Framing Effects on Metacognitive Monitoring and Control

Bridgid Finn

Columbia University
Abstract

Three experiments explored the contribution of framing effects on meta-memory judgments. In Experiment 1 participants studied word pairs. After each presentation they made an immediate judgment of learning (JOL) framed in terms of either remembering or forgetting. In the remember frame people made judgments about how likely it was that they would remember each pair on the upcoming test. In the forget frame people made judgments about how likely it was that they would forget each pair. Confidence differed as a result of the frame. Forget frame JOLs, equated to the remember frame JOL scale by a 1-judgment conversion, were lower and demonstrated a smaller overconfidence bias than remember frame JOLs. When judgments were made at a delay rather than immediately, framing effects did not occur. In Experiment 2 people chose to restudy more items when choices were made within a forget frame. In Experiment 3 people studied Spanish-English vocabulary pairs ranging in difficulty. The framing effect replicated with judgments and choices. Moreover, forget frame participants included more easy and medium items to restudy. These results demonstrated the important consequences of framing effects on assessment and control of study.
People have been shown to be fairly accurate at assessing how well they have learned something, however much research has shown that people’s metacognitive judgments about their memory can be miscalibrated (Benjamin, Bjork & Schwartz, 1998; Koriat, 1997; Koriat, Sheffer & Ma’ayan, 2002; Metcalfe, 1998; Zechmeister & Shaughnessy, 1980). For example, people’s initial judgments of learning (JOLs) about how much they think that they will remember on an upcoming test typically show an overconfidence bias, in which the judgments are higher, on average, than is subsequent test performance (Koriat, Lichtenstein, & Fischhoff, 1980; Lichtenstein, Fischhoff, & Phillips, 1982; Metcalfe, 1998). People have been shown to be so certain in their incorrect answers that they are even willing to bet money in the belief that they are correct (Fischhoff, Slovic & Lichtenstein, 1977).

Judgment accuracy is also thought have importance consequences for how people control their own learning. For example, an overconfident student may stop studying before actually mastering the material, resulting in a poor grade on the final test. As Nelson and Dunlosky (1991, p. 267) said, “The accuracy of JOLs is critical because if the JOLs are inaccurate, the allocation of subsequent study time will correspondingly be less than optimal.” Recently, Metcalfe and Finn (in press) provided evidence that people’s metacognitive judgments are directly linked to their choices for restudy, supporting the long held view that faulty metacognitive judgments can have unfavorably effects on study control (Benjamin, Bjork & Schwartz, 1998; Dunlosky & Hertzog, 1998; Koriat, 2002; Mazzoni & Cornoldi, 1993; Metcalfe, 2002; Nelson & Dunlosky, 1991; Pressley & Ghatala, 1990; Thiede, 1999). Metcalfe and Finn (in press) showed that when people’s
JOLs were manipulated independently of their recall performance, study choices were influenced by the judgment rather than performance. When the judgments were biased, the study choices reflected the same pattern. These results demonstrated a direct link between metacognitive monitoring and control of learning and underscored the importance of judgment accuracy in achieving effective self-guided learning.

Metacognitive overconfidence most likely arises through the use of memory-based processing heuristics, such as an evaluation of the fluency of information retrieved or cue or domain familiarity, that become available while making the judgment (Glenberg, Wilkinson, & Epstein, 1982; Koriat, 1993; Koriat & Bjork, 2006; Reder, 1987, 1988; Metcalfe, 1998; Metcalfe, Schwartz, & Joaquim, 1993; Tversky & Kahneman, 1974). According to Koriat et al. (1980) overconfidence occurs because people rely primarily on information that is consistent with the answer they have chosen and tend to neglect contradictory information. Of the various debiasing techniques that have been explored in an effort to reduce the overconfidence bias (Lichtenstein & Fischhoff, 1980; Yates, Veinott, & Patalano, 2003), one of the most successful techniques has been to ask people to change the way they make their judgments by generating counterfactual evidence for the answer they have just given (Hirt & Kardes & Markman, 2004; Hirt & Markman, 1995; Hoch, 1985; Koehler, 1991; Koriat, Lichtenstein & Fischhoff, 1980; Maki, 1998). Koriat et al. (1980) found improvements in the accuracy of confidence judgments when participants were asked to write down one reason contradicting the answer they had just given before rating their confidence in their answer. Judgments showed a smaller overconfidence bias after participants generated and considered reasons why their answers could be wrong.
More recently, Koriat, Bjork, Sheffer and Bar (2004) conducted an investigation testing people’s confidence in their memories across varying retention intervals. They tested whether people would give distinct judgments about how much they would recall on a later test that came after either a day, a week or even a one-year delay. Predictions were vastly overconfident. Performance judgments about a test following a week delay were about the same as predictions about performance on a test immediately following study. However, when people were asked about how much they thought they would forget either immediately, in a day, or in a week, judgments did show an effect of retention interval. As the retention interval increased, confidence about recall performance declined, as it should have. Forget judgments were sensitive to the retention interval whereas remember judgments were not.

