



HHS Public Access

Author manuscript

J Exp Psychol Learn Mem Cogn. Author manuscript; available in PMC 2018 October 01.

Published in final edited form as:

J Exp Psychol Learn Mem Cogn. 2017 October ; 43(10): 1581–1601. doi:10.1037/xlm0000395.

Free recall test experience potentiates strategy-driven effects of value on memory

Michael S. Cohen,

University of California, Los Angeles and Northwestern University

Jesse Rissman,

University of California, Los Angeles

Mariam Hovhannisyan,

University of California, Los Angeles

Alan D. Castel, and

University of California, Los Angeles

Barbara J. Knowlton

University of California, Los Angeles

Abstract

People tend to show better memory for information that is deemed valuable or important. By one mechanism, individuals selectively engage deeper, semantic encoding strategies for high value items (Cohen et al., 2014). By another mechanism, information paired with value or reward is automatically strengthened in memory via dopaminergic projections from midbrain to hippocampus (Shohamy & Adcock, 2010). We hypothesized that the latter mechanism would primarily enhance recollection-based memory, while the former mechanism would strengthen both recollection and familiarity. We also hypothesized that providing interspersed tests during study is a key to encouraging selective engagement of strategies. To test these hypotheses, we presented participants with sets of words, and each word was associated with a high or low point value. In some experiments, free recall tests were given after each list. In all experiments, a recognition test was administered 5 minutes after the final word list. Process dissociation was accomplished via Remember/Know judgments at recognition, a recall test probing both item memory and memory for a contextual detail (word plurality), and a task dissociation combining a recognition test for plurality (intended to probe recollection) with a speeded item recognition test (to probe familiarity). When recall tests were administered after study lists, high value strengthened both recollection and familiarity. When memory was not tested after each study list, but rather only at the end, value increased recollection but not familiarity. These dual process dissociations suggest that interspersed recall tests guide learners' use of metacognitive control to selectively apply effective encoding strategies.

Correspondence concerning this article should be addressed to: Michael S. Cohen, Department of Psychology, Northwestern University, 2029 Sheridan Rd., Evanston, IL 60208. michael.cohen@northwestern.edu.

Michael S. Cohen, Department of Psychology, University of California, Los Angeles, and Department of Psychology, Northwestern University, Evanston, IL; Jesse Rissman, Mariam Hovhannisyan, Alan D. Castel, and Barbara J. Knowlton, Department of Psychology, University of California, Los Angeles

Keywords

reward; value-directed remembering; metacognition; dopamine; test-potentiated learning

In recent years, there has been substantial interest in understanding how encoding processes are affected by the importance of a to-be-remembered item. In the neuroscience literature, a number of studies have focused on how dopamine-producing, reward-sensitive regions in the midbrain communicate with the hippocampus in anticipation of learning a high-value item, which is believed to strengthen hippocampal plasticity (e.g., Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006; Gruber, Gelman, & Ranganath, 2014; Shigemune, Tsukiura, Kambara, & Kawashima, 2014; Wolosin, Zeithamova, & Preston, 2012; see Miendlarzewska, Bavelier, & Schwartz, 2016, for a recent review). It seems clear that reward can strengthen memory via this mechanism even when there is no opportunity for learners to change intentional encoding strategies based on reward. For instance, Murayama and Kitagami (2014) manipulated whether or not a reward could be obtained in an unrelated task, presented after incidental memory encoding. On a delayed memory test, memory was still better on trials in which a reward could be obtained in the unrelated postencoding task than when no such reward was available. Thus, value-related differences in explicit motivation and/or attention are not necessary for producing putatively dopamine-driven enhancements in memory consolidation.

However, there are other conditions under which strategy-driven effects of value appear to be dominant. While the present research is focused only on healthy young adults, our recent work on older adults provides a particularly illustrative example of this point. Cohen, Rissman, Suthana, Castel, and Knowlton (2014, 2016) used fMRI to examine the neural mechanisms underlying value-related memory enhancement in a paradigm known as value-directed remembering (Castel, Benjamin, Craik, & Watkins, 2002; Castel, 2008). These studies found that in both young adults and older adults, the degree to which value affects brain activity during word encoding in brain regions related to strategic control of semantic processing correlates with individual differences in how strongly value affects memory on an immediate free recall test. These studies also found evidence that the mesolimbic dopamine system may have contributed to reward-driven memory in young adults. Older adults, however, showed no value-related changes in activity in dopaminergic reward-sensitive brain regions, and yet they still showed robust effects of value on memory. Thus, it is evident that while mesolimbic dopamine-driven effects on encoding and consolidation are important, strategic effects of value on encoding are also a key piece of the bigger picture.

There are a number of reasons why it is important to understand the conditions under which strategy-driven effects of value influence memory. For one, varying strategies as a function of value is an explicit process that draws upon metacognitive knowledge of learning. Thus, it could presumably be enhanced more easily via training than would automatic effects of value driven by midbrain dopamine release (although training the dopamine system via biofeedback is apparently possible; see MacInnes, Dickerson, Chen, & Adcock, 2016, for an example). In addition, dopamine-driven effects of value are more robust when memory is tested after some delay (e.g., Murayama & Kuhbander, 2011; Spaniol, Schain, & Bowen,

2014), while strategy-driven effects of value on memory are robust on immediate tests (e.g., Castel et al., 2002; Cohen et al., 2014, 2016). Such findings suggest that the impact of value on memory outcomes might qualitatively differ based on interactions between the mechanism being invoked and the time scale on which memory is being tested. Finally, as discussed above, effects of value on dopamine-producing regions during encoding may be weakened in older adults, even as strategy-driven value effects on memory are maintained with healthy aging (Cohen et al., 2016). It is thus important that we gain a better understanding of the interplay between strategy-driven and non-strategic, dopamine-driven effects of value in order to fully understand how value affects memory encoding processes.

Test-potentiated learning

There is reason to believe that the way in which a learning session is structured can impact the type of strategies that will be used to learn the items. Specifically, when tests are interspersed with study opportunities, people tend to engage metacognitive monitoring and control processes to assess the effectiveness of their learning strategies, try out new strategies, and optimize the use of those strategies to enhance performance on future tests. The benefits of testing on the effectiveness of subsequent study opportunities have been referred to as test-potentiated learning (Arnold & McDermott, 2013). These effects are distinct from the direct benefits of testing on learning, in which memory on subsequent tests is better for items that have been tested previously, relative to restudied items.

Most relevant to the present work are explanations of test-potentiated learning that focus on how tests provide an opportunity to use metacognitive monitoring to improve encoding strategies. For instance, Bahrick and Hall (2005) found that when the temporal delay between study sessions with practice tests was longer, more closely approximating the delay leading up to the final test, participants were more likely to engage effective strategies for learning foreign language word pairs than they were with a shorter delay. Bahrick and Hall proposed that retrieval failures in the practice tests, which are more likely to occur with a longer delay, led people to choose more effective encoding strategies on subsequent study opportunities.

Pyc and Rawson (2012) explicitly instructed participants to use “keyword” mediators to relate a foreign-language word with its English definition, and asked for reports of those keywords across multiple study opportunities or study-test cycles. They found a greater number of keyword shifts for participants in the condition that included tests after each study opportunity as compared to the study-only condition, and, within the study-test group, a greater number of keyword shifts following retrieval failures than following successful retrievals. This work provided the first direct evidence of shifts in encoding strategies stimulated by interspersed tests. Soderstrom and Bjork (2014) later examined how self-paced allocation of study time changed following a practice test, and they found that people devoted more time on a subsequent study opportunity to studying words that they had failed to recall on the practice test, compared to the average for participants in a restudy-only condition. It is also notable that for the self-paced restudy period, individuals in the test-restudy group were more likely to report using encoding strategies defined as effective (e.g.,

relating the words to something meaningful), and were less likely to report using ineffective strategies (e.g., rote repetition), compared to the restudy-only group.

The examples described above involve tests promoting more effective restudy of repeated items. However, other recent studies have shown that tests also lead to more effective subsequent encoding of new items. Such effects have been found using a variety of types of materials, including individual words (Szpunar, McDermott, & Roediger, 2008), text passages (Wissman, Rawson, & Pyc, 2011), face-name pairs (Weinstein, Gilmore, Szpunar, & McDermott, 2014), online course lectures (Szpunar, Khan, & Schacter, 2013), associated word pairs (Soderstrom & Bjork, 2014), and multimedia lessons (Yue, Soderstrom, & Bjork, 2015). These studies have provided further evidence to suggest that enhanced strategy use in subsequent study sessions is a key factor in test-potentiated learning. For instance, Wissman et al. found that testing benefitted learning of new text passages even when the topic was largely unrelated to the previously-tested texts, rendering unlikely some other non-metacognitive explanations such as spreading activation or reductions in proactive interference. Additionally, Soderstrom and Bjork found that when learners were given a practice test on some word pairs but not others, the non-tested items were allocated significantly more time than the items that were recalled correctly on the practice test. This result contrasts with what was found in individuals who were not given any practice tests; there, study time allocation was similar to that shown, among those in the group that did receive practice tests, for items that had already been recalled correctly. Thus, it again seems that the experience of being tested, rather than being tested on a particular item, helps people to realize the limitations of their learning and increase allocation of study time and other cognitive resources during subsequent study periods.

Other studies have shown that people can selectively allocate their study resources to items deemed as valuable. For instance, Ariel, Dunlosky, and Bailey (2009) showed that people devote more study time to items that are worth more points, regardless of item difficulty, and Toppino and Cohen (2010) showed that people are more likely to choose to space, rather than mass, a second study opportunity for high-value relative to low-value items. Finally, a recent study by Middlebrooks, Murayama, & Castel (in press) found that the degree to which value affects memory on free recall tests is greatly reduced when a recognition test, rather than a recall test, is expected, providing further evidence for the role of metacognitive control processes. We hypothesize that learners will be more likely to selectively apply effective but effortful study strategies to high value items when they have experienced interspersed recall tests. In comparison, without having experience with interspersed tests, we predict that learners will tend not to employ a selective encoding strategy that prioritizes certain items over others. Instead, under those conditions, we expect that enhancement of memory recall outcomes for high-value items would largely be attributable to increased engagement of reward mechanisms when learning those items.

Relating dual process models to value-directed remembering

A key question is whether and how memory quality differs depending on whether or not people used strategies during encoding to enhance memory for valuable items. We rely on the dual-process model described in detail by Yonelinas (2002), following earlier

forerunners by Mandler (1980), and Jacoby and Dallas (1981; see also Jacoby, 1991), among others, to address this question. The dual-process model assumes that there are two independent processes involved in explicit memory: recollection, which includes rich contextual detail, and familiarity, which is lacking in such contextual detail. Our core hypothesis is that the two putatively distinct mechanisms of value-directed memory enhancement described above, specifically, strategy-driven and non-strategic reward-driven mechanisms, will have differential effects on subsequent expressions of recollection and familiarity.

Prior literature strongly indicates that when words are learned in a way that directs more attention to the meaning of those words, by for instance using a cue to induce a deep vs. shallow level of processing (e.g., Craik & Lockhart, 1972), or by being asked to generate rather than read a study item (e.g., Slamecka & Graf, 1978), memory is strengthened, and both recollection and familiarity increase (Yonelinas, 2002). More recent work by Sheridan and Reingold (2011, 2012) has further shown that even when using a more precise variant of the Remember/Know procedure that allows for independent assessment of recollection and familiarity, manipulations of encoding strategy reliably enhance both processes. Given the overlap between the neural correlates of encoding via deep levels of processing (e.g., Kapur et al., 1994) and the neural correlates that we observed in what appear to be strategy-driven effects of value on memory (Cohen et al., 2014, 2016), we hypothesize that strengthening of encoding via value-related changes in strategy use should lead to value-related increases in both recollection and familiarity.

There is less prior work on the dual process correlates of more automatic, reward-driven effects of value on memory, but the evidence that is available largely suggests an enhancement of recollection but not familiarity. One reason to make such a prediction is mechanistic. Specifically, dopamine-driven effects of reward on memory have been shown to involve changes in brain activity in and connectivity with the hippocampus and parahippocampal cortex (Adcock et al., 2006; Shigemune et al., 2014; Wolosin et al., 2012). These areas are generally associated with recollection, while a separate region of medial temporal lobe, the perirhinal cortex, is typically associated with familiarity (Diana, Yonelinas, & Ranganath, 2007; Eichenbaum, Yonelinas, & Ranganath, 2007). Such findings support our prediction that if participants either choose not to vary strategies as a function of value, or if the task paradigm does not provide them with the opportunity to learn the importance of varying strategies as a function of value, value will only enhance recollection.

Prior literature also provides some empirical support for this hypothesis. One relevant study, by Shigemune et al. (2014), used an intentional encoding paradigm with a recognition test for items and source details. Memory for source details, a measure typically thought to reflect recollection, was enhanced for items in which correct recognition responses could lead to earning a reward, or could lead to avoiding punishment, relative to non-rewarded items. There was no effect of reward or punishment on the proportion of items correctly recognized without accurate source information, which can be considered a measure of familiarity. Neuroimaging results were fully consistent with engagement of a mesolimbic dopamine-driven enhancement of memory on both reward and punishment trials, relative to neutral trials. Thus, activation of the dopaminergic reward system seems to have increased

the likelihood of later recollection without a concomitant boost in familiarity-based recognition.

A recent study by Gruber, Ritchey, Wang, Doss, and Ranganath (2016) is also relevant. They presented participants with a series of images representing concrete objects, each with an associated background image. They then presented a question related to the foreground image, intended to evoke incidental deep encoding of the item (e.g., “Does this item weigh more than a basketball?”). This question was associated with either a high or low reward value. Gruber et al. found higher rates of self-reported recollection for high-reward items, and better memory for the background image on high-reward items. Importantly, there was no reliable difference in the likelihood of confident familiarity-based memory as a function of reward. Memory enhancement for high-reward items was associated with midbrain-hippocampal circuitry via a number of different neural measures. Thus, this study provides further evidence to suggest that dopamine-driven memory enhancement is likely to only enhance recollection.

