

Familiarity and Retrieval Processes in Delayed Judgments of Learning

Janet Metcalfe and Bridgid Finn
Columbia University

Two processes are postulated to underlie delayed judgments of learning (JOLs)—cue familiarity and target retrievability. The two processes are distinguishable because the familiarity-based judgments are thought to be faster than the retrieval-based processes, because only retrieval-based JOLs should enhance the relative accuracy of the correlations between the JOLs and criterion test performance, and because only retrieval-based judgments should enhance memory. To test these predictions, in three experiments, the authors either speeded people's JOLs or allowed them to be unspeeded. The relative accuracy of the JOLs in predicting performance on the criterion test was higher for the unspeeded JOLs than for the speeded JOLs, as predicted. The unspeeded JOL conditions showed enhanced memory as compared with the speeded JOL conditions, as predicted. Finally, the unspeeded JOLs were sensitive to manipulations that modified recallability of the target, whereas the speeded JOLs were selectively sensitive to experimental variations in the familiarity of the cues. Thus, all three of the predictions about the consequences of the two processes potentially underlying delayed JOLs were borne out. A model of the processes underlying delayed JOLs based on these and earlier results is presented.

Keywords: judgments of learning, cue familiarity, target retrievability, metacognition, dual-process model

People's judgments of learning (JOLs) have consequences for their subsequent study behavior (Finn, in press; Metcalfe & Finn, 2008). If JOLs are independently lowered, say, by framing the JOL question to participants to ask about whether they will remember the answer (resulting in high JOLs) or whether they will forget it (resulting in low JOLs), their study choice behavior is altered. They choose fewer items to study in the former case than in the latter, even though their learning, at the time of making the judgment, is the same (Finn, in press). Other manipulations that have altered people's JOLs in an illusory way also have been shown to have direct consequences for what they choose to study (Metcalfe & Finn, 2008). Given that people use these metacognitive judgments to control their subsequent behavior, it is important both that the judgments be accurate and that we understand the processes that underlie them. Delayed JOLs, in which the judgments are made using only the cue at some time after the study effort, appear to be among the most accurate ways of making a self assessment of one's own learning, both in terms of relative accuracy (Nelson & Dunlosky, 1991) and calibration (Finn & Metcalfe, 2007, 2008; Koriat & Bjork, 2005). For this reason, we were especially interested in understanding the mechanisms underlying delayed JOLs.

Research on delayed JOLs focuses on the postulate that the mechanism for making these judgments is an attempted retrieval of

the target (Nelson, Narens, & Dunlosky, 2004). Here, we tested the idea that although some delayed JOLs may, indeed, be based on a retrieval attempt, as most researchers have proposed, there is a second basis for these judgments—familiarity. We investigated whether these two mechanisms that may underlie delayed JOLs are separable, and also whether they may have different consequences for the accuracy of the JOLs and for people's subsequent memory.

The reasons many researchers have thought that delayed JOLs may be based on retrieval is that the relative accuracy of people's delayed judgments is substantially higher than when those judgments are made immediately after the study presentation (Begg, Duft, Lalonde, Melnick, & Sanvito, 1989; Benjamin & Bjork, 1996; Benjamin, Bjork, & Schwartz, 1998; Kimball & Metcalfe, 2003; Koriat, 1997; Nelson & Dunlosky, 1991, 1992; Nelson, Dunlosky, Graf, & Narens, 1994; Spellman & Bjork, 1992). There have been three main theories of why the delayed JOL accuracy advantage occurs, and each of the three implicates a retrieval attempt in the case of delayed JOLs. Indeed, only two studies (Benjamin, 2005; Son & Metcalfe, 2005) have suggested that something else may underlie some delayed JOLs.

The case for the postulate that people use an attempt to retrieve the target as the basis of their delayed JOLs comes primarily from studies and theories that have attempted to explain the difference in immediate and delayed JOL relative predictive accuracy with respect to the criterion test, that is, the "delayed JOL effect." The first proposal to explain this finding was the monitoring dual memories hypothesis given by Nelson and Dunlosky (1991), which states that immediate judgments are based on retrieval from both short-term memory (STM) and long-term memory (LTM). While making an immediate judgment the target item is still in STM and thus judgments made immediately will not entail a retrieval attempt from LTM and, hence, will be poor at discriminating between what will be remembered and what will be forgot-

Janet Metcalfe and Bridgid Finn, Department of Psychology, Columbia University.

This research was supported by National Institute of Mental Health Grant R01 MH60637. We thank Lisa Son and the MetaLab for their help and comments.

Correspondence concerning this article should be addressed to Janet Metcalfe, Department of Psychology, Columbia University, New York, NY 10027. E-mail: jm348@columbia.edu

ten when the test is delayed. By contrast, delayed JOLs rely only on retrieval from LTM, which is more diagnostic of what will happen at the final test.

The second explanation of the delayed JOL effect is the transfer appropriate processing view (Begg et al., 1989; Dunlosky & Nelson, 1992; Glenberg, Sanocki, Epstein, & Morris, 1987; Roediger, Weldon, & Challis, 1989), which states that the retrieval enacted while making delayed JOLs is more similar to the retrieval that the person will use at test than are the processes that people use to make immediate JOLs. Therefore, the delayed retrieval-related JOL will be more diagnostic of how people will do on the test. Although there are data mitigating against this theory (Dunlosky & Nelson, 1997; Dunlosky, Rawson, & Middleton, 2005; Weaver & Kelemen, 2003), our only point here is that the theory postulates that the reason for the delayed JOL to test accuracy is a retrieval attempt. By both of these views, if there were no target retrieval attempt, the correlations between JOLs and later test performance would be low rather than high. We make a similar assumption—that a target retrieval attempt should result in a high JOL to test correlation, but if no retrieval attempt is made, that correlation will be lower. We use this as a method to tease apart the hypothesized two processes in delayed JOLs.

The third view is the self-fulfilling prophecy explanation. By this view, the improvement in the relative accuracy of the delayed JOLs comes about because those judgments themselves—which involve retrieval, and retrieval, if successful, enhances memory—have an effect on the later memory test performance (Kimball & Metcalfe, 2003; Spellman & Bjork, 1992). This theory, like the others, states that people make their delayed JOLs by attempting to retrieve the target. If they are successful they give those items a high JOL; if unsuccessful they assign a low JOL. The critical difference between this theory and the two others is that these authors note (and demonstrate, in the case of Kimball & Metcalfe, 2003) that the act of successful retrieval at a delay enhances memory for those items that are brought to mind (see Roediger & Karpicke, 2006). Those retrieved items are not only given high JOLs but also get a memory boost. Thus, the JOL itself, insofar as it involves retrieval, should enhance memory. We return to this point shortly, as we look not only for higher relative accuracy, if the learner is retrieving to make his or her JOLs, but also for enhanced memory.

Despite the near consensus that delayed JOLs are based on an attempt at target retrieval, Son and Metcalfe (2005) have recently presented data suggesting that *some* delayed JOLs may not be based on target retrieval. Three experiments compared the reaction times (RTs) of people when making JOLs without any instructions to when they were told to retrieve and then make the JOLs. According to a retrieval-only hypothesis, people should attempt to retrieve the target in both cases: Telling them to do what they would do anyhow should not alter their behavior. If so, then the RT functions in these two cases should track one another. In both cases, the time needed to make the JOL should increase as the JOLs decrease and target retrieval becomes more difficult and time consuming.

However, Son and Metcalfe (2005) found that the RTs for the lowest JOL items did not follow this pattern: Some ‘don’t know’ judgments were made very quickly. The pattern of RT data followed the expectations of the retrieval hypothesis in the case where people were told to retrieve first and then make their JOLs:

Reaction times increased monotonically, with the lowest JOLs showing the longest reaction times. However, a different pattern was seen for the JOL-alone condition. It showed a nonmonotonic RT function, with the lowest JOLs being made very rapidly rather than very slowly. Indeed, a measure of the lowest JOLs in the JOL-alone condition showed that they were made faster than the time needed to make a retrieval attempt. When making the lowest JOLs, people seemed to know that they did not know without having to take the time needed to attempt to retrieve the target.

To make these very fast, low JOLs, Son and Metcalfe (2005) suggested that people might be evaluating how familiar they were with the cue, assessing it as low, and making their judgment based on this evaluation. The authors suggested that both cue familiarity and target retrievability may play a role in making JOLs. Fast, low JOLs arise because cue familiarity is assessed as low, and no attempt is made in these cases to retrieve the target. Thus, the judgment process can conclude rapidly. When cue familiarity is assessed as high, and the target is retrieved very quickly, a high JOL is given, but it is a somewhat slower judgment.

If Son and Metcalfe’s (2005) explanation of the RT data is correct, there are three testable consequences. First, there should be a beneficial memory effect of retrieval, but only when the JOLs are based on target retrieval and not when they are based only on cue familiarity. A number of research reports have shown that testing and retrieval have beneficial effects on later memory (e.g., Butler & Roediger, 2007; Karpicke & Roediger, 2008; McDaniel & Fisher, 1991; McDaniel, Kowitz, & Dunay, 1989; McDaniel & Masson, 1985; Roediger & Karpicke, 2006; Pashler, Cepeda, Wixted, & Rohrer, 2005; Pashler, Zarow, & Triplett, 2003). Whitten and Bjork (1977) have found similar memory benefits for retrieval practice. This enhancement presumably occurs only on the items that are retrieved (and not on the ones that fail to be retrieved). Nevertheless, some items should get a memory boost from the JOL procedure itself, as long as that JOL process involves retrieval. The finding that successful retrieval enhances memory can be used as a dependent measure to determine, retrospectively, whether one JOL condition was more likely to have involved retrieval than another.