Both studies reported above suggest that reframing the way a judgment is made can influence how people think about their memories and may serve to increase judgment accuracy. In addition, because of the link between monitoring and control, study behavior may also improve. To date, the vast majority of the research on framing effects has focused on people’s ethical and economic choice behavior. Research in these domains has demonstrated that across a variety of tasks people’s judgments and choice preferences about an identical situation can vary as a function of whether the choice has been positively or negatively framed (Tversky & Kahneman, 1981). In Tversky and Kahneman’s (1981) famous “Asian Disease Problem”, people are told that an outbreak of a disease in the United States is expected to kill 600 people. Participants are asked to choose between two programs that have been developed to combat the disease. They are told that if Program A is used 200 lives will be saved for sure, and if Program B is used
there is a one-third probability that 600 will be saved and a two thirds probability that no people will be saved. In this positive, gain frame, most choose Program A. However, when equivalent programs are described in terms of the number of people who will die (Program C: 400 will die for sure, Program D: one-third probability that no one will die, two-third probability that all 600 will die) the majority of people choose Program D despite the fact that C and D are simply reworded versions of A and B. The only difference between the contrasting programs is that A and B are framed in terms of number of lives that will be saved and C and D are framed in terms of the number of people who will die. Tversky and Kahneman described this finding as a shift from risk aversion and preference of a certain outcome when choices are framed in terms of gains to risk seeking when choices are framed in terms of loss.

A multitude of studies have demonstrated that framing effects have important implications for the kinds of social and economic decisions that people make (see Kühberger, 1998 for a review). Virtually no one (Koriat et al., 2004 excepted) has looked at the effect of framing on judgments about memory. The research presented here investigated the role of framing in metacognitive monitoring and control processes. In metacognition experiments participants typically make judgments based on whether they think they will remember each item on a later test. Of interest here was whether framing the JOL in terms of forgetting would debias people’s judgments about how well they had learned something, diminishing confidence and thus increasing the predictive accuracy about upcoming test performance.

The first research goal was to examine the role of framing on immediate and delayed JOLs. Immediate JOLs taken after an initial study presentation were important
judgments to investigate because they typically show a large overconfidence bias. In contrast judgments taken at a delay are usually more accurate, show a truncated overconfidence bias and are thought to rely on different heuristic information than immediate judgments. A test of both types of JOLs allowed a focused characterization of the effect of framing on metacognitive monitoring.

The second research goal was to investigate the effects of framing on the control of learning. The question was whether framing effects would arise at the level of the study choice both in terms of the number of items and the relative ease of the items people would select for restudy. If the forget frame reduces confidence then study choices should also reflect that debiasing. One possible outcome of reduced confidence was that people would choose to restudy more overall and, in particular, select more of the easy items to restudy.

Experiment 1a

Experiment 1a contrasted immediate JOLs framed in terms of remembering and forgetting. In the remember condition, people made typical JOLs in which they were asked to indicate how likely it was that they would remember each pair on the test coming up in a few minutes. In the forget frame participants were asked how likely it was that they would forget each pair. The hypothesis was that when people were asked to make immediate JOLs within the forget frame they would be less confident as compared to when JOLs were made within the remember frame.

Method

Participants, Design and Materials. The participants were 48 undergraduates at Columbia University and Barnard College. They participated for course credit or cash. In
this and in the experiments that follow participants were treated in accordance with APA ethical guidelines. The experiment was a between participants design. Participants were randomly assigned to either the remember frame or the forget frame condition. There were 24 participants in each condition.

Each participant studied 48 word pairs. The word lists were 48 cue target word pairs comprised of words taken from the Toronto Word Pool, a pool of 1,080 common English two-syllable words (Friendly, Franklin, Hoffman, & Rubin, 1982). Mean word length of cue and target was 6.24 letters. No word exceeded 8 letters. For each participant, the computer randomly combined the words into pairs.