The present studies

Here, we aim to dissociate two distinct mechanisms by which value enhances memory encoding. We hypothesize that the strategic differential encoding of valuable items leads to greater subsequent recollection and familiarity. In contrast, a more automatic, putatively dopamine-driven mechanism leads to enhanced binding of high value items to context, leading to an increase in recollection alone. We test how features of the value-directed remembering paradigm, particularly the inclusion of multiple study-test cycles with feedback, encourage people to selectively enhance their use of strategies for high-value items. If interspersed recall tests are in fact necessary to yield selective use of strategies, we expect to find that both recollection and familiarity will be enhanced for high-value items when people get practice and feedback with intervening free recall tests. We would expect such effects to reflect the simultaneous engagement of both a reward-driven mechanism, which putatively strengthens recollection, and of selective strategy use, which putatively enhances both recollection and familiarity. However, when participants are not provided with interspersed recall tests and feedback, we predict that value will enhance recollection, but not familiarity, as memory for high-value items is only being boosted via the more automatic, reward-driven mechanism.

Experiment 1

The encoding and recall tasks used in Experiment 1 were very similar to those used in our prior fMRI studies (Cohen et al., 2014, 2016), in which strategic modulation of semantic processing appeared to underlie effects of value on memory. In order to assess dual-process correlates of these memories, we added a new test at the end of the study session, a surprise yes-no recognition test that included all items studied after List 1. This test used a Remember-Know (R/K) procedure to determine the proportion of items that could be recognized using recollection-based memory, and the proportion recognized using familiarity. When correcting the raw proportions to assume independence, the Remember-

Know method typically yields reliable estimates of process contributions to memory (Sheridan & Reingold, 2012; Yonelinas & Jacoby, 1995).

In addition, while we did not directly manipulate the strategies that participants used during learning, we collected self-reports of how strongly participants believed that item value influenced their encoding process. Reporting that one approached the encoding process differently when learning high-value items, i.e., showing some degree of value sensitivity, would seem to be a prerequisite to explicitly changing strategies based on item value. Even under study conditions that encourage such changes, which we term selective strategy use, we would only expect to see evidence for selectivity, i.e., increases in both subsequent recollection and subsequent familiarity, in individuals who report that they approached high-value items differently. On the other hand, individuals who report being indifferent to value during encoding are unlikely to explicitly vary study strategies based on item value. Such individuals may still show effects of value on recollection due to non-strategic effects of reward, but we would not expect to find effects of value on familiarity in these individuals. To test this prediction, we used a questionnaire measure to assess participants' self-reported value sensitivity. The analyses based on this questionnaire measure are post hoc, and in some cases rely on relatively small numbers of participants, but they nevertheless provide a means for validating a central assumption of this work, that differences in how value affects dual-process correlates at retrieval can be driven by how value affected strategy use during encoding.

Methods

Participants—43 participants (31 female, 11 male, 1 gender not recorded, age range 18–23 years, $M_{\text{age}} = 20.10$ years) were recruited from the UCLA Department of Psychology undergraduate student subject pool, which includes students from psychology and linguistics courses, and were compensated with course credit for their participation.

Materials—Words used as study items in the value-directed remembering task, or as lures in the recognition test, were defined using the same criteria as in Cohen et al. (2014, 2016). Specifically, all words were drawn from clusters 6 and 7 of the Toggia and Battig (1978) word norms. All were 4–8 letter nouns, rated as highly familiar (range 5.5–7 on a 1–7 scale), moderate to high on concreteness and imagery (range 4–6.5 on a 1–7 scale), and moderate in pleasantness (range 2.5–5.5 on a 1–7 scale).

Procedure—All experiments followed a study protocol that was reviewed and approved by the UCLA Institutional Review Board. In all experiments, written informed consent was obtained from each participant prior to beginning the study.

After reading through the instructions on-screen, participants saw 6 practice items intended to familiarize them with the study phase of the task. Then, after the experimenter answered any questions that arose, seven complete study lists were presented. Each list included 24 items, half of which were randomly assigned to be low-value (worth 1, 2, or 3 points), and half of which were randomly assigned to be high-value (worth 10, 11, or 12 points), with the assignment of words to value level counterbalanced across subjects.

Each trial in the study phase began with an initial value cue, presented for 1 s, followed by a fixation period lasting 0.5 s. The value cue was presented in the form of a gold coin with a number inside indicating how many points the upcoming word would be worth (see Figure 1 of Cohen et al., 2014). The word was then shown for 2.5 s, followed by a 2 s blank screen before the next item was presented. After each list of 24 items was presented, participants were instructed to freely recall as many items as possible from the list that they just saw, and were given 60 seconds to do so verbally. The experimenter was in the room with the participant during the entirety of the encoding portion of the paradigm, and provided feedback as to how many points they earned at the end of each list.

Following all 7 study-test cycles, participants played the video game “Snood” for approximately 5 minutes. Then, they began an R/K recognition test that included all 144 words from study lists 2–7, intermixed with 144 lure words. Participants received careful instructions about the definition of R/K, which were adapted from those used by Rajaram (1993); see Appendix for details. After reading these instructions, participants were instructed to describe to the experimenter the difference between a Remember (R) and a Know (K) judgment. This was an added check to ensure that they had paid attention to the instructions, and an opportunity for the experimenter to correct any misunderstandings. Another important design feature was the use of two-stage R/K judgments with no “guess” option. Participants were first instructed to judge whether an item was “old” or “new”, and were told that they should only choose “old” if they are at least “fairly confident” that they saw the word, but should choose “new” if they either did not remember seeing the word, or if they were unsure. Then, only once they had chosen the “old” option did they make a judgment as to whether to classify their memory for the item as an R or a K. This procedure has been shown to reduce the use of K responses as a proxy for low-confidence judgments (Eldridge, Sarfatti, & Knowlton, 2002), which is important because a key assumption in the R/K paradigm is that the two judgments should be relatively equated in terms of confidence, yet vary in terms of the quality of the memories. After the recognition test was complete, participants were asked to write down the basis on which they made R/K judgments, as an additional check to confirm that they understood the procedure.

The post-study questionnaire also asked about what they did differently during the encoding procedure for high-value vs. low-value items, which we used to classify participants by value sensitivity. More specifically, we examined answers to the following open-ended question: “What strategy did you use to learn the words? Did you do anything differently to learn the high-value items?”. Two raters (M. S. Cohen and M. Hovhannisyan) made a subjective assessment of responses and assigned each participant to one of three categories. Ratings were made blind to the memory performance data, and discrepant ratings between the two raters were resolved by discussion. Individuals classified in the Weak value sensitivity group generally claimed to have been indifferent to value. Those classified in the Moderate group generally claimed to have “tried harder,” or something similar, for high-value items, but still seemed to apply some effort to low-value items as well. Finally, participants classified in the Strong group reported either ignoring low-value items completely, or having a specific encoding strategy that they only applied to the high-value items.

Results

Free recall tests—We begin by analyzing performance on the free recall tests (Table 1). A 2×7 (value \times list) repeated-measures ANOVA showed a main effect of value, $F(1, 42) = 47.80, p < .001, \eta_p^2 = .53$, as well as a main effect of list, $F(6, 252) = 4.77, p < .001, \eta_p^2 = .10$, and an interaction between list and value, $F(6, 252) = 2.36, p = .031, \eta_p^2 = .05$. Thus, high-value items were clearly remembered better than low-value items, and this effect appears to get stronger with practice, with notable increases in the effect of value on recall after the first and second lists. These findings replicate previous results from other studies using similar paradigms (e.g., Castel, 2008; Cohen et al., 2014).

The free recall data are also useful for assessing the validity of our post-hoc analysis of the post-study questionnaire responses. In Experiment 1, 13 individuals were classified as exhibiting Weak value sensitivity, 12 as Moderate, and 16 as Strong, while two additional participants were excluded from these analyses due to not providing an adequate response for us to assess value sensitivity. For this analysis, and for analyses throughout the paper except as noted, we combine items from all lists beginning with list 2, under the assumption that test-potentiated effects of strategy use would require exposure to at least one test. A 2×3 (item value \times value sensitivity) mixed ANOVA, with repeated measures on the first factor, showed a main effect of item value, $F(1, 38) = 78.84, p < .001, \eta_p^2 = .67$, no main effect of value sensitivity, $F(2, 38) < 1, \eta_p^2 = .03$, and an interaction between item value and value sensitivity, $F(2, 38) = 16.80, p < .001, \eta_p^2 = .47$. Tukey post-hoc tests showed that the effect of item value in the Weak group was significantly less than the effect of value in the Moderate group, $p = .013$, and less than the effect of value in the Strong group, $p < .001$. Additionally, the effect of value in the Moderate group was significantly less than that in the Strong group, $p = .041$. We also used paired-samples t-tests to probe the interaction, comparing the number of high-value vs. low-value items recalled within each value sensitivity group. In the Weak group, the proportion of items correctly recalled was equivalent between high-value ($M = .452, SE = .046$) and low-value items ($M = .403, SE = .060$), $t(12) = 1.07, p = .305, d = .30$. In the Moderate group, high-value items ($M = .535, SE = .043$) showed significantly better recall than low-value items ($M = .272, SE = .048$), $t(11) = 4.28, p = .001, d = 1.23$. Similarly, in the Strong group, recall was better for high-value ($M = .597, SE = .034$) than for low-value items ($M = .162, SE = .024$), $t(15) = 10.92, p < .001, d = 2.73$. Thus, the degree to which people reported being sensitive to item value during encoding clearly corresponded with how strongly value affected free recall performance.

Remember/Know recognition test—We turn next to the results from the R/K test. We limit these analyses to items that were not recalled during the study-test cycles. Because recalling an item would likely strengthen memory independently of processes active during initial encoding, and because more high-value items were recalled than low-value items, including these items would create a bias in favor of finding stronger memories for high-value items. Excluding such items is likely to bias us against finding significant effects of value, by eliminating the items that were mostly strongly encoded. Experiments 3 and 4 provide other ways to more directly circumvent this issue.

In order to calculate estimates for recollection and familiarity from the Remember/Know judgments, we adopted the approach advocated by Yonelinas and Jacoby (1995). Specifically, we computed familiarity estimates using the formula $F = ((K_{Hit}/(1 - R_{Hit})) - (K_{FA}/(1 - R_{FA})))$, where R is the proportion of items given “Remember” responses and K is the proportion of items given “Know” responses. Recollection estimates were computed using the formula $R = R_{Hit} - R_{FA}$. These formulas follow from an assumption that the two processes are independent, and also correct for false alarms.

We found that estimated recollection was greater for high-value items than for low-value items, $t(42) = 6.03, p < .001, d = .92$, and high-value items were also associated with greater familiarity than low-value items, $t(42) = 3.46, p = .001, d = .53$ (Figure 1).

We next examined how value affected performance on the R/K recognition test as a function of self-reported value sensitivity (Figure 2). A 2×3 (item value \times value sensitivity) mixed ANOVA on recollection found a main effect of item value, $F(1, 38) = 36.79, p < .001, \eta_p^2 = .49$, a main effect of value sensitivity, $F(2, 38) = 4.51, p = .017, \eta_p^2 = .19$, but no interaction between these factors, $F(2, 38) < 1, \eta_p^2 = .03$. Planned comparisons showed that in the Weak group, there was a significant effect of value on recollection, $t(12) = 3.51, p = .004, d = .97$. High-value items also showed better recollection in the Moderate group, $t(11) = 3.28, p = .007, d = .95$, and in the Strong group, $t(15) = 4.08, p = .001, d = 1.02$. Thus, value appears to robustly influence recollection regardless of self-reported sensitivity to value.

We also ran a 2×3 (item value \times value sensitivity) mixed ANOVA on the rate of familiarity. We found a main effect of item value, $F(1, 38) = 11.12, p = .002, \eta_p^2 = .23$, no main effect of value sensitivity, $F(2, 38) = 1.38, p = .265, \eta_p^2 = .07$, and, importantly, a significant interaction between item value and value sensitivity, $F(2, 38) = 4.61, p = .016, \eta_p^2 = .20$. Probing the interaction, Tukey post hoc tests showed a difference in the effect of value on familiarity between the Weak and Strong groups, $p = .012$, but no difference between the Weak and Moderate groups, $p = .223$, nor any difference between the Moderate and Strong groups, $p = .464$. Planned comparisons show that in the Weak group, there was no effect of value on familiarity, $t(12) < 1, d = -.08$. We did, however, find an effect of value on familiarity in the Moderate group, $t(11) = 2.55, p = .027, d = .73$, as well as in the Strong group, $t(15) = 3.67, p = .002, d = .92$.

Discussion

In this experiment, we found that high-value items, in the context of an encoding paradigm that included interspersed recall tests, showed robust increases in both recollection and familiarity relative to low-value items. However, our post-hoc analysis additionally showed that, while the effect of value on recollection did not depend on the degree to which participants reported that item value explicitly affected their approach to learning the items, the effect of value on familiarity did depend on this factor. Specifically, in those individuals who reported being sensitive to value, whether to a moderate or strong degree, high-value items showed both stronger recollection-based memory and stronger familiarity-based memory. In contrast, those individuals who reported not being sensitive to value, for whom any effects of value on memory were likely being driven by more automatic mechanisms,

showed a different pattern of results. High-value items still showed stronger recollection-based recognition than did low-value items, but familiarity estimates for high-value items were no higher than those for low-value items. This dissociation provides initial evidence in support of our hypothesis that value effects mediated by selective strategy use are likely to enhance both recollection and familiarity, while more automatic, putatively reward-driven value effects are likely to only enhance recollection, and not familiarity.

Experiment 2

A key question in the present set of studies is what effect the inclusion of free recall tests and feedback have on the mechanism by which value enhances memory. These tests are an important difference between the value-directed remembering paradigm used in Experiment 1 and many of the studies that have examined dopaminergic effects of value on memory. Generally, those studies either presented participants with a single large set of stimuli with a test at the end (e.g., Adcock et al., 2006; Spaniol et al., 2014), or interleaved encoding lists with recognition test lists (e.g., Shigemune et al., 2014; Wolosin et al., 2012). Neither task structure provides participants with experience comparable to the competitive dynamics of an interleaved free recall test with aggregate feedback. We hypothesized that such experience is critical for encouraging the engagement of metacognitive monitoring and control, which produces a test-potentiated selectivity in the application of effective encoding strategies. Thus, in Experiment 2, we used a paradigm identical to that used in Experiment 1, except that we removed the free recall tests and associated feedback. Here, the only test was an R/K recognition test presented after all words had already been encoded. We expected that this manipulation would eliminate any test-potentiated effects of value on memory that are related to the selective application of strategies, while leaving non-strategic effects of value on memory intact.