Second, all three dominant theories propose that the reason delayed JOLs accurately predict performance is because of the retrieval attempt. It follows that we would expect to see the very high JOL relative accuracy in the case where those JOLs are made primarily on the basis of target retrieval. JOL relative accuracy should be poorer if the JOLs were to be based mainly on cue familiarity without a retrieval attempt.

Third, we should be able to experimentally manipulate the two kinds of judgments rather than just relying on correlational evidence. If the cue-familiarity-based JOLs are made quickly, whereas the target-retrieval-based JOLs are made more slowly, we would expect that variables that selectively affect cue familiarity should impact more on the speeded JOLs whereas variables that affect retrieval should impact primarily on the unspeeded JOLs. In a study that manipulated cue and target familiarity, Benjamin (2005) found promising preliminary evidence in support of the second and third propositions. We explore this third prediction further as well.

Experiment 1

In the first experiment, we manipulated target retrievability by using multiple pictorial cue exemplars of a particular category (bear1, bear2, bear3, bear4) and either paired each category cue with a single target word—resulting in high retrievability, or paired each category cue with multiple targets—resulting in low target retrievability. An example of the pictorial cues used in this experiment is given in Figure 1. Using the pictorial variants of the category allowed us to be explicit about which target was specified in the multiple target condition while keeping the cue familiarity the same in the two conditions. Our two primary conditions were, therefore, A–B, A’–B, A’’–B, A’’’–B (which, for simplicity, we refer to as A–B A–B), and A–B, A’–C, A’’–D, A’’’–E (which we refer to as A–B A–C). A–B A–B is, of course, a positive transfer situation and should result in good recall of the target, whereas A–B A–C is a negative transfer situation and should result in poorer recall of the target.

We also varied whether the JOL that people made at a delay was speeded or unspeeded. In the speeded condition, participants had to respond in less than 0.75 s, or else they heard a voice (in the computer program we used) say “Hurry,” and a “Too slow! Data lost!” written message appeared onscreen. In the unspeeded condition the participants were told to take their time in making the judgments, and no voice ever intruded. In the judgment phase, we also included pictorial cues that had never been presented. We refer to this as the ‘new’ condition.

Our predictions were that in the speeded conditions, the JOLs would be lowest in the new cues condition (because of a lack of cue familiarity). They would be higher, but about the same, in the A–B A–C condition and in the A–B A–B condition (because of greater but equal cue familiarity, and little contamination from target retrieval). In the unspeeded condition, we expected low JOLs in the new condition as well (because of a lack of both cue familiarity and target retrievability). Here, however, we predicted higher JOLs in the A–B A–C condition than in the new condition

(because of higher target retrievability), and still higher JOLs in the A–B A–B condition (because the target would be easiest to retrieve in this condition).

We also predicted that the JOL gammas indexing the relative accuracy would be higher in the unspeeded than in the speeded JOL condition. The difference in gamma correlations was expected on the grounds that the JOLs would be based on attempted retrieval in the unspeeded JOL condition to a much greater degree than in the speeded condition. Finally, we predicted that recall would be better in the unspeeded JOL condition than in the speeded JOL condition. The purported retrieval attempt in the unspeeded JOL condition was expected to improve recall of those items that were retrieved. In the speeded JOL condition, a target retrieval was predicted much less frequently, and thus less recall enhancement was expected.

Method

Participants. The participants were 32 undergraduates at Columbia University and Barnard College. They participated for course credit or were paid at a rate of \$12 an hour for participating. Participants were treated in accordance with the ethical principles of the American Psychological Association, and the Columbia University Institutional Review Board approved all of the experiments in this article.

Design and materials. The experiment was a 2 (JOL: speeded vs. unspeeded) × 2 (encoding condition: A–B A–B vs. A–B A–C) × 12 (within-list repetitions of the basic design, over which the data were collapsed) within-participants design. Participants also made JOLs, in both the speeded and the unspeeded condition, on 12 new cues.

The picture cues were 4 distinct exemplars of a particular category, which shared a common name, as shown in Figure 1. These cues, each being slightly different from one another, allowed us to uniquely query a particular target in the JOL and memory tests.

Procedure. Participants were shown, one at a time, and instructed to remember, 48 picture–word pairs. The 48 cues represented six distinct categories, with four exemplars per category in each of the A–B A–B and the A–B A–C conditions, randomly mixed into a single list of items. Each picture–word pair was presented for 3 s of study on each presentation, and the entire list was shown twice. Participants were then asked for their JOLs for 12 cues from that list and 6 cues that were new. The 12 cues from the list were selected such that 6 were cues from the six categories in the just-studied A–B A–B condition, and 6 were from the A–B A–C condition. The cue used for the JOL was randomly selected from one of the four exemplar pictures that had been studied for each category. The JOL cue was the same as was then given in the test phase. The 6 new cues were randomly selected from other categories of pictures that each had 4 exemplars. After making their JOLs, participants were then tested for recall on the 18 items on which they had made JOLs.

There were two trials. The second trial was the same as the first (with different materials, of course), except that if the judgments had been speeded on the first trial, they were unspeeded on the second, and if they had been unspeeded on the first, they were speeded on the second. The speed of the first trial judgments was counterbalanced over participants.

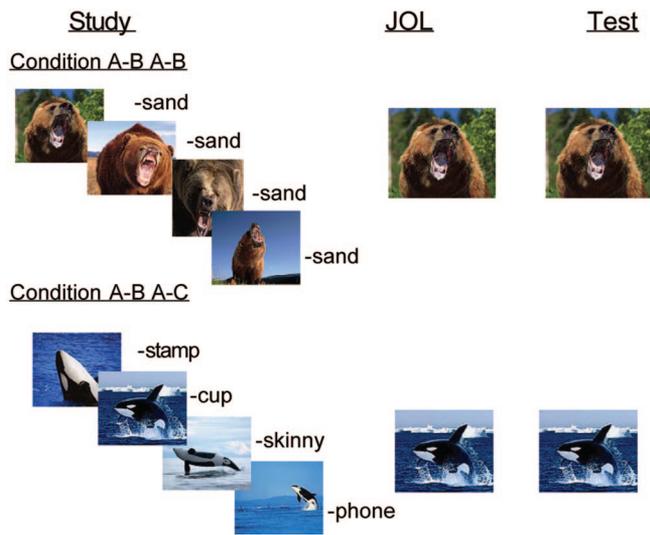


Figure 1. Pictorial cues and word targets used for Conditions A–B A–B and A–B A–C in Experiment 1. JOL = judgments of learning.

The procedure in making the JOLs was as follows. Participants were told, "After you are presented with the pairs, you will have an opportunity to give a JOL. A JOL is a judgment of learning which indicates how confident you are that in about 10 minutes from now you will be able to recall the target when prompted with the picture." Participants made their JOLs by pressing one of four keys that ranged in quarters from 0% to 100%. Keys were marked on the keyboard. In both conditions, there was a practice trial in which the judgments were made at the speed at which they would be made during the experiment and in which participants were told that for the upcoming trial they would be making either speeded or unspeeded judgments. This practice trial was especially important in the speeded conditions, because it gave participants the opportunity to practice with the JOL buttons as quickly as was necessary during the experiment, before data collection began. During the practice trial, as well as during the experiment, a prerecorded voice in the speeded conditions said "Hurry!" and a "Too slow! Data lost!" message appeared if the JOL response exceeded 0.75 s. This occurred during the experiment on 15% of the speeded trials. However, we included all of the items in the analyses below, even those that exceeded 0.75 s.

Results

Latencies. The mean time to make the speeded JOLs was 0.61 s, as compared with 1.48 s in the unspeeded condition, $t(31) = 7.37, p < .05$. (We also conducted a separate analysis excluding items that exceeded 0.75 s in the speeded condition. The pattern of results was the same as is shown below.)

Recall. As predicted, recall was better in the unspeeded JOL condition than in the speeded condition. Unsppeded judgments showed a recall advantage ($M = .69, SE = .04$) over those in the speeded condition ($M = .63, SE = .04$). This main effect was significant, $F(1, 31) = 5.53, MSE = .02, p < .05, \eta_p^2 = .15$ (effect size is reported with partial eta squared, η_p^2).

As was expected, encoding condition A-B A-B showed better recall performance ($M = .83, SE = .04$) than condition A-B A-C ($M = .48, SE = .05, F(1, 31) = 71.34, MSE = .06, p < .05, \eta_p^2 = .70$). The interaction between condition and judgment speed was not significant ($F < 1$). The recall means are shown in Figure 2.