**Procedure.** Participants were instructed that they would be learning 48 word pairs, making judgments and would take a cued recall test. At the beginning of the experiment participants in the remember frame condition were given the typical JOL instructions asking them to make their judgments based on what they thought their chances were that they would remember the second word when given the first word during a memory test that would happen in a few minutes. The forget frame instructions were identical except the word remember was replaced with the word forget. In both conditions participants were asked to use a scale from 0-100% to make their judgment. In the remember frame condition participants were told to use numbers closer to 100% to indicate that they were sure they would remember and numbers closer to 0% to indicate that they were sure they would not remember. In the forget frame condition they were told to use numbers closer to 100% to indicate that they were sure they would forget, and numbers closer to 0% to indicate that they were sure that they would not forget. They were told that at test they would be given the cue and would have to type in the target.
Pairs were presented once, for 3.5 s, and were immediately followed by a prompt to make the JOL. In the remember frame condition participants were asked to provide their judgment of remembering, and in the forget frame condition their judgment of forgetting, each time they were prompted to make a JOL. After all the pairs had been studied and given judgments, the pairs were reshuffled and tested. Each cue was presented and participants were asked to type in the target. There were no restrictions on the amount of time they could spend on the test.

**Results**

**Recall performance.** Recall performance was not expected to differ between the two conditions. Recall performance means were .17 (SE = .02) for the remember condition and .19 (SE = .03) for the forget condition. The two conditions were not significantly different from one another \( t < 1, p > .05 \), as evidenced by an independent samples t-test. A probability level of \( p < .05 \) was used as the criterion for statistical significance throughout.

**JOLs.** In this and in the experiments that follow, forget condition judgments were calculated as 1-judgment value so that the remember and the forget conditions could be compared on the same scale. As can be seen in Figure 1, judgments were significantly higher in the remember frame (\( M = .51, SE = .04 \)) than in the forget frame condition (\( M = .37, SE = .03 \)), by a difference of .14, \( t(46) = 2.83, p < .05, CI_{.95} = .04, .23 \). This result provided the first sign that framing effects occur with immediate JOLs.

A further analysis of the JOLs revealed significant differences between the remember and forget frame conditions in the number of items given low JOLs and the number of items given high JOLs. In this analysis a judgment of less than 50 was
classified as a low JOL, and a judgment of 50 or higher was classified as a high JOL.

People in the forget frame condition made a greater number of low JOLs ($M = 35.50$, $SE = 1.83$) than people in the remember frame condition ($M = 26.63$, $SE = 2.71$), $t(46) = 2.71$, $p < .05$, $CI_{.95} = 2.28,15.47$.

**Calibration.** An overconfidence bias was assessed by measuring calibration. A calibration score was calculated for each participant by subtracting mean recall performance from the mean judgment for each condition. Overconfidence was obtained if the score was significantly positive from zero. Of interest was whether the forget frame judgments would be more calibrated (i.e. less overconfident) than remember frame judgments. Participants in the remember frame condition were significantly more overconfident ($M = .34$, $SE = .04$) than the forget frame condition ($M = .19$, $SE = .03$), $t(46) = 3.07$, $p < .05$, $CI_{.95} = .05, .25$. Both were significantly different than zero (all $ts >1$, all $ps < .05$).

**Gammas.** For comprehensiveness, gamma correlations computed for each participant, for all three experiments are reported in Table 1. Gammas are also given between JOLs and restudy choice for Experiments 2 and 3. Gamma correlations are a non-parametric statistic indicating predictive metacognitive accuracy of the JOLs with respect to recall or restudy choice. This accuracy measure is also called resolution or relative accuracy. These data indicate that in all cases, as measured by independent sample t-tests, there were no differences in relative accuracy between the remember and forget conditions.

**Discussion**
The results of Experiment 1a show that framing effects occur when people make immediate JOLs. Whereas recall performance did not differ between the two conditions, judgments framed in terms of forgetting were less confident, and less overconfident, than the remember frame judgments. The only methodological difference between the two conditions was the substitution of one word, forget, for the word remember in the judgment instructions. This small change alone was enough to significantly reduce, though not eliminate, the persistent overconfidence bias shown with single study-test trial immediate JOLs.

JOLs made immediately after a study presentation are typically less accurate than judgments taken after even a short delay (Dunlosky & Nelson, 1992; 1994; Nelson & Dunlosky, 1991). This accuracy advantage is thought to be due to a difference in the types of cues used to make the judgment. Immediate JOLs are thought to be based on a range of cues, including information in short-term memory (Nelson & Dunlosky, 1991), normative ease (Koriat, 1997) or ease of encoding (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Hertzog, Dunlosky, Robinson, & Kidder; 2003; Koriat & Ma’ayan, 2005). In contrast, delayed JOLs typically involve a retrieval attempt, yielding a more accurate assessment of later eventual recall. This difference in cue utilization between immediate and delayed JOLs may modulate metacognitive biases, such as overconfidence (delayed JOLs show less overconfidence), and the underconfidence with practice effect, (see Finn & Metcalfe, 2007; in press). For example, immediate JOLs show underconfidence on and after a second study-judgment-test trial (Koriat et al., 2002), whereas delayed JOLs typically do not (Koriat & Ma’ayan, 2005; Koriat, Ma’ayan, Sheffer & Bjork, 2006; Meeter & Nelson, 2003; Scheck & Nelson, 2005; Serra & Dunlosky, 2005). According
to Finn & Metcalfe (2007; in press) this is because immediate JOLs are not made on the basis of a target retrieval and instead rely on other, less diagnostic information—such as memory for performance on the prior test—which produces underconfident second trial judgments. The approach adopted in Experiment 1b was to test whether the framing effect would generalize to delayed JOLs. The hypothesis was that framing effects would not arise in the case of delayed judgments, which have been shown to be less susceptible to confidence biases than immediate judgments.

