

Dissociations between learners' predicted and actual effects of level of processing and assigned point value on subsequent memory performance

Michael S. Cohen<sup>1</sup>, Michael C. Friedman<sup>1,2</sup>, Nicholas C. Soderstrom<sup>1</sup>, Alan D. Castel<sup>1</sup>,  
Robert A. Bjork<sup>1</sup>

<sup>1</sup>Dept. of Psychology, University of California, Los Angeles

<sup>2</sup>Harvard Initiative for Learning and Teaching, Harvard University

#### Author Note

We thank members of Cogfog for helpful suggestions related to this project, and we thank Neil Charness for providing a copy of the materials used in Experiment 3. We also thank Hillel Dodge Cogan, Jonathan Shapiro, Edrick Pirveysian, Zach Merrill, Brooke Hollyfield, Negan Boostani, Daniella Kaufman, Wesley Horng, Shruti Ullas, Yasmine Sherafat, Katie Swinnerton, James Mutter, and Andrea Del Castillo for help with data collection. Grant No. 29192G from the McDonnell Foundation supported this research. Portions of this research were presented at the 51<sup>st</sup> annual meeting of the Psychonomic Society, St. Louis, MO.

Correspondence should be addressed to Michael S. Cohen, Dept. of Psychology, University of California, Los Angeles, CA 90095. E-mail: mcohen1@ucla.edu.

### Abstract

Prior work (e.g., Koriat, 1997) has shown that when learners predict how well they will remember an item being studied (metacognitive monitoring), they account for certain factors that predict later memory, but ignore other relevant factors. In the present study, we examined how level of processing and assigned value affect metacognitive judgments. When the cue of a cue-target pair promoted the use of deeper processing at encoding, people predicted better cued-recall than they did for items associated with cues that related to the target on a more shallow level. However, when different levels of processing (LOP) were induced via questions at encoding, people predicted little to no benefit for deep levels of processing on a later free recall test. Thus, it appears that the degree to which the LOP condition is integrated with the to-be-studied item affects the degree to which LOP is incorporated into metacognitive predictions. In contrast to how level of processing affects JOLs, we find that items that are arbitrarily defined as being more valuable are given higher JOLs, even under conditions in which value does not reliably affect recall, and even when more diagnostic cues (e.g., LOP) are available. Thus, the present work refines our knowledge of the types of cues that people do or do not account for when making metacognitive monitoring judgments such as judgments of learning (JOLs).

*Keywords:* Level of processing, point value, metamemory, judgments of learning, cue utilization

Dissociations between learners' predicted and actual effects of level of processing and assigned point value on subsequent memory performance

Effective learning depends, in large part, on one's ability to accurately predict whether studied material will be remembered later. People act on such subjective assessments, and under certain conditions this metacognitive monitoring process is fairly accurate. When difficulty is based on relatedness of paired associates, for example, or on other intrinsic features of an item, people are reliably above chance at predicting which items they will be able to recall later (e.g., Arbuckle & Cuddy, 1969; Hart, 1965, 1967; Koriat, Bjork, Sheffer, & Bar, 2004; Koriat, 1997). However, in other cases, people appear to rely on cues that are not predictive of later memory performance (e.g., Rhodes & Castel, 2008), ignore cues that are predictive of later learning (Koriat et al., 2004), or otherwise fail to appreciate how aspects of the learning environment will affect later memory (e.g., Benjamin, Bjork, & Schwartz, 1998).

### **Levels of Processing**

One factor that has substantial effects on memory is level of processing (LOP) ( Craik & Lockhart, 1972). In a classic set of experiments, Craik and Tulving (1975) found that asking questions about semantic properties of a to-be-remembered word (e.g., "Is the word a type of fish?") led to the best performance on a later recall test; in other words, processing the words on a deep level was most effective. By contrast, phonetic questions (e.g., "Does the word rhyme with weight?") led to moderate levels of recall, and surface-level questions (e.g., "Is the word printed in capital letters?") led to the worst performance, indicating that shallow processing is

least beneficial for later recall. These results have been widely replicated, and have been very influential to our understanding of human memory.

Previous studies (e.g., Cutting, 1975; Shaw & Craik, 1989; Bieman-Copland & Charness, 1994; Matvey et al., 2002) have suggested that people are sometimes aware that deeper levels of processing lead to better memory, but participants' predictions generally underestimate such effects substantially. Cutting (1975), for example, found little to no difference in global estimates of recall performance for subjects who had performed "e-check" judgments (saying, simply, whether a given word contained the letter 'e') compared to those who had performed pleasantness judgments (deep LOP), even though the rate of recall was substantially higher for those in the deep LOP condition. Shaw and Craik (1989) found similar results using a paradigm in which item-by-item JOLs were compared to actual recall. Both at study and at test, the to-be-remembered target words were each paired with a specific cue stimulus that was associated with the target word via one of 3 levels of processing. Predictions of recall were approximately 15% higher for items in the deepest LOP condition relative to the shallowest condition, whereas actual recall was approximately 40% better for those in the deep LOP condition. Bieman-Copland and Charness (1994) refined the methods of Shaw and Craik, but still found that people underestimated the difference in recall between deep and shallow levels of processing when making item-by-item JOLs. The degree of underestimation of LOP effects was reduced on a second study-test cycle, however.

Notably, Matvey et al. (2002) found that when people are asked to make global predictions of recall for each of 3 levels of processing before having any

experience with the task, they tend to give the highest judgments to rhyming cues. However, similar to the earlier studies, learners' average item-by-item JOLs showed a tendency to predict better recall for the deepest level of processing (i.e., category), even on the first study list. In addition, retrospective confidence judgments and global postdictions taken after the first cued-recall test both show that people tended to accurately predict much higher recall for category-cued items. Consistent with the Bieman-Copland and Charness (1994) findings, Matvey et al. also showed that global predictions and item-by-item JOLs reflect a reliably stronger preference for deep levels of processing on a second list than they had on the first list, although these judgments still appeared to underestimate LOP effects. Thus, even when people do not start out with an appreciation of LOP effects, they do appear able to learn about the benefits of deep levels of processing with practice.

A related question, one that has not been examined prior to the present study, is whether participants' predictions differ as a function of whether the LOP orienting question matches, or does not match, the target word (i.e., "yes" or "no" answers). Craik and Tulving (1975) noted that the effects of LOP condition on recall are stronger when the answer to the question is "yes", presumably because a matching semantic or phonetic LOP question provides additional retrieval cues for recalling the target word. Prior studies of metacognition and LOP have used a manipulation in which the item is always associated with the LOP cue (following Shaw & Craik, 1989). In the present study, our decision to use the free-recall LOP manipulation introduced by Craik and Tulving was prompted, in part, by the opportunity to examine whether participants are sensitive to how later recall varies

with “yes” and “no” responding. As is discussed further below, this factor affects the degree to which LOP is incorporated as an intrinsic cue vs. an extrinsic cue in metacognitive monitoring judgments (following Koriat, 1997).

### **The Effect of Incentives or Value on Recall**

Similar to depth of processing, the importance, or value, that we place on information can have a substantial impact on how that information is processed and, in turn, how well it will be remembered. In particular, when each to-be-remembered word is assigned an arbitrary point value, the more valuable words (i.e., words paired with larger values) receive more attentional focus and are remembered better than the less valuable words (e.g., Castel, Benjamin, Craik, & Watkins, 2002; Loftus & Wickens, 1970; Watkins & Bloom, 1999). The degree to which value affects performance has also been shown to increase on later study-test cycles (Castel, 2008; Castel, Murayama, Friedman, McGillivray, & Link, 2013). Thus, once people have completed the first few study-test cycles, and presumably realized that they cannot recall all of the words, they are able to make better use of value to guide their learning, and thereby become more selective.

There is some evidence that people do account for the recall benefit afforded by higher point values when making metacognitive predictions of recall.

Soderstrom and McCabe (2011) varied both associative relatedness of word pairs and point value within-subjects. When point values were presented before the words, high-value words were remembered better than low-value words, and JOLs accounted for both this effect and for benefits of semantic relatedness. When point values were presented after the words, however, those values did not affect actual

recall, but participants still gave higher JOLs to high-value words. Similarly, Festini, Hartley, Tauber, and Rhodes (2012) found that when people were asked to remember nonwords that were labeled as either surnames or occupations, value effects on JOLs were larger than effects on actual recall. When the stimuli were words, but participants were required to use maintenance rehearsal for all items, value no longer had an effect on recall, but JOLs were still reliably higher for high-value items. Thus, unlike LOP effects, it appears that, if anything, people tend to overestimate the effect that value has on recall when making metacognitive judgments.

### **Experiment 1**

Experiment 1 represents an initial examination of how people jointly account for effects of levels of processing and value when making JOLs. Level of processing generally has a strong effect on recall, but, as noted above, people generally underestimate its effects when making JOLs. Past work (e.g., Matvey et al., 2002) has also shown that with practice, people begin to predict stronger effects of LOP than they do initially. On the other hand, people do tend to be sensitive to value when making JOLs (e.g., Soderstrom & McCabe, 2011). However, previous work has shown that effects of value on free recall only emerge after multiple study-test cycles. Given the prior work showing that people expect point value to affect recall even when it does not, it is of current interest how metacognitive predictions of the effects of both value and LOP change across lists.

