Title

Maximizing Memory: Improving Learning and Memory for Important Information

Author

Murphy, Dillon

Publication Date

2023

UNIVERSITY OF CALIFORNIA

Los Angeles

Maximizing Memory:
Improving Learning and Memory for Important Information

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Psychology

by

Dillon H. Murphy

2023

ABSTRACT OF THE DISSERTATION

Maximizing Memory:
Improving Learning and Memory for Important Information

by

Dillon H. Murphy
Doctor of Philosophy in Psychology
University of California, Los Angeles, 2023
Professor Alan Dan Castel, Chair

The present dissertation investigated how goals and information importance influence remembering and forgetting processes. Results revealed that individuals strategically forget less important information to optimize memory for valuable information. Moreover, the use of external memory aids (i.e., offloading) is influenced by the objective and subjective value of the information. When information differs in objective value, people are most likely to offload high-value items but if the external store is unreliable, people often forget this valuable information. In contrast, if the information differs in subjective value, people use memory for important information. I also observed that metacognition and other cognitive abilities impact optimal offloading.

Learners may strategically forget less important information to maximize memory for valuable information, which suggests a deliberate decision-making process regarding what information to offload or forget. Similarly, making decisions about what information to highlight may also involve a form of offloading, where learners decide to externalize or emphasize certain information by highlighting it for later reference, while potentially offloading or neglecting other information by not highlighting it. Moreover, metacognition, which involves monitoring and controlling one's own cognitive processes, can be reactive, meaning that decisions about what information to remember and forget can influence what is ultimately remembered. However, I did not find evidence that requiring learners to choose what information to highlight from a passage of to-be-tested content improves comprehension compared to traditional highlighting methods.

Lastly, I explored the relationship between value-directed remembering and desirable difficulties, which are cognitive strategies that can enhance learning and retention. However, the spacing effect, generation effect, and testing effect, which are known to benefit memory overall, did not interact with the value of the information being learned.

Overall, this dissertation contributes to our understanding of how goals, information importance, offloading, metacognition, and desirable difficulties influence memory processes. The findings highlight the complex interplay between objective and subjective value, external memory aids, metacognitive decision-making, and cognitive strategies in shaping memory performance. These findings have implications for educational settings, where understanding how learners prioritize and remember information of different values can inform instructional strategies and enhance learning outcomes.

The dissertation of Dillon H. Murphy is approved.

Barbara Knowlton

Elizabeth L. Bjork

Robert A. Bjork

Alan D. Castel, Committee Chair

University of California, Los Angeles

2023

Table of Contents

LIST of FIGURES

Chapter 2

Chapter 3

Chapter 4

LIST of TABLES

Chapter 2

VITA

EDUCATION

2022 C. Phil., Cognitive Psychology
University of California, Los Angeles

2020 M. A., Cognitive Psychology
University of California, Los Angeles

2019 B. S., Psychology
University of Oregon

PEER-REVIEWED PUBLICATIONS

1. **Murphy, D. H.**, Schwartz, S. T., & Castel, A. D. (in press). Value-directed retrieval: The effects of divided attention at encoding and retrieval on memory selectivity and retrieval dynamics. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*
2. **Murphy, D. H.**, Hoover, K. M., & Castel, A. D. (in press). The effect of video speed on mind wandering and comprehension in younger and older adults. *Memory.*
3. **Murphy, D. H.**, & Castel, A. D. (in press). Age-related differences in metacognition and memory when offloading important information. *Psychology and Aging.*
4. **Murphy, D. H.** (2023). Strategic offloading: How the value of to-be-remembered information influences offloading decision-making. *Applied Cognitive Psychology.*
5. **Murphy, D. H.**, & Castel, A. D. (in press). Age-related differences in overcoming interference when selectively remembering important information. *Experimental Aging Research.*
6. **Murphy, D. H.**, Hargis, M. B., & Castel, A. D. (2023). Younger and older adults' strategic use of associative memory and metacognitive control when learning foreign vocabulary words of varying importance. *Psychology and Aging.*
7. **Murphy, D. H.**, Halamish, V., Rhodes, M. G., & Castel, A. D. (2023). How evaluating memorability can lead to unintended consequences. *Metacognition and Learning.*
8. **Murphy, D. H.**, & Castel, A. D. (2022). Age-related differences in metacognitive predictions of semantic fluency and originality. *Aging, Neuropsychology and Cognition.*
9. **Murphy, D. H.**, & Castel, A. D. (2022). Differential effects of proactive and retroactive interference in value-directed remembering for younger and older adults. *Psychology and Aging.*
10. **Murphy, D. H.**, & Castel, A. D. (2022). Responsible attention: The effect of divided attention on metacognition and responsible remembering. *Psychological Research*.
11. **Murphy, D. H.**, & Castel, A. D. (2022). Selective remembering and directed forgetting are influenced by similar stimulus properties. *Memory, 30*, 1130-1147.
12. **Murphy, D. H.**, Bjork, R. A., & Bjork, E. L. (2022). Going beyond the spacing effect: Does it matter how time on a task is distributed? *Quarterly Journal of Experimental Psychology.*
13. **Murphy, D. H.**, Schwartz, S. T., & Castel, A. D. (2022). Serial and strategic memory processes in goal-directed selective remembering. *Cognition, 225*, 105178.
14. **Murphy, D. H.**, & Knowlton, B. J. (2022). Framing effects in value-directed remembering.

Memory & Cognition, 50, 1350-1361.

15. **Murphy, D. H.**, Hoover, K. M., & Castel, A. D. (2022). Strategic metacognition: Self-Paced study time and responsible remembering. *Memory & Cognition, 51*, 234-251.
16. **Murphy, D. H.**, & Castel, A. D. (2022). Responsible remembering and forgetting in younger and older adults. *Experimental Aging Research, 48*, 455-473.
17. **Murphy, D. H.**, Hoover, K. M., Agadzhanyan, K., Kuehn, J. C., & Castel, A. D. (2022). Learning in double time: The effect of lecture video speed on immediate and delayed comprehension. *Applied Cognitive Psychology, 36*, 69-82.
18. **Murphy, D. H.**, Silaj, K. M., Schwartz, S. T., Rhodes, M. G., & Castel, A. D. (2022). An own-race bias in the categorization and recall of associative information. *Memory, 30*, 190-205.
19. **Murphy, D. H.**, Huckins, S. C., Rhodes, M. G., & Castel, A. D. (2022). The effect of perceptual processing fluency and value on metacognition and remembering. *Psychonomic Bulletin & Review, 29*, 910-921.
20. **Murphy, D. H.**, & Castel, A. D. (2022). The role of attention and aging in the retrieval dynamics of value-directed remembering. *Quarterly Journal of Experimental Psychology, 75*, 954-968.
21. **Murphy, D. H.**, Friedman, M. C., & Castel, A. D. (2022). Metacognitive control, serial position effects, and effective transfer to self-paced study. *Memory & Cognition, 50*, 144-159.
22. **Murphy, D. H.**, & Castel, A. D. (2021). The dynamics of memory for United States presidents in younger and older adults. *Memory, 29*, 1232-1244.
23. **Murphy, D. H.**, & Castel, A. D. (2021). Tall towers: Schemas and illusions when perceiving and remembering a familiar building. *Applied Cognitive Psychology, 35*, 1236-1246.
24. **Murphy, D. H.**, Agadzhanyan, K., Whatley, M. C., & Castel, A. D. (2021). Metacognition and fluid intelligence in value-directed remembering. *Metacognition and Learning, 16*, 685-709.
25. **Murphy, D. H.**, & Castel, A. D. (2021). Metamemory that matters: Judgments of importance can engage responsible remembering. *Memory, 29*, 271-283.
26. **Murphy, D. H.**, & Castel, A. D. (2021). Responsible remembering and forgetting as contributors to memory for important information. *Memory & Cognition, 49*, 895-911.
27. **Murphy, D. H.**, & Castel, A. D. (2021). Age-related similarities and differences in the components of semantic fluency: Analyzing the originality and organization of retrieval from long-term memory. *Aging, Neuropsychology and Cognition, 28*, 748-761.
28. **Murphy, D. H.**, & Castel, A. D. (2020). Responsible remembering: How metacognition impacts adaptive selective memory. *Zeitschrift für Psychologie, 228*, 301-303.

CHAPTERS IN EDITED BOOKS

1. Whatley, M. C., Murphy, D. H., Silaj, K. M., & Castel, A. D. (2021). Motivated memory for what matters most: How older adults (selectively) focus on important information and events using schematic support, metacognition, and meaningful goals. In G. Sedek, T. M. Hess, & D. R. Touron (Eds.), *Multiple pathways of cognitive aging: Motivational and contextual influences*. Oxford University Press.

CHAPTER 1:

INTRODUCTION

Whether remembering items on a shopping list, children's allergies, or items to pack for a vacation, we are often exposed to more information than can be remembered. When attempting to retain large amounts of information, people should strategically focus on important information to maximize the likelihood that this information will be effectively encoded and later remembered. For example, if people fail to remember important information, the consequences of forgetting could have disastrous repercussions, such as giving a child food containing a known allergen or forgetting to pack your passport for a vacation getaway. Examining people's understanding of how their memory works (i.e., metamemory), what information they try to remember and forget, and adaptive mechanisms that contribute to memory for important information can help broaden our understanding of how and why people achieve memory for important information and avoid consequences for forgetting.