Experiment 1b

Method

Participants, Design and Materials. The participants were 40 undergraduates at Columbia University and Barnard College. They participated for course credit or cash. There were 20 participants in each condition.

The experiment was identical to Experiment 1a except that JOLs were made at a delay rather than immediately after study. After studying each pair the words were reshuffled and the cue was presented for a delayed JOL. After making delayed JOLs for each of the cues, the words were reshuffled again and the cue was presented for test.

Results

Recall Performance. Mean recall performance for the remember condition was .11 (SE = .03). The forget condition had a mean recall performance of .12 (SE = .03). The two conditions were not significantly different from one another, t<1, p > .05.

JOLs. In contrast to immediate JOLs, delayed judgments were not significantly different in the remember and forget frame conditions, t<1, p > .05. The mean JOL in the was .24 (SE = .03) in the remember frame condition and .29 (SE = .04) in the forget
frame condition. An analysis of the number of items given low JOLs revealed no significant differences, \( t(38) = 1.46, p > .05 \), between the remember (\( M = 29.60, SE = 1.51 \)) and forget frame conditions (\( M = 26.15, SE = 1.83 \)).

**Calibration.** There were no significant differences in calibration for the remember and forget frame conditions, \( t < 1, p > .05 \). The remember frame showed an overconfidence bias of .14 (\( SE = .03 \)), which was significantly different from zero \( t(19) = 5.11, p < .05, CI_{95} = .08, .20 \). The forget frame overconfidence bias was .17 (\( SE = .04 \)), also significant from zero, \( t(19) = 3.98, p < .05, CI_{95} = .08, .25 \).

**Discussion**

Results of Experiment 1b indicated that framing effects did not occur with delayed JOLs. In Experiment 1a forget framed immediate JOLs were less confident as compared to remember framed JOLs. Remember and forget framed delayed JOLs were not different. Whereas forget framed immediate JOLs showed a reduced overconfidence bias, and a larger number of low JOLs in comparison to remember frame JOLs no differences emerged when the judgments were delayed. To explore how framing effects influence restudy control subsequent experiments will focus on immediate judgments.

**Experiment 2**

An important implication of the findings in Experiment 1a was that restudy choices, because of the link between judgments and study control, should differ depending on the frame. Overconfidence during restudy selection may be particularly insidious since overconfident people would be expected to study less than they need to. The first objective of Experiment 2 was to replicate the results of Experiment 1a by
asking people to make either remember or forget framed JOLs. The second objective was to replicate the findings of Metcalfe and Finn (in press) and show that study choices, which immediately followed the judgments, also reflected judgment differences driven by the framing effect. The specific prediction was that reduced confidence in the forget frame would lead to the selection of more items for restudy.

Method

Participants. The participants were 46 undergraduates at Columbia University and Barnard College. They received course credit or cash for participation. There were 22 participants in the remember frame condition, and 24 participants in the forget frame condition.

Materials, Design and Procedure. Experiment 2 used the same materials and design as in Experiment 1a except the word list was shortened to 36 items, and participants made restudy choices. After each study presentation participants entered in a JOL. Instructions for the JOLs followed the same procedure in Experiment 1a; participants were instructed to enter in their judgment of remembering or judgment of forgetting. After making each judgment they were asked whether they would like to restudy that pair before the test. They made their study choices by hitting a ‘yes’ or ‘no’ button. Study choice framing always matched the judgment frame. In the remember frame participants were prompted to make each study choice with the instructions, “Would you like to restudy this pair to help you remember it before the test?” In the forget frame participants were prompted with the instructions, “Would you like to restudy this pair so that you don’t forget it before the test?” After their study choice they proceeded to the next item for study, JOL and choice. After all of the words had gone
through this procedure, the pairs were reshuffled for test. Participants were not allowed to restudy any items. At test each cue was presented and participants were asked to provide the corresponding target.