JOLs. The JOLs for the new items were included in this analysis, in both the unspeeded and the speeded JOL conditions. All of the relevant effects and interactions were still significant, however, when the data were reanalyzed with the new items eliminated. As predicted, when participants made speeded JOLs, their judgments followed the familiarity of the cue, whereas when they made unspeeded JOLs, the judgments followed the retrievability of the target. The interaction between JOL speed and encoding condition, $F(2, 62) = 16.82, MSE = .21, p < .05, \eta_p^2 = .35$, is shown in Figure 3. Both the speeded and the unspeeded JOLs showed low mean judgments on the new items. In the speeded condition, although both the A-B A-B and the A-B A-C conditions showed higher JOLs than those given to the new cues, $t(31) = 7.83, p < .05, t(31) = 8.74, p < .05$, respectively, there was no significant difference between them, $t(31) = 1.52, p > .05$. There was, however, a difference between the JOLs in the A-B A-B condition and the A-B, A-C condition in the unspeeded JOL condition, reflecting a similar difference in retrieval in these two conditions, $t(31) = 5.56, p < .05$.

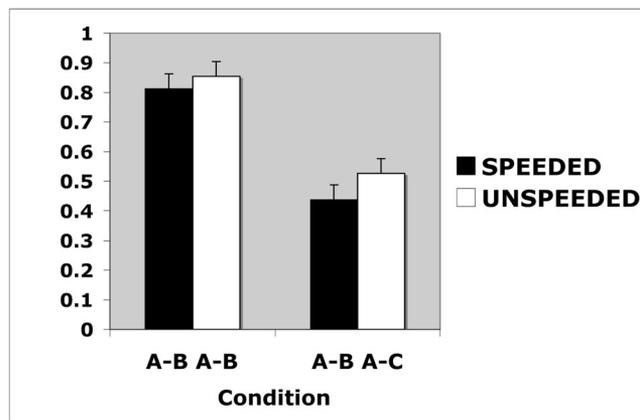


Figure 2. Mean recall performance for conditions A-B A-B and A-B A-C under speeded and unspeeded judgments of learning conditions in Experiment 1. Error bars indicate standard errors of the mean.

There was also, of course, a main effect of encoding condition, $F(2, 62) = 119.14, MSE = .51, p < .05, \eta_p^2 = .79$. In addition, there was a main effect of JOL speed, $F(1, 31) = 11.76, MSE = .22, p < .05, \eta_p^2 = .28$. However, these main effects were qualified by the interaction of interest.

Gamma correlations relating JOLs to recall. Gamma correlations between JOLs and recall index relative metacognitive accuracy. We computed gamma correlations collapsed over all conditions (including the new items) within the unspeeded and speeded JOL conditions. As predicted, the gammas were higher for the unspeeded condition ($M = .84, SE = .05$) than for the speeded JOL condition ($M = .61, SE = .08, t(30) = 2.60, p < .05$). We also eliminated the new items and recomputed the gammas only on items that had been presented for study. Once again, they were higher for the unspeeded JOL condition ($M = .58, SE = .11$) than for the speeded JOL condition ($M = .28, SE = .11, t(24) = 2.14, p < .05$). (The change in degrees of freedom occurred because some participants had either all answers wrong or all answers right, and thus a gamma could not be computed for them.)

Additional analyses. Using data only from the unspeeded JOL condition, we were able to investigate the RTs of participants making delayed JOLs when they were not constrained or subject to a time deadline. The data from this condition are comparable to the RT data of Son and Metcalfe (2005), where participants were simply asked to make delayed JOLs without further constraints. In addition, because we had used a condition in which the cues were new, we were able to investigate whether under unspeeded conditions people would spontaneously give very fast low JOLs selectively in this condition, presumably because of a lack of cue familiarity. The RT data for the three conditions, along with the proportion of responses in each condition at each of the four JOL levels, and the proportion correct at each of these four levels, are presented in Figure 4. As can be seen, most of the JOL responses in the new condition clustered in the lowest JOL category: Participants knew that they did not know. Moreover, they were very fast. In the A-B A-B condition, in contrast, most of the JOLs clustered in the highest JOL category. The proportion of responses in the highest JOL category was, appropriately, somewhat lower in the A-B A-C condition. The participants knew that they knew the

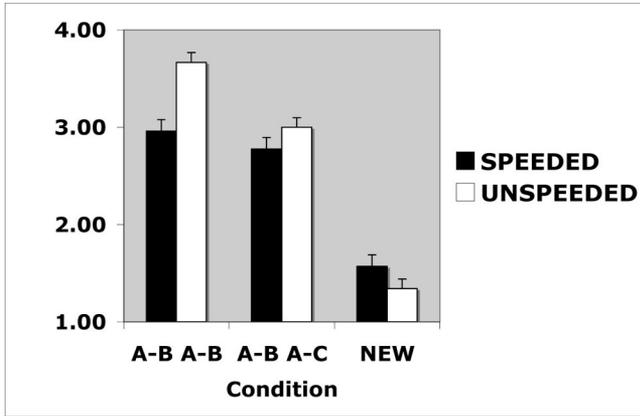


Figure 3. Mean judgments of learning (JOLs) for conditions A-B A-B, A-B A-C, and new under speeded and unspeeded JOL conditions in Experiment 1. Error bars indicate standard errors of the mean.

answers more often in the A-B A-B condition than in the A-B A-C condition. The ‘know,’ or highest JOL judgments, in both the A-B A-C and the A-B A-B conditions were made quickly, but numerically less quickly, than the ‘don’t know’ judgments in the new condition, consistent with the hypothesis. Medium-valued JOLs in the A-B A-B and A-B A-C conditions were made more slowly, just as Son and Metcalfe (2005) had shown.

We were unable to conduct an analysis of variance (ANOVA) combining both levels of JOLs and encoding conditions (new, A-B A-B and A-B A-C) on RTs, because there were many cases in which there were no responses at all in the new condition for the highest JOLs or in the A-B A-B condition for the lowest JOL category. Indeed, there was not a single participant in this experiment who had data in every cell of the full design. Thus, it was necessary to collapse the data. Accordingly, we conducted two separate one-way ANOVAs, the first comparing RTs across the three encoding conditions (collapsing over JOL levels) and the second comparing RTs over JOL levels (collapsing over encoding conditions). There was a significant effect of encoding condition with RT as the dependent variable, $F(2, 62) = 9.84, MSE = .39, p < .05, \eta_p^2 = .24$. Although numerically performance in the new condition (at 1.16 s) was faster than that in the A-B A-B condition (at 1.39 s), the post hoc test comparing these two conditions was not significant, $t(31) = 1.41, p > .05$. The post hoc tests comparing the new condition with the A-B A-C condition (at 1.84 s) and the A-B A-B condition with the A-B A-C condition were both significant, $t(31) = 3.86, p < .05$, and $t(31) = 3.69, p < .05$, respectively.

There was a main effect for JOL level when RT was the dependent measure, $F(3, 57) = 6.56, MSE = .61, p < .05, \eta_p^2 = .26$. All differences among means, except those between JOL Level 1 and JOL Level 4 and between JOL Level 2 and JOL Level 3, were significant, indicating an inverted U-shaped curve as a function of JOL level, with the collapsed RT data. Accordingly, we

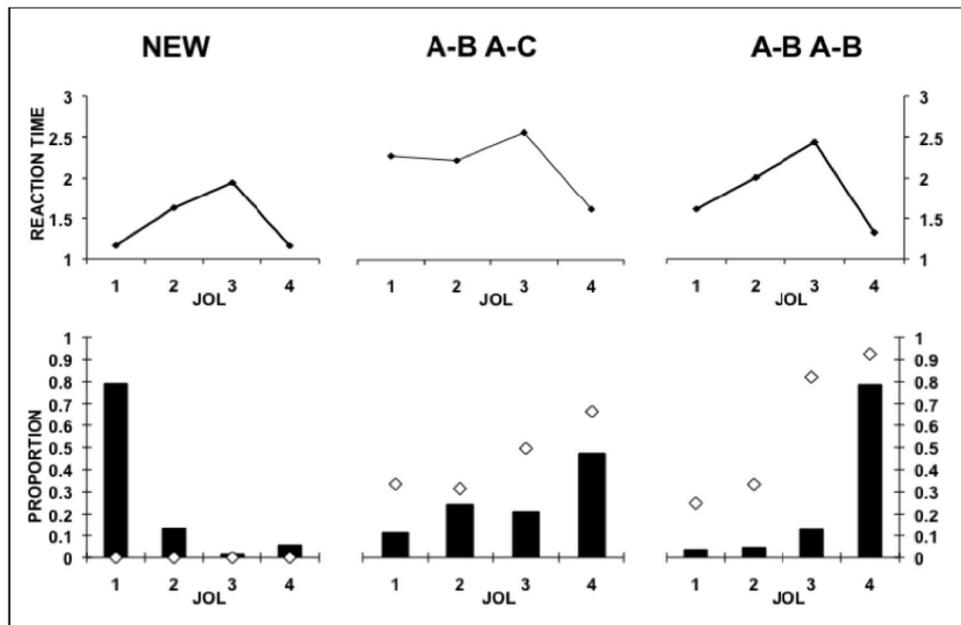


Figure 4. Reaction times at each of the four judgment of learning (JOL) levels are given for the new condition, the A-B A-C condition, and the A-B A-B condition in top left, center, and right graphs, respectively. A JOL of 1 indicates that the participant thought he or she did not know the response, whereas a JOL of 4 indicates that the participant thought he or she knew the response. The bottom graphs show the proportion of responses given at each JOL level, shown by the bars, and the proportion correct at each JOL level, represented by the diamonds, with the data from the new condition on the left, from the A-B A-C condition in the center, and from the A-B A-B condition on the right. All data are from Experiment 1.

tested for linear, quadratic, and cubic trends. Only the quadratic coefficient was significant, $t(19) = 2.90, p < .05$. These distributional and RT results extend and provide further support for the dual process hypothesis.