To examine memory for valuable information in the lab, Castel et al. (2002) presented participants with words paired with point values that count toward participants' scores if recalled. With their goal being to maximize their point scores, participants tend to optimize task performance by best recalling the most valuable information, often at the expense of low-value items (Ariel et al., 2009; Castel, 2008; Castel et al., 2007, 2013; Soderstrom & McCabe, 2011; see Knowlton & Castel, 2022; Madan, 2017 for review). Thus, people can use value to guide memory, but the evaluation and monitoring of one's memory processes may also play a role.

Metacognition refers to an awareness and understanding of one's cognitive processes (i.e., Dunlosky et al., 2016; Nelson, 1996), and in Nelson and Narens' (1990) framework of metamemory, they differentiate two major processes: metacognitive monitoring and control.

Measures of monitoring typically involve evaluating the likelihood of remembering something whereas metacognitive control often involves study decisions. For example, when a learner evaluates their learning, they are engaging in metacognitive monitoring (see Rhodes, 2016 for a review). When presented with information of varying value, prior work suggests that learners are generally metacognitively aware of the need to be selective. For example, Murphy et al. (2021a) found that participants rated high-value words as more likely to be remembered and low-value words as less likely to be remembered, demonstrating a metacognitive awareness of value-directed remembering. Furthermore, this sensitivity of participants' judgments of learning (JOLs) to item value was associated with greater selective memory for high-value words and increased as the task endured. Additionally, participants' metacognitive awareness of selectivity (i.e., the positive relationship between JOLs and item value) and metacognitive accuracy also increased with task experience, indicating that participants become more aware of their limited memory capacity and are more strategically selective with what they remember with increased task experience.

Accurately predicting recall exemplifies good metacognition, but people often have difficulty anticipating future forgetting (Koriat et al., 2004; Kornell et al., 2011). For example, most people have had experiences where they expected to remember important information, like names or anniversaries, but forgot it at an inopportune time and had to deal with the consequences like social embarrassment or an angry spouse. In some extreme cases, the consequences of unexpected forgetting can be deadly, such as distracted parents forgetting infants in the back seats of hot parked cars (see Anselmi et al., 2020)—something that seems unimaginable to those who have not experienced it.

Minor instances of forgetting can be inconvenient and failing to consider the potential (although sometimes minimal) consequences of forgetting can contribute to instances of inaccurate

JOLs (e.g., Serra & England, 2012). However, since the most important information is usually that which has the most severe outcomes if forgotten, situations involving negative consequences for forgetting often lead to improved metacognition and learning outcomes (e.g., McGillivray & Castel, 2011). How our memory adaptively functions to prioritize important information that will need to be remembered as well as how metacognitive processes may be more precise in situations involving consequences for forgetting, is a notion termed *responsible remembering* (Murphy & Castel, 2020).

Responsible remembering, a form of adaptive memory (see Nairne, 2010, 2013, 2015; Nairne & Pandeirada, 2008, 2010; Nairne et al., 2007), captures one's knowledge of selective memory processes and allows for the efficient use of memory in a variety of contexts. To demonstrate a situation where participants engage in responsible remembering, Murphy and Castel (2021a) presented participants with lists of children and their associated food preferences to remember for a later test (foods the children like, dislike, or are allergic to and must avoid). When participants were forced to consider the importance of remembering each child's food preferences (rather than passively monitoring their learning), information with consequences for forgetting (the children's allergies) was deemed most important and subsequently best remembered.

In a follow-up study, Murphy et al. (2023) used a similar paradigm but allowed participants to self-regulate their study of the children's food preferences. Results revealed that when participants were asked to consider the importance of remembering each child's food preferences, they subsequently spent more time studying the children's allergies, revealing a potential mechanism of their superior memory for the information with consequences if forgotten. Specifically, participants' reflection on the importance of remembering information resulted in a strategic allocation of study time towards this information to minimize the potential consequences of forgetting. Thus, if people learn to self-assess and prioritize information that will need to be remembered or have negative consequences if forgotten, the recall of said important information can be enhanced via strategic metacognitive mechanisms (cf. Koriat et al., 2006).

As demonstrated by participants' superior memory for important information or information with consequences if forgotten in various situations, when presented with more information than can be remembered, participants engage in responsible remembering. In addition to the strategic allocation of study time when evaluating the importance of remembering, other cognitive mechanisms may also contribute to memory for important information. For example, strategic forgetting or offloading may enhance a learner's ability to remember valuable information. Additionally, evaluating memorability and making decisions about to-be-learned information may critically influence how we remember. Moreover, techniques for increasing overall memory performance may also enhance one's ability to selectively remember important information and students may be able to harness the benefits of each of these potentially memory-enhancing techniques in the classroom.

CHAPTER 2:

STRATEGIC FORGETTING AND OFFLOADING

Portions of the following introductory comments, description of Experiments 1 and 2, and conclusion are taken directly from Murphy and Castel (2021b) and Murphy (2023)

When presented with to-be-remembered information, people's habitual response is often to attempt to remember as much information as possible. However, in some cases, it can be beneficial to forget less important information to facilitate memory for critical information and engage in responsible remembering. Cognitive control refers to the ability to engage in functional, goal-directed behavior that allows one to overcome previously learned habitual behaviors to accommodate competing task demands (Chiew & Braver, 2017; Diamond, 2013; Egner, 2017; Miller & Cohen, 2001). Applied to responsible remembering, cognitive control may be a critical component for remembering important information.

Unlike many memory tasks, directed forgetting tasks present participants with to-be-remembered as well as to-be-forgotten words (item method directed forgetting; see Bjork & Bjork, 1996 for list method). Largely pioneered by Bjork et al. (1968), directed forgetting tasks present participants with items one at a time and after each item, a cue indicates whether participants should remember or forget the item (see also Woodward & Bjork, 1971). Similarly, some directed forgetting tasks present words paired with either positive or negative point values that count towards participants' scores on the task if they later recall the word (see Castel et al., 2007). Because recalling the words with negative point values would reduce their score, participants should not be motivated to remember words associated with negative values and only remember the words resulting in gains if recalled.

Although often deemed an undesirable memory failure, forgetting can lead to memory benefits such that compared with baseline memory performance, recall for information not expected to be tested (or paired with negative values) tends to be poor (the costs of forgetting) while recall for information expected to be tested (or positively valued) tends to be enhanced, exemplifying the benefits of forgetting (e.g., Bjork & Bjork, 1996; Friedman & Castel, 2011; for reviews, see Basden & Basden, 1998; Bjork, 1998; MacLeod, 1998). Thus, responsibly forgetting information, perhaps unimportant or outdated information, may enhance the recall of target information and lead to responsible remembering.

Directed forgetting tasks can serve as an indicator of inhibitory control and exemplify the *retrieval inhibition theory* (Basden & Basden, 1998; Basden et al., 1993; Bjork, 1989; Geiselman & Bagheri, 1985; Geiselman et al., 1983; Weiner & Reed, 1969; see also Verde, 2012; see Racsmány & Conway, 2006 for episodic inhibition account; see Sahakyan & Kelley, 2002 for the context-change account), the theory that the inhibition of to-be-forgotten items facilitates retrieval of to-be-remembered items. Alternatively, researchers have also argued for selective rehearsal accounts of directed forgetting (Bjork, 1972; MacLeod, 1975; Sheard & MacLeod, 2005; Tan et al., 2020) such that participants strategically rehearse and encode to-be-remembered items rather than to-be-forgotten items. Specifically, selective rehearsal accounts posit that presented items are maintained in working memory until participants are cued to remember or forget the item, and to-be-remembered items are elaboratively rehearsed (to transfer to long-term memory) while rehearsal for to-be-forgotten items is stopped. Thus, functional forgetting due to inhibition or selective rehearsal may be used to strategically enhance memory.

Regardless of the mechanism behind item-method directed forgetting, forgetting may be a critical aspect of a functional memory system such that forgetting items that do not need to be remembered may facilitate memory for items that do need to be remembered. Exemplifying this effect, previous work has demonstrated that offloading to-be-remembered information (e.g., saving information to a computer, writing things down) facilitates memory for other to-be-remembered information by reducing the extent to which the offloaded information interferes with target information (Risko & Dunn, 2015; Risko & Gilbert, 2016; Sparrow et al., 2011; Storm & Stone, 2015). Thus, when remembering a list of items, selectively rehearsing goal-relevant items or inhibiting less important information may facilitate the retrieval of valuable items or items with negative consequences if forgotten.

When presented with information to remember (or forget), general representations or heuristics influence how we perceive and remember the world and impact how we value information (see McGillivray & Castel, 2017). This use of prior knowledge can enhance memory (a form of "schematic support," as described by Craik & Bosman, 1992) such that knowledge in a domain can facilitate memory for other information in that domain. Thus, schematic support may be beneficial for remembering important information; however, little research has investigated the role of schematic support and item importance in directed forgetting tasks.