**Results**

**Recall Performance.** The recall performance mean for the remember frame condition was .37 (SE = .06). The mean for the forget frame condition was .40 (SE = .04). The difference in recall performance was not significant (t<1, p >.05).

**JOLs.** The remember frame JOL mean was .56 (SE = .04). The forget frame JOL mean was .46 (SE = .02) a difference that was significant, \( t(44) = 2.61, p < .05, CI_{95} = .02, .18 \), and is shown in Figure 2. As in Experiment 1a, participants in the forget frame condition made a greater number of low JOLs (\( M = 17.79, SE = 1.79 \)) than people in the remember frame condition (\( M = 12.64, SE = 1.88, t(44) = 1.99, p = .05 \). These results replicated the findings in Experiment 1a, and provided additional evidence that a forget frame leads to the generation of lower immediate JOLs.

**Calibration.** There was a significant difference in calibration, \( t(44) = 2.07, p = .04, CI_{95} = .003, .25 \), with the remember frame condition showing significantly more overconfidence (\( M = .19, SE = .05 \)) than the forget frame condition. Participants in the remember frame condition were significantly overconfident, \( t(21) = 3.64, p = .002, CI_{95} = .08, .30 \), where as participants in the forget frame condition were calibrated (\( M = .06, SE = .03 \)), as indexed by a non-significant difference from zero, \( t(23) = 1.83, p > .05 \).

**Restudy Choices.** As can be seen in Figure 3, significantly fewer items were selected for restudy in the remember frame condition (\( M = .45, SE = .07 \)) than in the forget frame condition (\( M = .65, SE = .06 \)), by a difference of .20, \( t(44) = 2.22, p < .05 \),
This large and significant difference indicated that framing effects can alter people’s study behavior.¹

A further question was whether the ease of the items selected for restudy would change as people became less confident. One possible pattern of results was that as everything was perceived as more difficult within the forget frame, people would equally increase the number of easy, medium and difficult items for restudy. Another potential pattern was that when people’s confidence was lower their study choices would shift to include more of the easiest items. When people are overconfident, few of the easiest items are likely to be selected for restudy. As confidence shifts downward the number of easy items people feel they should study is likely to increase, while the number of difficult items selected for restudy might remain stable across the remember and forget frames, or even decrease. Individual’s JOLs were used to determine which items were easy, medium or difficult since no a priori measures were taken of objective difficulty. For each participant, JOLs ranging between 0-33% were labeled as difficult, 34-66% were labeled as medium, and 67-100% were labeled as easy. Study choice results were analyzed using a 2(remember or forget frame) X 3(easy, medium or difficult) repeated measures ANOVA, with vocabulary ease as a within-participants factor and judgment frame as a between-participants factor. There was a main effect of ease, \( F(2, 74) = 22.57, \ MSe = .07, p < .05, ES = .38 \), (effect size was computed using partial eta squared). People chose to study difficult items the most \( (M = .73, SE = .06) \), followed by medium items \( (M = .57, SE = .06) \) and easy items \( (M = .31, SE = .06) \). There was also a main effect of
frame condition, $F(2, 37) = 6.73, MSe = .24, p < .05, ES = .15$. As reported earlier, more items were chosen with in the forget frame than the remember frame. The interaction between ease and frame condition was not significant, $F(2, 74) = .88, MSe = .07, p > .05$, however because of an interest in how the ease of items selected for restudy would change or remain stable across frames, additional comparisons at each of the vocabulary ease levels were conducted. Significant differences emerged in the mean number of easy items selected for restudy (remember: $M = .19, SE = .07$, forget: $M = .48, SE = .09$), $t(37) = 2.82, p < .05$, CI$_{.95} = .09, .57$. There were no significant differences between the conditions in the number of medium items, $t(37) = 1.63, p > .05$, or difficult items, $t(37) = 1.63, p > .05$, selected for restudy.

Discussion

Across Experiments 1a and 2, a consistent pattern emerged. Forget framing reduced confidence and prompted people to select more items for restudy. Experiment 2 extended the findings of Experiment 1a by showing that the remember frame condition, the condition which demonstrated greater overconfidence, was also the condition in which fewer items were selected for restudy. An analysis that used people’s JOLs as an indication of which items were easy, medium or difficult hinted that as people became less confident the number of items that seemed difficult increased. People in the forget frame condition chose more easy items to restudy than people in the remember frame condition, suggesting that as confidence shifted downward, people modified the control of their learning to include more of the easiest items, in particular.

To more thoroughly investigate how people made study choices across a range of difficulty, item ease was systematically manipulated in Experiment 3. In Experiment 3
participants were asked to study easy, medium and difficult Spanish-English vocabulary and make judgments and study choices within either a remember or a forget frame.