Discussion

The predictions of the dual process model of delayed JOLs held up very well in the first experiment. The relative accuracy of the gamma correlations was higher with unspeeded than with speeded JOLs. This pattern was consistent with the idea that the slow process that people use in making delayed JOLs involves a target retrieval attempt, but the fast process involves something else. Memory was better when the JOLs were slow rather than fast, suggesting a benefit from retrieval practice that was greater in the unspeeded condition. The manipulation that affected target retrieval had an impact only on the unspeeded JOLs and did not show up on the speeded JOLs. These three results suggest that the two processes are different and dissociable. They also suggest that the slow process may be an attempt at target retrieval. The low JOLs in evidence in the condition in which the cues were new suggests that the fast process was probably cue familiarity, but this suggestion is equivocal because both the cue and the target were completely unfamiliar in this case. Not only was the cue unfamiliar, but the target was also unretrievable, because no target had been presented.

Experiment 2

Although the results of the first experiment were supportive of our hypothesis, we had only included a measure of cue familiarity during the judgment process and retrieval but not during encoding. Thus, in the second experiment, we used the same basic design as had been used in the first experiment, except that we added another condition in which the cue and target were presented only once. Thus, our three encoding conditions were A-B A-B, A-B A-C, and A-B, the latter being a condition in which the cue was presented only once, and hence, in which cue familiarity was expected to be lower than in the other two conditions.

Method

The participants were 42 undergraduates at Columbia University and Barnard College. They participated for course credit or were paid at a rate of \$12 an hour for participating. The method was identical to that of Experiment 1 except that an A-B condition was also included. In the A-B condition, pictorial cues were selected randomly from the same set as the other cues, and targets were drawn from the same set as the other targets and were presented only once during list presentation. Participants made speeded or unspeeded JOLs about four classes of cues in this experiment: those from the A-B A-B condition, those from the A-B A-C condition, those from the A-B condition, and cues that were new.

Results

Latencies. The mean time to make the speeded JOLs was 0.56 s. The mean time to make unspeeded JOLs was 1.24 s. This difference was significant, $t(41) = 9.64, p < .05$.

Recall. As predicted, recall was better in the unspeeded JOL condition ($M = .59, SE = .03$) than in the speeded JOL condition ($M = .53, SE = .03, F(1, 41) = 5.18, MSE = .03, p < .05, \eta_p^2 = .11$). In addition, the A-B, A-B condition showed the best recall performance ($M = .88, SE = .03$); condition A-B, A-C was in the middle ($M = .45, SE = .04$), and the A-B condition was the worst ($M = .35, SE = .04, F(2, 82) = 152.69, MSE = .06, p < .05, \eta_p^2 = .79$). The interaction between condition and speed was not significant. The means for recall are shown in Figure 5.

JOLs. Because of our manipulation, the familiarity of the cues was as follows: A-B A-B = A-B A-C > A-B > new. The pattern of JOLs in the speeded condition followed this ordering. In contrast, in the unspeeded JOL condition, the JOLs tracked the memorability of the targets: A-B A-B > A-B A-C > A-B > new. The interaction between JOL speed and encoding condition was significant, $F(3, 123) = 12.62, MSE = .25, p < .05, \eta_p^2 = .24$, as is shown in Figure 6. The pattern of judgments in the speeded condition showed the A-B A-B condition and the A-B A-C condition both being high but not significantly different from one another, $t(41) = 1.85, p > .05$; the A-B condition being lower and significantly different from both the A-B A-B condition, $t(41) = 5.75, p < .05$; and the A-B A-C condition, $t(41) = 4.61, p < .05$; and the new condition, in which the cue was not seen at all and, hence, was maximally unfamiliar, being lower yet and significantly lower than the speeded JOLs in the A-B condition, $t(41) = 3.61, p < .05$.

The post hoc comparisons for the unspeeded JOL conditions showed that the A-B A-B condition was higher than the A-B A-C condition, $t(41) = 6.49, p < .05$; the A-B A-C condition was higher than the A-B condition, $t(41) = 7.13, p < .05$; and the A-B condition was higher than the new condition, $t(41) = 8.83, p < .05$. This interaction, as before, was our main prediction concerning JOLs.

There was a main effect of encoding condition, $F(3, 123) = 129.49, MSE = .40, p < .05, \eta_p^2 = .76$. There was also a main effect of JOL speed, $F(1, 41) = 9.32, MSE = .36, p < .05, \eta_p^2 = .19$. However, these main effects were qualified by the interaction of interest.

Gamma correlations relating JOLs to recall. As in Experiment 1, gammas were predicted to be higher in the unspeeded than in the speeded condition. We computed gamma correlations collapsed over the three encoding conditions (excluding the new

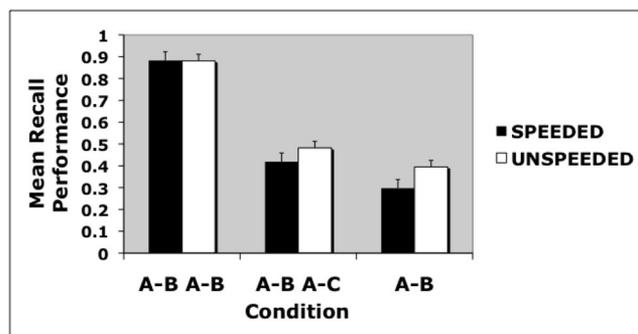


Figure 5. Mean recall for conditions A-B A-B, A-B A-C, and A-B under speeded and unspeeded judgment of learning (JOL) conditions in Experiment 2. Error bars indicate standard errors of the mean.

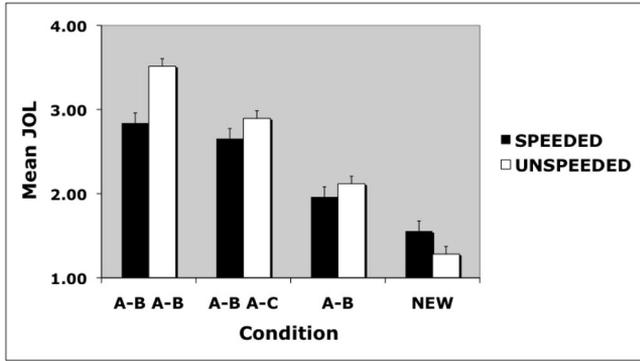


Figure 6. Mean judgments of learning (JOLs) for conditions A-B A-B, A-B A-C, A-B, and new under speeded and unspeeded JOL conditions in Experiment 2. Error bars indicate standard errors of the mean.

items) within the unspeeded and speeded JOL conditions. As predicted, they were higher for the unspeeded condition ($M = .51$, $SE = .07$) than for the speeded JOL condition ($M = .19$, $SE = .06$), $t(41) = 3.87$, $p < .05$.

Additional analyses. The RT data for the four conditions in this experiment, and the proportion of responses in each condition at each of the four JOL levels, as well as the proportion correct at each of these four levels, are presented in Figure 7. Most JOL responses in the new condition were found to be in the lowest ‘don’t know’ JOL category. These responses were very fast. As in

the first experiment, in the A-B A-B condition, most of the JOLs clustered into the highest JOL category, and they were also fast but not quite as fast as the ‘don’t know’ responses in the new condition. The proportion of responses in the highest JOL category was lower in the A-B A-C condition and in the A-B condition. Medium JOLs in the conditions where targets had been presented were made more slowly than when high JOLs were given, as Son and Metcalfe (2005) showed. These distributional and RT results are consistent with those of the first experiment and provide further support for the dual-process hypothesis.

We were unable to conduct an ANOVA combining both levels of JOLs and conditions (new, A-B A-B, A-B A-C, and A-B) on RTs, because again there were no participants in this experiment who had data in every cell of the full design. Thus, we had to collapse into two separate one-way ANOVAs, the first comparing RTs across the four encoding conditions (collapsing over JOL levels), and the second comparing RTs over JOL levels (collapsing over encoding conditions). There was a significant effect of encoding condition, with RT as the dependent variable, $F(3, 123) = 6.44$, $MSE = .17$, $p < .05$, $\eta_p^2 = .14$. Although numerically the new condition (at 1.03 s) was faster than the A-B A-B condition (at 1.20 s), the post hoc test comparing these two conditions was just shy of significance, $t(41) = 1.94$, $p = .06$. The post hoc tests comparing the new condition with the A-B A-C condition (at 1.38 s) and with the A-B condition (at 1.36 s) were both significant, $t(41) = 3.11$, $p < .05$, and $t(41) = 4.53$, $p < .05$, respectively.