Experiment 1

Responsible remembering and forgetting

Previous work has modified memory through explicit instruction to forget (directed forgetting tasks). However, in Experiment 1, a cue indicated whether the participant ("You") or a hypothetical friend ("Friend") was responsible for remembering each item (as opposed to a transactive memory system, see Hollingshead, 2001; Wegner, 1986 or collaborative memory, see Rajaram & Pereira-Pasarin, 2010) on a list of to-be-remembered words (unassociated words or items to pack for a camping trip). Specifically, I presented participants with a number of items (20) greater than the typical memory span of an individual (Cowan, 2001; Unsworth & Engle, 2007) to determine how participants allocate cognitive recourses to maximize memory utility. I expected participants to engage in responsible forgetting by selectively rehearsing and remembering goal-relevant information and forgetting goal-irrelevant information from the list. Thus, the current paradigm could result in responsible forgetting such that participants prioritize their cognitive resources for items on the list that they are responsible for remembering.

I was also interested in whether responsible forgetting occurs in situations offering schematic support. Specifically, while I expected participants to show enhanced recall for the items they were responsible for remembering, if participants use schematic support to enhance their memory, then they may be able to remember goal-relevant, important items to pack for a camping trip (see McGillivray & Castel, 2017) even if their friend was responsible for remembering them. Thus, I expected responsible remembering to be enhanced in conditions offering schematic support such that items of greatest importance, or biggest consequences if forgotten, are best recalled.

To further investigate memory for items that participants were responsible for remembering compared with items that their friend was responsible for remembering, I had participants complete a surprise cue pairing test where they were presented with all words from the study phase and asked to indicate whether they or their friend were responsible for remembering each item. I expected participants to demonstrate enhanced accuracy for items they were responsible for remembering.

Method

Participants. Participants were 60 undergraduate students (M_{age} = 20.52, SD_{age} = 1.64) recruited from the University of California Los Angeles (UCLA) Human Subjects Pool and received course credit for their participation. A sensitivity analysis using G*Power (Faul et al.,

2007) indicated that for a repeated-measures, between-subjects ANOVA with 2 groups (stimulus type: schematic support, unassociated) and 2 measurements (cue: Friend, You), with a low correlation between repeated measures (recall for Friend and You items, $r = -.16$), assuming alpha = .05, power = .80, the smallest effect size the design could reliably detect is $\eta^2 = .05$. Participants were tested individually or in groups of up to 8 individuals in a laboratory session lasting approximately 1 hour.

Materials and Procedure. Participants were informed that they would be presented with a list of words that they and a (hypothetical) friend needed to remember and that after each word was presented, a cue would indicate whether they (You) or their friend (Friend) was responsible for remembering the word. Participants were randomly assigned to either be presented with unassociated words ($n = 30$) or asked to imagine that they and a friend were going camping (items offering schematic support; $n = 30$). If presented with items to remember for a camping trip, participants were told that they would be presented with a list of items that they and their friend needed to remember to bring on the trip.

For each participant, half of the words were randomly designated as to-be-remembered words for participants and half were designated as words their friend was responsible for remembering. Each word was preceded by a 1-second fixation cross, then appeared on the screen, one at a time, in random order, for 3 seconds followed by the cue for an additional 2 seconds. After the presentation of all 20 words, participants were given a 1-minute free recall test in which they were asked to recall all the words that both they and their friend needed to remember from the just-presented list.

Following the recall test, participants completed a surprise cue pairing test where they were shown the words from the just-presented list (in random order) and asked to indicate whether they or their friend were responsible for remembering the word. Participants also provided confidence judgments on a scale from 0 to 100 (with 0 being not at all confident and 100 being very confident) and were given as much time as they needed for this portion of the task.

Lastly, to evaluate the general importance of each item on the list of items to bring on a camping trip, I recruited a separate sample of undergraduate students ($n = 60$) from the UCLA Human Subjects Pool. These participants were shown all 20 items and then asked to rate each item on a scale from 0 (not important) to 100 (very important).

Results

The results are divided into five primary sections: recall, output order, importance, cue pairing performance, and confidence on the cue pairing test. In each section, I investigated differences as a function of cue (Friend, You) and stimulus type (words offering schematic support, unassociated words). To reinforce each effect, I computed Bayes Factor (a ratio of the marginal likelihood of the null model and a model suggesting group differences). The data will be reported as either in favor of the null hypothesis (which would be supported by a large BF_{01}) or the alternative hypothesis (which would be supported by a large BF_{10}; see Kass & Raftery, 1995 for more information).

Recall. A 2 (cue: Friend, You) x 2 (stimulus type: schematic support, unassociated) repeated-measures ANOVA on recall performance revealed a main effect of cue [$F(1, 58) = 43.46$, $p < .001$, $\eta2 = .41$, $BF_{10} > 100$] such that participants recalled more You items ($M = .56$, $SD = .21$) than Friend items ($M = .30$, $SD = .20$). Additionally, results revealed a main effect of stimulus type [$F(1, 58) = 11.81$, $p = .001$, $\eta2 = .17$, $BF_{10} > 100$] such that participants presented with items offering schematic support ($M = .48$, $SD = .12$) recalled a greater proportion of items than participants presented with unassociated words (*M* = .37, *SD* = .13). Additionally, cue interacted with condition [*F*(1, 58) = 4.83, *p* = .032, η2 = .05, BF_{10} > 100].

To investigate differences in recall between You items for each stimulus type, an independent samples *t*-test was conducted but Levene's test of equality of variances indicated a violation of the equal variance assumption (*p* = .026). Welch's *t*-test did not reveal differences in recall of You items between the schematic support (*M* = .57, *SD* = .17) and unassociated word conditions (*M* = .54, *SD* = .25), [*t*(51.72) = .42, *p* = .675, *d* = .11, BF_{01} = 3.54]; however, participants recalling words offering schematic support recalled more Friend items (*M* = .39, *SD* = .19) than those recalling unassociated words (*M* = .20, *SD* = .17), [*t*(58) = 4.29, *p* < .001, *d* = 1.11, BF_{10} > 100]. Thus, participants demonstrated similar recall of You items across conditions but elevated recall of Friend items in the schematic support condition.

Output order. To examine how participants organized retrieval, a Gamma correlation between the output position of each correct item (with larger numbers indicating later output) and the corresponding cue (You coded as 1, Friend, coded as 0) was computed across participants. A strong positive correlation would indicate that participants recalled Friend items before You items while a negative correlation would indicate recalling You items before Friend items. A correlation near 0 would indicate no organization of recall according to cue. Results revealed that, overall, participants recalled You items before Friend items (γ = -.25, *p* < .001). Additionally, I computed Gamma correlations at the participant level in each condition (schematic support: *M* = -.15, *SD* = .50; unassociated words: *M* = -.38, *SD* = .66) but an independent samples *t*-test did not reveal group differences in the organization of recall [*t*(50) = 1.46, *p* = .151, *d* = .41, BF_{01} = 1.51].

Importance. In the schematic support condition, to determine if participants prioritized recall for important items, a Gamma correlation between recall accuracy and item importance was computed across participants. A strong positive correlation would indicate that participants prioritized recall for important items while a negative correlation would indicate that participants prioritized recall of unimportant items. A correlation near 0 would indicate no sensitivity to importance. Results revealed that, overall, participants' recall was sensitive to importance ($\gamma = .14$, $p = .006$) indicating that participants recalled important items better than less important items. I also computed Gamma correlations between recall accuracy and item importance at the participant level for each cue (You and Friend items) and these correlations (You: $M = .37$, $SD = .44$; Friend: $M = -.01$, $SD = .50$) served as the dependent variable in a paired samples t-test. Results revealed that participants were more sensitive to importance (see Figure 1) for the items they were responsible for remembering compared with items their friend was responsible for remembering [$t(26) = 3.74$, $p < .001$, $d = .72$, $BF_{10} = 37.03$].

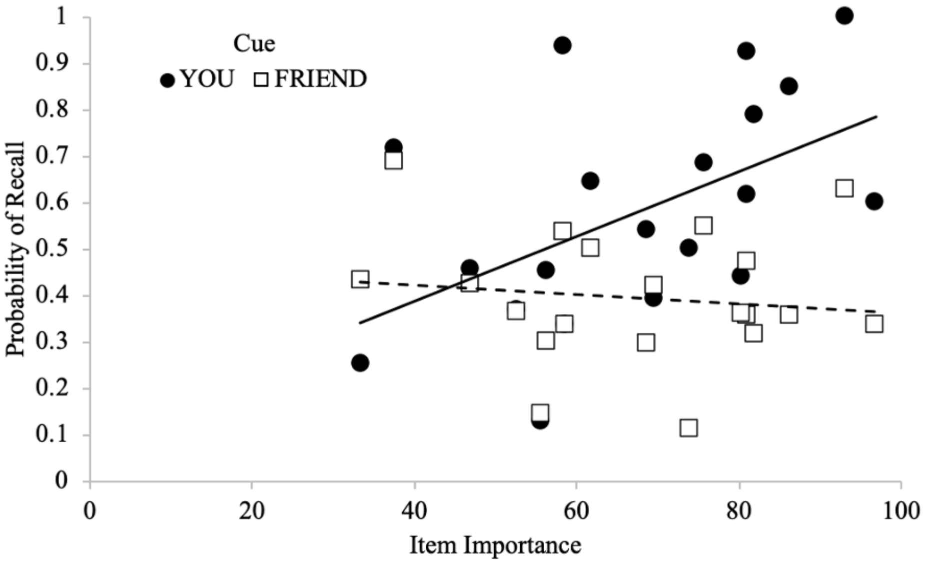


Figure 1. Probability of recall as a function of cue and item importance with regression lines in Experiment 1.

