final test was 5 minutes in the immediate condition and one week in the delayed condition. In this experiment, the delay before the final test was two weeks in the delayed condition. Everything else was the same.

Results and Discussion

To test the hypothesis that there would be an interaction between study strategy and delay, a two-way mixed model ANOVA was conducted (see Figure 5). There was a significant interaction, $F(1,31) = 15.206$, $p < .05$, with the restudy unknown condition outperforming the restudy known condition on the immediate test but not on the delayed test. As in the first two experiments, there was a significant main effect of both study strategy and test delay, $F(1,31) = 11.344$, $p < .05$, and $F(1,31) = 115.314$, $p < .05$, respectively. A paired samples t-test was also conducted to compare performance in the restudy known and restudy unknown conditions on the delayed test. The results of this test were not significant, $t(14) = .413$, $p > .05$. (See Table 1 for means.)

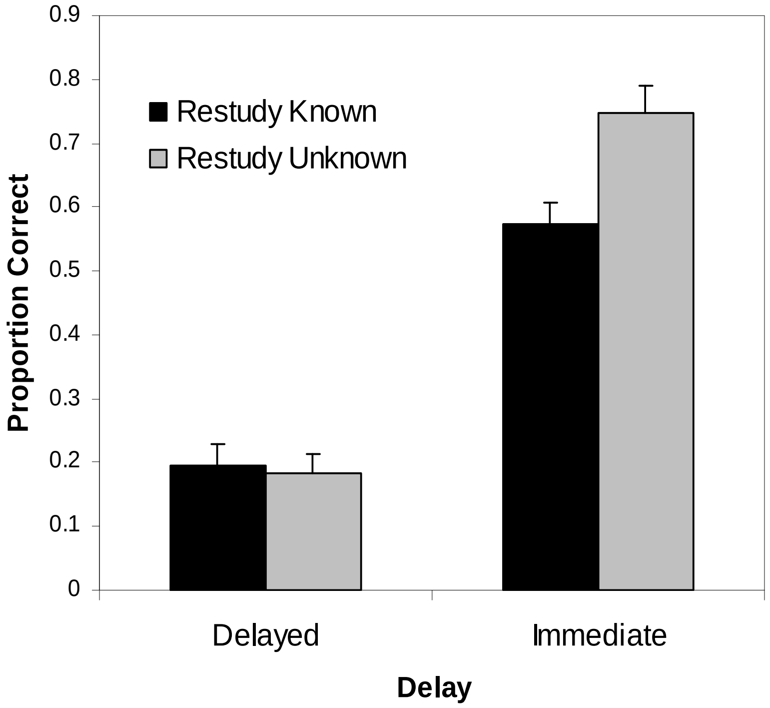

Figure 5. **Proportion of word pairs correctly recalled on the final test in Experiment 3 as a function of test delay and study strategy (restudy known or restudy unknown).**

As Figure 5 shows, these results confirm the finding in Experiment 2 that the restudy unknown strategy appears to be associated with more forgetting (57 percentage point drop in accuracy) than the restudy known strategy (37 percentage point drop in accuracy). They also provide support for my hypothesis that restudy known can actually be the optimal strategy for long-term learning, as this strategy outperformed restudy unknown on the delayed test (although the difference was not significant).

Experiment 4

In this experiment participants were asked to make study choices for themselves, rather than assuming that they would choose one way or another based on item knowledge. This was done in an attempt to draw a more concrete connection between learners' actions and the consequences of those actions.

Method

Participants

One hundred twenty-seven participants were recruited from Amazon's Mechanical Turk. Of these, 92 were excluded from analysis for not completing the experiment, serious technical problems, or self-reported familiarity with the materials. (The number of excluded participants for this experiment is high because the program failed to record essential information for many of them.) Of the 35 included in the analysis, 23 were female, and ages ranged from 18 to 57. All participants lived in the United States, and all were fluent in English. Participants were paid $1.00 for completing Part One of the experiment, which took about 25 minutes, and another $1.00 for completing Part Two, which took about 5 minutes. Nineteen participants were randomly assigned to the delayed condition; the other 16 were in the immediate condition.