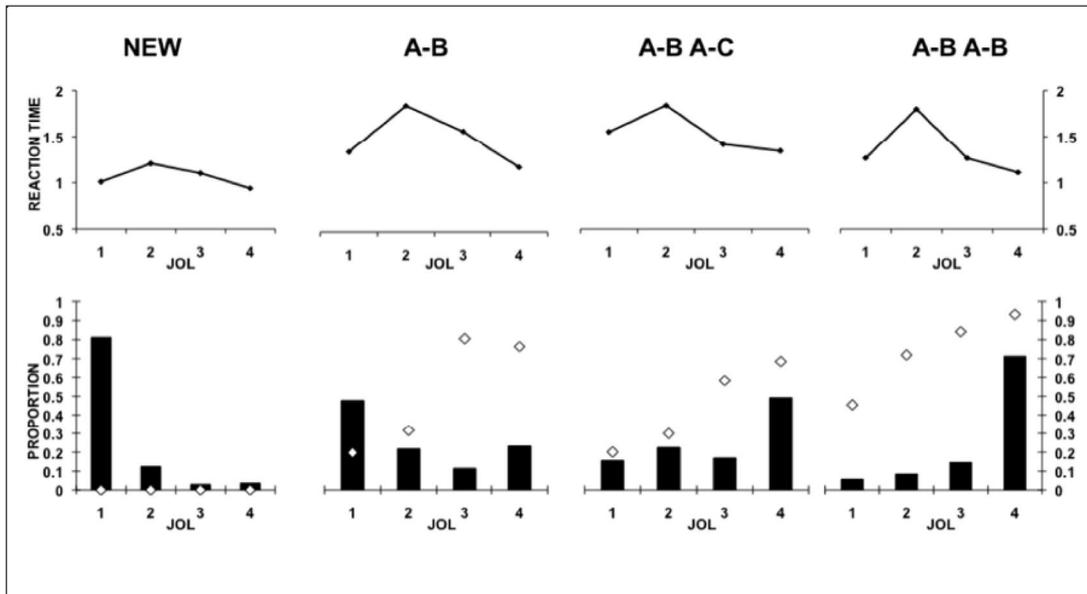


Figure 7. Reaction times at each of the four judgment of learning (JOL) levels are given for the new condition, the A-B condition, the A-B A-C condition, and the A-B A-B condition, in the top left, center left, center right, and right graphs, respectively. A JOL of 1 indicates that the participant thought he or she did not know the response, whereas a JOL of 4 indicates that the participant thought he or she knew it. The bottom row presents the proportion of responses given at each JOL level, shown by the bars, and the proportion correct at each JOL level, represented by the diamonds, with the data from the new condition on the far left, from the A-B condition in the center left graph, from the A-B A-C condition in the center right graph, and from the A-B A-B condition on the far right. All data are from Experiment 2.

There was a main effect for JOL level when RT was the dependent measure, $F(3, 63) = 5.13, MSE = .18, p < .05, \eta_p^2 = .20$. All differences among means except that between JOL Level 1 and JOL Level 4 and between JOL Level 2 and JOL Level 3 were significant—indicating an inverted U-shaped curve as a function of JOL level. We tested for linear, quadratic, and cubic trends. Only the quadratic coefficient was significant, $t(21) = 2.68, p < .05$.

Experiment 3

In Experiment 3 we again varied the target retrievability and cue familiarity, as well as the speed of the JOLs, in a single crossed design. The results of the first two experiments were supportive of the idea that slow JOLs were based on retrieval and that cue familiarity was what drove the fast JOLs, especially fast ‘don’t know’ JOLs. However, the fact that the recall shown for the A–B condition in Experiment 2 was lower than in the other conditions in which the target had been presented (A–B A–B and A–B A–C) made our results equivocal. We had intended the A–B condition to vary only in terms of cue familiarity. We thought it unlikely, but it was nevertheless possible, that target recall, rather than only cue familiarity, could have been a factor in the difference in the fast JOLs between the A–B condition and the A–B A–C and the A–B A–B conditions. Here, we sought to devise a manipulation that would allow us to better isolate cue familiarity.

Specifically, we wanted to eliminate the possibility that target recall could be a contaminant of cue familiarity (or vice versa). To do so, we attempted to construct a zero retrieval condition, in which cue familiarity was still varied. If retrieval were zero in both high and low cue familiarity conditions, then the only thing that could affect JOLs would be cue familiarity (if that were, in fact, what drove the fast JOLs). If the magnitude of the cue-familiarity JOL effect with fast JOLs was the same when retrieval was zero and when retrieval was higher, then we could be more confident in attributing the effect to cue familiarity itself. Thus, by negating the possibility of target retrieval, the effect of the cue familiarity variation could be isolated.

To vary target retrievability, from retrievable to not retrievable, we could, of course, simply either present a target or not. However, a no-target condition posed other problems in terms of the sensibility of the JOL question we were asking our participants. If no target were given following a cue, but just a blank space, the participant might remember that nothing at all was present following a particular cue. What would the correct answer be, then, to the JOL question of how likely is it that you will be able to recall what was paired with the cue, in a few minutes? If nothing had been presented, and the participant knew that nothing had been presented, he or she might be justified in answering the question with a very high JOL, and then later, correctly, answering “nothing.” Would he or she be right or wrong to do this? We did not know, but this did not seem to be a good solution.

To get around this conundrum, we needed to present something, but something that would not be retrievable. Therefore, in the no-target condition, we presented scrambled letters for 16 ms, followed immediately by a pattern mask. Participants saw something rather than simply nothing. They were, however, unable to retrieve anything from this presentation. No participant reported that there had not been words presented in the no-target condition.

It simply seemed to them that whatever had been presented had gone by too quickly for them to process—they could retrieve nothing. (In a pilot experiment, a word, rather than scrambled letters, had been presented in this condition followed by a pattern mask. We had presented the word for 21 ms, supposedly below the threshold of word recognition. However, our participants remembered the words presented in this manner about 10% of the time, and when we presented them three times—to vary target retrievability—their recall performance was about 30%. For this reason, we resorted to presenting scrambled letter strings rather than masked words.) This procedure of presenting something, however unable the participant was to process it, allowed us to ask the JOL question in a way that made sense.

Cue familiarity was varied by altering the duration of the cue. Cues were presented for either 0.5 s or 8.0 s. However, we sought cues in which a difference in duration would make a large difference in their familiarity or fluency. Some stimulus items may be fully processed within a very short time interval, in which case even a large difference in cue duration might not be effective in altering cue familiarity. We also wanted cues for which, especially at the fast rate, we could be fairly sure that processing would be closely limited to the presentation time, as we did not want participants to continue to process the cue even after it had been removed from the perceptual field. If we had used words, for example, so long as participants could read the word in 0.5 s, they could have continued to elaborate and think about it during the seconds that followed, which would not have allowed us to construct a clean design. Participants could time steal to further encode the cue, after its own nominal presentation interval.

To get around this problem, we used materials that made this possibility unlikely—fractal patterns. These patterns, two of which are shown in Figure 8, are exceedingly difficult, if not impossible, to verbalize when presented for only 0.5 s. They could be fairly well encoded, and, for some participants, verbalized, when they were exposed for 8.0 s. Thus, these particular cues afforded a large difference in familiarity, usability, and fluency as a function of presentation duration, which is what we wanted.

The third factor we varied was JOL speed, either speeded or unspeeded, as in the previous experiments. We predicted that recall would be better in the unspeeded JOL condition than in the speeded JOL condition as before (though, of course, only when a target word had actually been presented). We also predicted, as before, that the gammas relating JOLs to recall would be higher in the unspeeded condition than in the speeded condition. In addition, here we predicted a three-way interaction. Cue familiarity alone

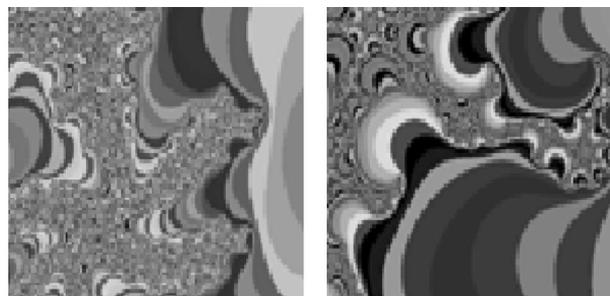


Figure 8. Examples of fractal stimuli used in Experiment 3.

was predicted to selectively affect the speeded JOLs, with the 8.0-s cues giving rise to higher speeded JOLs than the 0.5-s cues. We expected no effect of target retrievability on the speeded JOLs. Target retrievability was predicted to affect the unspeeded JOLs, with the retrievable targets giving rise to higher JOLs than the unretrievable targets. This three-way interaction would provide firmer evidence, not only that the slow process was attempted target retrieval but that the fast process was an assessment of cue familiarity.

Method

Participants were 35 Columbia University or Barnard College students who received course credit or cash. The design was a $2 \times 2 \times 2$ factorial within-participants design, where the variables were speed of judgments (either speeded, 0.75 s or less, or unspeeded, as long as they wanted), cue familiarity (fractal shown for either 0.5 s in the unfamiliar condition or 8.0 s in the familiar condition), and target retrievability (target or no target). In the target condition, the words were presented, following the cue, for 3.0 s. In the no-target condition, scrambled letter strings were presented for 16 ms, followed immediately by a pattern mask for 250 ms. The procedure was basically the same as that of Experiments 1 and 2. The dependent variables were recall performance, JOLs, and gammas between JOLs and recall performance.