Cue pairing test. To examine differences in performance (scored as proportion correct) on the cue pairing test, a 2 (cue: Friend, You) x 2 (stimulus type: schematic support, unassociated) repeated-measures ANOVA did not reveal a main effect of stimulus type [*F*(1, 58) = 3.01, *p* = .088, η2 = .05, BF$_{10}$ = 1.02] such that participants presented with words offering schematic support (*M* = .71, *SD* = .18) paired a similar proportion of words with the correct cue as participants presented with unassociated words (*M* = .78, *SD* = .17). Additionally, results did not reveal a main effect of cue [*F*(1, 58) = 1.57, *p* = .216, η2 = .03, BF$_{01}$ = 2.54] such that participants correctly paired a similar proportion of You items (*M* = .76, *SD* = .19) as Friend items (*M* = .73, *SD* = .21). Moreover, cue did not interact with condition [*F*(1, 58) = .11, *p* = .743, η2 < .01, BF$_{01}$ = 9.69]. Thus, participants' associative memory for items was not sensitive to cue suggesting that participants had generally accurate associative memory for who was responsible for remembering each item.

To investigate whether cue pairing accuracy was sensitive to item importance, a Gamma correlation was computed across participants. Results revealed that, overall, participants' cue pairing accuracy was not sensitive to importance (γ = -.04, *p* = .505). Additionally, I computed Gamma correlations at the participant level for each cue and these correlations (You: *M* = .08, *SD* = .40; Friend: *M* = .00, *SD* = .54) served as the dependent variable in a paired samples *t*-test. Results revealed that participants were similarly insensitive to importance for the items they were responsible for remembering compared with items their friend was responsible for remembering [*t*(21) = .82, *p* = .420, *d* = .18, BF$_{01}$ = 3.31].

Confidence. To determine if participants' confidence on the cue pairing test differed as a function of cue and stimulus type, a 2 (cue: Friend, You) x 2 (stimulus type: schematic support, unassociated) repeated-measures ANOVA did not reveal a main effect of stimulus type [*F*(1, 58) < .01, *p* = .992, η2 < .01, BF$_{01}$ = 2.64] such that participants presented with words offering schematic support (*M* = 75.85, *SD* = 17.56) were similarly confident as participants presented with unassociated words (*M* = 75.81, *SD* = 16.42). However, results revealed a main effect of cue [*F*(1, 58) = 15.96, $p < .001$, $\eta 2 = .22$, $BF_{10} > 100$] such that participants were more confident in You items (*M* = 79.26, *SD* = 19.10) than Friend items (*M* = 72.40, *SD* = 17.03) but cue did not interact with condition [*F*(1, 58) = .03, $p = .856$, $\eta 2 < .01$, $BF_{10} = 13.68$].

Discussion

In Experiment 1, I examined how participants prioritized memory for items they and a friend were responsible for remembering. Additionally, I investigated how the prioritization of memory was influenced by schematic support and item importance. Similar to previous work on directed forgetting (e.g., Bjork & Bjork, 1996; MacLeod, 1998), recall for information not expected to be tested (Friend items) was poorer than recall for information expected to be tested ("You" items). Additionally, items offering schematic support were recalled better than unassociated words such that participants recalled a similar number of "You" items in each condition but fewer Friend items when the words were unassociated, consistent with schema-related encoding (e.g., Bransford & Johnson, 1972; Mandler, 1984) such that pre-existing knowledge can enhance memory for to-be-remembered information in this domain.

Moreover, the enhanced recall of camping trip items compared with unassociated words is consistent with the congruity effect (Craik & Tulving, 1975) whereby memory is enhanced when the context (going on a camping trip) forms an integrated unit with the words presented (i.e., tent, matches, water, etc.). Furthermore, individual-item processing (the encoding of item-specific information) and relational processing (the encoding of similarities among a category; Einstein & Hunt, 1980; Hunt & McDaniel, 1993) may have aided memory performance via the encoding of both types of information (see Hunt & Einstein, 1981). Thus, participants appear to have allocated their cognitive resources in favor of items they were responsible for remembering but recalled additional items they were not responsible for remembering as a result of the benefits of schematic support (cf., Castel, 2005; Craik, 2002; Craik & Bosman, 1992; Golding et al., 1994; McGillivray & Castel, 2017).

Although there were no differences in performance between the groups or cues (and participants were not sensitive to importance) on the cue pairing test, when participants were responsible for remembering an item for the camping trip ("You"), their recall was sensitive to importance. Specifically, participants better remembered important items compared with unimportant items that they were responsible for remembering while showing no sensitivity for importance of items their friend was responsible for remembering. These findings suggest that both You and Friend items were still accessible in memory (they had generally accurate associative memory for who was responsible for remembering each item) and that participants may have inhibited Friend items to engage in responsible forgetting to enhance memory of goal-relevant, important information, potentially fitting inhibition accounts of directed forgetting (see also Aguirre et al., 2017).

Experiment 2

Strategically offloading valuable information

Whenever someone takes notes during class, uses Alexa or Siri to make a shopping list, or sets a reminder on their phone, they have engaged in cognitive offloading: the use of an external mechanism to record thoughts and memories to reduce cognitive load. Cognitive offloading can affect how we think (Barr et al., 2015; Dror & Harnad, 2008), is employed from an early age (e.g., Armitage et al., 2020), and has obvious benefits: one can remember more information with less effort. Unfortunately, there are also costs associated with offloading. For example, when someone's phone has a full battery and they urgently need to call a friend, accessing their friend's phone number and completing the call is easy and efficient. However, if someone’s phone were to run out of battery and they needed to use a payphone or a friend’s phone, accessing phone numbers from memory may be difficult or unsuccessful.

In terms of the benefits of offloading, externally storing information can facilitate memory for other, not-offloaded information by reducing potential interference from offloaded information (Risko & Dunn, 2015; Risko & Gilbert, 2016; Sparrow et al., 2011; Storm & Stone, 2015; see Carter, 2018; Dawson, 2020 for educational implications). For example, Storm and Stone (2015) asked participants to remember lists of information but allowed them to offload some of the information. Results revealed that when participants expected to be able to access offloaded information, the need to encode offloaded information was reduced and participants subsequently prioritized memory for not-offloaded information. Similarly, Henkel (2014) allowed participants to take pictures of objects on a museum tour. When participants’ memory for the objects was later tested, objects participants observed but did not take pictures of (not offloaded) were better remembered than photographed (offloaded) objects.

Although there are obvious memory benefits of offloading, there are also drawbacks. Specifically, when offloaded information becomes unavailable, memory for that information is impaired (e.g., Eskritt & Ma, 2014; Kelly & Risko, 2019a, 2019b; Lu et al., 2020; Marsh & Rajaram, 2019; Sparrow et al., 2011). For example, Sparrow et al. (2011) had participants study trivia questions and either told them that they would later have access to the information (offloaded) or that they would not have access to the information later (not offloaded). On a subsequent memory test for the trivia questions (without access to the information), participants who did not expect to have access to the information performed better than participants who thought they would have access to the stored information. Together, previous work indicates that externally storing information can benefit memory for not-offloaded information but memory for offloaded information is poorer if participants expect to have access to the saved information. Thus, if important information is offloaded and access to this external store became unexpectedly unavailable, this could lead to the forgetting of valuable information and potentially negative outcomes.

In many offloading tasks, participants study lists of words for a later test. On the first few lists, all participants are allowed to externally store each of the presented words and access the stores during the recall test. As a result, participants largely rely on the external store to recall most of the presented words and subsequently develop trust in the external store. However, on a later list, some participants are surprised by not having access to their external store on the memory test while other participants are warned that they will be unable to access any offloaded words during the recall test (thus they should encode the information internally). When both groups do not have access to any external stores during recall, the group that was aware that they would not be able to access offloaded words generally better recalls the words due to intentional encoding (e.g., Kelly & Risko, 2021; Risko et al., 2019). Thus, learners may strive to maximize memory utility and reduce interference by engaging in minimal encoding of offloaded information and prioritizing memory for not-offloaded information but if the offloaded words are inaccessible, this information may not be remembered.

While offloading can be helpful by reducing the cognitive load on memory and can assist in achieving task goals (Weis & Wiese, 2019), there may be instances where offloading everything is difficult or not feasible (e.g., taking notes during a long, fast-paced lecture). Therefore, when presented with information of varying importance, selectively offloading important information may be an effective strategy to minimize the consequences of forgetting. In contrast, remembering important information may be a critical aspect of a functional memory system if access to external stores is unreliable (i.e., we should remember important information in case we are unable to access the external store). Thus, while it may be beneficial to offload valuable information in case one experiences retrieval failure, it may be of adaptive benefit to encode this information in case the external store is unavailable.

In Experiment 2, I was interested in whether learners are strategic in their offloading decisions when the external store has a limited capacity. Specifically, when presented with to-be-remembered information of various values, I examined participants' offloading decisions as well as memory for information that is not offloaded. Generally, I expected participants to offload high-value information, but this may come at the expense of memory selectivity for not-offloaded information. Additionally, when the external store is surprisingly unavailable, I expected participants to forget the important information that was offloaded, illustrating the potential dangers of offloading.