## Method

**Participants.** We tested 48 participants from the undergraduate student subject pool at the University of California, Los Angeles (UCLA). An additional 2 participants were run but were replaced, one for failure to follow instructions, the other for being a clear outlier in terms of age.

**Materials.** Items were drawn from clusters 6 and 7 of the Toglia and Battig (1978/2009) “Colorado” word norms. All target words were 4-7 letter, 1-2 syllable nouns, rated as highly familiar (range 5.5-7 on a 1-7 scale), moderately high on concreteness and imagery (range 4-6 on a 1-7 scale), and moderate in pleasantness (range 2.5-5.5 on a 1-7 scale).

For each item, point value (1, 5, or 10), question answer (yes or no), and case (uppercase or lowercase) were manipulated within-subjects, such that each list had one word at each combination of LOP, value, and yes/no response. In addition, half of the words on each list were in caps while half were in lowercase, with capitalization balanced across other conditions as much as possible. Assignment of item to condition was counterbalanced across subjects, producing 12 counterbalanced conditions.<sup>1</sup> Assignment of items to lists and order within lists was

---

<sup>1</sup> There were two errors that caused deviation from the intended procedures in Experiment 1. On the phonetic condition of one word list for all subjects, the assignment of two words to answer type condition was reversed. Thus, instead of having one word at each combination of LOP x yes/no x value, there were two “yes” words at one value and two “no” words at another value for this word list. The precise point values involved were counterbalanced across participants. In addition, a programming error on one of the 12 counterbalancing conditions caused the case (uppercase vs. lowercase) to be reversed on all items for one of the four lists, which meant that the counterbalancing of case was not precisely even across other factors for one out of every 12 participants. Note that because list order was randomized, there was not a regular distribution of either error through list positions.



held constant, and order of list presentation was randomized independently for each subject.

In constructing the lists, we produced an LOP question for each word at each level. The questions were in the form of the case (surface), rhyme (phonetic), and category (semantic) cues used by Craik and Tulving (1975) (see Appendix). False rhyme and category LOP questions, in which the correct answer was to be “no,” were taken from true LOP questions for other items. Words from the initial pool for which a suitable semantic or phonetic question could not be generated were discarded. Note that in Experiment 1, each item was always associated with a particular level of processing (i.e., level of processing was not counterbalanced).

**Procedure.** After reading through the instructions on-screen, 6 sample items were presented to familiarize participants with the structure of the task. Participants then studied 4 lists of 18 words each, with a free recall test after each list. For each word, an LOP question (either semantic, phonetic, or surface) was shown for 3 seconds prior to the word appearing on the screen. A point value (either 1, 5, or 10 points) also appeared with the LOP question. The word was then shown for 1 sec, after which participants answered the LOP question by entering “y” or “n” on the keyboard and then entered a judgment of learning (JOL) on a scale from 0 to 100 in numeric form on the keyboard.

At the end of each list, participants completed a 30 second geography-based distractor task (e.g., name U.S. states, name countries in Europe, etc.). They were then given 60 seconds to recall as many words as possible from the preceding list. Responses were made verbally and were recorded by an experimenter who was in

the room. The experimenter immediately computed a point total for all words recalled on that list, and participants were given this score at the conclusion of each recall test, before going on to the next list. At the end of the study, participants were given a questionnaire, which included global judgments about which LOP condition seemed to be best. Responses were grouped for analysis as preferring semantic, phonetic, or surface/no preference.

## Results

**Levels of Processing.** We first examined effects of levels of processing on the proportion of items recalled. Because effects of LOP differ based on whether the given question matches the to-be-remembered word (i.e., answer type, whether the correct answer is “yes” or “no”), we also included this factor in all analyses. A  $4 \times 3 \times 2$  (List  $\times$  LOP  $\times$  Answer type) repeated-measures ANOVA on percentage recall, collapsing across point values<sup>2</sup>, yielded a main effect of LOP,  $F(2, 94) = 7.55$ ,  $MSE = 666.29$ ,  $p = .001$ ,  $\eta_p^2 = .14$  (Figure 1A), with no list  $\times$  LOP interaction,  $F(6, 282) < 1$ . This pattern of effects indicated a consistent effect of LOP condition in the expected direction, with deeper levels of processing yielding better recall across lists. We also observed a main effect of answer type,  $F(1, 47) = 13.89$ ,  $MSE = 765.67$ ,  $p = .001$ ,  $\eta_p^2 = .23$ , with “yes” questions producing higher recall than “no” questions. In addition, there was a significant interaction between answer type and LOP condition,  $F(2, 94) = 3.61$ ,  $MSE = 701.59$ ,  $p = .031$ ,  $\eta_p^2 = .07$ , showing that LOP effects were stronger

---

<sup>2</sup> It was not possible to run an omnibus analysis of variance (ANOVA) that included all factors (LOP, value, list, and answer type) because of the error in list construction described above.

when the correct answer was “yes” (Figure 1A). There was no 3-way (List x LOP x Answer type) interaction, meaning that the previously described interaction is consistent across lists. Finally, there was a main effect of list,  $F(3, 141) = 5.46$ ,  $MSE = 695.74$ ,  $\eta_p^2 = .10$ ,  $p = .001$ , such that recall tended to improve on later lists.

Probing the interaction between answer type and LOP condition using separate one-way repeated measures ANOVAs for “yes” and “no” items revealed a significant LOP effect in the predicted direction for “yes” items,  $F(2, 94) = 9.95$ ,  $MSE = 182.10$ ,  $p < .001$ ,  $\eta_p^2 = .18$ . While there was a trend towards an LOP effect in the predicted direction for “no” items as well, this effect, surprisingly, was not significant,  $F(2, 94) = 0.49$ ,  $p = .613$ . Thus, while there was a trend for LOP to affect recall in the expected direction for both “yes” and “no” items in both experiments, the effect of LOP condition on recall in Experiment 1 was reliable only with “yes” questions.

We also examined effects of levels of processing on metacognitive judgments using a 4 x 3 x 2 (List x LOP x Answer type) repeated-measures ANOVA on judgments of learning (JOLs). This analysis showed no main effect of LOP,  $F < 1$  (Figure 1B), and no interaction between list and LOP,  $F < 1$ . Thus, the main effect of LOP that was present in the recall data was not predicted in JOLs. There was, however, a main effect of answer type,  $F(1, 47) = 59.63$ ,  $MSE = 190.40$ ,  $p < .001$ ,  $\eta_p^2 = .56$ , with “yes” items receiving higher JOLs than “no” items. In addition, there was an interaction between LOP and answer type (yes/no),  $F(2, 94) = 7.12$ ,  $MSE = 150.21$ ,  $p = .001$ ,  $\eta_p^2 = .13$ , with no list x LOP x answer type interaction,  $F < 1$ . Finally, a main effect of list was observed on JOLs,  $F(3, 141) = 49.14$ ,  $MSE = 444.96$ ,  $p < .001$ ,  $\eta_p^2 =$

.51, but this effect went in the opposite direction of the effect of list on recall, such that mean JOLs decreased across lists.

Probing the LOP x answer type interaction with separate one-way ANOVAs for “yes” items and “no” items yielded significant effects in opposite directions (Figure 1B). On “yes” items, there was a significant effect of level of processing in the expected direction,  $F(2, 94) = 4.45, MSE = 39.34, p = .014, \eta_p^2 = .09$ . Thus, people do incorporate effects of LOP into their JOLs on “yes” items. However, on “no” items, there was actually an effect in the opposite direction,  $F(2, 94) = 4.00, MSE = 25.98, p = .022, \eta_p^2 = .08$ . This finding means that people appear to have believed that shallow “no” items were remembered better than deep “no” items, while there was not an effect in that direction for actual recall.

To compare more directly JOLs to recall, we ran a  $2 \times 4 \times 3 \times 2$  (measure x list x LOP x answer type) repeated measures ANOVA that included both JOLs and recall data. Note that this analysis is consistent with Koriat’s (1997) suggestion that JOLs collected in terms of a 0-100 probability scale can be compared directly with the proportion of items recalled, in order to assess calibration. We saw a main effect of measure,  $F(1, 47) = 6.52, MSE = 4513.48, p = .014, \eta_p^2 = .12$ , such that learners were overconfident on average. We also found an interaction between measure and LOP,  $F(2, 94) = 6.59, MSE = 353.79, p = .002, \eta_p^2 = .12$ , which confirmed that there was a main effect of LOP on recall that was not apparent in JOLs. There was also an interaction between measure and list,  $F(3, 141) = 39.29, MSE = 531.99, p < .001, \eta_p^2 = .46$ , as JOLs tended to decrease across lists while actual recall tended to increase. Other interactions with measure were not significant,  $F_s < 1.45$ .

Because there was an interaction between answer type and LOP condition on recall and on JOLs, we also ran separate 2 x 4 x 3 (measure x list x LOP) repeated-measures ANOVAs comparing JOLs with recall on “yes” items and on “no” items. For “yes” items, there was a significant measure x LOP interaction,  $F(2, 94) = 4.03$ ,  $MSE = 454.56$ ,  $p = .021$ ,  $\eta_p^2 = .08$ . Thus, while there was an effect of LOP on JOLs for “yes” items, it is significantly underestimated. For “no” items, there was a trend towards a measure x LOP interaction, but it was not significant,  $F(2, 94) = 2.04$ ,  $p = .136$ .