Experiment 3

Method

Participants. The participants were 42 undergraduates at Columbia University and Barnard College. They received course credit or cash for participation. None of the participants were native Spanish speakers. There were 21 participants in each condition.

Materials, Design and Procedure. The materials were Spanish-English vocabulary pairs. There were 12 pairs in each of the three (easy, medium and difficult) conditions, for a total of 36 pairs. The vocabulary pairs came from Metcalfe (2002) and had been created to encompass three difficulty conditions. Participant recall performance in the four experiments in Metcalfe (2002) established these difficulty levels. The experiment was a mixed design with vocabulary ease as a within-participants factor and framing condition as a between-participants factor. Participants were randomly assigned to either the remember frame or the forget frame condition, and all participants studied easy, medium and difficult pairs. With the exception of the materials, the procedure was identical to Experiment 2. All 36 pairs were randomly shuffled and presented for 3 seconds of study. Immediately following the presentation of each pair participants made either a remember or forget frame judgment followed by a restudy choice. Word pairs were randomly shuffled by the computer before the start of the test. At test the Spanish word was presented and participants were asked to type in the English translation. When they had entered in their answer they hit return and the next Spanish cue appeared. There were no time constraints on the amount of time they could spend on each item at test.
Results

Recall performance. There was a main effect of ease, $F(2, 80) = 141.88, MSe = .03, p < .05, ES = .78$. Recall performance was best for the easy items ($M = .98, SE = .01$) followed by the medium items ($M = .68, SE = .04$) and the difficult items ($M = .39, SE = .04$). All ease conditions were significantly different from one another (all $ts > 1, ps < .05$). No other main effect or interaction was significant.

JOLs. Mean judgments for remember and forget frame conditions at each level of difficulty are plotted in Figure 4. There was a main effect of ease, $F(2, 80) = 232.09, MSe = .01, p < .05, ES = .85$. Easy items were given the highest JOLs ($M = .94, SE = .01$), followed by medium items ($M = .64, SE = .03$), and difficult items ($M = .38, SE = .04$, all $ts > 1, ps < .05$). There was also a main effect of condition $F(2, 40) = 4.61, MSe = .07, p < .05, ES = .10$. As in Experiments 1a and 2, people in the remember frame condition made higher JOLs overall ($M = .70, SE = .03$) than people in the forget frame condition ($M = .60, SE = .03$). The interaction between item ease and frame condition was not significant. Planned comparisons revealed significant differences for the easy items, (remember frame: $M = .97, SE = .01$, forget frame: $M = .91, SE = .02$, $t(40) = 2.26, p < .05$) and medium items, (remember frame: $M = .71, SE = .04$, forget frame: $M = .57, SE = .04$, $t(40) = 2.28, p < .05$), but no difference for the difficult items, (remember frame: $M = .43, SE = .06$, forget frame: $M = .33, SE = .04$, $t(40) = 1.48, p > .05$).

Calibration. There were no significant main effects of ease or condition on calibration, nor was the interaction between ease and frame condition significant. Calibration values for both groups across item ease were not significantly different from zero (remember frame: $M = -.02, SE = .04$, $t < 1, p > .05$, forget frame: $M = -.05, SE = .04$).
Planned comparisons revealed significant differences only for the easy items (remember frame: $M = -0.02, SE = 0.01$, forget frame: $M = -0.07, SE = 0.03$).

Remember frame calibration for the easy items was not different from zero, $t(20) = 1.13, p > .05$, but forget frame calibration was significantly underconfident, $t(20) = 2.65, p < .05$.

Calibration for all of the other levels was not significantly different from zero.

**Restudy choices.** Restudy choices showed a main effect of ease, $F(2, 80) = 83.73, MSerror = 0.01, p < .05, ES = .68$. Across frame conditions people selected to study difficult items the most ($M = 0.71, SE = 0.05$), followed by medium items ($M = 0.49, SE = 0.05$) and easy items ($M = 0.12, SE = 0.03$). There was also a main effect of condition, $F(2, 40) = 5.11, MSerror = 0.17, p < .05, ES = .11$. This replicated the findings in earlier experiments showing that people in the remember frame condition chose to restudy fewer items overall ($M = 0.36, SE = 0.05$) than people in the forget frame condition ($M = 0.52, SE = 0.05$). The item ease by frame condition interaction was not significant. However, planned comparisons revealed that people in the forget frame condition chose to study more easy items ($M = 0.21, SE = 0.06$) than people in the remember frame condition ($M = 0.03, SE = 0.02$), $t(29) = 2.44, p < .05, CI_{95} = .08, .20$. Indeed, the number of easy items selected in the remember frame condition did not significantly differ from zero, $t(20) = 1.66, p > .05$. A comparison of the choices for medium difficulty also revealed a significant difference between the frame conditions, $t(40) = 1.81, p < .05, CI_{95} = .10, 18$. People in the forget frame condition selected more medium difficulty items for restudy ($M = 0.62, SE = 0.07$) than people in the remember frame condition ($M = 0.44, SE = 0.06$). Remember and forget frame did not show any difference in number of difficult items selected for restudy ($M = 0.04, t = 1.34, p > .05$).
.69, \( SE = .08, M = .79, SE = .06 \) accordingly, \( t < 1, p > .05 \). These results are plotted in Figure 5.