Method

Participants—We tested 46 individuals (36 female, 10 male, age range 18–23 years, $M_{\text{age}} = 20.02$ years) from the UCLA Psychology department undergraduate student subject pool in this study.

Materials and Procedure—The materials and procedure in this study were identical to those used in Experiment 1, except that no free recall tests were administered during the encoding phase. Words were still presented in distinct lists of 24 items; however, at the end of each list, instead of having a recall test, participants were merely told that they had reached the end of the current list, and they could press a key to continue on to the next list when they were ready. In addition, during the initial instructions for this experiment, participants were told that they would be given a yes/no recognition test later on the words that they were learning. They were also told that on the later recognition test, they would receive points for each studied word that they correctly recognized, with the number of points determined by the value cues that were initially paired with each word. Additionally, they were informed that they would lose one point for any incorrect “yes” responses during the recognition test. No feedback regarding scores was given during the recognition test, however. Note that in Experiment 1, the final recognition test was never mentioned prior to

the beginning of that test. However, given that we did still want the encoding task in Experiment 2 to evoke intentional encoding, we believed it was necessary to indicate that there would be such a test at the end.

Results

Across all participants, there was an effect of value on recollection estimates, $t(45) = 3.50$, $p = .001$, $d = .52$, but no effect of value on familiarity estimates, $t(45) < 1$, $d = .05$ (Figure 3). These results clearly differ from those obtained on the recognition test in Experiment 1.

As in Experiment 1, we also examined how individual differences in self-reported value sensitivity affected value-related changes in process estimates. In this experiment, we classified 25 individuals as reporting Weak value sensitivity and 18 individuals as part of a combined Moderate/Strong group, with a single combined group used because only one individual reported Strong value sensitivity. An additional 3 individuals were excluded due to insufficient self-reports. A 2×2 (item value \times value sensitivity) mixed ANOVA, with repeated measures on the first factor, showed that for recollection, there was a significant effect of item value, $F(1, 41) = 19.64$, $p < .001$, $\eta_p^2 = .32$, no significant effect of value sensitivity, $F(1, 41) < 1$, $\eta_p^2 = .01$, and a significant item value \times value sensitivity interaction, $F(1, 41) = 11.60$, $p = .001$, $\eta_p^2 = .22$, indicating that the effect of value on recollection was significantly larger in the Moderate/Strong group (Figure 4). Planned comparisons show that for individuals in the Weak group, there was no effect of value on recollection, $t(24) = 1.11$, $p = .278$, $d = .22$, but for individuals in the Moderate/Strong group, there was an effect of value on recollection, $t(17) = 3.95$, $p = .001$, $d = .93$. For familiarity, there was no significant main effect of item value, $F(1, 41) < 1$, $\eta_p^2 = .00$, but there was a significant main effect of value sensitivity, $F(1, 41) = 5.23$, $p = .027$, $\eta_p^2 = .11$, and a marginal item value \times value sensitivity interaction, $F(1, 41) = 2.94$, $p = .094$, $\eta_p^2 = .07$ (Figure 4). Planned comparisons show no effect of value on familiarity in the Weak group, $t(24) = -1.40$, $p = .175$, $d = -.28$, nor was there such an effect in the Moderate/Strong group, $t(17) = 1.06$, $p = .304$, $d = .25$.

Discussion

When participants were presented with items of different value at encoding, but were not preparing for free recall tests or given any sort of feedback to encourage the development of strategies for utilizing those values, higher values led to increased recollection, while not increasing familiarity.

Those individuals who reported being indifferent to value, i.e., the Weak group, showed no reliable effects of value on either process measure, however. This result was contrary to our expectation that some value-related enhancement of recollection would occur via relatively automatic processing of value, as we believe occurred for the Weak group in Experiment 1. It may be that in Experiment 1, even people who claimed that their encoding process was not affected by item value were still implicitly sensitive to the point values because of their experience with interspersed tests and associated feedback, and this led to non-strategy-driven effects of value on memory. Under the conditions of Experiment 2, however, value was never made motivationally salient, and thus, individuals who claimed to be insensitive to

value may have in fact been ignoring value entirely. We can then speculate that value may need to be motivationally salient for the effects of value that we describe as automatic to emerge. In other words, non-strategic effects of value on recollection do not appear to be obligatory, but may instead depend on attention to value during encoding.

Individuals who claimed to encode high and low value items differently, i.e., the Moderate and Strong groups, showed reliable effects of value on recollection but not on familiarity. This finding reflects an expected difference from Experiment 1, supporting our hypothesis that when tests are not available to potentiate selective strategy use, selective encoding strategies are unlikely to be consistently engaged, even when people claim to be sensitive to effects of value. To elaborate further on a possible mechanism, it may be that with feedback from interspersed recall tests, as was present in Experiment 1, subjects become more aware of the limitations in their ability to recall items on the list, i.e., that recall of all items on every list is impossible for most participants (cf., Soderstrom & Bjork, 2014; Middlebrooks et al., in press). Thus, when presented with low-value items, they may be more likely to refrain from applying explicit strategies that would enhance encoding, in addition to trying harder to successfully encode the high value items. Without this feedback, as in Experiment 2, participants may simply attend more to high-value items. Such an attentional shift seems to be sufficient to produce non-strategic effects of value on subsequent recollection, but not to evoke the selective use of encoding strategies that would lead to stronger familiarity, as well as to further strengthening of recollection.

Experiment 3

One potential concern with interpreting the results of Experiments 1 and 2 is that the critical results are dependent on the R/K procedure. The R/K procedure relies entirely on self-report measures, and in computing estimates of familiarity, we must make a strong assumption about the two processes being independent (following Yonelinas & Jacoby, 1995). Thus, in Experiments 3–5, we attempt to gain converging evidence for our hypotheses by assessing recollection and familiarity using two additional approaches.

One alternative method is to use a task dissociation procedure, presenting two different recognition tests that are differentially sensitive to the two processes. This approach has the advantage of not requiring an assumption that the two processes are fully independent. We do assume, however, that familiarity can be assessed using a speeded forced-choice test between old items and unrelated lure words, with a limited enough response window that it is unlikely for participants to be able to access recollection. This approach is supported by prior literature demonstrating that familiarity-based memories, indicating that a given word was presented in some form, can influence performance more quickly than does recollection (Curran, Tepe, & Piatt, 2006; Hintzman & Curran, 1994; Mandler, 1980; but see Dewhurst, Holmes, Brandt, & Dean, 2006, for a contradictory viewpoint). More specifically, Hintzman and Curran (1994) showed that there was an initial increase in familiarity when the lag between stimulus onset and response was in the 550–700 ms range, while recollection seemed to influence responses made at longer response intervals. Based on these earlier findings, and taking into account the additional time necessary to process two words on-

screen rather than one, we assume that performance on a forced-choice recognition test with a 750 ms response deadline will primarily reflect familiarity.

To assess recollection, we used a different type of recognition test, designed to assess memory for whether words had been presented in plural or singular form at encoding. There is evidence to suggest that the ability to remember the plurality status in which a given word was presented relies on recollection. Such a result was originally shown in behavioral experiments by Hintzman, Curran, and Oppy (1992) and by Hintzman and Curran (1994). A subsequent event-related potential (ERP) study by Curran (2000) found that the late parietal old/new effect, considered indicative of recollection-based retrieval, was greater for items in which the plurality status was correctly identified relative to switched-plurality lures. However, the frontal N400 effect, often considered a signature of familiarity-based memory, did not differ between these two item types. These and other findings (e.g., Malmberg, Holden, & Shiffrin, 2004; Quamme, Weiss, & Norman, 2010; Rotello, Macmillan, & Van Tassel, 2000) support our assumption that making this distinction requires recollection, as participants must remember a specific detail associated with the item.

It should be noted, however, that the prior studies upon which we based this assumption used yes-no recognition tests, while we used a forced-choice recognition test in order to maintain consistency with the speeded forced-choice test that we used to measure familiarity. There is evidence to suggest that when recollection is unavailable in memory impaired patients, or when its use is discouraged by task instructions, familiarity can be used to successfully distinguish between old items and similar lures on a forced-choice test (Holdstock et al., 2002; Migo, Montalidi, Norman, Quamme, & Mayes, 2009; Westerberg et al., 2006). Thus, we cannot be certain that our forced-choice test of singular/plural form memory relies entirely on recollection. Still, in healthy young adults without strict time constraints or other instructional manipulations, it seems likely that this test is primarily measuring recollection. In addition, we should note that none of our hypotheses depend on a finding of familiarity without recollection. Thus, even if familiarity made a small contribution to performance on the plurals recognition task in addition to recollection, it would not impact the interpretation of our findings.

An additional measure by which we can address our questions of interest in Experiment 3 is to examine how value impacts the likelihood that items will be freely recalled with the correct plurality status, as opposed to being recalled but with the incorrect plurality. We assume that recall with correct plurality requires recollection, based on the aforementioned work showing that distinguishing between the singular and plural forms of a word on yes/no recognition tests requires recollection. It thus seems likely that in order to recall items with singular/plural form accuracy greater than chance, participants must bind that root word with its plurality status in addition to remembering the root word. We then assume that formation of and access to these high-fidelity memories is analogous to other forms of item/context binding, which typically depend on recollection (Yonelinas, 2002; Diana et al., 2007).

Recall of an item in the incorrect singular/plural form, in contrast, indicates the presence of a memory trace strong enough to be freely recalled, yet for which the original plurality status was not successfully incorporated into the trace. It is not entirely clear whether such a

memory should properly be considered recollection, familiarity, or something else. Recent work by Mickes, Seale-Carlisle, and Wixted (2013) found evidence that free recall of an item without contextual detail is possible, contrary to the common assumption that the act of free recall necessitates recollection (e.g., Tulving, 1985). Indeed, this type of recall appears to be distinct from both recollection with context and from familiarity (Brainerd, Gomes, and Moran, 2014; Mickes et al., 2013).

It seems plausible to assume that under conditions in which value only enhances recollection (e.g., with a dopamine-driven strengthening of hippocampal processing), the binding of items with contextual information, such as plurality status, would be preferentially strengthened (cf., Diana et al., 2007). In contrast, we assume that recalling items without contextual detail depends on processing more like that underlying familiarity-based memory. In other words, when value enhances recall both with and without contextual detail, we can assume that the value-related benefit to encoding is not limited to item-context bindings, but affects item memory as well. We would expect such a result to be produced by selective strategy use during encoding.

In addition to providing another opportunity for a conceptual replication of Experiment 1, the free recall data in Experiment 3 also allow us to address two additional issues. First, we are able to examine how value affects the quality of memory under conditions in which the measures are not biased either by testing some items that were already recalled on a previous test, or by the need to discard such items from the analysis, as in Experiment 1. The free recall measures are also of interest because they provide an opportunity to directly compare value effects on the first list, prior to any test-potentiated effects on the encoding process, with subsequent lists on which such effects do have the opportunity to emerge. We expect that if having test experience leads to enhanced strategy use, then effects of value should be strengthened between the first list and subsequent lists, both for recall with correct singular/plural form and for item-only recall. If effects of value are mediated by some other mechanism, they should be relatively constant across lists.

Method

Participants—For the free recall tests, we report data from 112 individuals recruited from the UCLA Department of Psychology undergraduate student subject pool. For the recognition tests, we report data from a subset of 64 individuals from that larger sample. The latter group of individuals (48 female, 16 male, age range = 18–33 years, $M_{age} = 20.37$ years) received a recognition test consistent with the procedure described below. Demographic data are not available for the 48 individuals who were only in the free recall sample. Those participants performed a recognition test with a longer response deadline on the speeded test, which was not sufficiently speeded to reflect primarily familiarity; thus, their recognition data could not be used. Their free recall data were valid, though, as the procedure during the encoding period and recall tests was identical to what was experienced by other participants in Experiment 3. The effects described below are largely similar whether or not these additional 48 participants are included, but we chose to include the additional free recall data to provide increased power.

Materials—The words used in this set of studies met the same psychometric criteria as the items used in Experiments 1 and 2. However, it was also necessary that all words that were either learned during the encoding task or used as lures have a reasonable plural form; thus, some of the specific words used in this task were different from prior experiments. Note that while most words that we included in the study have a plural form that could be generated by adding “s” to the end, we did include a few words for which the plural form is produced by adding “es”, as well as one word requiring replacement of a “y” at the end of the word with “ies”. In order to gain more statistical power, Experiments 3–5 included 8 lists of words, rather than 7 as in Experiments 1 and 2. Items from list 1 were again excluded from the recognition tests, but all 168 words from lists 2–8 were tested during the later recognition tests, half in the plurals test and half in the speeded item recognition test. An additional 84 words, meeting the same psychometric criteria as the studied words, were used as lures for the speeded test.

Procedure—The procedure for each trial was essentially the same as that used in Experiment 1. However, words were presented in either singular or plural form, and participants were instructed that on the free recall tests, they would be required to recall each item in the correct plural or singular form in order to get credit for that item. Indeed, when giving feedback, we only counted items that were recalled in the correct singular/plural (S/P) form as correct. However, items that were recalled in the incorrect form were indicated as such on the scoring sheet to allow for analysis of these items. Additionally, in any analyses in which recalled items were excluded, items that were recalled with incorrect S/P status were also excluded.

Following the proposal by Brainerd et al. (2014) that item-only recall is a distinct process from recollection, we apply an independence correction to the item-only data. The goal of this correction is to account for the fact that items that are recalled in their correct S/P form are ineligible to also be designated as exhibiting item-only memory. Thus, a proper index of item-only memory should be conditionalized on the absence of correct S/P form recall. We thereby computed the rate of item-only recall, for each condition on each list, by dividing the number of items recalled with incorrect S/P form by the sum of that quantity and the number of items not recalled at all.

During the recognition test, half of the participants were given the plurals test first, while half were given the speeded test first. Each test began with instructions and included 4 practice items. After the practice items, participants were given an opportunity to ask the experimenter questions; then, the experimenter typically left the room. Each test included 84 pairs of words, with one word presented on the right side of the screen and one word presented on the left side of the screen. Participants were instructed to press the “m” key, on the right side of the keyboard, if they had previously studied the word appearing on the right side of the screen, and to press the “z” key, on the left side of the keyboard, if they had studied the word appearing on the left side of the screen.