**Value.** We then examined effects of point value on percentage of items recalled, collapsing across level of processing, and ignoring effects of list described above. A 4 x 3 (list x value) within-subjects ANOVA on percentage recall showed a main effect of value,  $F(2, 94) = 10.69$ ,  $MSE = 493.18$ ,  $p < .001$ ,  $\eta_p^2 = .19$ , and a list x value interaction,  $F(6, 282) = 2.87$ ,  $MSE = 309.31$ ,  $p = .010$ ,  $\eta_p^2 = .06$  (Figure 2A). Probing the interaction with one-way repeated measures ANOVAs revealed no significant effect of value on list 1,  $F(2, 94) = 1.71$ , or list 2,  $F(2, 94) = 2.21$ , but there were significant effects of value in the expected direction on list 3,  $F(2, 94) = 5.38$ ,  $MSE = 410.14$ ,  $p = .006$ ,  $\eta_p^2 = .10$ , and on list 4,  $F(2, 94) = 11.68$ ,  $MSE = 385.02$ ,  $p < .001$ ,  $\eta_p^2 = .20$ . Thus, higher value items were remembered better than were lower value items, but this effect was only reliable on lists 3 and 4.

We also examined effects of point value on metacognitive judgments, collapsing across level of processing, using a 4 x 3 (list x value) repeated-measures ANOVA on JOLs (Figure 2B). This analysis showed a main effect of value,  $F(2, 94) = 14.41$ ,  $MSE = 104.39$ ,  $p < .001$ ,  $\eta_p^2 = .24$ , but no interaction between list and value,

$F(6, 282) = 1.193$ . Thus, participants appear to have predicted a consistent effect of value across lists.

We then ran a  $2 \times 4 \times 3$  (measure  $\times$  list  $\times$  value) repeated measures ANOVA to more directly compare value effects in recall to the corresponding JOLs. The most notable result of this analysis was a significant measure  $\times$  list  $\times$  value interaction,  $F(6, 282) = 2.61$ ,  $MSE = 153.00$ ,  $p = .018$ ,  $\eta_p^2 = .05$ , which reflected a list  $\times$  value interaction in recall that was not observed in JOLs (Figure 2).

**Interactive effects.** We also examined possible interactions between value and LOP condition on recall and metacognitive judgments. A  $4 \times 3 \times 3$  (list  $\times$  value  $\times$  LOP) repeated-measures ANOVA on percentage recall, collapsing across answer type, showed no significant interaction between value and LOP,  $F < 1$ , and no 3-way list  $\times$  value  $\times$  LOP interaction,  $F < 1$ . A similar  $4 \times 3 \times 3$  (list  $\times$  value  $\times$  LOP) within-subjects ANOVA on JOLs also showed no interaction between value and LOP,  $F < 1$ , and no 3-way list  $\times$  value  $\times$  LOP interaction,  $F < 1$ .

**Global judgments.** Finally, we examined global judgments of levels of processing effects from the post-study questionnaires. Surprisingly, we observed that global judgments about which LOP condition was best differed reliably from what would be expected based on the actual recall data,  $\chi^2 = 13.40$ ,  $df = 2$ ,  $p = .001$ , with fewer individuals than would be expected believing that deep LOP was best (Table 1). Instead, we found a preference for rhyming questions, which was reliably different from a random distribution of choices,  $\chi^2 = 6.13$ ,  $df = 2$ ,  $p = .047$ . This is similar to a pattern previously observed by Matvey et al. (2002), as discussed in the introduction, but under different circumstances.

### Discussion

Overall, we found that participants largely fail to appreciate the benefits of deep levels of processing, even after multiple study-test cycles. The tendency to prefer phonetic rather than semantic orienting questions when making global judgments is particularly striking, given that global judgments are made after 4 study-test cycles. Matvey et al. (2002) did previously find a tendency for people to give higher global predictions for items processed at a phonetic level, but this trend was only apparent prior to them gaining any experience with the task.

Participants in the present study also had difficulty appreciating the benefits of deep levels of processing in item-by-item judgments. They do show some awareness of LOP effects when we only look at items for which the question matches the to-be-remembered word (“yes” items), but this effect is greatly underestimated. Moreover, when the orienting question does not match the to-be-remembered word (“no” items), the effect of LOP on JOLs is actually in the reverse direction from how LOP affects recall. While the precise reason for this reversal is not clear, one speculative explanation is that the distinction between “yes” and “no” items in the deep and phonetic conditions is more salient than the distinction between levels of processing for “no” items. Thus, for “no” items in the deep and phonetic LOP conditions, JOLs may be decreased relative to the “yes” items in that LOP condition. For the surface items, orienting questions are identical for “yes” and “no” items, so the “yes”/“no” distinction is less meaningful, and the predicted difference between “yes” and “no” items should be smaller in the other conditions. The apparent reversal in effects of LOP on JOLs for “no” items could be explained in

this context if we suppose that participants are exaggerating the degree to which memory for “yes” vs. “no” items differs across different LOP conditions.

We also ran a variant of Experiment 1 with an additional 108 participants, which was identical to Experiment 1 except that participants were not asked to provide JOLs on each item. We do not report those data in full here due to space considerations; however, the data from global post-study judgments is notable. Only 35% of participants appreciated that a deep level of processing was best, even though the deep level of processing actually yielded the highest recall for 69% of participants. In contrast, the largest number, 47% of participants, thought a phonetic level of processing was best, even though it was actually best for only 20% of individuals.

We observed these results despite the fact that the benefit of deep LOP over shallower levels was significantly stronger in this variant than in Experiment 1, most likely due to the fact that making a JOL can induce deeper processing on all items, including those intended to be studied at only a shallow level of processing (cf., Dougherty, Scheck, Nelson, and Narens, 2005; Fujita, 2010). These data provide further evidence that people do not tend to appreciate the benefit of levels of processing, even when those benefits are greater than those observed in Experiment 1.

At the same time, the participants in Experiment 1 did not appreciate the fact that point value only affects recall on later lists; instead, they predict a consistent effect of value beginning on the very first list. This finding is consistent with



previous work (e.g., Soderstrom & McCabe, 2011) showing that JOLs are sensitive to point value even when actual recall is not.

### **Experiment 2**

In previous studies (e.g., Shaw & Craik, 1989; Bieman-Copeland & Charness, 1994; Matvey et al., 2002), JOLs reflected more of an appreciation for effects of levels of processing than we saw in Experiment 1. Our preferred explanation is that the key difference between Experiment 1 and previous studies is that we manipulated levels of processing in a way that was less integrated with the to-be-learned words—which presumably makes LOP more of an extrinsic cue (cf., Koriat, 1997), and thus is harder for learners to appreciate. At the same time, there are two other ways in which Experiment 1 differed from earlier studies, which could provide alternate explanations for the differing results. The first is that words varied on two factors, LOP and point value, rather than merely varying on LOP. In addition, to avoid having an unwieldy number of counterbalancing conditions, we chose not to counterbalance level of processing across items in Experiment 1. However, this design choice meant that the observed effects of levels of processing could have been due to item effects. Experiment 2 was designed to replicate Experiment 1, but with the design modified to focus only on levels of processing and to rule out item effects.

### **Method**

**Participants.** We tested 60 participants from the undergraduate student subject pool at UCLA. An additional 2 participants were tested but were replaced owing to experimenter error during testing.<sup>3</sup>

**Materials.** The to-be-remembered word lists were the same as those used in Experiment 1. The LOP orienting questions were from the same set as those used in Experiment 1 (see Appendix); however, some of the specific orienting questions were not used in Experiment 1. With the new orienting questions we were able to counterbalance LOP condition, answer type (yes/no), and capitalization (upper/lower case), for a total of 12 counterbalancing conditions.

**Procedure.** The procedure was largely identical to Experiment 1, except that no point values were shown. Thus, the feedback given at the end of each list was in terms of number of items remembered, rather than in terms of point total.

## Results

**Recall.** We first examined effects of level of processing on percentage of items recalled using a 4 x 3 x 2 (list x LOP x answer type) within-subjects ANOVA. We found a main effect of level of processing,  $F(2, 118) = 14.19$ ,  $MSE = 717.97$ ,  $p < .001$ ,  $\eta_p^2 = .19$ , such that a deep level of processing led to better recall than shallower levels of processing (Figure 3A). There was also a main effect of answer type,  $F(1, 59) = 62.57$ ,  $MSE = 629.89$ ,  $p < .001$ ,  $\eta_p^2 = .52$ , with “yes” items remembered better than “no” items (Figure 3A). In addition, we found a main effect

---

<sup>3</sup> We eliminated 2 individuals for which erroneous feedback was given on two lists, but retained 7 individuals for whom feedback was off by one item on a single list. Eliminating those 7 individuals did not change any of the major results, although some effects of LOP on JOLs for “no” items, and the answer type x list interaction for JOLs, become only marginally significant.

of list,  $F(3, 177) = 3.42$ ,  $MSE = 620.53$ ,  $p = .019$ ,  $\eta_p^2 = .06$ , such that recall improved on later lists. There was a trend towards an interaction between answer type and LOP (Figure 3A), such that LOP effects were greater for “yes” items, but this effect was not significant,  $F(2, 118) = 1.56$ ,  $p = .214$ . No interactions with list approached significance, all  $F_s < 1$ .