Materials

The materials used in this experiment were the same as those used in Experiments 1 and 3.

Design

The design of this experiment was a 2 x 2 mixed design, with test delay as the between subjects factor, as in the first three experiments. The within subjects factor was whether a given restudy choice was honored or dishonored.

Procedure

The procedure was similar to that of Experiments 2 and 3, but after the initial study and initial test phases, we added a study choice phase. This was followed by the restudy and final test phases. During this study choice phase, each cue from every word pair was displayed simultaneously, and participants were instructed to choose half of the pairs for restudy. They were not allowed to continue until they had chosen exactly half. Then they proceeded to the restudy phase as in the other experiments, where half of their study choices were honored and the other half were dishonored. After this, participants in the immediate condition took a final test and all participants took a test after a two week delay, as in the previous experiments.

Results and Discussion

To test the hypothesis that there would be an interaction between test delay and whether or not a participant's study choice was honored, a two-way mixed model ANOVA was conducted (see Figure 6). The interaction was not significant, $F(1,35) = .613$, $p > .05$. There was no significant main effect of honoring participants' study choices either, $F(1,35) = .154$, $p > .05$, with participants performing only negligibly better in the honor condition than in the dishonor condition overall. As expected, there was a significant main effect of test delay on performance, $F(1,35) = 66.277$, $p < .05$. A paired samples t-test was also conducted to compare performance in the dishonor and honor conditions on the delayed test. The results of this test were not significant, $t(18) = .371$, $p > .05$. (See Table 1 for means.)

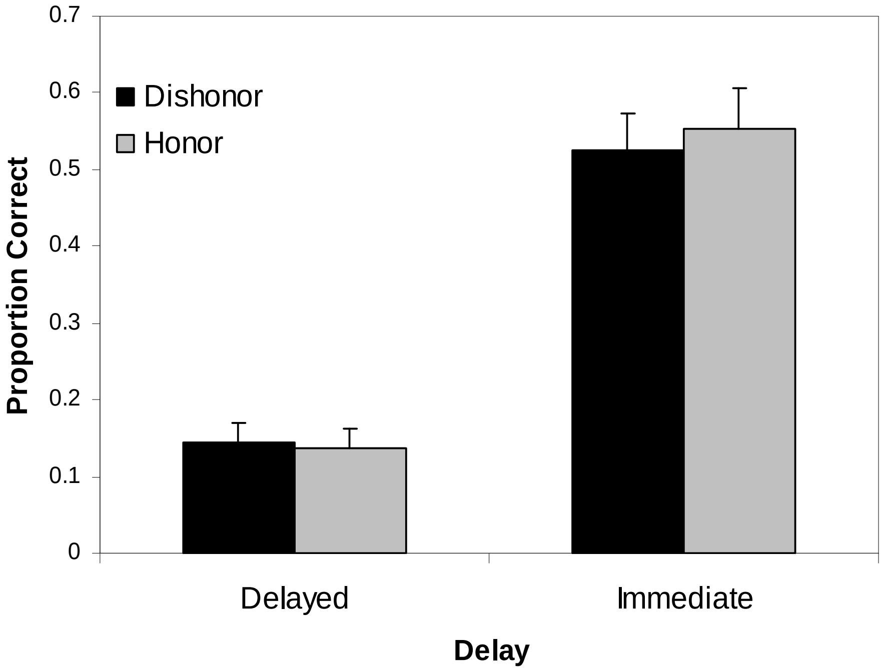

Figure 6. **Proportion of word pairs correctly recalled on the final test in Experiment 4 as a function of test delay and whether or participants' study choices were honored.**

If we use these data to perform the same analysis as in Experiments 2 and 3, however, we do see a significant crossover interaction between study strategy and test delay, $F(1,35) = 18.595$, $p < .05$ (see Figure 7). There was also a significant main effect of study strategy, $F(1,35) = 5.342$, $p < .05$, with "restudy unknown" outperforming "restudy known" across delay conditions. The results of a paired samples t-test comparing performance in the restudy known and restudy unknown conditions on the delayed test were significant, $t(18) = 2.094$, $p = .05$. (See Table 1 for means.) This secondary analysis confirms the results of Experiment 3 and supports the hypothesis that the restudy known strategy can actually be better for retention after a sufficient delay.