**Discussion**

Participants gave lower JOLs and chose to restudy more items overall in the forget than remember frame, despite identical recall performance. The forget frame led people to chose more easy and medium items to restudy in particular. The results also showed that JOLs and study choices were tightly linked. Choices followed the same pattern as JOLs across the easy medium and difficult items in the two conditions. JOLs were lower for the easy and the medium items in the forget condition, and the easy and medium items were also selected for study more often in the forget condition. With difficult items JOLs did not differ between the remember and forget conditions and neither did choices.

**General Discussion**

The most important finding in the experiments presented here was that reframing immediate JOLs in terms of forgetting, in contrast to remembering, changes the way people monitor and control their learning. The forget frame reduced confidence resulting in judgments that showed a smaller overconfidence bias. In addition, the forget frame led to the selection of more items for restudy than the remember frame.

In Experiments 2 and 3 people who were less confident chose to restudy the easiest items. This pattern is generally consistent with the region of proximal learning model of study time allocation (e.g. Metcalfe, 2002). The region of proximal learning model suggests that people should study items that are the closest to being learned. These are the items in the learner’s region of proximal learning and are what Atkinson
(1972a; 1972b) called transitional items. As people’s confidence dropped so too did their
perception about what items could benefit most from additional study time. The forget
frame contributed to the number of items people perceived as difficult overall, but, as the
region of proximal learning model would predict, they opted to increase the number of
easy (Experiment 2 & 3) and medium items (Experiment 3) for restudy.

These experiments show clearly that framing effects have an impact on
metamemory judgments and restudy choices. What are the processes that underlie these
effects? One possibility is that the forget frame reduced confidence by making people
more sensitive to the fallibility of their memories. According to Koriat et al. (2004, p.
654) “…the mere mention of forgetting can activate people’s knowledge about the
decline in memory performance that is expected to occur with time…” When people
make a typical remember framed immediate JOL they can potentially rely on several
different types of cues (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Hertzog et al.
here suggest that a forget frame can prompt people to utilize additional information about
how their memories may potentially fail, and that making use of this cue yields
judgments that are more accurate. Interestingly, these framing effects were confined to
immediate JOLs. In Experiment 1b, delayed remember and forget framed JOLs were not
different. These results indicate that when diagnostic information about whether the item
can be retrieved from memory is available, as is the case while making a delayed JOL,
other cues do not play as large a role in making the judgment.

Within the judgment and decision making literature there are a number of
potential explanations for why framing effects occur (see Levin, Schneider & Gaeth,
According to prospect theory, Kahneman and Tversky’s (1979) widely accepted descriptive model of decision making, people are more sensitive to losses than they are to comparable gains, that is, losses “loom larger” than gains. Framing effects, whereby people—and capuchin monkeys for that matter (Chen, Lakshminaryanan, Santos, 2006)—respond differently depending on whether a choice or scenario is framed in terms of a gain or a loss, can occur because of loss aversion. The characterization of loss aversion corresponds well with the idea that the forget frame can activate cues about memory failure producing a shift in people’s meta-memory judgments and restudy choices. When people were asked to evaluate their memories the forget frame may have pointed toward potential loss of memory. This loss frame may have worked to shift people toward a more conservative criterion for endorsing something as learned or not, as well as to the selection of more, as well as easier items to restudy in an effort to avoid the potential risk of losing those items from memory.

People who do not exhibit overconfidence have been shown to perform better on a subsequent test (Maki, Shields, Wheeler & Zachilli, 2005). It may be that successful learners are better at making metacognitive predictions, but the demonstrated link between monitoring accuracy and test performance has been mixed. Several studies have shown a positive correlation between metacognitive ability and test performance (Bisanz, Vesonder & Voss, 1978; Maki & Berry, 1984; Yan, 1994), while others have shown that learning ability is not related to metacognitive accuracy (Kearney & Zechmeister, 1989; Kelemen, Frost & Weaver, 2000; Lovelace, 1984; Maki & Swett, 1987; Underwood, 1966). The results here support the proposal that lower confidence may change the way people evaluate and plan their study, which may play an important role in later test
accuracy. Performance may improve for people who are less confident because they are restudying items that benefit most from additional study, in particular the easy and medium items. Another possible outcome is that lower confidence influences the amount of time people initially spend trying to learn the word. These are possibilities that are currently under investigation.