Given the interaction between answer type and LOP observed in Experiment 1, we also conducted separate one-way ANOVAs to look for LOP effects on “yes” items and “no” items as planned comparisons. There was a significant effect of LOP on “yes” items,  $F(2, 118) = 9.26$ ,  $MSE = 208.55$ ,  $p < .001$ ,  $\eta_p^2 = .14$ . The effect of LOP on “no” items was also significant,  $F(2, 118) = 7.27$ ,  $MSE = 115.62$ ,  $p = .001$ ,  $\eta_p^2 = .11$ . Regardless of answer type, deep levels of processing yielded better recall than shallower levels of processing.

**Judgments of Learning.** We then examined effects of levels of processing on metacognitive judgments, using a 4 x 3 x 2 (list x LOP x answer type) within-subjects ANOVA on JOLs. We found a main effect of answer type,  $F(1, 59) = 35.50$ ,  $MSE = 627.83$ ,  $p < .001$ ,  $\eta_p^2 = .38$ , such that people predicted better recall for “yes” items than for “no” items (Figure 3B). In addition, there was a main effect of list,  $F(3, 105) = 69.81$ ,  $MSE = 786.92$ ,  $p < .001$ ,  $\eta_p^2 = .54$ , such that JOLs were lower on later lists than on earlier lists; this effect of list on JOLs was opposite in direction from the actual effect of list on recall, as in Experiment 1. Importantly, there was not a reliable effect of LOP on JOLs,  $F(2, 70) = 1.61$ ,  $p = .205$  (Figure 3B). There was an interaction between answer type and LOP,  $F(2, 118) = 13.24$ ,  $MSE = 163.74$ ,  $p < .001$ ,  $\eta_p^2 = .18$  (Figure 3B). A one-way repeated-measures ANOVA, collapsing across lists,

showed a significant main effect of LOP in the correct direction for “yes” items,  $F(2, 118) = 10.93$ ,  $MSE = 41.29$ ,  $p < .001$ ,  $\eta_p^2 = .16$ . For “no” items, however, as in Experiment 1, we found a main effect of LOP in the opposite direction,  $F(2, 70) = 3.31$ ,  $MSE = 38.20$ ,  $p = .040$ ,  $\eta_p^2 = .05$ , such that people predicted better recall for shallower LOPs. Finally, there was an interaction between list and answer type,  $F(3, 177) = 3.60$ ,  $MSE = 157.44$ ,  $p = .015$ ,  $\eta_p^2 = .06$ , with the effect of answer type getting smaller across lists. Other interactions with list were not significant,  $F_s < 1.62$ .

We then made a direct comparison between JOLs and percentage of items recalled using a  $2 \times 4 \times 3 \times 2$  (measure  $\times$  list  $\times$  LOP  $\times$  answer type) within-subjects ANOVA. In addition to the main effects described above, we found a main effect of measure,  $F(1, 59) = 19.27$ ,  $MSE = 4632.54$ ,  $p < .001$ ,  $\eta_p^2 = .25$ , with JOLs higher than recall on average. We also found an interaction between measure and LOP,  $F(2, 118) = 9.73$ ,  $MSE = 420.73$ ,  $p < .001$ ,  $\eta_p^2 = .14$ , but no interaction between measure, answer type, and LOP,  $F < 1$ . Planned comparisons on separate  $2 \times 4 \times 3$  (measure  $\times$  list  $\times$  LOP) within-subjects ANOVAs for both “yes” and “no” items showed a reliable interaction between measure and LOP for “yes” items,  $F(2, 70) = 4.52$ ,  $MSE = 501.98$ ,  $p = .013$ ,  $\eta_p^2 = .07$ , and also for “no” items,  $F(2, 118) = 6.69$ ,  $MSE = 304.09$ ,  $p = .002$ ,  $\eta_p^2 = .10$  (Figure 3). Thus, people’s JOLs reliably underestimated the degree to which LOP affected memory for “yes” items. Predicted effects of LOP were also reliably different from recall, as well as being in the wrong direction, for “no” items. In addition, there was an interaction between measure and list,  $F(3, 177) = 54.28$ ,  $MSE = 716.96$ ,  $p < .001$ ,  $\eta_p^2 = .48$ . As noted above, JOLs tended to decrease across

lists while actual recall increased across lists. No other interactions with measure were significant, all  $F$ s < 2.21.

**Global judgments.** As in Experiment 1, we examined global postdictions that were made after four study-test cycles (Table 1). We again found that people's postdictions about which LOP was better differed reliably from what would be expected based on the actual distribution of LOP effects,  $\chi^2 = 13.70$ ,  $df = 2$ ,  $p = .001$ . We also found that the largest number of individuals tended to prefer a phonetic level of processing, but unlike in Experiment 1, this pattern did not differ reliably from a random distribution of preferences,  $\chi^2 = 0.70$ ,  $df = 2$ ,  $p = .70$ .

### Discussion

It seems clear that the lack of appreciation for levels of processing effects, which we observed in Experiment 1, generally persists even when level of processing is the only obvious difference between items. This conclusion appears to hold true whether we consider global preferences about LOP conditions or differences in item-by-item JOLs as a function of level of processing.

### Experiment 3

Experiment 3 was intended primarily to rule out additional alternate explanations as to why our results differed so strongly from prior studies investigating metacognitive awareness of levels of processing effects. Those earlier studies used cued-recall items in which the level of processing was determined by the nature of the cue, unlike the materials that we used in Experiments 1 and 2. However, there were also other differences that were introduced in the course of blending the LOP manipulation with the value-directed remembering paradigm. For

instance, prior studies did not include feedback on recall performance after each study-test cycle. We also used shorter word lists, as well as a distractor task that was arguably more taxing than those used in previous studies. In order to rule out these and other idiosyncratic factors, we adapted our paradigm to use items taken from the set of cued-recall LOP items created by Bieman-Copland & Charness (1994).

### Method

**Participants.** We tested 36 individuals from the student subject pool at the University of California, Los Angeles. An additional 12 individuals were run but were excluded. Of these, 7 were excluded due to experimenter errors in feedback provided during testing, and 2 were excluded because the experimenter noted serious difficulties with the English language. Three were excluded because they had too much prior knowledge of levels of processing, either because they had taken a particular research methods course that extensively covers the topic, or based on the experimenter's subjective assessment of responses to a question on the post-study questionnaire about prior study of levels of processing. All critical results were the same regardless of whether these individuals were excluded.<sup>4</sup>

**Materials.** We selected 120 of the cued-recall LOP items used by Bieman-Copland & Charness (1994), and re-sorted them into 5 lists of 24 items each. As Bieman-Copland & Charness had done in constructing the original lists, we ensured that only one item from a particular list would be a correct response to a cue on that

---

<sup>4</sup> The only results that differed based on whether these 12 individuals were excluded is that the main effect of list on recall, and the effect of LOP on JOLs during list 1, were both only marginally significant when these individuals were included.

list. Cues were based on semantic features (e.g., “something slippery: ice”), rhyming (e.g., “rhymes with dice: ice”), or initial letters (e.g., “starts with ‘ic’: ice”). Each word could be matched with any one of the three cue types, with the cue type counterbalanced between subjects. The items on each list were divided evenly between the three types of LOP cues.

In addition, each item was assigned a value of either 1 point or 5 points. Note that while Experiment 1 used 3 levels of value, value effects appeared to be largely based on differences between 10 point items and lower-valued items. Thus, we streamlined the design in this experiment to use only two levels of value. Prior studies (e.g., Ariel, Dunlosky, & Bailey, 2009; Toppino & Cohen, 2010) have manipulated value in this way and have found resulting differences in strategy use; thus, we assumed that two values would be sufficient to affect recall in the present experiment. The two values were divided evenly between items in each LOP condition, and were also counterbalanced between subjects.

**Procedure.** The basic structure was similar to Experiment 1. After reading the instructions on-screen, participants were given 6 practice items, and a practice cued recall test. This practice session was followed by 5 full lists, with 24 items in each list. Each trial began with the point value being displayed for 1.5 seconds. Then, we added both the LOP cue and the target word, which were displayed together with the value cue for 4 seconds. Note that unlike in Experiments 1 and 2, there was no question to which to respond when the items were presented. Participants were then allowed up to 8 seconds to enter a JOL on the computer. At the end of each list, after a 30 second geography distractor task, participants

completed a cued recall test. Items appeared in the same order in the cued-recall test as during encoding. For each item on the cued-recall test, the cue appeared alone on the screen for 5 seconds, and participants were to say out loud the corresponding word that had been studied with that cue. The experimenter recorded the answer on paper. As in Experiment 1, participants were given feedback as to their point total before going on to the next list.

### Results

We examined effects of list, LOP cue type, and value on the percentage of items recalled using a 5 x 3 x 2 (list x LOP x value) within-subjects ANOVA. We found a main effect of LOP condition,  $F(2, 70) = 179.76$ ,  $MSE = 927.76$ ,  $p < .001$ ,  $\eta_p^2 = .84$ , with deep LOP cues yielding better recall than rhyme or letter cues (Figure 4A). We also saw a list x LOP interaction,  $F(8, 280) = 2.77$ ,  $MSE = 419.37$ ,  $p = .006$ ,  $\eta_p^2 = .07$  (Figure 4A), but this interaction seems to have reflected fluctuations across lists in recall of items with rhyme and letter cues, and does not appear to be theoretically meaningful. We did not find a main effect of value,  $F < 1$ , nor did we see an interaction between list and value,  $F < 1$  (Figure 5A). We did see a main effect of list,  $F(4, 140) = 2.93$ ,  $MSE = 463.24$ ,  $p = .023$ ,  $\eta_p^2 = .08$ , reflecting improvement across the first four lists, but poorer recall on list 5. Finally, we did not find an interaction between LOP and value, nor did we find a 3-way interaction between list, LOP and value,  $F_s < 1.06$ .