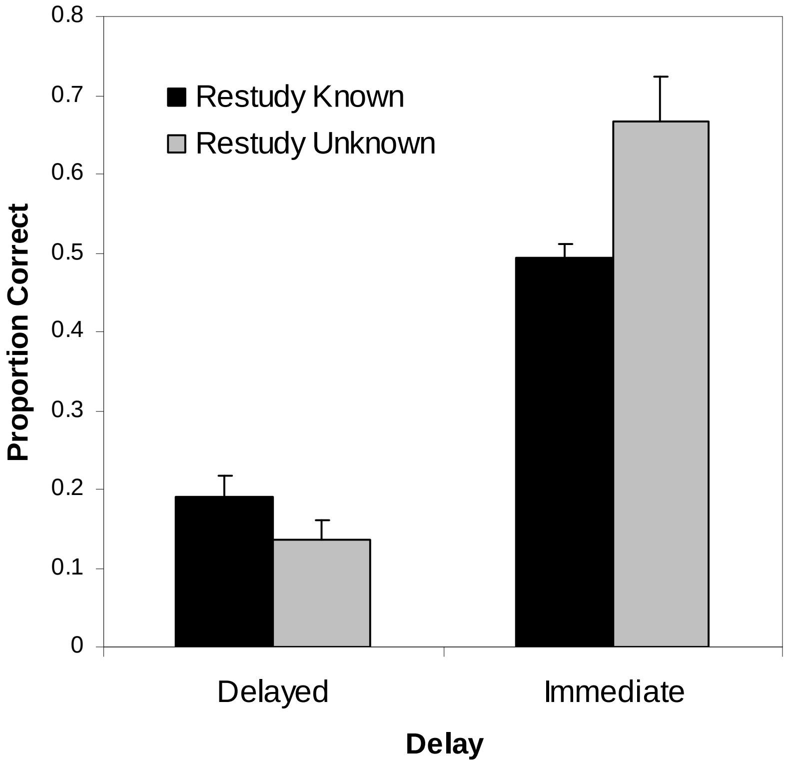

***Figure 7*. Proportion of word pairs correctly recalled on the final test in Experiment 4 as a function of test delay and study strategy (restudy known or restudy unknown).**

These results indicate some inconsistency between our participants' actual restudy choices and those we anticipated they would make. As Figures 6 and 7 show, the honor and dishonor conditions did not correspond to the restudy unknown and restudy known conditions. It appears the learners in our experiment were not choosing to drop items they knew and restudy items they did not. As addressed in the introduction of this paper, however, previous research shows that learners do choose to study this way (Nelson et al., 1994; Son and Metcalfe, 2000; Son, 2004). This inconsistency could be due to any of multiple factors interfering with metacognitive judgments. To begin with, restudy choices were presented at the end of the initial test phase. Participants may have had some difficulty remembering which items they correctly recalled during the initial test and so may have been unable to use this knowledge in making their decisions. In addition to this, there were 60 word pairs that participants were asked to learn. Being presented with all 60 pairs at once at the end of the initial test phase could have been overwhelming, and it is also very different from traditional restudy in which learners usually make decisions about individual items immediately after they have tested themselves.

Experiment 5

In this experiment we attempted to reduce the negative influence of the above-mentioned factors on metacognitions so that our participants would be better equipped to make their restudy decisions.

Method

Participants

Ninety-seven participants were recruited from Amazon’s Mechanical Turk. Of these, 29 were included in the analysis. Participants were excluded if they experienced technical problems, if they reported being familiar with the experiment or materials, or if they failed to complete both parts of the experiment. Seventeen of the participants were female, and ages ranged from 18 to 60 years old. All participants reported living in the United States and being fluent in English. Part one of the experiment took about 15 minutes, and part two took about five minutes. Participants were paid $1.00 for completing part one and another $1.00 for completing part two. Fourteen of the participants were assigned to the delayed condition and fifteen to the immediate condition.