For the plurals test, both the singular form and the plural form of the word were presented on-screen for up to 6 seconds. For the speeded item test, the presented item and an unrelated lure were presented for up to 750 ms, with the lure word always presented in the same S/P

form as the corresponding studied word. In both tests, the response needed to be made while the item was still on the screen. If the allocated presentation time passed without a response being entered, the screen displayed the message, “Too slow! Please respond faster next time” for 2 s. After a response was made, the words immediately disappeared from the screen. Following either a response or the appearance of the “Too slow” screen, a blank screen was displayed for 1.5 s, after which the next word would be presented. The order of items within each test was randomized independently by the computer for each individual participant. After each third of the test (i.e., after each 28 items) a screen came up that allowed the participant to take a short break; they could then press a key to resume the test. After the first full recognition test was complete, instructions were provided on-screen for the second test, along with 4 additional practice items. Then, the participant went on to complete the second recognition test. Finally, a post-study questionnaire was completed at the end, which would again allow us to divide participants by their self-reported sensitivity to value. Note that the questions providing the critical information for determining value sensitivity were phrased in a slightly different way in this and subsequent experiments. Specifically, participants were asked, as separate open-ended questions, “What strategy did you use for encoding the words?”, and, as the next question, “Did you do anything differently for the high-value items?”.

The presentation duration for the speeded test was chosen to be just fast enough to allow for some recognition by familiarity, while being too short to allow for recollection. Indeed, accuracy on this test was relatively low, and participants also often complained that they had great difficulty answering within the allotted time. Thus, it seems that we were successful in choosing a response deadline at the limit of young adults’ capabilities.¹

The paradigm used in this and the following experiments included 16 counterbalancing conditions. The following factors were counterbalanced across participants: assignment of items to value groups (high or low) at encoding, the plurality of a given word (singular or plural), the assignment of item to the type of recognition test (plurals or speeded item test), and which recognition test was presented first (plurals or speeded item test). In addition, across all items, the correct item was equally likely to be on the left side or the right side of the screen, although the assignment of item to side of the screen during the test was not fully independent of all other factors. Finally, the same 84 words were used as lures on the speeded item test across conditions, while the assignment of studied items to either the speeded test or the plurals test was counterbalanced. Thus, each lure word on the speeded test was paired with one old word for half of the participants, and with a different old word for the other half of the participants.

¹Although performance on the speeded item test was below 50% in some conditions, and only two response options were available, these results should not be interpreted as being below chance. There were a substantial number of non-responses on the speeded test, owing to the difficulty of responding in such a short period of time. 50% is only an appropriate chance performance level when only items for which a response was made in the specified amount of time are analyzed. When only items with valid responses were analyzed, accuracy was above 50% for all reported conditions that included the full samples, in each of Experiments 3, 4, and 5. However, as noted in the main text, we chose to characterize non-responses as valid data points reflecting memory failures for the main analysis, rather than excluding them.

Results

Free recall—First, we examine how value affected performance on the initial free recall test. A 2×8 (value \times list) repeated-measures ANOVA on the proportion of items recalled in the correct singular/plural (S/P) form (Table 2) showed a main effect of value, $F(1, 111) = 327.36, p < .001, \eta_p^2 = .75$, and a main effect of list, $F(7, 777) = 14.04, p < .001, \eta_p^2 = .11$. There was also a value \times list interaction, $F(7, 777) = 16.61, p < .001, \eta_p^2 = .13$, as the effect of value on high-fidelity memory became stronger with practice. We also ran an analogous repeated-measures ANOVA on the rate of items recalled in the incorrect S/P form, conditionalized on not being recalled in the correct S/P form, which we refer to as item-only recall. Note that five participants were excluded from this analysis because they recalled all 12 high-value items in the correct S/P form on at least one list, and thus, the critical measure could not be computed. Here, we again found a main effect of value, $F(1, 106) = 86.67, p < .001, \eta_p^2 = .45$, a main effect of list, $F(7, 742) = 2.33, p = .024, \eta_p^2 = .02$, and a value \times list interaction, $F(7, 742) = 2.55, p = .014, \eta_p^2 = .02$, showing that effects of value on item-only memory also became stronger with practice.

To better understand these effects, we examined list group as a factor in subsequent analyses, directly comparing memory on list 1 with memory performance collapsed across lists 2–8. The assumption behind this comparison is that recall on list 1 will not show test-potentiated effects of value, while test experience is available to potentially motivate selective strategy use on subsequent lists. We also examined self-reported value sensitivity as a factor modulating how tests change the effect of value on memory. For this analysis, there were 14 participants in the Weak value sensitivity group, 38 participants in the Moderate group, and 58 participants in the Strong group, with 2 additional participants excluded because their questionnaire responses could not be classified.

We first examined items that were recalled in the correct S/P form (Figure 5, top panels). A $2 \times 2 \times 3$ (item value \times list group \times value sensitivity) mixed ANOVA with repeated measures on the first two factors showed a main effect of item value, $F(1, 107) = 142.04, p < .001, \eta_p^2 = .57$, a main effect of list group, $F(1, 107) = 50.92, p < .001, \eta_p^2 = .32$, and a main effect of value sensitivity, $F(2, 107) = 6.09, p = .003, \eta_p^2 = .10$. There was no interaction between list group and value sensitivity, $F(2, 107) = 1.36, p = .260, \eta_p^2 = .02$, but there was an interaction between item value and value sensitivity, $F(2, 107) = 13.59, p < .001, \eta_p^2 = .20$, with stronger self-reported value sensitivity associated with stronger effects of value on recall with correct form. Importantly, there was also an interaction between list group and item value, $F(1, 107) = 19.83, p < .001, \eta_p^2 = .16$, and a 3-way interaction between list group, item value, and value sensitivity, $F(2, 107) = 4.18, p = .018, \eta_p^2 = .07$, which we probed using separate 2×2 (item value \times list group) repeated measures ANOVAs for each level of value sensitivity. In the Weak group, there was a main effect of item value, $F(1, 13) = 5.50, p = .036, \eta_p^2 = .30$, a main effect of list group, $F(1, 13) = 8.60, p = .012, \eta_p^2 = .40$, but, critically, no interaction between these factors, $F(1, 13) < 1, \eta_p^2 = .00$. Thus, although participants in the Weak group had better high-fidelity memory for high-value items than for low-value items, there was no trend for these effects to get stronger with practice. In contrast, in the Moderate group, there was a main effect of item value, $F(1, 37) = 56.38, p < .001, \eta_p^2 = .60$, a main effect of list group, $F(1, 37) = 25.37, p < .001, \eta_p^2 = .41$, as well as an

interaction between these factors, $F(1, 37) = 12.08, p = .001, \eta_p^2 = .25$. Finally, in the Strong group, there was a main effect of item value, $F(1, 57) = 295.88, p < .001, \eta_p^2 = .84$, a main effect of list group, $F(1, 57) = 22.85, p < .001, \eta_p^2 = .29$, and an interaction between the two, $F(1, 57) = 47.60, p < .001, \eta_p^2 = .46$. Thus, in individuals who reported varying strategies as a function of value, whether in the Moderate or Strong group, the effect of value on high-fidelity memory does appear to have become stronger with practice. Planned comparisons further decomposing these analyses are reported in the Supplemental Material.

We also ran an analogous analysis on the rate of item-only recall (Figure 5, bottom panels). A $2 \times 2 \times 3$ (item value \times list group \times value sensitivity) mixed ANOVA with repeated measures on the first two factors showed a main effect of item value, $F(1, 107) = 35.69, p < .001, \eta_p^2 = .25$, and an interaction between item value and value sensitivity, $F(2, 107) = 3.27, p = .042, \eta_p^2 = .06$, again showing a greater overall effect of value in the Moderate and Strong groups relative to the Weak group. There was also a marginal trend towards an interaction between list group and value sensitivity, $F(2, 107) = 2.42, p = .094, \eta_p^2 = .04$, and a marginal trend towards a 3-way interaction, $F(2, 107) = 2.62, p = .078, \eta_p^2 = .05$. No other effects are significant, all $F < 1.70$, all $p > .188$, all $\eta_p^2 < .03$. Still, based on the marginal 3-way interaction, and also to allow a comparison between item-only recall and the results above for the correct-plurality recall, we also examined the results from the separate 2×2 (item value \times list group) ANOVA for each level of value sensitivity. For the Weak group, there was no main effect of item value, $F(1, 13) < 1, \eta_p^2 = .06$, no significant effect of list group, $F(1, 13) = 3.00, p = .107, \eta_p^2 = .19$, nor was there an interaction, $F(1, 13) = 1.04, p = .327, \eta_p^2 = .07$. Note also that the numeric trends were for the effect of value on item-only recall to be reduced in this group following the first test. For the Moderate group, however, there was a main effect of item value, $F(1, 37) = 26.88, p < .001, \eta_p^2 = .42$, but no main effect of list group, $F(1, 37) < 1, \eta_p^2 = .01$. The list group \times value interaction was not significant, $F(1, 37) = 2.22, p = .144, \eta_p^2 = .06$, but the apparent trend for this group was for the effect of value to be larger following the first list. For the Strong group, there was also a main effect of item value, $F(1, 57) = 56.12, p < .001, \eta_p^2 = .50$, no main effect of list group, $F(1, 57) = 1.85, p = .179, \eta_p^2 = .03$, and a significant interaction between list group and value, $F(1, 57) = 5.26, p = .025, \eta_p^2 = .08$, showing a significantly stronger value effect after the first list. Because the critical item value \times list group interaction effect was only significant for the Strong group, but there was also a trend in the same direction in the Moderate group, we also ran a $2 \times 2 \times 2$ (item value \times list group \times value sensitivity) mixed ANOVA on data from these two groups. This analysis showed a main effect of item value, $F(1, 94) = 74.45, p < .001, \eta_p^2 = .44$, and an interaction between list group and item value, $F(1, 94) = 6.80, p = .011, \eta_p^2 = .07$. There was no trend whatsoever towards a 3-way interaction, $F(1, 94) < 1, \eta_p^2 = .00$, and no other effects in this analysis were significant, all $F < 1.86$, all $p > .176$, all $\eta_p^2 < .02$. Thus, we can assume that the effects of interest were similar across both the Moderate and Strong groups, with both groups showing more item-only recall for high-value items than for low-value items overall, and, importantly, the effect becoming stronger after the first list across both groups. This analysis is also broken down further in the Supplemental Material.

Recognition data—Similar to Experiment 1, we only scored the recognition data in this experiment obtained from non-recalled items. Note that items with a reaction time (RT) less than 50 ms were excluded from the analysis, while items for which no response was provided in the allowed amount of time were counted as incorrect. We felt that it would be appropriate to classify such trials as memory failures given that time constraints are an integral part of the speeded recognition task (see Goldstone & Medin, 1994, for a similar approach to this issue).

Across all 64 participants with valid recognition data, we found a significant effect of value on the plurals test, $t(63) = 3.21, p = .002, d = .40$, and a trend for an effect of value on the speeded item test that approaches significance, $t(63) = 1.95, p = .056, d = .24$ (Figure 6). We also analyzed effects of self-reported value sensitivity on recognition results (Table 3). On the plurals test, a 2×3 (item value \times value sensitivity) mixed ANOVA shows a main effect of item value, $F(1, 61) = 10.38, p = .002, \eta_p^2 = .15$, a main effect of value sensitivity, $F(2, 61) = 4.75, p = .012, \eta_p^2 = .13$, but no interaction, $F(2, 61) = 1.98, p = .146, \eta_p^2 = .06$. We also performed a similar set of analyses on performance on the speeded test. A 2×3 (item value \times value sensitivity) mixed ANOVA, with repeated measures on the first factor, found a main effect of item value, $F(1, 61) = 4.62, p = .036, \eta_p^2 = .07$, but no main effect of value sensitivity, $F(2, 61) < 1, \eta_p^2 = .02$, and no interaction between item value and value sensitivity, $F(2, 61) < 1, \eta_p^2 = .02$. Planned comparisons separately examining effects of value in each value sensitivity group are reported in the Supplemental Material.

Discussion

The results of Experiment 3 replicate and extend our findings from Experiment 1 in several ways. First, we focus on the results from the free recall tests across all participants. High-value items were more likely to be recalled with correct plurality than were low-value items, and when items were not recalled with the correct plurality, the item alone was still more likely to be recalled, in the incorrect S/P form, for high-value items. If we accept the assumption that recall with plurality requires recollection, while item-only recall is akin to familiarity, then this result constitutes a replication of one of our key findings from Experiment 1.

It is also notable how effects of value on free recall differed with respect to participants' self-reported value sensitivity. Individuals in the Moderate and Strong groups showed better recall with the correct S/P form, and better item-only recall, for high-value items. In the Weak group, i.e., people who self-reported being indifferent to value, there was still a significant effect of value on recall of items with accurate plurality information. However, there was no effect of value on the rate of item-only recall, particularly after the first list. These results conceptually replicate another key finding from Experiment 1, which is that individuals who are not explicitly sensitive to value show an enhancement only in recollection, likely driven by more automatic mechanisms, while individuals who are explicitly regulating their encoding based on item value show a broader-based increase in memory for high-value items.

It is worth noting that the effect of value on recall with correct S/P form was reduced in the Weak group compared to the Moderate and Strong groups, which contrasts with the lack of

an item value \times value sensitivity interaction on recollection in Experiment 1. We would expect non-strategic and strategy-driven mechanisms to additively enhance recollection, which would imply a reduction in the effect of value on recollection in the Weak group. While we did not find evidence of additive effects of both mechanisms on the measure of recollection in Experiment 1, it is likely that the need to exclude previously-recalled items from the analysis of recognition data accounts for the difference. Indeed, in a supplementary analysis that we do not formally report, we found that when previously-recalled items were not excluded, there was an item value \times value sensitivity interaction on recollection data in Experiment 1, with only a marginal simple effect of value on recollection in the Weak group, but robust effects in the Moderate and Strong groups. Thus, we can assume that value most likely does, in fact, affect recollection more strongly under conditions consistent with selective strategy use, as we observed in Experiment 3.

Experiment 3 also provided an opportunity to test how effects of value change with test experience. In the Moderate and Strong groups, high value enhanced both types of recall memory, with and without correct S/P form, more strongly after the first test. In contrast, while the Weak group did show increased recall with correct S/P form for high value items overall, this effect did not change with test experience. Additionally, for item-only recall in the Weak group, we saw, if anything, a reverse pattern from what other participants showed, as the numeric trend towards a value-related benefit that was present on list 1 disappeared on later lists. These results provide further evidence that test experience can potentiate value-related strengthening of memory in a manner consistent with selective strategy use, but only in people whose self-reports indicate that they intentionally varied their encoding process as a function of the value of the items being learned.