We also examined metacognitive judgments using a 5 x 3 x 2 (list x LOP x value) within-subjects ANOVA on item-by-item JOLs. We also found a main effect of LOP condition,  $F(2, 70) = 52.33$ ,  $MSE = 482.69$ ,  $p < .001$ ,  $\eta_p^2 = .60$ , as people tended



to predict better recall for deeper LOP questions (Figure 4B). In addition, there was a significant interaction between LOP condition and list,  $F(8, 280) = 7.60$ ,  $MSE = 117.49$ ,  $p < .001$ ,  $\eta_p^2 = .18$ , indicating that predictions of LOP effects were stronger on later lists (Figure 4B). Probing the interaction with separate one-way within-subjects ANOVAs for each list showed a reliable effect of LOP in the correct direction even on list 1,  $F(2, 70) = 3.50$ ,  $MSE = 68.67$ ,  $p = .035$ ,  $\eta_p^2 = .09$ . The effect remains reliable on list 2,  $F(2, 70) = 29.00$ ,  $MSE = 90.44$ ,  $p < .001$ ,  $\eta_p^2 = .45$ ; list 3,  $F(2, 70) = 30.18$ ,  $MSE = 93.70$ ,  $p < .001$ ,  $\eta_p^2 = .46$ ; list 4,  $F(2, 70) = 40.79$ ,  $MSE = 99.68$ ,  $p < .001$ ,  $\eta_p^2 = .54$ ; and list 5,  $F(2, 70) = 37.72$ ,  $MSE = 123.45$ ,  $p < .001$ ,  $\eta_p^2 = .52$ . Returning to the omnibus ANOVA, we also found a main effect of value,  $F(1, 35) = 6.11$ ,  $MSE = 292.53$ ,  $p = .018$ ,  $\eta_p^2 = .15$ , as participants predicted better recall for more valuable items (Figure 5B). There was no interaction between list and value,  $F < 1$  (Figure 5B). We additionally observed a main effect of list, with JOLs generally reduced on later lists,  $F(4, 140) = 7.99$ ,  $MSE = 221.24$ ,  $p < .001$ ,  $\eta_p^2 = .19$ . We did not see an interaction between LOP and value, nor did we find an interaction between list, LOP, and value,  $F_s < 1.16$ .

We also compared JOLs to actual percentage recall using a  $2 \times 5 \times 3 \times 2$  (measure  $\times$  list  $\times$  LOP  $\times$  value) within-subjects ANOVA. We saw a main effect of measure,  $F(1, 35) = 52.17$ ,  $MSE = 3221.63$ ,  $\eta_p^2 = .60$ ,  $p < .001$ , such that JOLs were lower, on average, than actual recall. However, we did see an interaction between measure and LOP condition,  $F(2, 70) = 61.02$ ,  $MSE = 513.13$ ,  $\eta_p^2 = .64$ ,  $p < .001$ , such that people predicted a smaller benefit for deep levels of processing than actually occurred (Figure 4). We also saw an interaction between measure, list, and LOP

condition,  $F(8, 280) = 5.47$ ,  $MSE = 242.12$ ,  $\eta_p^2 = .14$ ,  $p < .001$ , in that predicted LOP effects increase across lists while actual LOP effects do not. We did not find a reliable interaction between measure and point value,  $F(1, 35) = 1.08$ ,  $p = .306$ , nor did we see an interaction between measure, list, and point value,  $F < 1$  (Figure 5). We did see an interaction between measure and list,  $F(4, 140) = 6.57$ ,  $MSE = 351.01$ ,  $p < .001$ ,  $\eta_p^2 = .16$ , as the previously-described main effects of list differed between recall and JOLs. There was no interaction between measure, LOP, and value, nor was there a 4-way interaction between measure, list, LOP, and value,  $F_s < 1.30$ .

Finally, we examined global judgments about which LOP condition is best. Unlike in the previous experiments, we could not compare global LOP preferences with which LOP condition was actually better because deep LOP cues yielded better actual recall for all participants. It appears, on a qualitative level, that most people were aware of this benefit for deep LOP items, as 31 out of 36 participants thought deep LOP cues were better (Table 1). We also found that this pattern of choices differs reliably from a random distribution of choices,  $\chi^2 = 45.17$ ,  $df = 2$ ,  $p < .0001$ .

### Discussion

When using a cued-recall LOP manipulation, a very different pattern of metacognitive results emerged as compared to Experiments 1 and 2. That is, people did tend to show a consistent awareness of the benefits of deep levels of processing, both on an item-by-item basis and in global judgments. In addition, people predicted a larger effect of LOP after the initial study-test cycle in their item-by-item JOLs. These findings are consistent with prior work examining metacognitive awareness of the benefits of deep levels of processing using cued-recall LOP items

(e.g., Shaw & Craik, 1989; Bieman-Copland & Chaness, 1994; Matvey et al., 2002). Still, it is also notable that, as in those prior studies, the magnitude of LOP effects predicted by people's metacognitive judgments was much less than the actual effects of LOP condition on recall.

Contrary to Experiment 1, we did not find an effect of value on recall in this paradigm. The null result was somewhat surprising, since value effects have been observed in cued recall when paired associates are used (e.g., Soderstrom & McCabe, 2011). Still, the cues for the items in the present study seem qualitatively different from ordinary paired associates, in that these cues direct attention to a specific feature of each target word during study. This feature may be what prevents the differential processing that would lead to superior encoding of valuable items. It is interesting to note that even when high value did not actually lead to better memory, people still predicted better recall for more valuable items. The effect of value on JOLs was small in magnitude, and in a direct comparison, it was not reliably different from the lack of value effects found in actual recall. Still, this result provides further evidence that people do not underestimate the effects of value on recall as they do for levels of processing effects. Instead, if anything, they overestimate the effects of point value on recall.

### **General Discussion**

We find that people tend to underestimate the effects of levels of processing when making JOLs, and under some circumstances, show a complete lack of awareness of LOP effects. More specifically, in Experiments 1 and 2, JOLs are slightly but reliably higher for items with deep orienting questions than for items

with shallow orienting questions for items in which the LOP orienting question is congruent with the to-be-remembered word (“yes” items). By contrast, when the LOP orienting question is not congruent with the word (“no” items), JOLs are actually lower for items associated with deep LOP questions. While the benefits of deep LOP questions are reduced when the question is not congruent with the answer, it is clear that participants are incorrect in predicting a reversal of LOP effects under those conditions. Even on global judgments given after multiple study-test cycles, a large number of participants in Experiments 1 and 2 failed to appreciate that a deep level of processing leads to better recall. It was only in Experiment 3, using cued-recall LOP items, that people were generally aware that a deep level of processing benefits recall.

Thus, it seems that the degree to which people are able to appreciate levels of processing effects depends on how closely the LOP manipulation is integrated with the cue. When the answer to an LOP orienting question is “no”, learners still process the word on the level indicated by the question, but that question is essentially unrelated to the word itself. When the answer to an LOP orienting question is “yes”, in contrast, the question provides an additional retrieval cue for the to-be-remembered word, particularly for deeper levels of processing. Still, the question is not present during the free recall test, so it will only benefit recall if the learner happens to remember the question. With the cued-recall LOP paradigm, by contrast, the LOP cue is presented at the time of the test, and it is always related to the to-be-remembered target word. In addition, cues that yield a deeper level of processing would potentially yield a smaller set of possible answers than the cues

that induce shallower levels of processing, further aiding recall of deep LOP items. The cued-recall LOP paradigm provides the strongest integration between the LOP cue and the to-be-recalled item at test, and it also produces the strongest effects of LOP on metacognitive judgments.

Thus, we might conclude that when people do partially appreciate effects of LOP, they do so because deeper LOP cues provide more direct retrieval routes to the item at the time of test. In terms of Koriat's (1997) cue-utilization approach, levels of processing might be considered to be a partially intrinsic cue under these circumstances. By contrast, the benefits of merely thinking about the item on a deeper level, as was done for deep "no" items in Experiment 1 and 2, go completely unappreciated. This component of LOP effects could be considered an extrinsic cue in Koriat's framework. The fact that people fail to appreciate one major contributor to LOP effects is likely why, even when participants do correctly predict better recall for items processed using a deeper LOP, they greatly underestimate those effects.

In contrast to the failure to fully appreciate levels of processing effects, participants' JOLs do tend to account for effects of point value. However, our participants did not seem to fully understand the dynamics of when value has an effect and when it does not. They predicted a consistent effect of value across lists in Experiment 1, even though, in actuality, value effects were absent in list 1, began to emerge on list 2, and then reliably emerged on lists 3 and 4. Similarly, in Experiment 3, even though value did not reliably affect recall, we found that JOLs were slightly but reliably higher for high-value items. Thus, while people do

appreciate that point value affects recall, they seem to have an overly simplistic theory about how it does so.