Materials

Thirty-two of the 60 word pairs used in Experiments 1, 3, and 4 were randomly chosen for use in this experiment. With fewer word pairs, participants were making their restudy decisions closer in time, and I expected the choices to be less overwhelming and their decisions more closely aligned with actual study behaviors.

Design

The design of this experiment was the same as that of Experiment 4.

Procedure

The experiment proceeded in the same way as Experiment 4, with two exceptions. First, as mentioned above, participants were asked to learn fewer word pairs in this experiment. Second, when participants were presented with restudy choices, they were told which pairs they had recalled incorrectly on the initial test (these items were displayed in red text). This information was provided to fill in any gaps where they could not remember whether or not they had recalled an item correctly. If learners use prior recall to decide whether or not to restudy items, this manipulation should have assisted them in making those decisions. If they prefer to base their restudy decisions on something else, this manipulation should have had little or no effect on participants' choices.

Results and Discussion

A two-way mixed model ANOVA was conducted to test the hypothesis that there would be an interaction between test delay and honor condition (see Figure 8). The interaction was significant, $F(1,27) = 4.275$, $p < .05$. There was a significant main effect of both test delay and honor, $F(1,27) = 59.476$, $p < .05$, and $F(1,27) = 6.302$, $p < .05$, respectively. A paired samples t-test was conducted to compare performance in the dishonor and honor conditions on the delayed test. The results of this test were not significant, $t(13) = -.308$, $p > .05$. (See Table 1 for means).

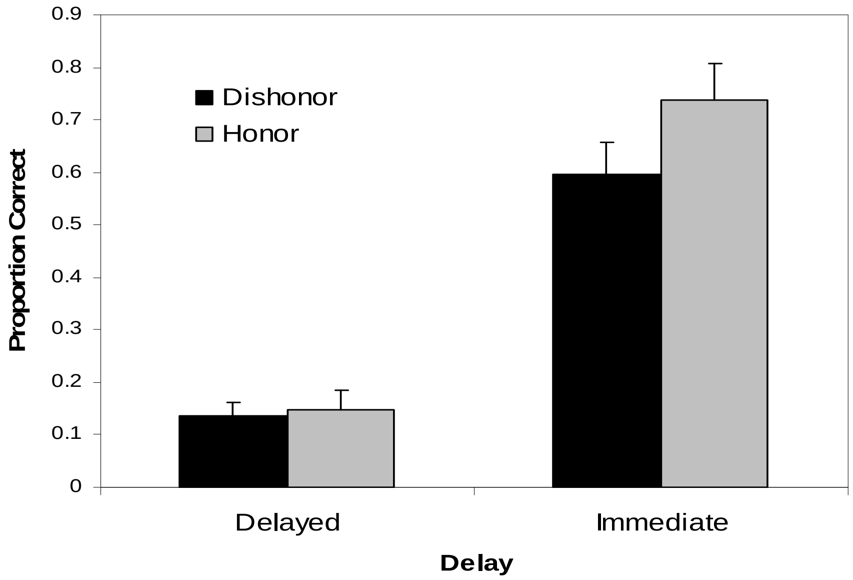

Figure 8. **Proportion of word pairs correctly recalled on the final test in Experiment 5 as a function of test delay and whether or not participants' study choices were honored.**

The same secondary analysis as in Experiment 3 was performed, comparing "restudy known" and "restudy unknown" rather than honor and dishonor as a function of test delay (see Figure 9). A two-way mixed model ANOVA produced a significant interaction, $F(1,25) = 16.464$, $p < .05$. There was also a significant main effect of study strategy, $F(1,25) = 10.321$, $p < .05$, with "restudy unknown" outperforming "restudy known." A paired samples t-test was conducted to compare the performance of the restudy known and restudy unknown strategies on the delayed test. The results were not significant, $t(13) = .573$, $p > .05$.