When presented with words paired with point values to remember for a test (where points are scored by recalling words), an obvious strategy may be to offload the five highest valued words and use your memory system to recall the remaining words. While this seems like the most efficient strategy, if the external store were unreliable, this strategy results in frequent forgetting of high-value words. However, in situations where the external store is unreliable, the best strategy may be to prioritize memory for the highest valued words and avoid redundancy by offloading medium- or low-value words. The drawback of this strategy is that memory is fallible and may result in the forgetting of high-value words. Alternatively, one could both prioritize memory for and offload high-value words, but this redundancy may limit overall task performance via the number of words remembered.

In Experiment 2, I presented participants with six lists of words to remember for a later test, with each list containing 15 words. Each word was paired with a point value counting towards participants' score if recalled and I allowed participants to offload five words of their choice per list (offloaded words remained on the screen during the study phase). On the first four lists, the offloaded words were available to participants during the test phase (and participants were told that they would be available, establishing trust in the external store). However, on List 5, the offloaded words were not given to participants during the test phase and one group of participants was not aware that the words would be unavailable (decreasing trust in the external store); the other group of participants was aware that they would not have access to any offloaded words. On the final list, participants were told that they may or may not be given access to the offloaded words during the recall test (and the offloaded words were not available).

On the first four lists, I expected participants to be selective in their offloading such that they strategically offload high-value words. However, this offloading of high-value words may come at the expense of memory selectivity for not-offloaded words. Particularly, compared with the group who was aware that they would not have access to offloaded words on List 5, I expected participants presuming access to the external store to be less selective for high-value words (but no differences in total recall). Finally, on the sixth list where the reliability of the external store was unknown, I expected participants to be more cautious with their offloading decisions and opt for the more redundant strategy.

Participants. Participants were 192 undergraduate students (M_{age} = 20.30, SD_{age} = 2.39) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in two exclusions. With this sample size, I had an 80% chance of detecting a *medium* (Cohen's *d* = .45) effect between groups.

Materials. The to-be-remembered words (unrelated) were between 4 and 7 letters (*M* = 4.99, *SD* = .98), and on the log-transformed Hyperspace Analogue to Language frequency scale (with lower values indicating lower frequency in the English language and higher values indicating higher frequency), ranged from 5.48-12.65 and averaged a score of 8.81 (*SD* = 1.57). In terms of concreteness (with lower values indicating lower concreteness and higher values indicating higher concreteness), words ranged from 2.50-5.00 and averaged a score of 4.52 (*SD* = .46). Words were classified according to the English Lexicon Project website (Balota et al., 2007).

Procedure. Participants were presented with six lists of unique, randomly selected to-be-remembered words with each word paired with a unique, randomly assigned value between 1 and 15 indicating how much the word was "worth." Each point value was used only once within each list and the order of the point values within lists was randomized. The stimulus words were presented for 3 seconds each. After the presentation of all 15 word-number pairs in each list, participants were given a 1-minute immediate free recall test in which they had to recall as many words as they could from the list (they did not need to recall the point values). Immediately following the recall period, participants were told their score (the sum of the values of the words they recalled) on the list but were not given feedback about specific items. Participants then proceeded to the next list when ready.

On each list, participants were allowed to offload up to five words of their choice and the offloaded words remained on the screen during the study phase. To offload a word, participants clicked a button to add it to their external store. On the first four lists, the offloaded words were available to participants during the test phase (and participants were told that they would be available). However, on List 5, the offloaded words were not given to participants during the test phase and one group of participants (*n* = 64) was aware that the words would not be available (they were told “On the next list, you will not have access to the words you save during the study phase when you take the test”) while another group of participants (*n* = 68) was not aware that they would not have access to any offloaded words (they were simply told to begin the next list when ready). On the final list, participants were told that they may or may not have access to offloaded words on List 6, and participants were not given access to the offloaded words on this list. As a baseline, I also included a condition in which participants did not engage in offloading (*n* = 60).

Results

To examine differences in offloading behavior and recall, I computed multilevel models (MLMs) using Jamovi where I treated the data as hierarchical or clustered (i.e., multilevel) with items nested within individual participants. Since offloading and recall at the item level was binary (offloaded or not offloaded; correct or incorrect), I conducted logistic MLMs. In these analyses, the regression coefficients are given as logit units (i.e., the log odds of offloading/correct recall). I report exponential betas (e^B), and their 95% confidence intervals ($CI_{95\%}$), which give the coefficient as an odds ratio (i.e., the odds of offloading/correctly recalling a word divided by the odds of not offloading/recalling a word). Thus, e^B can be interpreted as the extent to which the odds of offloading/recalling a word changed. Specifically, values greater than 1 represent an increased likelihood of offloading/recall while values less than 1 represent a decreased likelihood of offloading/recall. In our analysis of each group, I grouped participants as: participants who were aware that they would not have access to offloaded words on List 5 (“aware”), participants who were not aware that they would not have access to offloaded words on List 5 (“unaware”), and the participants who did not do any offloading (“control”).

To examine offloading behavior on Lists 1-4 (see Figure 2a), I conducted a logistic MLM with item-level offloading modeled as a function of value with awareness of the surprising inaccessibility of the offloaded words on List 5 (aware, unaware) as a between-subjects factor. Results revealed that value significantly predicted offloading decisions [$e^B = 1.37$, $CI_{95\%} = 1.34 – 1.39$, $z = 37.49$, $p < .001$] such that high-value words were offloaded more than low-value words. However, there were no group differences in total offloading [$e^B = .85$, $CI_{95\%} = .71 – 1.02$, $z = -1.78$, $p = .075$] and value did not interact with awareness of List 5 [$e^B = 1.01$, $CI_{95\%} = .98 – 1.04$, $z = .66$, $p = .508$].

To examine recall on Lists 1-4 (see Figure 3a), I conducted a logistic MLM with item-level recall modeled as a function of value with group (the control group was the reference group) as a between-subjects factor. Results revealed that value significantly predicted recall [$e^B = 1.15$, $CI_{95\%} = 1.14 – 1.16$, $z = 28.46$, $p < .001$] such that high-value words were better recalled than low-value words. Additionally, aware participants recalled a greater proportion of words (*M* = .66, *SD* = .14) than control participants (*M* = .51, *SD* = .13), [$e^B = 2.00$, $CI_{95\%} = 1.57 – 2.55$, $z = 5.59$, $p < .001$]. Similarly, unaware participants recalled a greater proportion of words (*M* = .62, *SD* = .14) than control participants [$e^B = 1.73$, $CI_{95\%} = 1.36 – 2.19$, $z = 4.47$, $p < .001$]. Both comparisons interacted with value [both *p*s < .001] such that participants who could offload words were more selective than participants not offloading any words. Specifically, an analysis of the simple effects revealed that value was a better predictor of recall for aware [$e^B = 1.15$, $CI_{95\%} = 1.13 – 1.17$, $z = 16.2$, $p < .001$] and unaware participants [$e^B = 1.21$, $CI_{95\%} = 1.19 – 1.23$, $z = 21.1$, $p < .001$] compared with control participants [$e^B = 1.10$, $CI_{95\%} = 1.09 – 1.12$, $z = 11.8$, $p < .001$].

I was also interested in the recall of words that participants did not offload. As such, I conducted a logistic MLM with item-level recall on Lists 1-4 modeled as a function of value with group (the control group was the reference group) as a between-subjects factor. Results revealed that value significantly predicted recall [$e^B = 1.09$, $CI_{95\%} = 1.08 – 1.10$, $z = 14.82$, $p < .001$] such that high-value words were better recalled than low-value words. Additionally, aware participants recalled a similar proportion of not-offloaded words ($M = .53$, $SD = .18$) as control participants ($M = .51$, $SD = .13$), [$e^B = 1.21$, $CI_{95\%} = .91 – 1.60$, $z = 1.33$, $p = .183$]. However, unaware participants recalled a similar proportion of words ($M = .48$, $SD = .20$) than control participants [$e^B = .98$, $CI_{95\%} = .74 – 1.29$, $z = -.17$, $p = .868$]. The comparison between aware and control participants interacted with value [$e^B = .96$, $CI_{95\%} = .94 – .99$, $z = -2.86$, $p = .004$] but the comparison between unaware and control participants did not significantly interact with value [$e^B = 1.00$, $CI_{95\%} = .97 – 1.02$, $z = -.34$, $p = .732$]. Specifically, an analysis of the simple effects revealed that value was a better predictor of recall for control participants [$e^B = 1.10$, $CI_{95\%} = 1.09 – 1.12$, $z = 11,86$, $p < .001$] than aware participants [$e^B = 1.06$, $CI_{95\%} = 1.04 – 1.08$, $z = 5.63$, $p < .001$] but value was similarly predictive of recall for unaware participants [$e^B = 1.10$, $CI_{95\%} = 1.08 – 1.12$, $z = 9.08$, $p < .001$].

To examine offloading behavior on List 5 (see Figure 2b), I conducted a logistic MLM with item-level offloading modeled as a function of value with awareness of the surprising inaccessibility of the offloaded words on List 5 (aware, unaware) as a between-subjects factor. Results revealed that value significantly predicted offloading decisions [$e^B = 1.37$, $CI_{95\%} = 1.32 – 1.42$, $z = 17.06$, $p < .001$] such that high-value words were offloaded more than low-value words. Additionally, unaware participants offloaded more than aware participants [$e^B = 2.14$, $CI_{95\%} = 1.47 – 3.11$, $z = 4.00$, $p < .001$] but value did not interact with awareness of List 5 [$e^B = 1.06$, $CI_{95\%} = .99 – 1.14$, $z = 1.63$, $p = .103$].