**Concluding Comments.** The most notable finding from this series of studies is that people show more of an appreciation for the fact that a deeper level of processing leads to better recall under conditions in which the LOP manipulation is more closely bound to the target word. When the LOP manipulation is not bound to the target word (i.e., for the “no” items in Experiments 1-2), people tend to be entirely unable to appreciate how LOP affects recall. However, when the LOP manipulation is more closely related to the target word (i.e., for the “yes” items in Experiments 1-2, and the cued-recall items in Experiment 3), people do begin to show some appreciation for how LOP affects recall. Our findings should therefore help to refine understanding of what constitutes an intrinsic vs. extrinsic cue in metacognitive monitoring (Koriat, 1997). That refinement helps us, more broadly, to better understand what types of cues people do or do not use when making metacognitive judgments, and thus, the circumstances under which people will have accurate metacognitive information to guide further study.

### References

- Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81*, 126-131.
- Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General, 138*, 432-447.
- 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.
- Bieman-Copland, S., & Charness, N. (1994). Memory knowledge and memory monitoring in adulthood. *Psychology and Aging, 9*, 287- 302.
- Castel, A. D., Benjamin, A. S., Craik, F. I. M., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition, 30*, 1078-1085.
- Castel, A. D. (2008). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.), *The psychology of learning and motivation* (Vol. 48). San Diego: Academic Press.
- Castel, A. D., Murayama, K., Friedman, M. C., McGillivray, S. & Link, I. (2013). Selecting valuable information to remember: Age-related differences and similarities in self-regulated learning. *Psychology and Aging, 28*, 232-242.

- Craik, F. I. M., & Lockhart, R.S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-684.
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294.
- Cutting, J.E. (1975). Orienting tasks affect recall performance more than subjective impressions of ability to recall. *Psychological Reports*, *36*, 155-158.
- Dougherty, M. R., Scheck, P., Nelson, T. O., & Narens, L. (2005). Using the past to predict the future. *Memory and Cognition*, *33*, 1096-1115.
- Festini, S. B., Hartley, A. A., Tauber, S. K., & Rhodes, M. D. (2012). Assigned value improves memory of proper names. *Memory*. Advance online publication.
- Fujita, T. (2010). There is more to judgments of learning than prediction: They improve memory. Paper presented at the 51<sup>st</sup> annual meeting of the Psychonomic Society, St. Louis, MO.
- Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*, 208-216.
- Hart, J. T. (1967). Memory and the memory-monitoring process. *Journal of Verbal Learning and Verbal Behavior*, *6*, 685-691.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*, 349-370.



- Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*, 643–656.
- Loftus, G. R., & Wickens, T. D. (1970). Effect of incentive on storage and retrieval processes. *Journal of Experimental Psychology*, *85*, 141-147
- Matvey, G., Dunlosky, J., Shaw, R. J., Parks, C., & Hertzog, C. (2002). Age-related equivalence and deficit in knowledge of cue effectiveness. *Psychology and Aging*, *17*, 589–597.
- Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General*, *137*, 615–625.
- Shaw, R. J., & Craik, F. I. M. (1989). Age differences in predictions and performance on a cued recall task. *Psychology and Aging*, *4*, 131–135.
- Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1236-1242.
- Toglia M. P., & Battig W. F. (1978). *Handbook of semantic word norms*. Republished online as supplement to: Toglia, M.P. (2009). Withstanding the test of time: The 1978 semantic word norms. *Behavior Research Methods*, *41*, 531-533.
- Toppino, T.C., & Cohen, M.S. (2010). Metacognitive control and spaced practice: clarifying what people do and why. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1480-1491.

Watkins, M. J., & Bloom, L. C. (1999). *Selectivity in memory: An exploration of willful control over the remembering process*. Unpublished manuscript.

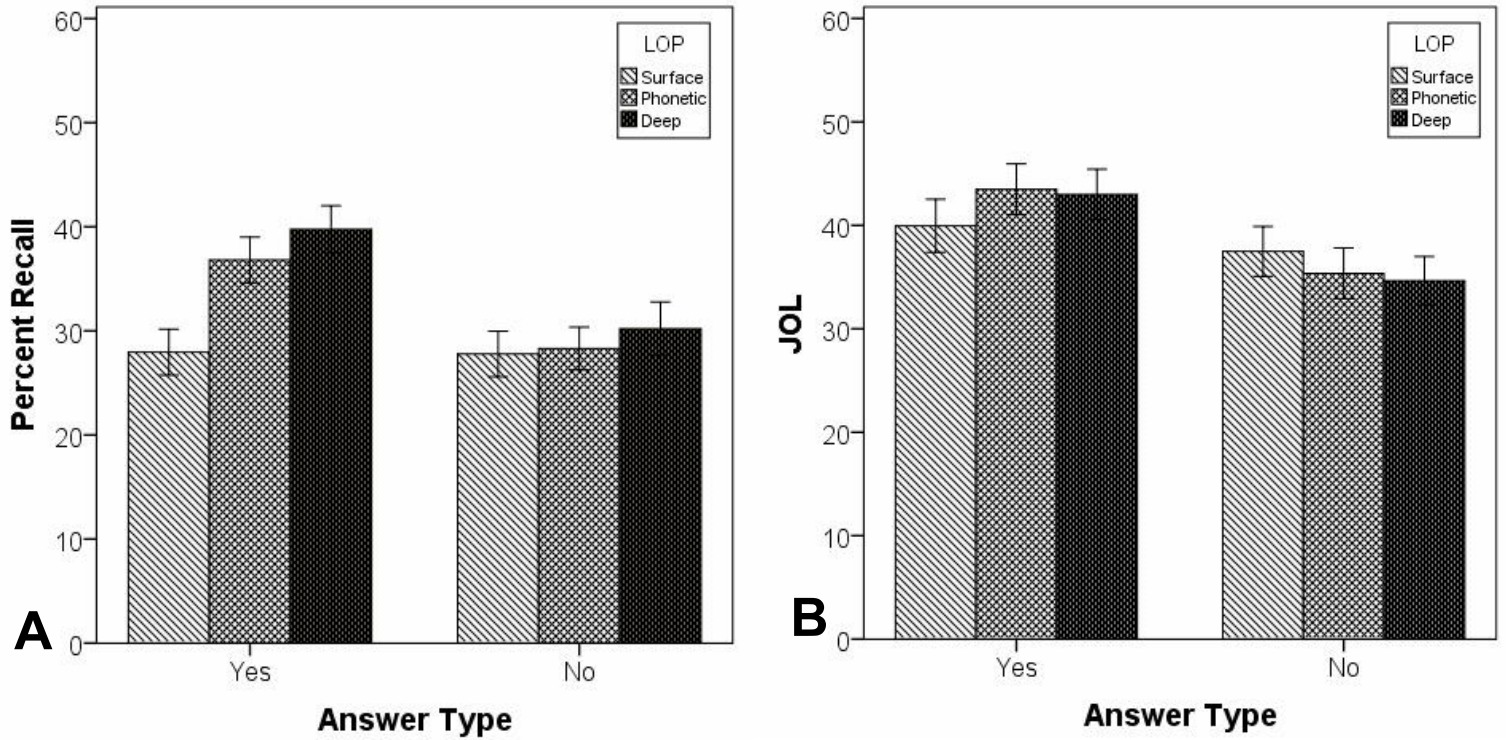


Figure 1. Effects of answer type and LOP condition on recall (A) and JOLs (B) in Experiment 1. Error bars represent +/- 1 SE.