Finally, we also used a task dissociation method to isolate expressions of recollection and familiarity-based memory in recognition. When computed across the entire sample, these results largely replicated Experiment 1. High-value items showed better performance on the plurals test, assumed to reflect largely recollection, and also showed marginally better performance on the speeded test, assumed to reflect familiarity. The fact that we obtained such results even when it was not necessary to make a strong assumption of independence, and even when items that were recalled on the free recall tests were excluded from the analysis, should help to strengthen confidence in the veracity of our broader pattern of results.

It is necessary to note, however, that unlike the results obtained when data from the free recall test were split by self-reported value sensitivity, the analysis of recognition data did not show any interactions between item value and value sensitivity. One possible explanation for the inconclusive results is that effects of value on recognition, particularly in the speeded test, were small to begin with, so reducing power by dividing the sample may have had a particularly detrimental effect. A second possible explanation follows from the fact that this experiment had an unusually large number of counterbalancing conditions. Given that the sorting of participants by value sensitivity was necessarily post hoc, and thus it was not possible to balance the groups across counterbalancing conditions, there could be interactions between counterbalancing condition (i.e., which specific words were associated with high and low value cues, and which ended up being tested in each of the two

recognition tests), value sensitivity level, and effects of value on the memory test that would overshadow the true effects of the manipulations of interest. Indeed, a close examination of data from Experiment 3 found evidence consistent with the presence of such confounding interactions. Nevertheless, this should not overshadow the fact that when data from the recognition tests were collapsed across all participants, allowing both for increased power and for full counterbalancing of relevant factors, the recognition results were largely consistent with other findings from Experiments 1 and 3.

Experiment 4

The primary aim of Experiment 4 was to address a potential source of bias found in the recognition data in both Experiment 1 and Experiment 3. Specifically, as described above, we excluded items that had been previously recalled from the analyses of recognition data because of the possibility that subsequent recognition performance would be influenced by memory for the successful recall event, rather than the initial encoding event. This exclusion does not, however, rule out the possibility that value effects were enhanced by the mere attempt to recall items on the free recall test, rather than by our proposed mechanism of selective strategy use during encoding. In addition, we were concerned that excluding all previously-recalled items might have distorted the true pattern of effects. To resolve these issues, we ran a modified procedure in which participants were only given interspersed free recall tests on 3 of the 8 lists. This meant that items from the other 5 lists could be analyzed without contamination from prior recall tests. If having prior experience with a test potentiates selective use of strategies during encoding, high-value items should still show stronger recollection and stronger familiarity than low-value items, even when those items were not previously tested via free recall.

Method

Participants—Data from 48 students (35 female, 13 male, age range = 18–36 years, M_{age} = 20.57 years) from the UCLA Department of Psychology undergraduate student subject pool are reported in this study.

Materials and Procedure—The materials and procedure were identical to those used in Experiment 3, except that, as noted above, free recall tests were only presented on 3 of the 8 lists. The first list, for which items were not included in the recognition test, was always given a free recall test. After that, the computer randomly chose one list of lists 2–4 to get the second free recall test, and randomly chose one of lists 5–8 to get the third free recall test. Participants were not informed about how the tested lists would be chosen, but were told that some lists would have a recall test and some lists would not. They were also reminded to always study the words as if they were going to have a recall test on that list. Participants were not told whether there would be a test on a given list until presentation of that list was complete. If there was to be a test, the instructions for the test would be displayed, otherwise a message would be displayed saying that “you will not be tested on this list,” and the participant could then press a key to continue to the next list. Participants were also not told about the recognition test in this experiment until immediately before it began.

Results

Free recall—We first examined whether subsequent performance on the free recall tests differed based on which lists were randomly chosen to be tested. Test 2 could be positioned after list 2, 3, or 4, while test 3 could be positioned after list 5, 6, 7, or 8. We ran 2×3 (item value \times test 2 position) mixed ANOVAs, with repeated measures on the first factor, on recall with S/P form and on item-only recall, for test 2 and test 3. In addition, we ran 2×4 (item value \times test 3 position) mixed ANOVAs for the same performance measures for test 3. All analyses showed a main effect of value, both on recall with correct S/P form, all $F(1, 45) = 71.22$, all $p < .001$, all $\eta_p^2 = .61$, and on item-only recall, all $F(1, 45) = 15.77$, all $p < .001$, all $\eta_p^2 = .26$. However, there were no main effects of test position on recall with correct S/P form, all $F < 1.07$, all $p > .351$, all $\eta_p^2 < .06$, nor were there any such effects on item-only recall, all $F < 1.62$, all $p > .209$, all $\eta_p^2 < .09$. Finally, there were no reliable interactions between value and test position, either for recall with correct S/P form, all $F < 1$, all $\eta_p^2 < .03$, or for item-only recall, all $F < 1.43$, all $p > .247$, all $\eta_p^2 < .09$. Thus, we collapsed across test position in all further analyses.

Next, we examined overall effects of value on the free recall tests, and how these effects changed with test experience (Table 4). A 2×3 (item value \times list group) repeated-measures ANOVA on the proportion of items recalled in the correct S/P form showed a main effect of value, $F(1, 47) = 119.16$, $p < .001$, $\eta_p^2 = .72$, no main effect of list group, $F(2, 94) = 1.96$, $p = .147$, $\eta_p^2 = .04$, and an interaction between these factors, $F(2, 94) = 12.20$, $p < .001$, $\eta_p^2 = .21$, showing stronger value effects on later tests. We also ran an analogous analysis on the corrected rate of item-only recall. This analysis showed a main effect of value, $F(1, 47) = 71.97$, $p < .001$, $\eta_p^2 = .60$, but no main effect of list group, $F(2, 94) < 1$, $\eta_p^2 = .01$, nor was there a significant interaction between value and list group, $F(1, 47) = 1.31$, $p = .274$, $\eta_p^2 = .03$, although there was a numeric trend for value effects to get stronger with practice.

We next compared performance on items from list 1 to the average performance on items from the 2 lists that were tested out of the final 7 lists, in an attempt to replicate our findings from Experiment 3. Because there were only 3 participants in the Weak group of this experiment, too few to produce meaningful inferential statistics, and because we did not find theoretically relevant differences between the Moderate and Strong groups in previous experiments, we do not report detailed analyses as a function of self-reported value sensitivity in the main text. Analyses comparing the latter two groups, with 11 participants in the Moderate group and 34 in the Strong group, are reported in the Supplemental Material. Note as well that in Experiment 3, value effects did not get stronger between List 1 and subsequent lists for individuals in the Weak group, contrary to what we observed for participants in the Moderate and Strong groups. Thus, in the interest of focusing on comparisons where we expect to observe the critical effects, we exclude Weak value sensitivity individuals from the analyses that follow.

We first ran a 2×2 (item value \times list group) mixed ANOVA, with repeated measures on the first two factors, on the rate of recall with correct S/P form for individuals showing Moderate or Strong value sensitivity (Supplemental Table 1). This analysis showed a main effect of value, $F(1, 44) = 111.15$, $p < .001$, $\eta_p^2 = .72$, but no main effect of list group, $F(1,$

44) = 2.20, $p = .145$, $\eta_p^2 = .05$. There was, however, a significant interaction between value and list group, $F(1, 44) = 22.29$, $p < .001$, $\eta_p^2 = .34$, indicating that value effects became stronger with practice. We also ran a 2×2 (item value \times list group) repeated-measures ANOVA on the corrected rate of item-only recall for the same participants (Supplemental Table 1). This analysis found a main effect of item value, $F(1, 44) = 70.93$, $p < .001$, $\eta_p^2 = .62$, no main effect of list group, $F(1, 44) < 1$, $\eta_p^2 = .00$, and a marginal item value \times list group interaction, $F(1, 44) = 4.04$, $p = .051$, $\eta_p^2 = .08$, reflecting a trend for value effects to be strengthened by test potentiation in item-only recall as well.

Recognition test—Examining recognition performance for items from the 5 lists that were not previously tested, we found a reliable effect of value on the plurals test, intended to assess recollection, $t(47) = 4.73$, $p < .001$, $d = .68$, and also a significant effect of value on the speeded test, intended to assess familiarity, $t(47) = 2.70$, $p = .010$, $d = .39$ (Figure 7). As was the case with the free recall data, it was not particularly informative to examine effects of value sensitivity because there were not enough participants in the Weak group to support use of inferential statistics, and we did not expect to see theoretically relevant differences between the Moderate and Strong groups. A comparison of the latter two groups is, however, reported in the Supplemental Material.

Discussion

In this experiment, we largely replicated the key findings from Experiments 1 and 3. Most notably, across the entire sample, high-value items were remembered significantly better than low-value items on both a plurals test, assumed to reflect primarily recollection, and on a speeded test, assumed to reflect familiarity. In this experiment, unlike in the prior experiments, recognition performance for the critical items was not biased by having attempted to recall them on an earlier test, nor was it biased by the need to discard items that were correctly recalled earlier. In addition, when separating effects of value on recollection and familiarity, it was not necessary to assume that recollection and familiarity are independent processes, nor to rely on participants' self-reports of recollective experience. Thus, our primary pattern of results does not appear to depend on those particular assumptions.

The results from the free recall tests also largely replicate what we observed in Experiment 3. Specifically, recall with correct plural form was reliably greater for high-value items, and this effect clearly became stronger with practice. Item-only recall also was better for high-value items, and this effect also tended to get stronger with practice. If we accept the assumption that these measures reflect recollection and familiarity-like memory, respectively, then these results replicate our findings of a value-related increase in both types of memory after a test, providing further evidence consistent with our hypothesis that interspersed tests stimulate value-related selectivity in the use of study strategies during the encoding phase of subsequent lists.

Experiment 5

While Experiments 3 and 4 largely replicate the findings of Experiment 1, another key question is whether we can replicate the findings from Experiment 2 using a task

dissociation procedure rather than an R/K test to assess dual process correlates. We used the same recognition procedure as in Experiments 3 and 4, but eliminated the opportunities for practice and feedback by removing all free recall tests. We expected to find effects of value on recollection but not familiarity, even for individuals who do report being sensitive to value, similar to what we found in Experiment 2.

Method

Participants—Data collected from 64 students (45 female, 19 male, age range = 18–34 years, $M_{age} = 20.56$ years) who participated for course credit via the UCLA Department of Psychology undergraduate student subject pool are included in this experiment.

Materials and Procedure—The items in this study were identical to those used in Experiment 3 and 4. The procedure was similar as well, except that, as in Experiment 2, instead of having a free recall test at the end of each 24-item list, participants were instructed that they “had finished learning this set of words,” and were to press a key to continue on to the next set. During the initial instructions, participants were informed that they would be completing a recognition test later, in which they would have to choose between a word that they saw and a word that they did not see, and they would get the points associated with a given word if they chose correctly. They were also told that they would need to know whether the word was plural or singular when taking the later test, in order to motivate paying attention to the singular/plural status during encoding.

Results

Across all participants (Figure 8), we found a significant effect of value on the plurals test, $t(63) = 2.41, p = .019, d = .30$, but no effect of value on the speeded item test, $t(63) < 1, d = -.03$. As in the preceding experiments, we also examined effects of self-reported differences in value sensitivity. In this experiment, there were 23 individuals in the Weak value sensitivity group, 21 individuals in the Moderate group, and 19 individuals in the Strong group, with one participant excluded who could not be reliably classified. A 2×3 (item value \times value sensitivity) mixed ANOVA, examining performance on the plurals test, found a main effect of item value, $F(1, 60) = 10.02, p = .002, \eta_p^2 = .14$, a main effect of value sensitivity, $F(2, 60) = 3.91, p = .025, \eta_p^2 = .12$, and a significant interaction, $F(2, 60) = 7.34, p = .001, \eta_p^2 = .20$ (Figure 9). Planned comparisons showed no effect of value on plurals test performance in the Weak value sensitivity group, $t(22) = -1.51, p = .144, d = -.32$, but participants in the Moderate group did show better memory for high value items, $t(20) = 3.13, p = .005, d = .68$, as did participants in the Strong group, $t(18) = 2.94, p = .009, d = .67$. An analogous 2×3 (item value \times value sensitivity) ANOVA examining speeded test performance found no main effect of item value, $F(1, 60) < 1, \eta_p^2 = .00$, no main effect of value sensitivity, $F(1, 60) = 1.94, p = .152, \eta_p^2 = .06$, and no interaction between these factors, $F(2, 60) < 1, \eta_p^2 = .01$ (Figure 9). Planned comparisons show that there was no effect of value on familiarity in the Weak group, $t(22) < 1, d = -.19$, nor was there such an effect in the Moderate group, $t(20) < 1, d = .04$, or in the Strong group, $t(18) < 1, d = .00$.

Discussion

In this experiment, we largely replicated the pattern of effects observed in Experiment 2. Specifically, when participants were not given an opportunity for practice and feedback, we saw an effect of value on the plurals test, measuring recollection, but not on the speeded test, measuring familiarity. The effect of value on recollection appears to be smaller than it was under similar conditions in which participants did gain experience with a free recall test (e.g., Experiment 4), but there was still a significant effect present. For familiarity, however, it seems that there was no hint of an effect. The effects shown here may represent the degree to which relatively automatic, putatively dopamine-driven effects of value can improve memory in this type of paradigm.

It is also notable that we replicated the relationship between self-reported value sensitivity and item value that we observed in Experiment 2. Specifically, individuals who claimed to be insensitive to value showed no effect of value on either the recollection-based or familiarity-based test. In contrast, both groups of participants who did report being sensitive to value showed significant effects of value on recollection, but not on familiarity. These results reinforce the idea that value does not necessarily enhance memory if participants are not motivated to attend to it.