Next, I examined recall on List 5 (see Figure 3b). I conducted a logistic MLM with item-level recall modeled as a function of value with group (the control group was the reference group) as a between-subjects factor. Results revealed that value significantly predicted recall [$e^B = 1.07$, $CI_{95\%} = 1.05 – 1.09$, $z = 7.16$, $p < .001$] such that high-value words were better recalled than low-value words. Additionally, aware participants recalled a similar proportion of words ($M = .48$, $SD = .20$) as control participants ($M = .51$, $SD = .19$), [$e^B = .86$, $CI_{95\%} = .64 – 1.16$, $z = -.97$, $p = .333$]. However, unaware participants recalled a smaller proportion of words ($M = .43$, $SD = .17$) than control participants [$e^B = .69$, $CI_{95\%} = .51 – .92$, $z = -2.52$, $p = .012$]. The comparison between aware and control participants did not interact with value [$e^B = .98$, $CI_{95\%} = .93 – 1.02$, $z = -1.03$, $p = .304$] but the comparison between unaware and control participants significantly interacted with value [$e^B = .84$, $CI_{95\%} = .81 – .88$, $z = -7.46$, $p < .001$]. Specifically, an analysis of the simple effects revealed that value was a positive predictor of recall for aware [$e^B = 1.12$, $CI_{95\%} = 1.08 – 1.15$, $z = 6.64$, $p < .001$] and control participants [$e^B = 1.14$, $CI_{95\%} = 1.10 – 1.18$, $z = 7.72$, $p < .001$]. However, value was a negative predictor of recall for unaware participants [$e^B = .96$, $CI_{95\%} = .93 – .99$, $z = -2.56$, $p = .011$].

To examine offloading behavior on List 6 (see Figure 2c), I conducted a logistic MLM with item-level offloading modeled as a function of value with awareness of the surprising inaccessibility of the offloaded words on List 5 (aware, unaware) as a between-subjects factor. Results revealed that value significantly predicted offloading decisions [$e^B = 1.11$, $CI_{95\%} = 1.08 – 1.13$, $z = 8.86$, $p < .001$] such that high-value words were offloaded more than low-value words. However, there were no group differences in total offloading [$e^B = .90$, $CI_{95\%} = .67 – 1.21$, $z = -.68$, $p = .497$] and value did not interact with awareness of List 5 [$e^B = .99$, $CI_{95\%} = .94 – 1.03$, $z = -.65$, $p = .518$].

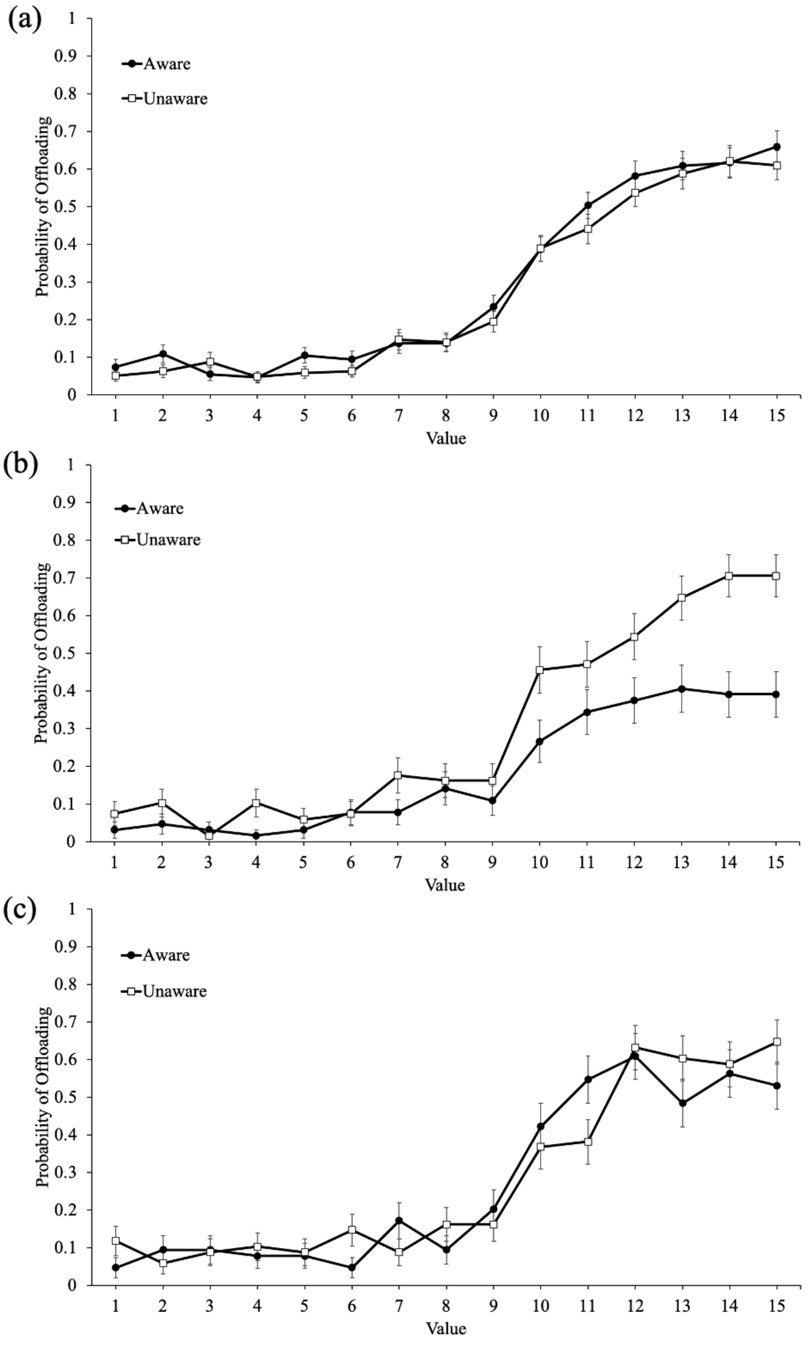


Figure 2. Probability of offloading a given word on Lists 1-4 (a), List 5 (b), and List 6 (c) as a function of participants' awareness about losing access to offloaded words on List 5 and word value in Experiment 2. Error bars reflect the standard error of the mean.

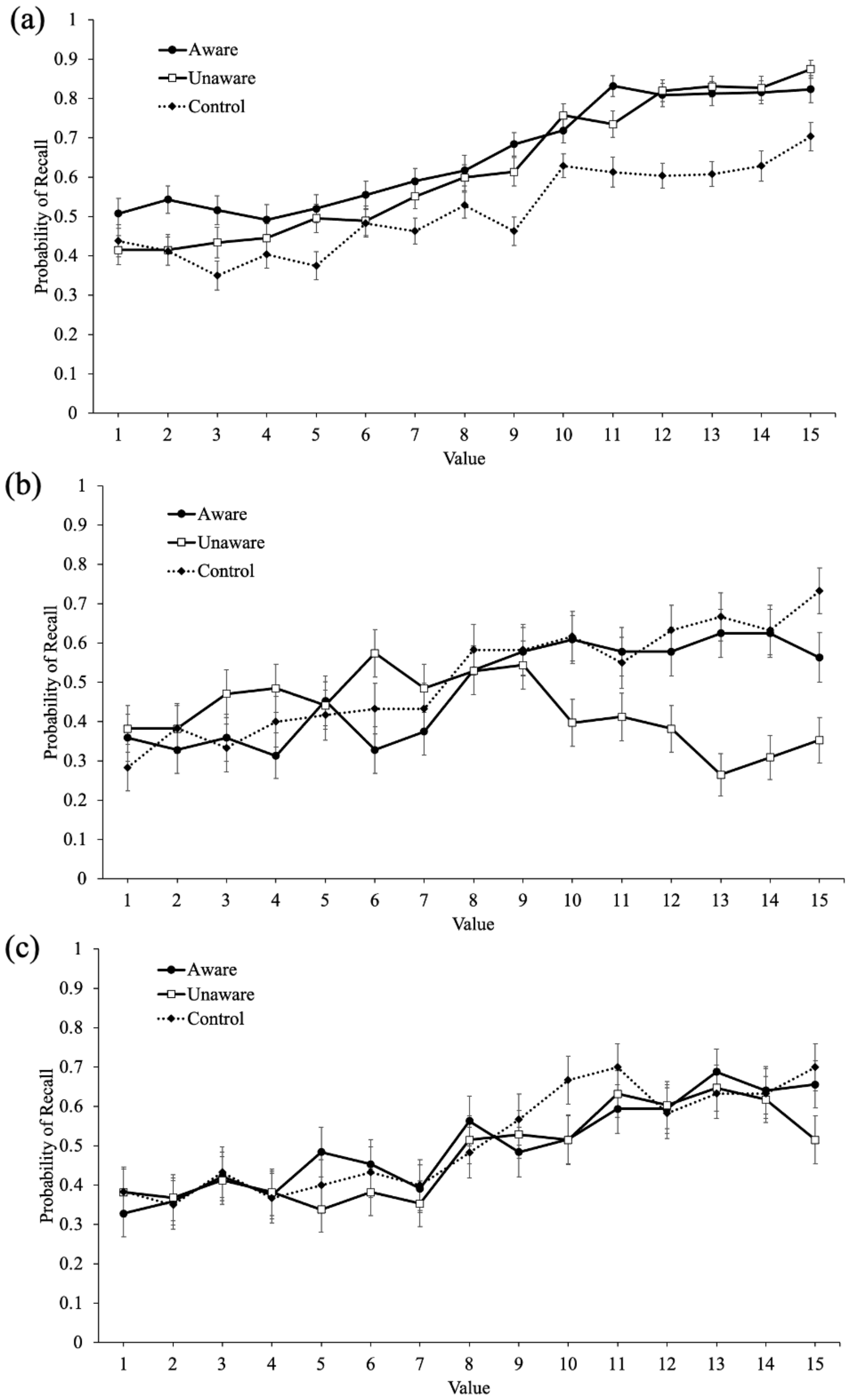


Figure 3. Probability of recall for Lists 1-4 (a), List 5 (b), and List 6 (c) as a function of participants' awareness about losing access to offloaded words on List 5 (as well as participants not offloading any words) and word value in Experiment 2. Error bars reflect the standard error of the mean.