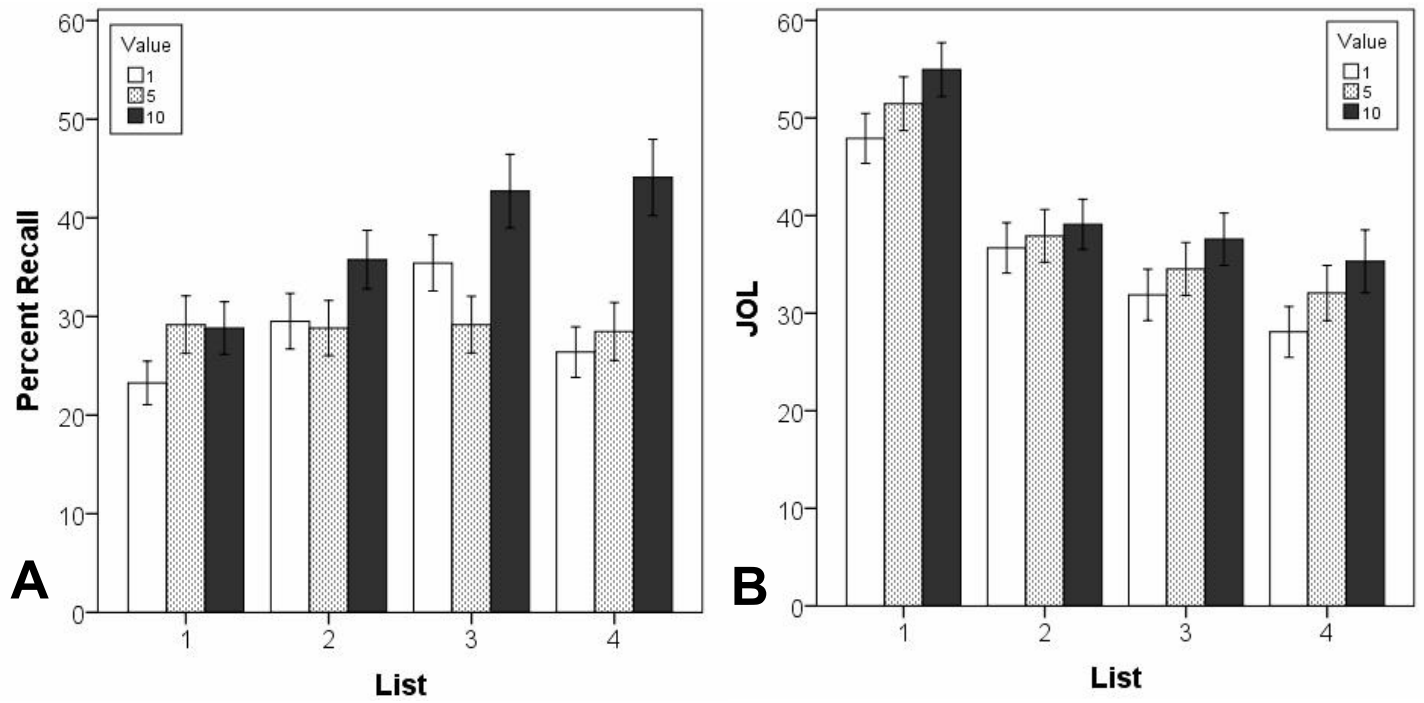


Figure 2. Effects of list and value on recall (A) and JOLs (B) in Experiment 1. Error bars represent +/- 1 SE.

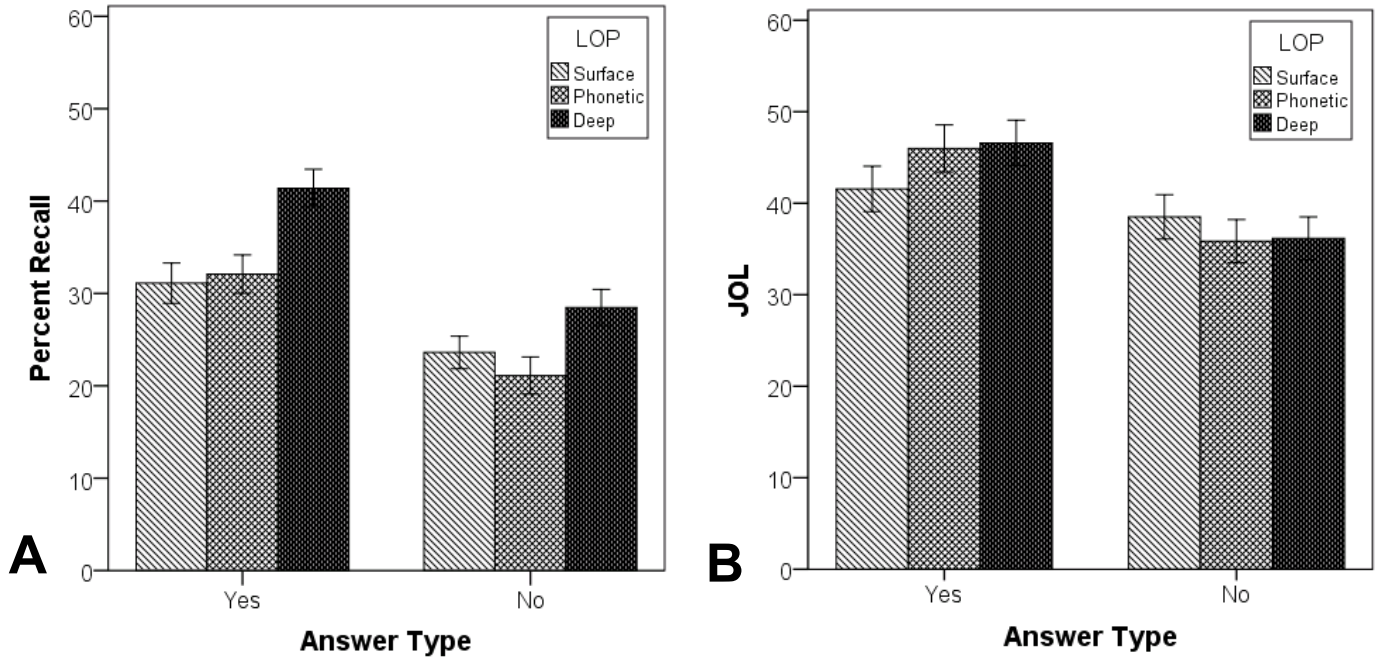


Figure 3. Effects of answer type and LOP condition on recall (A) and JOLs (B) in Experiment 2. Error bars represent +/- 1 SE.

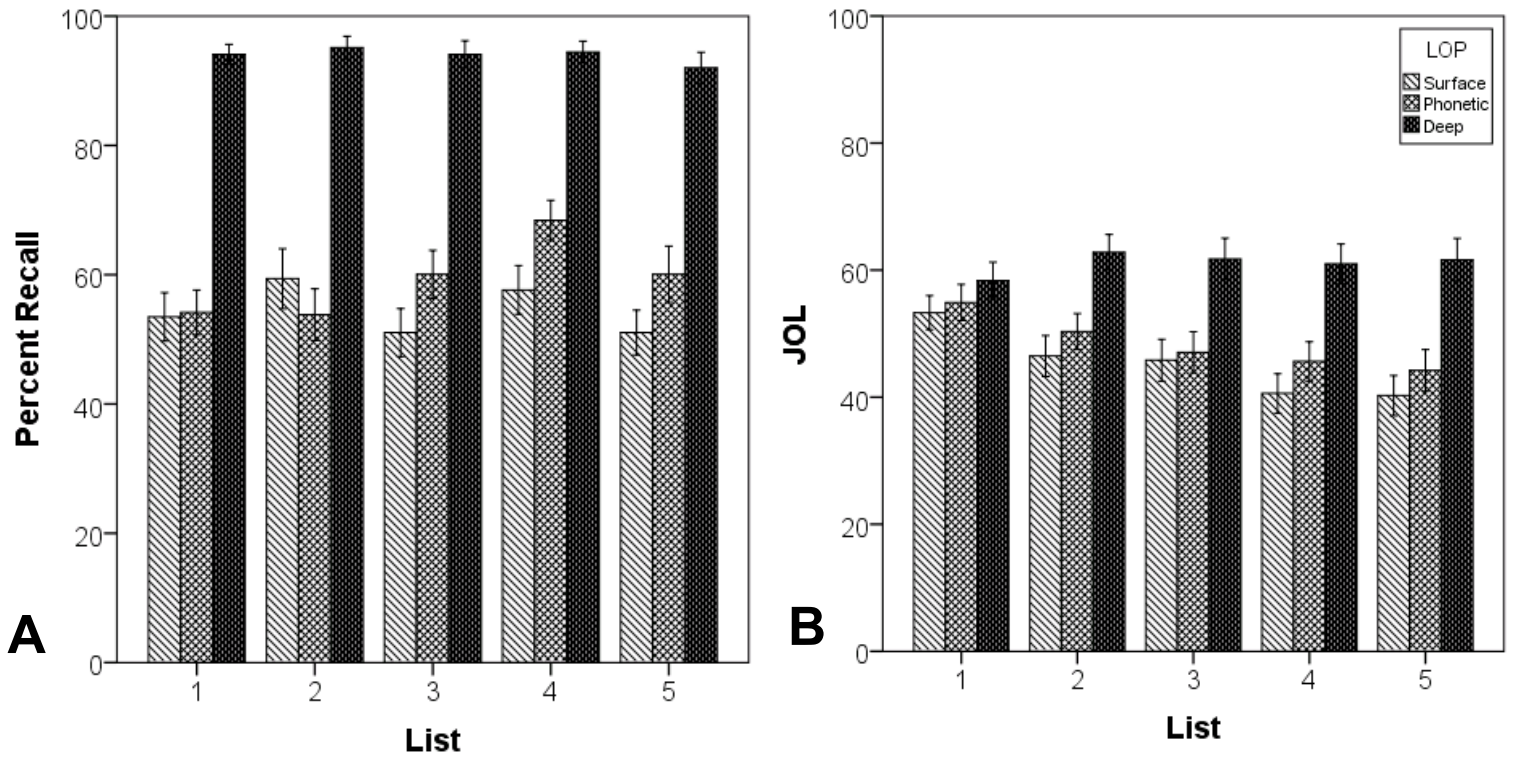


Figure 4. Effects of list and LOP condition on recall (A) and JOLs (B) in Experiment 3. Error bars represent +/- 1 SE.