Results

Latencies. The mean time to make the speeded JOLs was 0.49 s. The mean time to make unspeeded JOLs was 1.74 s. This difference was significant, $t(44) = 9.17, p < .05$.

JOLs. As predicted, when the JOLs were manipulated to be speeded, cue familiarity had an effect on the JOLs (with high-familiarity cues producing higher JOLs than low-familiarity cues), and target retrievability had no effect. When judgments were unspeeded, target retrievability had an effect. The three-way interaction, shown in Figure 9, was significant, $F(1, 34) = 9.26, MSE = .01, p < .05, \eta_p^2 = .21$. With the speeded JOLs, the high-familiarity cues resulted in higher JOLs than did the low-familiarity cues, $t(34) = 2.54, p < .05$. At the same time, target retrievability had no effect, $t(34) = 1.68, p > .05$. When the JOLs

were unspeeded, target retrievability had an effect, such that having had a target presented resulted in much higher JOLs than having had no target, $t(34) = 8.76, p < .05$. Furthermore, in the unspeeded JOL condition, the effect of cue duration only had an effect when this could have had an impact on retrieval, that is, when there was a target present, $t(34) = 6.69, p < .05$. When no target was present, then the familiarity of the cue produced no difference in JOLs, $t(34) = 1.05, p > .05$, and the JOLs were close to the lowest possible value of 1. In summary, then, this significant three-way interaction indicates that the fast JOLs were driven by cue familiarity, with little or no influence of target retrievability, whereas the slow JOLs depended on retrieval of the target.

All of the other main effects and interactions in this experiment were significant, but they are all explained by (and qualified by) the pattern of data shown in the three-way interaction. There was a main effect of JOL speed, such that unspeeded JOLs were, on average, higher than speeded JOLs, $F(1, 34) = 3.78, MSE = .06$, one tailed $p < .05, \eta_p^2 = .10$. There was an effect of target condition, such that JOLs were higher when there was a target than when there was not, $F(1, 34) = 73.86, MSE = .02, p < .05, \eta_p^2 = .69$. There was an effect of cue condition, such that JOLs were higher when the cue was presented for a long time rather than a short time, $F(1, 34) = 55.92, MSE = .01, p < .05, \eta_p^2 = .62$. There was an interaction between JOL speed and whether or not a target was given, such that presentation of the target mattered much more for the unspeeded JOL conditions than for the speeded JOL conditions, $F(1, 34) = 25.76, MSE = .02, p < .05, \eta_p^2 = .43$. An interaction between JOL speed and cue condition was obtained, such that the duration of the cue mattered more in the unspeeded condition than in the speeded condition, $F(1, 34) = 6.88, MSE = .01, p < .05, \eta_p^2 = .17$. This interaction perhaps deserves comment, as, at first blush, it would seem to counter the idea that cue familiarity matters for the speeded and not the unspeeded judgments. The significant two-way interaction collapses over both of the cases where there was a target and where there was not. There was a large difference in unspeeded JOLs as a function of cue presentation when there was a target, and this large difference was responsible for the double interaction. This occurred because the duration of the cue was important only when a target followed and when the cue was presented for a long enough time to allow the

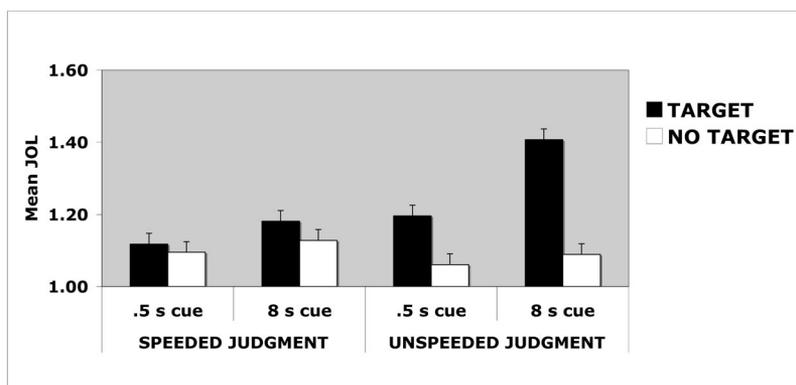


Figure 9. Mean judgments of learning (JOLs) for the interaction among cue familiarity, target retrievability, and JOL speed in Experiment 3. Error bars indicate standard errors of the mean.

presented target to be retrieved. When there was no target presented (as shown in the figure for the three-way interaction) there was no effect of cue duration whatsoever at the slow speed. Thus, taking this two-way interaction at face value without considering the significant three-way interaction, which qualifies it, would be mistaken.

Finally, the interaction between cue duration and whether or not a target was presented was significant, $F(1, 34) = 11.33$, $MSE = .02$, $p < .05$, $\eta_p^2 = .25$, such that the change in duration of the cue mattered much more when a target had been presented than when no target had been presented. All of these main effects and interactions are qualified by the significant three-way interaction, which really tells the whole story.

Recall. Because recall in the no-target conditions was necessarily zero, we dropped this condition from all of the analyses on recall. Performance was better in the unspeeded JOL conditions (.25) than the speeded JOL conditions (.19), $F(1, 34) = 6.54$, $MSE = .02$, $p < .05$, $\eta_p^2 = .16$. There was an effect of cue condition, $F(1, 34) = 70.17$, $MSE = .02$, $p < .05$, $\eta_p^2 = .67$, such that recall was better with the long presentation of the cues than with the short presentation of the cues. This effect is important because it mirrors the effect of cue familiarity seen in the three-way interaction in which the JOLs are the dependent measure. The difference in recall underscores the idea that this effect in the JOLs is very likely due to differential retrievability. The main finding of interest, however, in the recall data was the finding that making the JOLs slowly improved recall more than making the JOLs quickly, as predicted if the slow JOLs involve a memory-enhancing retrieval process, whereas the fast JOLs do not use such a process.

Gamma correlations relating JOLs to recall. Gamma correlations were computed separately for speeded and unspeeded JOL conditions. Within these conditions, they were computed by taking the JOL values given for all cues compared with whether the person gave the word that had been presented with that cue. Thus, a 1 was assigned for recall of the word. A zero was assigned if there had been a word presented and it was not recalled or if no word had been presented (and, of course, it could not be recalled). Thus, a participant's assignment of a low JOL value to cues with which a target word had not been presented would contribute to the goodness of the resultant gamma, increasing its positive value. We predicted higher gammas in the unspeeded than in the speeded JOL condition, as before. True to prediction, the gammas were higher in the unspeeded ($M = .87$, $SE = .08$) than in the speeded ($M = .56$, $SE = .13$) condition, but the effect was significant only by a one-tailed (though predicted, and therefore justified) test, $t(26) = 1.92$, $p < .05$.

General Discussion

These experiments provide support for the conclusion that two processes underlie people's delayed judgments of learning. The first of these processes is the recognition of the cue. The second uses the recognized cue in an attempt to retrieve the target. The first process—the recognition of the cue—may, if the cue fails to be recognized, give rise to fast 'don't know' judgments. In such a situation, where the individual does not even recognize the cue, he or she will not go on to the second process of trying to retrieve the target. Instead, a quick and decisive low JOL will be given and further processing stopped. Furthermore, because there has been

no attempt to retrieve the target, no beneficial memory enhancement, attributable to retrieval, ensues.

But all low JOLs are not fast. It is also possible to obtain slow, low JOLs. These, however, come about after a successful recognition of the cue coupled with an unsuccessful attempt at retrieval. Thus, at the low to middle end of the JOL scale, there is a mix between fast 'don't know' JOLs and fairly slow JOLs that come about because of retrieval failure.

A flowchart outlining the two processes that we propose underlie spontaneous delayed judgments of learning is given in Figure 10. As is shown in this figure, upon receiving the cue, the first process is to determine whether the cue, itself, is recognized. If it is recognized, then the way is clear to move on to the second stage. If not, then there is an endpoint fast 'don't know' JOL. There are, of course, a number of compelling and well-elaborated models of recognition and, for purposes of determining this first stage of JOLs, differences among these are most likely inconsequential. However, for the sake of illustration, consider how the processes in random walk or diffusion models of recognition (see Ashby, 2000; Luce, 1986; Ratcliff, 1978; Ratcliff, Van Zandt, & McKoon, 1999) would map onto fast JOLs. In such models, there are two criteria: a lower match boundary and an upper match boundary. The lower boundary results in a decision, in old–new recognition experiments, that the cue is new (i.e., “no”). In the case of JOLs, reaching this criterion results in the lowest JOL being output and in further processing stopping. The upper boundary in old–new recognition tasks results in an “old” or “yes” decision. In the case of JOLs, this positive recognition triggers the next phase, an attempt at target retrieval. The amount of time it takes to reach these two boundaries generates the RT functions in recognition experiments. In the JOL situation, the time to reach the “no” boundary generates the RT for the fast 'don't know' judgments. RTs for fast, high JOLs, in this simple model, are the sum of the time to reach the “yes” cue-recognition boundary plus the time taken to retrieve the target following cue recognition.