To examine recall on List 6 (see Figure 3c), I conducted a logistic MLM with item-level recall modeled as a function of value with group (the control group was the reference group) as a between-subjects factor. Results revealed that value significantly predicted recall [$e^B = 1.11$, $CI_{95\%} = 1.09 – 1.13$, $z = 11.24$, $p < .001$] such that high-value words were better recalled than low-value words. However, aware participants recalled a similar proportion of words ($M = .50$, $SD = .18$) as control participants ($M = .52$, $SD = .19$), [$e^B = .94$, $CI_{95\%} = .69 – 1.27$, $z = -.40$, $p = .691$]. Similarly, unaware participants recalled a similar proportion of words ($M = .48$, $SD = .20$) as control participants [$e^B = .85$, $CI_{95\%} = .63 – 1.15$, $z = -1.07$, $p = .286$]. Neither comparison interacted with value [both ps > .276] such that participants who could offload words were similarly selective in their recall as participants not offloading any words. Specifically, an analysis of the simple effects revealed that value was a similar predictor of recall for aware [$e^B = 1.11$, $CI_{95\%} = 1.08 – 1.15$, $z = 6.63$, $p < .001$], unaware [$e^B = 1.10$, $CI_{95\%} = 1.06 – 1.13$, $z = 5.95$, $p < .001$], and control participants [$e^B = 1.13$, $CI_{95\%} = 1.09 – 1.16$, $z = 6.97$, $p < .001$].

Discussion

With increased access to technology, offloading is easier than ever (see Grinschgl et al., 2020; Storm et al., 2017). While offloading may offer some memory benefits, there can also be costs (e.g., Grinschgl et al., 2021b; Kelly & Risko, 2019a, 2019b). For example, although offloading can help store large amounts of information, there are instances where we do not have access to external stores such as closed-note exams or if technology runs out of power or malfunctions. Thus, learners may need to be strategic with what information is offloaded and what information is stored in memory.

When able to offload certain information, the responsible remembering framework (Murphy & Castel, 2020) suggests that if learners are guaranteed access to offloaded information, they should prioritize the offloading of high-value information. However, if access to the offloaded information is fallible, learners should encode and remember valuable information to avoid consequences for forgetting. In Experiment 2, I was interested in whether learners are strategic in their offloading decisions when the external store has a limited capacity and how the surprising unavailability of offloaded information influences memory selectivity. I presented learners with words paired with point values counting towards their scores if recalled and allowed participants to offload some of the words. Generally, results revealed that learners selectively offloaded the highest valued words but were also selective in their memory for the words that they did not offload. However, when access to the external store was surprisingly unavailable, learners frequently forgot high-value words, demonstrating the potential dangers of offloading.

When choosing what information to offload, there are three possible strategies for using offloading to enhance memory for valuable information. First, a learner could offload the most valuable information and use memory to remember medium-value information. If the external store is reliable, this would ensure the availability of high-value information, but the learner would be prone to forgetting high-value words if the external store is unreliable. Second, a learner could offload medium-value information and use memory for high-value words. However, with this technique, the learner would be prone to forgetting high-value words if the encoding of the information is insufficient or in the case of retrieval failure. Finally, a learner could both offload and use memory for high-value information, although this redundancy could reduce total capacity.

Each of the strategies has advantages and disadvantages and which strategy a learner should employ depends on 1) the reliability of the external store, 2) one’s memory abilities, and 3) the variability of the value of the to-be-remembered information. For example, imagine that a given participant is presented with 15 words to remember (and the values range from 1 to 15) and they can offload five words, but the reliability of the external store is 80% (as seen on List 6). Imagine that, of the 10 not-offloaded words, this given participant focuses on remembering the five highest valued words but the probability of recall for a given word is 60%. In this case, the expected score for the offloaded words would be 52 (.80 x 65) and the expected score for the medium-value words encoded in memory would be 24 (.60 x 40). This amounts to a total score of 76. Imagine that a different participant opts to use their memory to remember the five highest-valued words and to offload the medium-valued words. In this case, the expected score for the words in memory would be 39 (.60 x 65) and the expected score for the offloaded words would be 32 (.80 x 40 = 32). This amounts to a total score of 71. If a participant employed the redundancy strategy (both offloaded and using memory for high-value words), even if this resulted in a 100% chance of recall, I would expect this participant to score 65 points (1.00 x 65 + 0.00 x 40). Thus, given these probabilities, offloading high-value words and using memory for the other words would be the optimal strategy.

In the previous example, imagine that the probability of recall for a given word is 90%. Now, if a participant opts to offload the five highest valued words and use memory for the medium-valued words, their expected score would be 88 (.80 x 65 + .90 x 40). However, if a participant uses memory for the five highest valued words, their expected score would be 90.5 (.90 x 65 + .80 x 40). Therefore, the optimal strategy would be to use memory for high-value words and offload medium-value words. Together, this example illustrates that, when deciding what to offload and what to encode in memory, learners should store valuable information in whichever store is more reliable.

In the previous example, the value of the to-be-remembered information contained relatively little variability. If there were an instance where some information was of extreme importance (e.g., your social security number), the redundancy (both offloading and remembering information) strategy may be superior to offloading some information and remembering other information. For example, imagine that a given participant is presented with 15 words to remember but the values are 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 55, 55, 55, 55, and 55. Again, imagine that the participant can offload five items but the reliability of the external store is 80% and that of the 10 remaining words, the learner focuses on the five highest valued words but the probability of recall for a given word is 60%. If the learner offloads the five highest valued words, their expected score would be 244 (.80 x 275 + .60 x 40). If the learner aims to remember the five highest valued words, their expected score would be 197 (.60 x 275 + .80 x 40). However, if the redundancy strategy is employed (and assume that this yields 100% success), the expected score would be 275 (1.00 x 275 + 0.00 x 40). Thus, when some information is vastly more important than other information, it may be optimal to both offload and encode this important information.

Together, this suggests that if memory is more reliable, using memory for high-value words would be better but if the external store is more reliable, using the external store for high-value words would be better; in the case of extreme values, the redundancy strategy would be best. Thus, there may be a metacognitive component to maximizing the efficiency of offloading. Future work could examine the implications of offloading strategies in the classroom. For example, if an instructor told their students that they may or may not allow the use of a notecard or cheat sheet during an important exam, but that the students would not know whether they would be able to use it until the exam begins, students may opt to both offload and remember the most important information. Additionally, offloading important information to enhance exam performance could come at the expense of learning. Future work should examine the classroom implications of offloading.

In sum, responsible rememberers should maximize memory for important information to minimize the consequences of forgetting (Murphy & Castel, 2020) but an effective strategy to achieve memory for important information may be to strategically offload it. At the same time, the potential dangers of offloading suggest that it may be responsible to store important information internally rather than offloading it. Here, learners appeared to prefer the former strategy (i.e., offloading high-value words) and attempted to maximize performance by also being selective for words stored internally. However, important information was frequently forgotten when the external store was surprisingly unreliable. Thus, while responsible remembering may be enhanced by offloading important information, it may be beneficial to remember some important information (i.e., a guardian's phone number in case of an emergency) to prepare for instances of fallible external stores.

Experiment 3

Offloading important, goal-relevant information

In Experiment 3, I was interested in how learners utilize offloading when information differs in subjective value rather than objective value. Specifically, using a similar procedure as Experiment 2, I presented participants with lists of words that were semantically related (i.e., items to bring on a camping trip). I expected people to offload items they considered important. Alternatively, learners may prioritize these important items in memory and utilize the external store for less important items to maximize total output.

Method

Participants. After exclusions, participants were 52 undergraduate students ($M_{age} = 20.06$, $SD_{age} = 1.96$) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. No younger adults were excluded for cheating but 11 younger adults were excluded because they did not offload at least 10 words throughout the task.