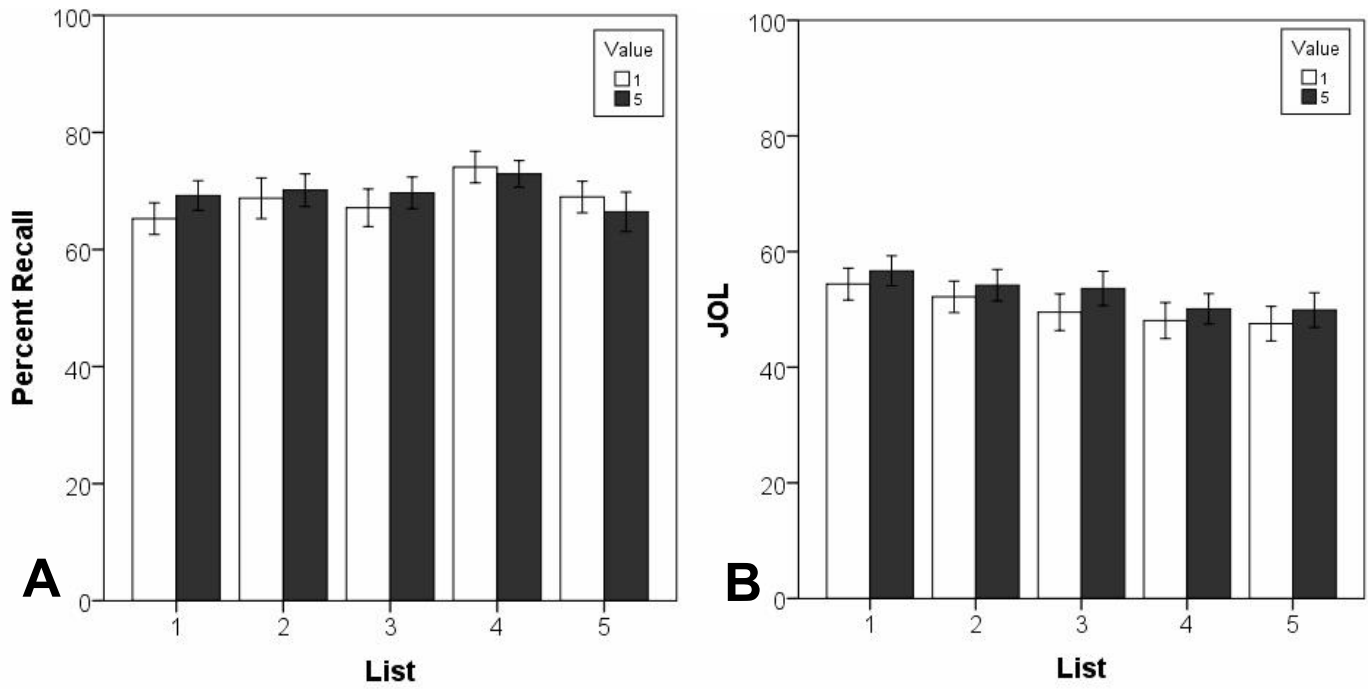


Figure 5. Effects of list and value on recall (A) and JOLs (B) in Experiment 3. Error bars represent +/- 1 SE.

Table 1.

*Number of individuals who preferred a particular LOP condition, compared to the number of individuals for which a particular LOP condition was actually best.*

Experiment	Predicted			Actual		
	Deep	Phonetic <sup>5</sup>	Surface/None <sup>6</sup>	Deep	Phonetic <sup>7</sup>	Surface/None <sup>8</sup>
1	11	24	13	23	18	7
2	18	23	19	32	14	14
3	31	3	2	36	0	0

<sup>5</sup> Includes 1 individual in Exp. 1, 2 individuals in Exp. 2, and 3 individuals in Exp. 3 who believed that deep and phonetic LOP conditions were equally beneficial.

<sup>6</sup> Includes 1 individual in Exp. 1 who believed that phonetic and surface LOP conditions were equally beneficial. It also includes 2 individuals in Exp. 1 and 3 individuals in Exp. 2 who believed that capitalized words were better, but did not express a preference on LOP conditions.

<sup>7</sup> Includes 2 individuals in Exp. 1 and 3 individuals in Exp. 2 for whom deep and phonetic LOP conditions were equally beneficial.

<sup>8</sup> Includes 1 individual in Exp. 1 and 1 individual in Exp. 2 for whom rhyming and surface LOP conditions were equally beneficial.



Appendix

Word stimuli, semantic, and phonetic LOP questions used in Experiments 1-2

Word	List group	Counter-balance group	Semantic		Phonetic	
			Yes	No	Yes	No
ARCH	1	1	a structural form	a type of meat	starch	leave
HEAT	1	2	a type of energy	something people walk on	greet	rent
BONE	1	3	a part of a skeleton	a fashion accessory	loan	wage
GERM	1	2	a type of microorganism	a way to group objects	term	punch
PURSE	1	1	a fashion accessory	a type of camping gear	verse	brief
POWDER	1	1	a type of makeup	an underground passageway	chowder	machine
CHART	1	2	a way to display information	a type of garden tool	part	hide
SHAPE	1	3	a type of geometric object	an item used for pulling things	grape	part
HOOK	1	2	something that can hold a hat	an element of a rainstorm	book	crush
MOLD	1	2	a type of fungus	an element of a story	fold	sport
CHIEF	1	3	a type of leader	an element of a flower	brief	board
TUBE	1	1	something that carries liquids	a type of furniture	cube	pig
PLOT	1	3	an element of a story	something that carries liquids	knot	flirt
SILK	1	3	a type of fabric	a type of breakfast food	milk	bring
TWIG	1	1	a part of a tree	a role in a business	pig	peel
BRICK	1	2	a type of building exterior	an item that one wears	wick	book
MAYOR	1	3	a political office	a type of geometric object	player	grape
SALUTE	1	1	a type of greeting	something found on a door	astute	player
RAKE	2	3	a type of garden tool	a side of an object	shake	lamp

DIRT	2	1	something plants grow in	a place of law	flirt	stork
CORD	2	3	a part of an electrical appliance	a part of a skeleton	board	loan
JUDGE	2	2	a presiding official	a type of makeup	nudge	melt
BUNCH	2	1	a way to group objects	a part of a house	punch	pool
NOTE	2	2	a type of message	a type of sports equipment	coat	host
BELT	2	3	an item that one wears	a marine animal	melt	nudge
CRUMB	2	3	an unwanted bit of food	a way to display information	numb	sleuth
CANTEEN	2	1	a type of camping gear	a part of a camera	machine	flipper
METAL	2	1	a building material	a type of message	settle	chowder
STRING	2	3	something found on a guitar	a type of leader	bring	mink
PAPER	2	2	an item you can write on	a part of a highway	taper	custard
SCALE	2	1	a tool for weighing	a type of fabric	rail	plunge
MUSTARD	2	2	a type of condiment	a type of microorganism	custard	taper
OWNER	2	3	a role in a business	a part of a shirt	toner	socket
RAMP	2	1	a part of a highway	a part of a tree	lamp	size
COURT	2	2	a place of law	something that holds liquids	sport	starch
DUSK	2	2	a time of day	an unwanted bit of food	husk	stain
FLASH	3	2	a part of a camera	something that can hold a hat	cash	milk
PORK	3	1	a type of meat	a structural form	stork	fold
TUNNEL	3	3	an underground passageway	an object used for grooming	funnel	blunder
FRONT	3	2	a side of an object	a type of fungus	hunt	cube
TANK	3	3	something that holds liquids	a part of a song	bank	numb
DITCH	3	1	something found on a road	a type of condiment	glitch	bank

SPONGE	3	2	an item used for washing dishes	a presiding official	plunge	cheek
THUNDER	3	1	an element of a rainstorm	something found on a bird	blunder	settle
DRINK	3	3	a consumable liquid	a form of light	mink	sound
POCKET	3	1	a part of a pair of pants	a place to sit at a restaurant	socket	stellar
PRIZE	3	2	a type of reward	a type of laugh	size	shake
BEAK	3	1	something found on a bird	a type of joke	cheek	rail
KNOB	3	3	something found on a door	a type of energy	blob	greet
BRUSH	3	1	an object used for grooming	a type of information	crush	wick
SLEEVE	3	3	a part of a shirt	a type of reward	leave	hunt
SCENT	3	2	an element of a flower	a tool for weighing	rent	blob
NEWS	3	2	a type of information	a type of building exterior	choose	cash
CAVE	3	3	a geologic feature	a political office	shave	prune
TOAST	4	3	a type of breakfast food	something found on a guitar	host	funnel
SLIDE	4	2	a type of playground equipment	something plants grow in	hide	knot
STOOL	4	1	a type of furniture	something found on a road	pool	husk
CHAIN	4	3	an item used for pulling things	an item you can write on	stain	toner
GROUND	4	2	something people walk on	a consumable liquid	sound	coat
CHUCKLE	4	1	a type of laugh	a part of a pair of pants	buckle	rocker
SEAL	4	3	a marine animal	a type of shoe	peel	flop
BEAM	4	1	a form of light	a part of a book	dream	verse
TIRE	4	3	an item found on a car	a part of an electrical appliance	wire	middle
BOOTH	4	2	a place to sit at a restaurant	a time of day	sleuth	term
CELLAR	4	1	a part of a house	a type of playground	stellar	buckle

PAGE	4	2	a part of a book	equipment an item found on a car	wage	choose
RACQUET	4	3	a type of sports equipment	an item used for washing dishes	packet	glitch
RIDDLE	4	1	a type of joke	a type of business	middle	wire
LOCKER	4	2	an area used for storage	a geologic feature	rocker	packet
TUNE	4	3	a part of a song	an area used for storage	prune	shave
SHOP	4	1	a type of business	a building material	flop	dream
SLIPPER	4	2	a type of shoe	a type of greeting	flipper	astute