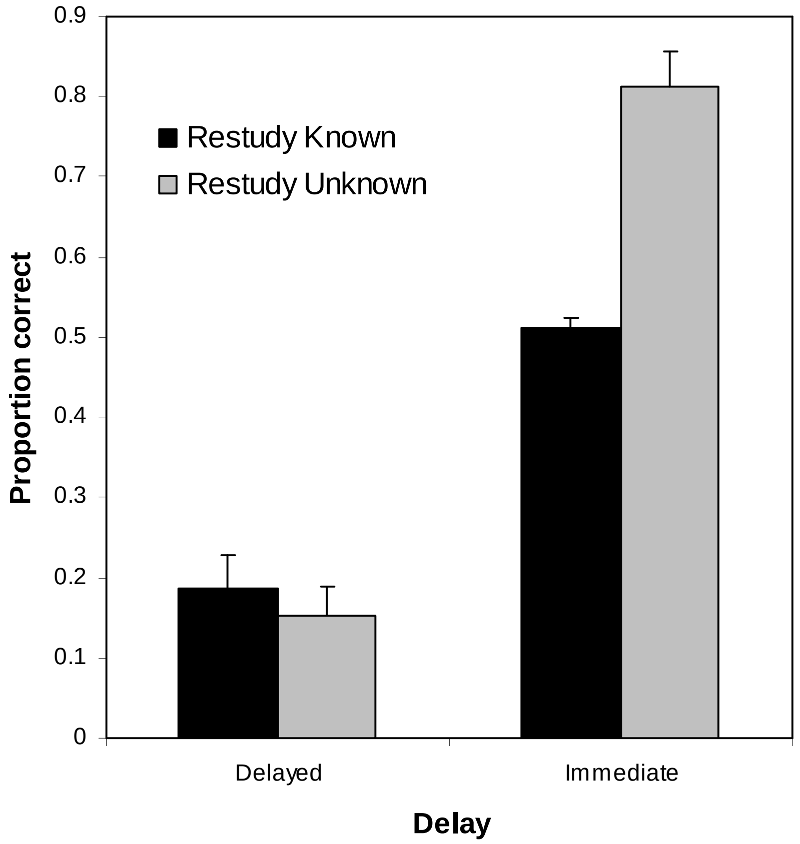

Figure 9. **Proportion of word pairs correctly recalled on the final test in Experiment 5 as a function of test delay and study strategy (restudy known or restudy unknown).**

As Figure 8 shows, these results demonstrate that more forgetting occurred in the honor condition over the two-week delay than in the dishonor condition. (A comparison to Figure 9 also shows that the honor and dishonor conditions now correspond more closely to the restudy unknown and restudy known conditions used in previous experiments.) This finding supports my hypothesis that restudying known items increases their strength in memory so that they are not as quickly forgotten as the unknown items are when they are restudied. The results also provide a tangible connection between learners' actual study decisions and the consequences of those decisions.

Because the results are certainly stronger than those in Experiment 4 and more in line with my hypothesis, the changes that were made to make the restudy decision-making process easier likely had an effect on participants' decisions. All the same, the results are still weaker than those from Experiment 3. The most likely explanation for this is that learners do not actually choose exclusively unknown items to restudy, and thus when they were allowed to make the decisions it may not correspond exactly to the restudy unknown condition from Experiments 2 and 3. And when I looked at the data to see which types of items participants chose to restudy, I saw why their decisions did deviate somewhat from those that were expected.

Recall that in this experiment there was a measure of participants' knowledge of each item: accuracy on the initial test that occurs right after the initial study phase. If participants were adhering strictly to the restudy unknown strategy, they would only choose to restudy the items that they got wrong on this initial test. So if we look at the initial test accuracy for items that participants chose to restudy, it should be zero percent or close to it. Likewise, if we look at the initial test accuracy for the items which participants chose not to restudy, it should be 100 percent or close to it. What I actually saw was about 20 percent accuracy for the items chosen for restudy and about 80 percent for the items not chosen. This shows that learners in this experiment did not actually drop all known items and exclusively restudy unknown items, implying that learners in general may not adhere strictly to this strategy.