General Discussion

The experiments reported here provide insight into the mechanisms by which value can impact the efficacy of memory encoding. Specifically, our results suggest that having experience with free recall tests with feedback, interspersed with learning, leads people to apply strategies selectively during encoding of high-value items relative to low-value items. Without such experience, or with this experience but without the explicit intention to encode high-value items more effectively, value appears to have only non-strategic effects, which may be driven by activity in the dopaminergic reward system. Finally, when no experience with interspersed recall tests was available, and participants reported being indifferent to value during encoding, there was no effect of value on subsequent memory performance. These results suggest a novel dissociation between the different ways in which value can affect the memory encoding process, contrasting strategy-driven effects with more automatic, non-strategic effects. These results also provide important context for our prior fMRI work (Cohen et al., 2014, 2016), and for relating those findings with other work on reward-driven learning (e.g., Adcock et al., 2006; Gruber et al., 2014, 2016; Shigemune et al., 2014; Shohamy & Adcock, 2010; Wolosin et al., 2012).

These conclusions emerge from applying a dual process analysis to the data from the five experiments described herein. The fact that different measures, with different sets of underlying assumptions, generally show converging results should allow for confidence that the effects we observed are independent of the assumptions being made by any single approach used to test recollection and familiarity processes. In Experiment 1, high-value items in the value-directed remembering paradigm tended to show enhanced recollection and familiarity, as measured by independence-corrected R/K judgments. In contrast, in Experiment 2, in which the paradigm was modified to remove the interspersed free recall tests, value only strengthened recollection on this same measure. In Experiments 3, 4, and 5,

we found similar results as in Experiments 1 and 2, while using different methods to assess recollection and familiarity. In Experiment 3, with a similar encoding paradigm as in Experiment 1, we saw a strong trend towards value strengthening both recollection and familiarity on later recognition tests, as assessed by, respectively, plurals recognition and speeded item recognition, which we assume to be differentially sensitive to recollection and familiarity, respectively. During the free recall phase, we also saw significant value-related increases in both recall with correct plurality, and recall of the item alone, which we suggest to be an additional means of assessing recollection and familiarity-like memory, respectively. These effects also became stronger with test experience, supporting a further prediction of our hypothesis about test-potentiated effects on learning. In Experiment 4, in which we removed the potential bias of the free recall test on memory for the critical recognition items, but participants were still given some exposure to free recall tests during encoding, we again saw reliable effects of value on both recollection and familiarity using the same measures as in Experiment 3. However, in Experiment 5, as in Experiment 2, when the free recall tests were removed entirely, value affected performance on the putatively recollection-driven plurals test, but not on the presumably familiarity-driven speeded test. Thus, there was a clear tendency, replicating across multiple methodologies, for value to improve both recollection and familiarity when recall tests were interspersed at encoding, but to only improve recollection without such tests.

As noted above, engaging deep semantic strategies has been shown in prior literature (e.g., Yonelinas, 2002) to enhance estimates of both recollection and familiarity, while more automatic effects of reward on memory tend to exclusively benefit recollection (e.g., Gruber et al., 2016; Shigemune et al., 2014). However, another piece of evidence supporting this interpretation comes from the way in which self-reported value sensitivity impacted dual process correlates of memory in the current work. Specifically, in Experiment 1, people who reported doing something different to encode high-value relative to low-value items showed value-related improvement in both recollection and familiarity, as assessed by an R/K test, while people who reported being indifferent to value showed improvement only in recollection. If we accept that recall with correct S/P form relies on recollection, while item-only recall without correct S/P form is similar to familiarity, the free recall data from Experiment 3 provide an important replication of this finding from Experiment 1. People who were indifferent to value only showed a value-related enhancement in recall in the correct S/P form, while other participants who reported being more sensitive to value showed an increase in both recall measures. In addition, the value-related memory increase that was present in recollection for individuals who were indifferent to value did not become stronger with test experience, while in individuals who were more sensitive to value, value effects on both free recall measures became stronger after the first test. While retrospectively reporting a sensitivity to value does not necessarily mean that one's strategies were being explicitly varied as a function of value, such sensitivity would seem to be a prerequisite to doing so. In other words, it is unlikely that individuals who reported being indifferent to item value would have explicitly varied their strategies based on those values. These findings thus converge with the prior literature to support our hypothesis that strategy-driven effects of value tend to enhance both recollection and familiarity, while more automatic, non-strategic effects of value only enhance recollection.

Self-reported value sensitivity seems to have had a different pattern of effects on dual-process correlates in Experiments 2 and 5, when free recall tests were not interspersed with encoding. In both of those experiments, people who reported trying to do something different to encode high-value items showed effects of value on recollection alone. These results suggest that without interspersed recall tests with feedback, value is not causing people to use strategies selectively, and instead only enhances memory via more automatic processes such as the dopamine-driven strengthening of hippocampal processing. We interpret the fact that value did not also enhance familiarity-based memory to mean that subjects in these experiments tend to not vary their use of deep semantic strategies as a function of value, even when they report that they are encoding high and low value items differently. At the same time, those who reported being indifferent to value showed no effect of value on either process measure. A plausible post hoc explanation for this difference from Experiments 1 and 3 is that interspersed tests help to make the point values more salient for non-strategic mechanisms, even when there is no specific intention to modulate attention at encoding based on value. In other words, it may be that interspersed tests prevent learners from ignoring value entirely, as they may be able to do when such tests are not present. It may also be the case that point values are not as effective in enhancing memory as the monetary rewards that have been used in previous studies. Participants do appear to be motivated by points when feedback is provided, but in the absence of such feedback, the rewards may be too abstract for some individuals.

Our findings suggest a role for metacognition in the response to value, in that participants can become aware of limitations on memory during interspersed recall tests, and adjust their encoding strategies to strengthen important items in memory at the expense of less important items. This effect seems analogous to the idea of test-potentiated learning, but instead of enhancing memory for all items, tests in this context potentiate increased selectivity in how encoding strategies are applied. This strategy-driven mechanism should be seen as distinct from, and complementary to, other mechanisms by which reward can affect memory. For instance, strategy-driven effects are maintained with healthy aging while dopamine-driven effects of value may not be (e.g., Cohen et al., 2016), and thus, our findings may have important implications for how older adults can be trained to remain sensitive to the importance of studied items. In addition, using strategies to enhance memory for high-value items appears to lead to a much stronger effect of value on memory than does the more automatic, reward-driven mechanism, which could have important practical consequences for learning. Thus, the apparent dissociation between strategy-driven versus non-strategic, reward-driven effects of value should provide an important framework for further work in this domain.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the National Institutes of Health (grant F31 AG047048 to Michael S. Cohen, and grant T32 NS047987, via a training position awarded to Michael S. Cohen). We thank Courtney Clark for a key

suggestion related to task design, and thank Yasmine Sherafat, Katie Swinnerton, James Mutter, Andrea Del Castillo, and Brent Amiri for help running participants.

References

- Adcock RA, Thangavel A, Whitfield-Gabrieli S, Knutson B, Gabrieli JDE. Reward-motivated learning: Mesolimbic activation precedes memory formation. *Neuron*. 2006; 50:507–517. [PubMed: 16675403]
- Ariel R, Dunlosky J, Bailey H. Agenda-based regulation of study-time allocation: When agendas override item-based monitoring. *Journal of Experimental Psychology: General*. 2009; 138:432–447. [PubMed: 19653800]
- Arnold KM, McDermott KB. Test-potentiated learning: Distinguishing between direct and indirect effects of tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2013; 39:940–945.
- Bahrick HP, Hall LK. The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language*. 2005; 52:566–577.
- Brainerd CJ, Gomes CFA, Moran R. The two recollections. *Psychological Review*. 2014; 121:563–599. [PubMed: 25347309]
- Castel, AD. The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In: Benjamin, AS., Ross, BH., editors. *The psychology of learning and motivation*. Vol. 48. San Diego: Academic Press; 2008.
- Castel AD, Benjamin AS, Craik FIM, Watkins MJ. The effects of aging on selectivity and control in short-term recall. *Memory and Cognition*. 2002; 30:1078–1085. [PubMed: 12507372]
- Cohen MS, Rissman J, Suthana NA, Castel AD, Knowlton BJ. Value-based modulation of memory encoding involves strategic engagement of fronto-temporal semantic processing regions. *Cognitive, Affective, and Behavioral Neuroscience*. 2014; 14:578–592.
- Cohen MS, Rissman J, Suthana NA, Castel AD, Knowlton BJ. Effects of aging on value-directed modulation of semantic network activity during verbal learning. *NeuroImage*. 2016; 125:1046–1062. [PubMed: 26244278]
- Craik FIM, Lockhart RS. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*. 1972; 11:671–684.
- Curran TC. Brain potentials of recollection and familiarity. *Memory and Cognition*. 2000; 28:923–938. [PubMed: 11105518]
- Curran, T., Tepe, KL., Piatt, C. ERP explorations of dual processes in recognition memory. In: Zimmer, HD, Mecklinger, A., Lindenberger, U., editors. *Binding in Human Memory: A Neurocognitive Approach*. Oxford: Oxford University Press; 2006. p. 467–492.
- Dewhurst SA, Holmes SJ, Brandt KR, Dean GM. Measuring the speed of the conscious components of recognition memory: remembering is faster than knowing. *Consciousness and Cognition*. 2006; 15:147–162. [PubMed: 16019226]
- Diana RA, Yonelinas AP, Ranganath C. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends in Cognitive Science*. 2007; 11:379–386.
- Eichenbaum H, Yonelinas AP, Ranganath C. The medial temporal lobe and recognition memory. *Annual Reviews of Neuroscience*. 2007; 30:23–52.
- Eldridge LL, Sarfatti S, Knowlton BJ. The effect of testing procedure on remember-know judgments. *Psychonomic Bulletin & Review*. 2002; 9:139–145. [PubMed: 12026946]
- Goldstone RL, Medin DL. Time course of comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1994; 20:29–50.
- Gruber MJ, Gelman BD, Ranganath C. States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*. 2014; 84:486–496. [PubMed: 25284006]
- Gruber MJ, Ritchey M, Wang SF, Doss MK, Ranganath C. Post-learning hippocampal dynamics promote preferential retention of rewarding events. *Neuron*. 2016; 89:1110–1120. [PubMed: 26875624]
- Hintzman DL, Curran T. Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language*. 1994; 33:1–18.

- Hintzman DL, Curran T, Oppy B. Effects of similarity and repetition on memory: Registration without learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1992; 18:667–680.
- Holdstock JS, Mayes AR, Roberts N, Cezayirli E, Isaac CL, O'Reilly RC, Norman KA. Under what conditions is recognition spared relative to recall after selective hippocampal damage in humans? *Hippocampus*. 2002; 12:341–351. [PubMed: 12099485]
- Jacoby LL. A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*. 1991; 30:513–541.
- Jacoby LL, Dallas M. On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*. 1981; 110:306–340. [PubMed: 6457080]
- Kapur S, Craik FIM, Tulving E, Wilson AA, Houle S, Brown GM. Neuroanatomical correlates of encoding in episodic memory: Levels of processing effect. *Proceedings of the National Academy of Sciences USA*. 1994; 91:2008–2011.
- MacInnes JJ, Dickerson K, Chen NK, Adcock RA. Cognitive neurostimulation: Learning to volitionally sustain ventral tegmental area activation. *Neuron*. 2016; 89:1331–1342. [PubMed: 26948894]
- Malmberg KJ, Holden JE, Shiffrin RM. Modeling the effects of repetitions, similarity, and normative word frequency on old-new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, Cognition*. 2004; 30:319–331.
- Mandler G. Recognizing: The judgment of previous occurrence. *Psychological Review*. 1980; 87:252–271.
- Mickes L, Seale-Carisle TM, Wixted JT. Rethinking familiarity: Remember/Know judgments in free recall. *Journal of Memory and Language*. 2013; 68:333–349. [PubMed: 23637470]
- Middlebrooks CD, Murayama K, Castel AD. Test expectancy and memory for important information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. (in press).
- Miendlarzewska EA, Bavelier D, Schwartz S. Influence of reward motivation on human declarative memory. *Neuroscience and Biobehavioral Reviews*. 2016; 61:156–176. [PubMed: 26657967]
- Migo M, Montaldi D, Norman KA, Quamme J, Mayes A. The contribution of familiarity to recognition memory is a function of test format when using similar foils. *Quarterly Journal of Experimental Psychology*. 2009; 62:1198–1215.
- Murayama K, Kitagami S. Consolidation power of extrinsic rewards: Reward cues enhance long-term memory for irrelevant past events. *Journal of Experimental Psychology: General*. 2014; 143:15–20. [PubMed: 23421444]
- Murayama K, Kuhbandner C. Money enhances memory consolidation—But only for boring material. *Cognition*. 2011; 119:120–124. [PubMed: 21292249]
- Pyc MA, Rawson KA. Why is test-restudy practice beneficial for memory? An evaluation of the mediator shift hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2012; 38:737–746.
- Quamme JR, Weiss DJ, Norman KA. Listening for recollection: a multi-voxel pattern analysis of recognition memory retrieval strategies. *Frontiers in Human Neuroscience*. 2010; 4:61. [PubMed: 20740073]
- Rajaram S. Remembering and knowing: Two means of access to the personal past. *Memory & Cognition*. 1993; 2:89–102.
- Rotello C, Macmillan NA, Van Tassel G. Recall-to-reject in recognition: evidence from ROC curves. *Journal of Memory and Language*. 2000; 43:67–88.
- Sheridan H, Reingold EM. Recognition memory performance as a function of reported subjective awareness. *Consciousness and Cognition*. 2011; 20:1363–1375. [PubMed: 21605985]
- Sheridan H, Reingold EM. Levels of processing influences both recollection and familiarity: Evidence from a modified remember-know paradigm. *Consciousness and Cognition*. 2012; 21:438–443. [PubMed: 22051555]
- Shigemune Y, Tsukiura T, Kambara T, Kawashima R. Remembering with gains and losses: Effects of monetary reward and punishment on successful encoding activation of source memories. *Cerebral Cortex*. 2014; 24:1319–1331. [PubMed: 23314939]

- Shohamy D, Adcock RA. Dopamine and adaptive memory. *Trends in Cognitive Sciences*. 2010; 14:464–472. [PubMed: 20829095]
- Slamecka NJ, Graf P. The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*. 1978; 4:592–604.
- Soderstrom NC, Bjork RA. Testing facilitates the regulation of subsequent study time. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2014; 73:99–115.
- Spaniol J, Schain C, Bowen HJ. Reward-enhanced memory in younger and older adults. *Journal of Gerontology, Series B: Psychological Sciences and Social Sciences*. 2014; 69:730–740.
- Szpunar KK, Khan NY, Schacter DL. Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences USA*. 2013; 110:6313–6317.
- Szpunar KK, McDermott KB, Roediger HL III. Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2008; 34:1392–1399.
- Toglia MP, Battig WF. *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*. 1978; 41:531–533.
- Toppino TC, Cohen MS. Metacognitive control and spaced practice: Clarifying what people do and why. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2010; 36:1480–1491.
- Tulving E. Memory and consciousness. *Canadian Psychology*. 1985; 26:1–12.
- Weinstein Y, Gilmore AW, Szpunar KK, McDermott KB. The role of test expectancy in the build-up of proactive interference in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 2014; 40:1039–1048.
- Westerberg CE, Paller KA, Weintraub S, Mesulam MM, Holdstock JS, Mayes AR, Reber PJ. When memory does not fail: Familiarity-based recognition in mild cognitive impairment and Alzheimer’s disease. *Neuropsychology*. 2006; 20:193–205. [PubMed: 16594780]
- Wissman KT, Rawson KA, Pyc MA. The interim test effect: Testing prior material can facilitate the learning of new material. *Psychonomic Bulletin & Review*. 2011; 18:1140–1147. [PubMed: 21796496]
- Wolosin SM, Zeithamova D, Preston AR. Reward modulation of hippocampal subfield activation during successful associative encoding and retrieval. *Journal of Cognitive Neuroscience*. 2012; 24:1532–1547. [PubMed: 22524296]
- Yonelinas AP. The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*. 2002; 46:441–517.
- Yonelinas AP, Jacoby LL. The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language*. 1995; 34:622–643.
- Yue CL, Soderstrom NC, Bjork EL. Partial testing can potentiate learning of tested and untested material from multimedia lessons. *Journal of Educational Psychology*. 2015; 107:991–1005.