These results are important for several reasons. Overconfidence is a persistent problem with immediate JOLs. Simply framing immediate judgments in terms of forgetting was able to reduce the overconfidence bias underscoring the importance of absolute accuracy study choice behavior. The results presented here then are not only exciting because of the finding that a word can significantly change the way people to think about their memories, but more importantly because of the influence of that change can have on improving people’s choices about their learning behavior.
References


Finn, B., & Metcalfe, J. (in press). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language*.


and Human Decision Processes, 76, 149–188.


Serra, M. J., & Dunlosky, J. (2005). Does Retrieval Fluency Contribute to the


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.
Footnotes

1 It is generally agreed that when people make study choices they are doing so after having made an assessment about the current learned state of the item they are evaluating for restudy, with or without an explicit judgment phase (Nelson & Narens, 1990; Son & Metcalfe, 2005; and see Kornell, Son & Terrace, 2007 for hint taking results with monkeys). An additional experiment was conducted in which study choices were made without an immediately preceding judgment. While recall performance for the remember and forget frame conditions was not different ($t(46) = 1.20, p > .05$), restudy choices did show a .23 difference, $t(46) = 2.70, p < .05$, $CI_{95} = .06, .40$. Even without an explicit JOL phase, forget framing led to the selection of more items (.62) than in the remember frame condition (.39).
Table A1

*Gamma correlations*

<table>
<thead>
<tr>
<th></th>
<th>JOL – Recall</th>
<th>JOL - Study Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Experiment 1a</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remember</td>
<td>.17(.10)</td>
<td>.42(.07)*</td>
</tr>
<tr>
<td>Forget</td>
<td>.42(.07)*</td>
<td>--</td>
</tr>
<tr>
<td><strong>Experiment 1b</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Delayed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remember</td>
<td>.91(.07)*</td>
<td>.75(.11)*</td>
</tr>
<tr>
<td>Forget</td>
<td>.75(.11)*</td>
<td>--</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remember</td>
<td>.31(.10)*</td>
<td>-.71(.10)*</td>
</tr>
<tr>
<td>Forget</td>
<td>.21(.10)</td>
<td>-.51(.16)*</td>
</tr>
<tr>
<td><strong>Experiment 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remember</td>
<td>.79(.02)*</td>
<td>-.76(.11)*</td>
</tr>
<tr>
<td>Forget</td>
<td>.52(.13)*</td>
<td>-.56(.14)*</td>
</tr>
</tbody>
</table>

Numbers in parentheses are standard error of the mean. *Denotes significantly different from zero. No significant differences between conditions were found for JOL-Recall gammas or JOL-Study Choice gammas in any experiment (all $p > .05$).
Figure Captions

Figure 1. Mean JOLs for remember and forget frame conditions in Experiment 1.

Figure 2. Mean JOLs for remember and forget frame conditions in Experiment 2.

Figure 3. Proportion of items chosen for restudy for remember and forget frame conditions in Experiment 2.

Figure 4. Mean JOLs for remember and forget frame conditions at each of easy, medium and difficult levels of vocabulary difficulty in Experiment 3.

Figure 5. Proportion of items chosen for restudy for remember and forget frame conditions at each of easy, medium and difficult levels of vocabulary difficulty in Experiment 3.
Figure 1. C352

![Mean JOL for Remember and Forget Frame Conditions](image_url)
Figure 2. C352

MEAN JOL FOR REMEMBER AND FORGET FRAME CONDITIONS

- Remember
- Forget
Figure 3. C352

PROPORTION OF ITEMS CHOSEN FOR RESTUDY FOR REMEMBER AND FORGET FRAME CONDITIONS
Figure 4. C352
Figure 5. C352

PROPORTION OF ITEMS CHOSEN FOR RESTUDY FOR REMEMBER AND FORGET FRAME CONDITIONS

- EASY
- MEDIUM
- DIFFICULT

PROPORTION CHOSEN FOR RESTUDY

- REMEMBER
- FORGET
Author Note

Bridgid Finn, Department of Psychology, Columbia University.

This research was supported by NIMH grant R01 MH60637. I would like to thank the MetaLab as well as Lisa Son & Nate Kornell for their help and comments.

Correspondence should be addressed to Bridgid Finn, Department of Psychology, Columbia University, New York, NY 10027.

Email: bmf2003@columbia.edu