Relative to a recall process, which may sometimes take seconds to complete, the recognition process is fast. Reber, Alvarez, and Squire (1997), for example, reported recognition RT functions with short retention intervals for correct “yes” decisions that peaked at around 0.68 s, with about two thirds of the responses being under 0.75 s. Nearly all “yes” responses had been made within the first second of processing. “No” responses are often a bit longer, but not much. In the current study, mean RT for making 'don't know' judgments in the unspeeded new condition—which may best reflect a relatively pure cue-recognition process in which the lower “no” boundary is reached—were 1.16 s in Experiment 1 and 1.01 s in Experiment 2. The latencies are about right for this first process to be a cue-recognition process.

This cue-recognition stage of processing accounts, in a natural way, for the fast 'don't know' JOL responses seen in Son and Metcalfe's (2005) data. We assume, in the model shown in Figure 10, that the recognition process will normally run to completion and the process will either result in a fast 'don't know' judgment or lead to Stage 2, in which target retrieval is attempted. What about in our own deadline paradigm experiment, presented here, in which processing was truncated in the speeded conditions? It is straightforward to see that if the participant in the experiment is forced to give a very fast JOL—supposedly not to exceed 0.75 s—then the first stage of the JOL's normal processing may not

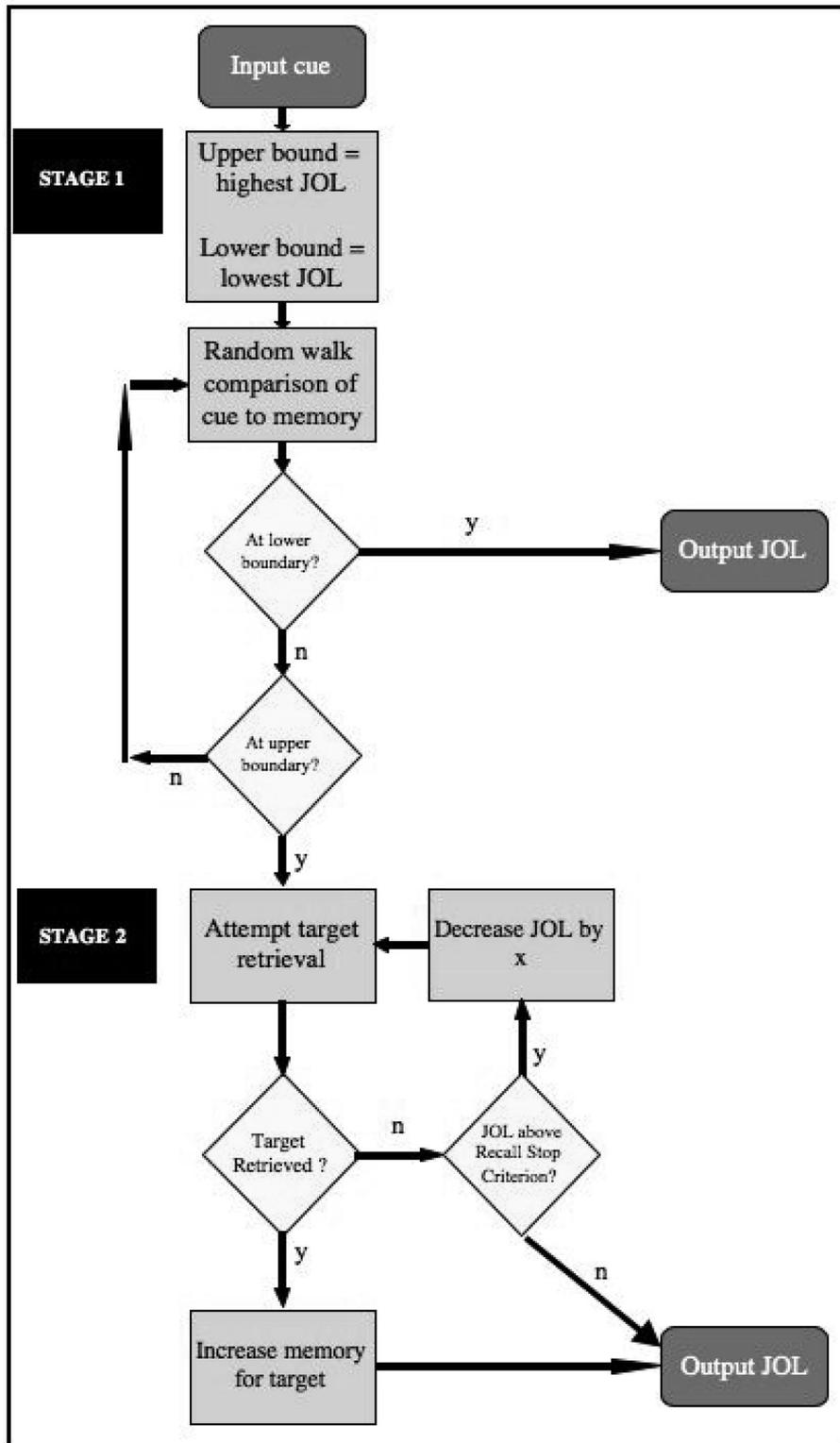


Figure 10. A dual-process model of the processes underlying delayed judgments of learning (JOLs).

always run to completion. We assume that under these conditions, the person assesses the state of the recognition random walk itself at the time of the deadline. If the person did that, then the cues that were more familiar would have shown, on average, at the time of the deadline, greater drift toward the positive boundary than would the cues with less familiarity. This would result in JOLs that would be sensitive to the familiarity manipulation alone, as was shown in the experiments presented here.

The second stage of processing in the model is an attempt at target retrieval. This stage is predicated on successful recognition of the cue. Once the cue is recognized, it is used to attempt to retrieve the target. How long will the person persist with this retrieval attempt, and how does the time to retrieve the target relate to the person's JOL? We propose that the dynamic JOL values themselves are instrumental in determining how long the person will attempt retrieval before giving up and giving the metacognitive judgment that they do not know. As is shown in Figure 10, following successful cue recognition, the person starts the retrieval process with a very high setting on the JOL counter. This counter will remain high if the retrieval process is successful nearly immediately, resulting in fast high JOLs. Because the attempt at retrieval will take some time after successful recognition, these fast high JOLs might be slightly slower than the fast 'don't know' JOLs (though the time to reach the "no" boundary, giving rise to 'don't know' JOLs is often slower than the time to reach the "yes" boundary, which would trigger the second stage of JOL processing. Accordingly, some fast 'know' responses—even though two processing stages are recruited—might be faster than some fast 'don't know' responses). This overall result was shown in both Experiments 1 and 2, in which the RTs for the 'don't know' judgments in the new condition were the same or slightly faster than the high 'know' judgments given in the conditions in which the cues and targets had been presented and the items were given high fast 'know' responses.

According to the model, whenever target retrieval is successful—no matter how long it takes—a memory strengthening process should be enacted. If retrieval is not successful on the first attempt, the retrieval attempts will continue, taking time, of course, with each try. On each successive attempt, the JOL counter decreases. (The model is neutral as to the exact nature of these attempts at retrieval and, to our knowledge, there are no data on what happens during the time it takes someone to recall. Perhaps more and more features are retrieved in succession, eventually resulting in an interpretable item, or perhaps different memory images or echoes are successively retrieved as a whole, through different epochs of retrieval attempts. But however it occurs, we assume that there is a counter that is decrementing the JOL value as the process takes more and more time). If retrieval is successful, at any point in this process, then the current JOL—whatever it is—will be given as the output. This loop results in decreases in JOLs with increases in retrieval time and maps well onto the findings, not only of Son and Metcalfe (2005) but also of Benjamin, Bjork, and Schwartz (1998). These authors showed that decreases in retrieval fluency, as indicated by increased retrieval times, resulted in increasingly lower JOLs.

The stop rule, in this iterative retrieval process, is a predetermined value of the JOL (presumably, for most participants, the lowest JOL value which indicates that they do not know). So long as the JOL is above that lower criterion that the person has set as

the value at which they say that their JOL is so low that they definitely do not know the item, they will continue to attempt to retrieve. Once the JOL becomes too low, that is, it hits the lower JOL criterion, no further retrieval attempts ensue, and the model exits the cycle with a low slow JOL.

What about the frequency distributions of delayed JOLs? As was shown by Kelemen and Weaver (1997), the frequency distributions of delayed JOLs over the range of possible JOL ratings is bimodal (and different from immediate JOLs, which are unimodal, and centered in the midrange but relatively flat). There is a large preponderance of very low and very high JOLs, but few observations in the mid-range. Notice that in the frequency distribution data that we presented with Experiments 1 and 2, in Figures 4 and 7, the overall data are also bimodal. However, in our data they are bimodal in an analyzable way—the lowest JOLs are selectively attributable to the new cues. The highest JOLs are attributable to presented materials that people subsequently recall with a very high probability. This bimodality, seen in delayed JOL data, falls out of the proposed model in a natural way. Many fast, low JOLs result simply because the participant fails to recognize the cue. If they do recognize the cue, however, they will then be automatically set to give the highest JOLs for those items that are retrieved. Insofar as most recall is fast, and only a few straggler items will be retrieved slowly, most of the retrievable items are likely to meet with success quickly and be assigned high JOLs. There will be a few stragglers, however. It is these that are expected to be produced increasingly slowly and with decreasing JOL values. Thus, the model makes the prediction, consistent with the quadratic RT functions of Son and Metcalfe (2005) and the data presented here, that the slow judgments should be those that are neither very high nor very low, but rather in the middle.