Appendix. Remember/Know instructions from Experiments 1 and 2

“You should make a Remember judgment if you can consciously recollect what you experienced when you studied the word earlier. This may include aspects of the physical appearance of the item, of something that happened in the room, or of what you were thinking or doing at the time. You should make a Know judgment if you recognize the item as being one that you studied, but you cannot consciously recollect what you experienced while studying it. In other words, choose “Know” when you are fairly certain that you recognize the item, but it fails to evoke any specific conscious recollection of your experience learning that word.

Consider the following examples. If I asked you to remember eating breakfast this morning, you'd likely be able to recollect where you were, what you ate, and what you were thinking about. You would thus give a "Remember" response. However, in another situation, you may see someone on campus and know that you've met that person before, but you have no idea where and can't remember anything else about him or her. In this situation, you would give a "Know" response."

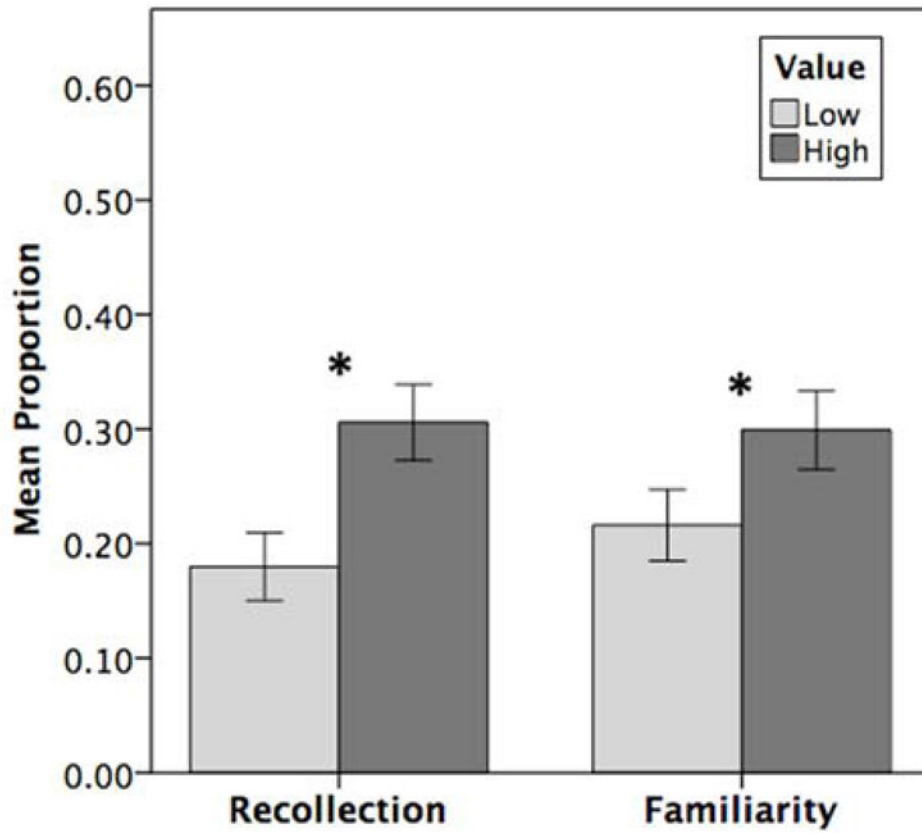


Figure 1.

Effects of value on recollection and familiarity estimates in Experiment 1. Error bars indicate ± 1 SE.

In all figures, * indicates $p < .05$, ~ indicates $.05 < p < .10$, for t-tests comparing high-value and low-value items.

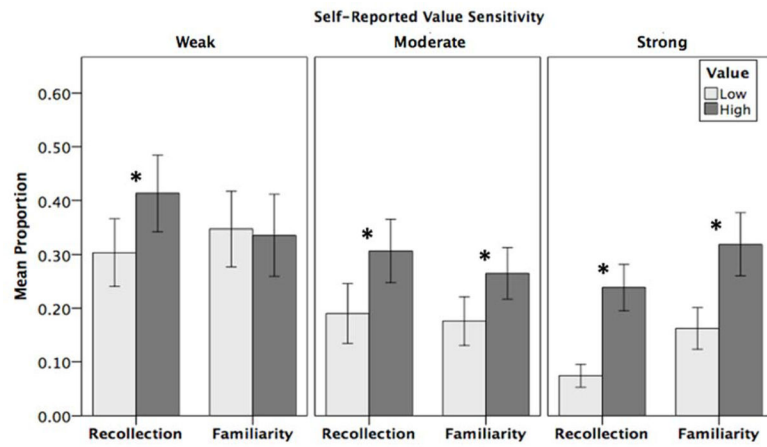


Figure 2. Effects of value on recollection and familiarity estimates, split based on self-reported value sensitivity, in Experiment 1. Error bars represent ± 1 SE.

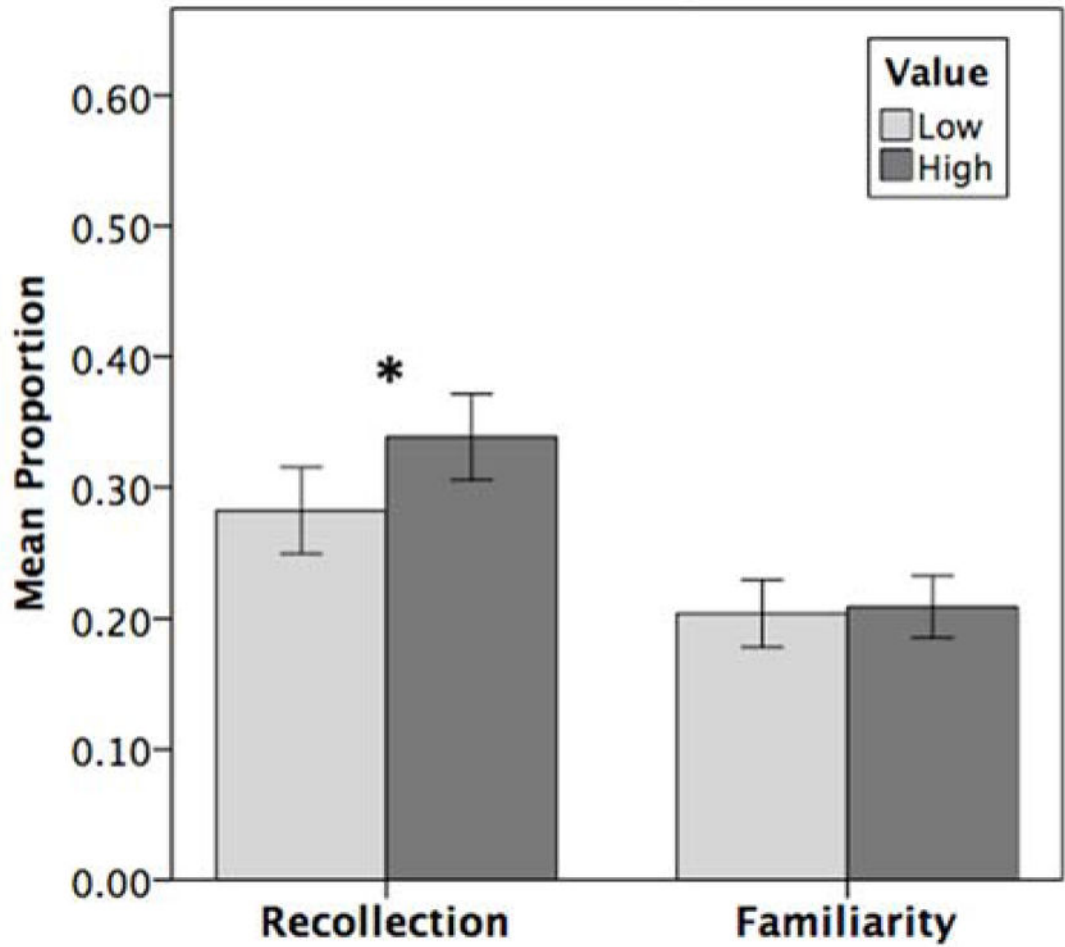


Figure 3. Effects of value on recollection and familiarity estimates in Experiment 2. Error bars represent ± 1 SE.

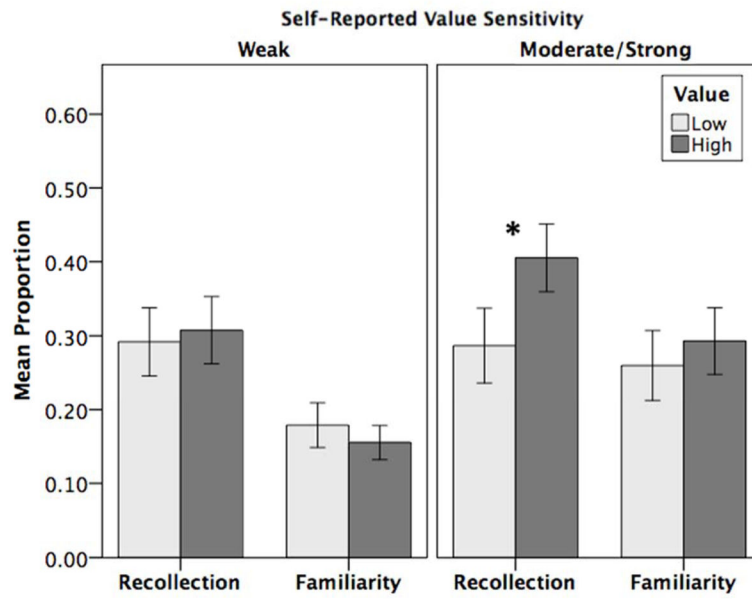


Figure 4. Effects of value on recollection and familiarity estimates in Experiment 2, split by self-reported value sensitivity. Error bars represent ± 1 SE.

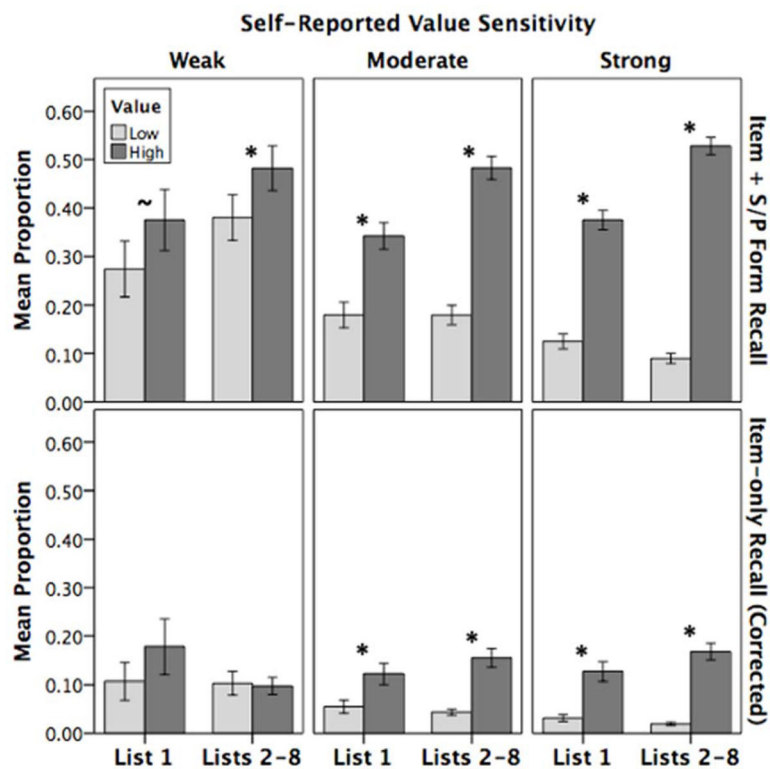


Figure 5. Effects of value on the proportion of items correctly recalled on the free recall test, split by list group, type of recall, and self-reported value sensitivity. For type of recall, “Item + S/P Form Recall” refers to the proportion of all items recalled in the correct singular/plural form. “Item-only Recall (Corrected)” refers to the proportion of items recalled in the incorrect singular/plural form, conditional on not being recalled in the correct singular/plural form. Error bars represent +/- 1 SE.