Upon closer examination, however, I found that each one of our participants followed the restudy unknown strategy to a tee. The problem derives from the constraint placed on participants of choosing exactly half of the items for restudy. (I believed this constraint necessary, because if it were not present people would choose different numbers of items to restudy. This could lead to advantages for some participants and disadvantages for others on the final test.) Participants rarely got exactly half of the items correct on the initial test, but when they did they chose to restudy all of the ones they got wrong and none of the ones they got right. More often, participants got more or less than half correct on the initial test. When they answered more than half correctly, they chose to restudy all of the ones they got wrong, plus a few that they got right because they had to meet our quota. When they answered fewer than half of the items correctly, they dropped all of the items they got right, plus a few that they got wrong because there was no room for them in the restudy group.

If we such a constraint had not been placed on participants' choices, it seems probable they would have conformed precisely to the restudy unknown strategy that the vast majority of learners use, in which case the interaction likely would have been more pronounced. As it stands, this experiment serves to support an explicit connection between how learners choose to study and what the consequences are for short- and long-term learning.

General Discussion

My results have generally supported the hypothesis that dropping known items from restudy can be simultaneously a good learning strategy for short-term retention and not particularly beneficial for long-term retention. Further, they suggest the validity of the bifurcation model introduced at the beginning of this paper explaining why this should happen (Kornell, Bjork & Garcia, 2011). As this model predicted, when participants used the counterintuitive "restudy known" strategy (or its equivalent in Experiments 4 and 5, the dishonor condition), they did not forget word pairs as quickly as when they used the more typical “restudy unknown” (honor condition) strategy. This is probably surprising to most people (including learning and memory researchers), because it challenges the assumption that if a study strategy is optimal for short-term learning it is also optimal for long-term learning.

This finding can be added to the ever-increasing list of inconsistencies between how we think memory works and how memory actually works. Previous research is filled with examples of these errors: overconfidence in one’s memory (Metcalfe, 1998; Koriat & Bjork, 2005), underconfidence in one’s memory (Butler, Karpicke, & Roediger, 2008; Koriat, Sheffer, & Ma’ayan, 2002), overestimation of the ability to detect change (Levin, Momen, Drivdahl, & Simons, 2000), underestimation of the beneficial effects of studying (Kornell & Bjork, 2009), and so on. The process of identifying and reporting such metacognitive errors is an interesting pursuit for academics, but it is also important in that it can lead to the correction of these errors in the metacognitive beliefs of students.

Students use metacognition to make study choices all the time, so the consequences of metacognitive errors for learning can be severe. As addressed in the introduction of this paper, the resources that students have for learning are limited, and many of their study choices deal with how to allocate these resources most effectively. When learners base their decisions on metacognitive beliefs that are incorrect, they risk losing some of that effectiveness during study.

Similar decisions must be made in the classroom by teachers, decisions that are very much influenced by the teachers’ beliefs about how memory works. They must choose the amount of time they are going to spend on each subject, and they need to decide what, if any, material they are going to re-teach. Some teachers like to spend more time on those subjects in which the students are struggling: If the students appear to be doing well in math but poorly in science, they devote more time to science. Other teachers are concerned with teaching the students as much material as possible, regardless of their ability to retain the information; summaries of material that has already been taught are seen as a waste of time that could be spent on introducing new material.

Valverde and Schmidt (1997) show that this is in fact a common difference between how students in the United States learn and how students learn in other countries. It seems teachers in the U.S. tend to prefer a curriculum that is a mile wide but only an inch deep, and cross-country comparisons of academic achievement indicate this might be one reason American students are not as academically advanced as some of their international counterparts (Peak, Caldwell, Owen, Stevenson, Suter, Frase, Schmidt, Stigler and Williams, 1996). Still other teachers might recognize the value of reinforcing already-learned concepts, and look for quality rather than quantity in their students' learning.

The contribution I would like to make to this complicated process of picking and choosing what and how much to study and teach is just that: the process is complicated. There really is no single right answer to the questions "How should I study?" and "How should I teach?" The answers will likely change depending on the materials involved, the context in which they are learned, how learning is evaluated, or even the learner's mood. Students and teachers should be aware of these complexities and use them to improve the effectiveness of their learning and teaching. If students and teachers have a goal of remembering or instilling in memory something for years to come, the most effective approach may be to emphasize the material that the students know best.

References

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

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 918-928.

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 801-812.

Karpicke, J.D., & Roediger, H.L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151-162.

Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 187-194.

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.

Koriat, A., Sheffer, L., & Ma'ayan, H. (2002). Comparing objective and subjective learning curves: Judgments of learning exhibit increased underconfidence with practice. *Journal of Experimental Psychology: General*, *131(2)*, 147-162.

Kornell, N., & Bjork, R. A. (2008). Optimizing self-regulated study: The benefits--and costs--of dropping flashcards. *Memory, 16*, 125-136.

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

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language, 65*, 85-97.

Kornell, N., & Metcalfe, J. (2006). Study efficacy and the region of proximal learning framework. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 32*, 609–622.

Levin, D. T., Momen, N., Drivdahl, S. B. & Simons, D. J. (2000). Change blindness blindness: The metacognitive error of overestimating change-detection ability. *Visual Cognition, 7*, 397–412.

Metcalfe, J. (1998). Cognitive optimism: self-deception or memory-based processing heuristics? *Personality and Social Psychological Review, 2*, 100-110.

Metcalfe, J. (2002). Is study time allocated selectively to a region of proximal learning? *Journal of Experimental Psychology: General, 131*, 349 –363.

Nelson, T.O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study time during multitrial learning. *Psychological Science, 5,* 207-213.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

Peak, L., Caldwell, N., Owen, E., Stevenson, H., Suter, L., Frase, M., Schmidt, W., Stigler, J., & Williams, T. (1996). *Pursuing Excellence: A Study of U.S. Eighth-Grade Science and Mathematics Teaching, Learning, Curriculum, and Achievement in International Context.* Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.

Roediger, H. L., & Karpicke, J. D. (2006a). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255.

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

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.

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.

Thiede, K. T., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1024-1037.

Tullis, J. G. & Benjamin, A. S. (in press). The effectiveness of updating metacognitive knowledge in the elderly: Evidence from metamnemonic judgments of word frequency. *Psychology and Aging*.

Valverde, G.A., & Schmidt, W. H. (1997). Refocusing U.S. math and science education. *Issues in Science and Technology*, *14(2)*, 60-67.

Wheeler, M.A., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory, 11,* 571–580.

Appendix

***Table 1.* Mean Accuracy on Final Test (standard deviation in parentheses).**

		Easy/ Restudied	Easy/ Not Restudied	Hard/ Restudied	Hard Not/ Restudied	Easy Focus	Hard Focus
Experiment 1	Immediate	.755 (.199)	.324 (.210)	.418 (.299)	.155 (.230)	.455 (.186)	.371 (.213)
	Delayed	.415 (.233)	.215 (.245)	.155 (.219)	.062 (.199)	.239 (.194)	.185 (.213)
		Known/ Restudied	Known/ Not Restudied	Unknown/ Restudied	Unknown/ Not Restudied	Restudy Known	Restudy Unknown
Experiment 2	Immediate	.976 (.044)	.922 (.041)	.885 (.159)	.137 (.188)	.557 (.101)	.904 (.087)
	Delayed	.545 (.187)	.409 (.080)	.303 (.193)	.042 (.055)	.294 (.092)	.356 (.118)
Experiment 3	Immediate	.955 (.069)	.659 (.236)	.819 (.160)	.193 (.287)	.574 (.134)	.748 (.170)
	Delayed	.327 (.242)	.103 (.097)	.261 (.180)	.062 (.082)	.195 (.129)	.182 (.125)
		Chose Not Restudy/ Dishonor	Chose Not Restudy/ Honor	Chose Restudy/ Honor	Chose Restudy/ Dishonor	Dishonor	Honor
Experiment 4	Immediate	.774 (.230)	.377 (.282)	.726 (.218)	.278 (.203)	.526 (.107)	.552 (.113)
	Delayed	.252 (.214)	.113 (.086)	.158 (.159)	.039 (.060)	.145 (.107)	.136 (.113)
Experiment 5	Immediate	.882 (.193)	.762 (.278)	.723 (.310)	.302 (.359)	.595 (.235)	.737 (.274)
	Delayed	.244 (.217)	.170 (.144)	.128 (.141)	.027 (.053)	.135 (.100)	.149 (.131)