In summary, then, these experiments provide evidence that there are two successive processes that underlie delayed JOLs. The first process is recognition of the cue, and it occurs quickly. This process accounts for the observed very fast RTs given to some 'don't know' responses. The third experiment showed that when the JOLs were made under a deadline procedure, these fast JOLs were responsive only to variations in the familiarity of the cue, as would be expected if they were based on cue recognition. The relative accuracy of these fast JOLs is above chance—if the person does not recognize the cue, they have virtually no chance of recalling the target, and this alone produces above-chance JOL to recall gammas. However, there is no discrimination among the cues that are recognized, so more fine-grained predictions about future recall performance are not possible from this first stage. The attempt at retrieval, as is postulated to occur in the second stage, should increase the JOL relative accuracy further. This stage indicates whether the recall process is successful. If it is, then presumably it is likely to also be successful later, and hence the results of this second stage are highly diagnostic of whether the target item will be retrievable later. Thus, the second, attempted-retrieval stage results in higher relative accuracy than the first stage alone. This prediction was confirmed in the present experiments and has previously been observed by Benjamin (2005). The second stage, which is an attempt at retrieval of the target, is sensitive to experimental variations in target retrievability, as was shown here in all three experiments. All three experiments confirmed the prediction that memory enhancement should obtain primarily with slow JOLs—which presumably entail retrieval of

the target, and not with fast JOLs—which are less likely to entail target retrieval. This simple dual-process model, then, can account for these findings in the delayed JOL paradigm and provides a foundation for further understanding of how people make such metacognitive judgments.

Decades ago, Kolers and Palef (1977) raised the question of “knowing not”: How could people know that they do not know? Furthermore, how could they know that they do not know quickly? At that time in the history of psychology, search models of memory retrieval were popular, though what puzzled Kolers and Palef (1977) may apply even without recourse to a search metaphor. Should the person not have to laboriously exhaust all of their memory knowledge store, coming up with nothing, to reach the conclusion that the desired information is not there? And should that process, which allows the conclusion that they do not know, not take a long time? How could it be possible that a person could answer that they did not know very quickly—even more quickly, sometimes, than that they knew something? To use an analogy based on a search metaphor of memory, if a person is asked to say whether she knows where she left her iPhone, should she not have to search until she either finds it (to say she knows) or search a long time, and perhaps exhaustively, and eventually give up (to say that she does not know)? It should take less time to find than not find, because at the time the iPhone is found, there are still a large (maybe infinite) number of places where the person could still look if it had not yet been found. Each will take some time to explore. By this rationale, knowing not should be a long and tedious process. And yet, people are often quick to say they don’t know. The answer to this dilemma, given substance in the results of the present article, is that there is another process that precedes the search. To revert to the analogy, she asks herself, “Hmmm, iPhone?” And if the answer is, “I don’t have an iPhone,” she gives a quick ‘don’t know’ response and does not search at all.

References

- Ashby, F. G. (2000). A stochastic version of general recognition theory. *Journal of Mathematical Psychology, 44*, 310–329.
- Begg, I., Duft, S., Lalonde, P., Melnick, R., & Sanvito, J. (1989). Memory predictions are based on ease of processing. *Journal of Memory and Language, 28*, 610–632.
- Benjamin, A. S. (2005). Recognition memory and introspective remember/know judgments: Evidence for the influence of distractor plausibility on “remembering” and a caution about purportedly nonparametric measures. *Memory & Cognition, 33*, 261–269.
- Benjamin, A. S., & Bjork, R. A. (1996). Retrieval fluency as a metacognitive index. In L. M. Reder (Ed.), *Implicit memory and metacognition: The 27th Carnegie Symposium on Cognition* (pp. 309–338). Hillsdale, NJ: Erlbaum.
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*, 55–68.
- Butler, A. C., & Roediger, H. L., III. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology, 19*, 514–527.
- Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgment of learning (JOL) and the delayed JOL effect. *Memory & Cognition, 20*, 374–380.
- Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory and Language, 36*, 34–49.
- Dunlosky, J., Rawson, K. A., & Middleton, E. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language, 52*, 551–565.
- Finn, B. (in press). Framing effects on metacognitive monitoring and control. *Memory & Cognition*.
- Finn, B., & Metcalfe, J. (2007). The role of memory for past test in the underconfidence with practice effect. *Journal of Experimental Psychology: Learning Memory and Cognition, 33*, 238–244.
- Finn, B., & Metcalfe, J. (2008). Judgments of learning are influenced by memory for past test. *Journal of Memory and Language, 58*, 19–34.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General, 116*, 119–136.
- Karpicke, J. D., & Roediger, H. L. (2008). The critical importance of retrieval for learning. *Science, 319*, 966–968.
- Kelemen, W. L., & Weaver, C. A. (1997). Enhanced metamemory at delays: Why do judgments of learning improve over time? *Journal of Experimental Psychology: Learning, Memory and Cognition, 23*, 1394–1409.
- Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition, 31*, 918–929.
- Kolers, P. A., & Palef, S. R. (1977). Knowing not. *Memory & Cognition, 5*, 553–558.
- Koriat, A. (1997). Monitoring one’s knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*, 349–370.
- 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.
- Luce, R. D. (1986). *Response times*. New York: Oxford University Press.
- McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16*, 192–201.
- McDaniel, M. A., Kowitz, M. D., & Dunay, P. K. (1989). Altering memory through recall: The effects of cue-guided retrieval processing. *Memory & Cognition, 17*, 423–434.
- McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 371–385.
- Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin and Review, 15*, 174–179.
- Nelson, T. O., & Dunlosky, J. (1991). When people’s judgments of learning (JOL) are extremely accurate at predicting subsequent recall: The delayed-JOL effect. *Psychological Science, 2*, 267–270.
- Nelson, T. O., & Dunlosky, J. (1992). How shall we explain the delayed-judgment-of-learning effect? *Psychological Science, 3*, 317–318.
- Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multi-trial learning. *Psychological Science, 5*, 207–213.
- Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological Methods, 9*, 53–69.
- Pashler, H., Cepeda, N. J., Wixted, J. T., & Rohrer, D. (2005). When does feedback facilitate learning of words? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 3–8.
- Pashler, H., Zarow, G., & Triplett, B. (2003). Is temporal spacing of tests helpful even when it inflates error rates? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1051–1057.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85*, 59–108.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review, 106*, 261–300.
- Reber, P. J., Alvarez, P., & Squire, L. R. (1997). Reaction time distribu-

tions across normal forgetting: Searching for markers of memory consolidation. *Learning and Memory*, 4, 284–290.

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

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving*. (pp. 3–39). Hillsdale, NJ: Erlbaum.

Son, L. K., & Metcalfe, J. (2005). Judgments of learning: Evidence for a two-stage model. *Memory & Cognition*, 33, 1116–1129.

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality:

Judgments of learning may alter what they are intended to assess. *Psychological Science*, 3, 315–316.

Weaver, C. A., III, & Kelemen, W. L. (2003). Processing similarity does not improve metamemory: Evidence against transfer-appropriate monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 1058–1065.

Whitten, W. B., & Bjork, R. A. (1977). Learning from tests: Effects of spacing. *Journal of Verbal Learning and Verbal Behavior*, 16, 465–478.

Received November 12, 2007
 Revision received April 12, 2008
 Accepted April 16, 2008 ■



**AMERICAN PSYCHOLOGICAL ASSOCIATION
 SUBSCRIPTION CLAIMS INFORMATION**

Today's Date: _____

We provide this form to assist members, institutions, and nonmember individuals with any subscription problems. With the appropriate information we can begin a resolution. If you use the services of an agent, please do **NOT** duplicate claims through them and directly to us. **PLEASE PRINT CLEARLY AND IN INK IF POSSIBLE.**

PRINT FULL NAME OR KEY NAME OF INSTITUTION _____		MEMBER OR CUSTOMER NUMBER (MAY BE FOUND ON ANY PAST ISSUE LABEL) _____
ADDRESS _____		DATE YOUR ORDER WAS MAILED (OR PHONED) _____
CITY _____	STATE/COUNTRY _____	ZIP _____
YOUR NAME AND PHONE NUMBER _____		ISSUES: <input type="checkbox"/> MISSING <input type="checkbox"/> DAMAGED
TITLE _____	VOLUME OR YEAR _____	NUMBER OR MONTH _____
_____	_____	_____
_____	_____	_____
_____	_____	_____

 PREPAID CHECK CHARGE
 CHECK/CARD CLEARED DATE: _____

(If possible, send a copy, front and back, of your cancelled check to help us in our research of your claim.)

Thank you. Once a claim is received and resolved, delivery of replacement issues routinely takes 4–6 weeks.

(TO BE FILLED OUT BY APA STAFF)

DATE RECEIVED: _____	DATE OF ACTION: _____
ACTION TAKEN: _____	INV. NO. & DATE: _____
STAFF NAME: _____	LABEL NO. & DATE: _____

Send this form to APA Subscription Claims, 750 First Street, NE, Washington, DC 20002-4242

PLEASE DO NOT REMOVE. A PHOTOCOPY MAY BE USED.