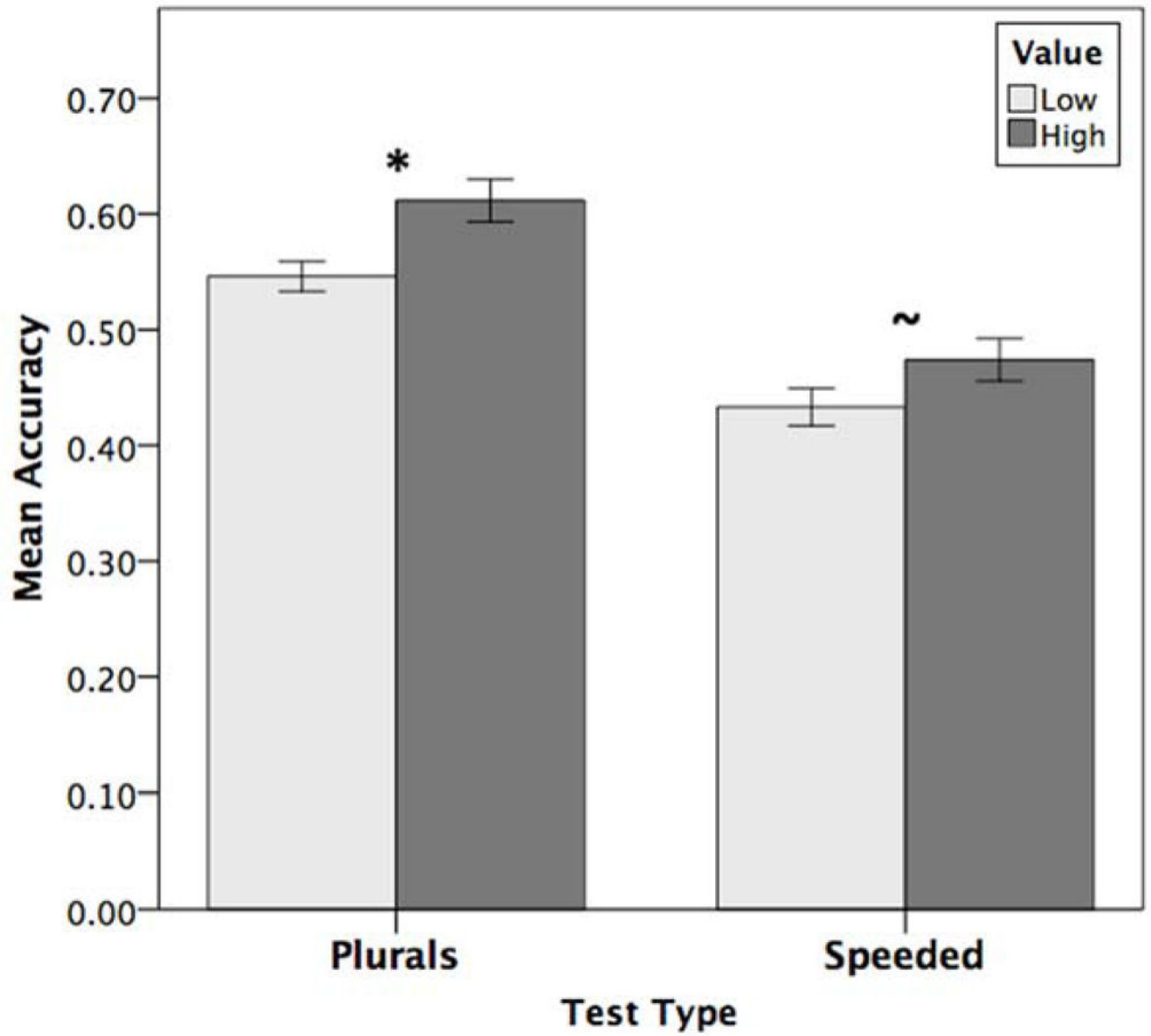


Figure 6. Effects of value on accuracy during the plurals and speeded recognition tests in Experiment 3. Error bars represent ± 1 SE.

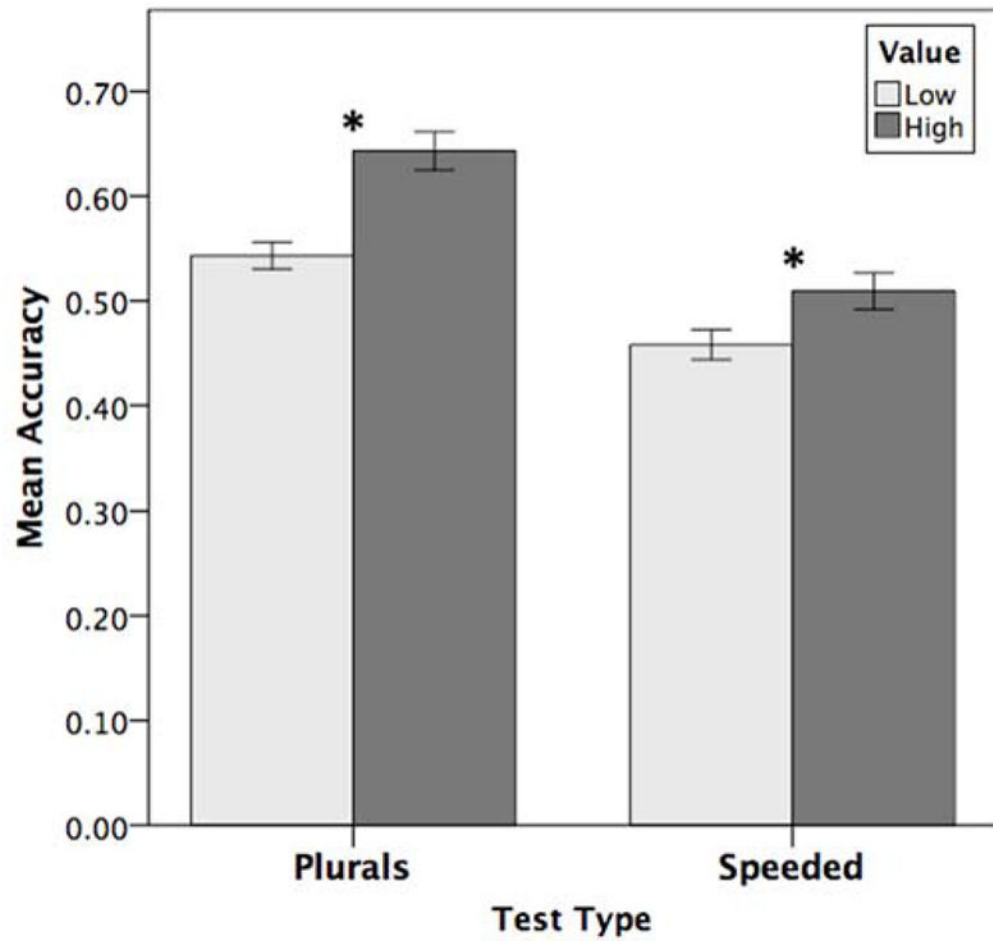


Figure 7. Effects of value on accuracy during the plurals and speeded recognition tests in Experiment 4. Error bars represent +/- 1 SE.

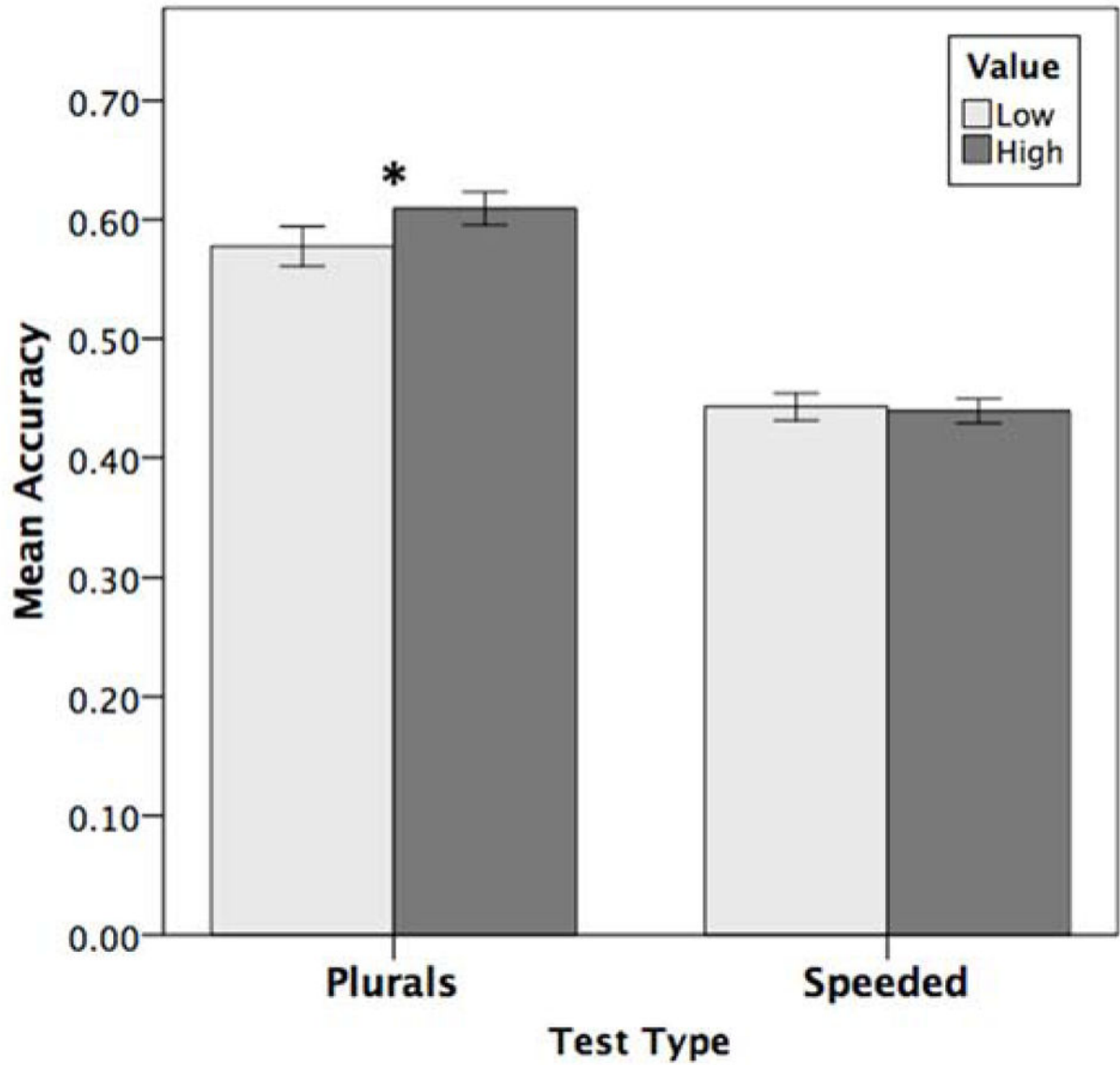


Figure 8. Effects of value on accuracy during the plurals and speeded recognition tests in Experiment 5. Error bars represent ± 1 SE.

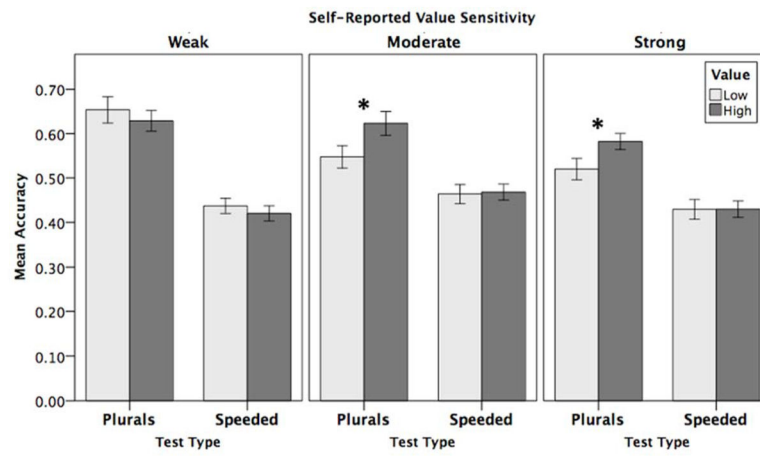


Figure 9. Effects of value on accuracy during the plurals and speeded recognition tests in Experiment 5, split by self-reported value sensitivity. Error bars represent ± 1 SE.

Table 1

Mean proportion correct (and standard error) on recall test, split by list and item value, in Experiment 1.

	1	2	3	4	5	6	7
High Value	.419 (.027)	.465 (.027)	.529 (.032)	.556 (.032)	.556 (.031)	.521 (.037)	.508 (.030)
Low Value	.287 (.028)	.262 (.029)	.267 (.032)	.295 (.031)	.281 (.035)	.266 (.035)	.252 (.031)

Table 2

Mean proportion correct (and standard error) on recall test, split by list, item value, and recall type, in Experiment 3.

	List								
	1	2	3	4	5	6	7	8	
Item + S/P Form Recall	High Value	.361 (.016)	.435 (.016)	.484 (.018)	.488 (.018)	.525 (.019)	.557 (.018)	.522 (.019)	.526 (.019)
	Low Value	.164 (.015)	.173 (.016)	.161 (.015)	.150 (.015)	.151 (.016)	.152 (.017)	.154 (.017)	.152 (.015)
Item-only Recall (Corrected)	High Value	.128 (.014)	.152 (.017)	.165 (.018)	.176 (.021)	.114 (.016)	.195 (.022)	.146 (.016)	.165 (.023)
	Low Value	.050 (.008)	.045 (.009)	.035 (.007)	.034 (.006)	.038 (.009)	.042 (.007)	.029 (.006)	.039 (.010)

Table 3

Mean proportion correct (and standard error) on plurals recognition test (recollection) and speeded item recognition test (familiarity), split by self-reported value sensitivity, in Experiment 3.

Value Sensitivity	Weak (n = 11)		Moderate (n = 21)		Strong (n = 32)	
	High	Low	High	Low	High	Low
Plurals Recognition	.639 (.048)	.586 (.036)	.683 (.029)	.556 (.023)	.563 (.024)	.528 (.017)
Speeded Recognition	.469 (.054)	.378 (.055)	.477 (.036)	.448 (.027)	.474 (.023)	.443 (.019)

Table 4

Mean proportion correct (and standard error) on recall test, split by test (list group) and item value, in Experiment 4.

	Test	1	2	3
Item + S/P Form Recall	High Value	.340 (.026)	.443 (.029)	.476 (.033)
	Low Value	.155 (.017)	.115 (.018)	.078 (.018)
Item-only Recall (Corrected)	High Value	.135 (.017)	.124 (.021)	.152 (.030)
	Low Value	.051 (.011)	.030 (.008)	.012 (.005)

Note: Test 1 occurs after list 1, Test 2 occurs after one of lists 2, 3, or 4, and test 3 occurs after one of lists 5, 6, 7, or 8.