Materials and Procedure. Participants were told that they would be presented with lists of words to remember for a later test with each list being along a theme and that they should try to imagine themselves in that situation. Participants were then presented with five lists of 15 words, with each list containing items along a theme (going camping, going on vacation, throwing a child's party, going to class, and going on a picnic; stimuli were adapted from McGillivray & Castel, 2017 are available on OSF). Each word was presented one at a time, for 3 seconds each, and in random order; list themes occurred in a fixed order.

On each list, participants were allowed to offload up to five words of their choice. To offload a word, participants clicked a button to add it to their external store. During the test, offloaded words appeared at the top of participants' screens, and they were reminded to retype the offloaded words into the text box. I scored all offloaded words as correct even if participants did not type them into the box. On the first four lists, the offloaded words were available to participants on the test. However, on List 5, the offloaded words were surprisingly unavailable to participants (during the test, they were told "Sorry, you will not have access to the words you saved on this list").

After the presentation of all 15 words on each list, participants were given a self-paced free recall test in which they were asked to recall all the words from the just-presented list. Following each study-test cycle, participants were shown the words from that list, one at a time (in alphabetical order), and asked to rate the words from that list on a scale of how important it would be to remember them from 1 (not at all important to remember) to 7 (very important to remember).

Results

I conducted logistic MLMs with item-level offloading modeled as a function of each participant's own importance ratings (see Figure 4). Results revealed that importance ratings significantly predicted offloading [$e^B = .90$, $CI_{95\%} = .86 – .94$, $z = -4.69$, $p < .001$] such that items rated as less important to remember were offloaded more than items rated as more important.

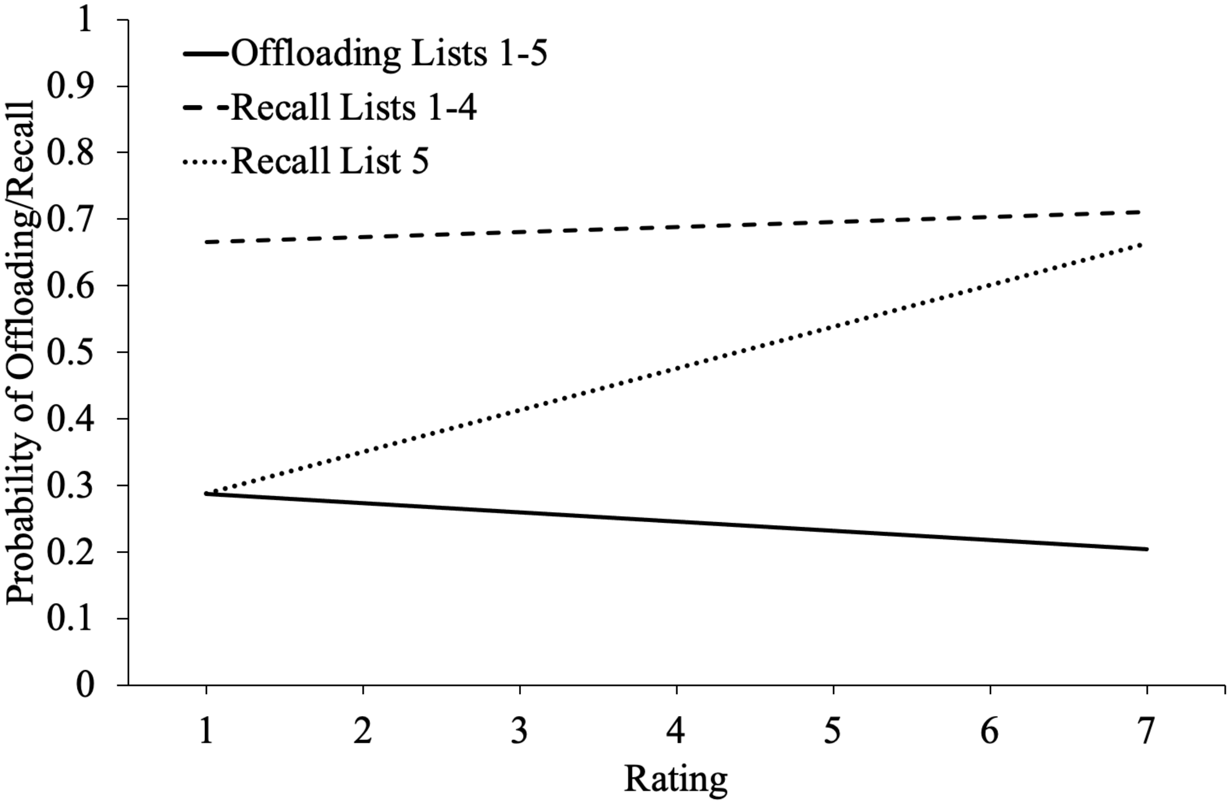


Figure 4. Linear trendlines for the probability of offloading on Lists 1-5, recall on Lists 1-4, and recall on List 5 as a function of importance ratings in Experiment 3.

Next, I examined recall as a function of participants' ratings and whether they had access to the external store on the test (i.e., Lists 1-4 vs List 5). Results from our model revealed that importance ratings significantly predicted recall [$e^B = 1.22$, $CI_{95\%} = 1.15 – 1.29$, $z = 6.99$, $p < .001$] such that items rated as important were better recalled than items rated as less important. Additionally, list type significantly predicted recall [$e^B = 2.20$, $CI_{95\%} = 1.85 – 2.61$, $z = 8.98$, $p < .001$] such that participants recalled more words when given access to the external store. Critically, ratings interacted with whether participants had access to the external store on the test [$e^B = .81$, $CI_{95\%} = .72 – .90$, $z = -3.90$, $p < .001$] such that importance ratings were a better predictor of recall on List 5 [$e^B = 1.36$, $CI_{95\%} = 1.23 – 1.50$, $z = 6.09$, $p < .001$] than Lists 1-4 [$e^B = 1.09$, $CI_{95\%} = 1.04 – 1.15$, $z = 3.57$, $p < .001$].

Discussion

In Experiment 3, learners offloaded more items they rated as less important compared to items they judged as more important. However, they better recalled important items relative to less important items, particularly on List 5 when the external store was surprisingly unavailable. Thus, learners may be tuned to the need to use memory for important items and thus use the external store to remember the less important items that are less likely to be remembered. In sum, when information differs in subjective importance, in the absence of the external store, important information is still remembered better than less important information such that under these conditions (goal-based memory of semantically related items), learners can use memory efficiently.

Experiment 4

How the Costs of Offloading Influence Offloading Behavior and Memory Reliance

To maximize the utility of offloading, one should prioritize the offloading of valuable information (Murphy, 2023). Specifically, assuming the external store is reliable, offloading information provides a greater likelihood of information accessibility due to the fallibility of memory (e.g., if we write something down on our phone, we are probably more likely to have access to that information than if we were to try to remember it). However, if the external store is unreliable, or at least less reliable than memory (e.g., if your phone is low on battery), one should prioritize important information in memory to ensure the greatest likelihood of access to this information. As such, prior work has demonstrated that learners do not selectively encode valuable information when told they can rely on an external store to remember information and if the external store is surprisingly unavailable, this often leads to the forgetting of valuable information (Park et al., 2022).

To maximize the accessibility of information by both remembering as much information as possible and offloading the most valuable information, people may need to rely on their metacognition. When making offloading decisions, learners need to monitor what information has been successfully encoded and what information will not be remembered later. The information that has been successfully encoded can be retrieved from memory and thus does not need to be offloaded (assuming memory is reliable). However, information that will not be remembered could be offloaded to provide later accessibility (assuming the external store is reliable). As such, metacognition should play a crucial role in offloading behavior by monitoring learning but also in the strategic control of cognitive and offloading resources. In sum, the optimal offloader should be aware of what information will be remembered and offload information that will not be remembered.

To further maximize the utility of an external store, one should offload the most difficult information to remember. For example, when studying lists of words, there are many stimulus characteristics that contribute to a given item's memorability such as concreteness, frequency, emotionality, word length, and animacy (see Murphy & Castel, 2022 for a recent demonstration of these effects). Narrowing in on animacy, animate items (e.g., animals) tend to be better remembered than inanimate items (e.g., objects), which may be a form of adaptive memory (Bonin et al., 2014, 2015; Leding, 2019; Nairne et al., 2013; Popp & Serra, 2016, 2018; Serra, 2021; see Nairne, 2016; Nairne et al., 2017 for a review). In the context of offloading, using an external store to remember the most difficult items to remember and memory to remember the easiest items should help increase the total amount of information accessible. Specifically, if the external store has a limited capacity or there is a cost of offloading, if a learner uses memory for just the easiest items and offloads the hardest items to remember, they may be able to access more items than if they were to use memory for the hardest items to remember.

In Experiment 4, I was interested in the role of metacognition in optimizing the use of offloading. Specifically, I employed a novel metacognitive offloading optimization task (MOOT) where I used optimal offloading as a measure of effective metacognition by having participants "pay" to offload an item. In this task, participants studied lists of words to remember for a later test and were told that each word was worth 10 points such that if they recalled five words on the test, they would score 50 points (with the participants' goal being to maximize their score). However, I allowed participants to offload as many words as they wanted (offloaded words were available to them during the test and thus did not need to be retrieved from memory) but included a cost for offloading each word such that offloading a word reduced its value, and I varied the severity of this cost.

For example, suppose participants are presented with a list of 20 words to remember for a test and participants score 10 points for every item recalled from memory on the test but "pay" 5 points to offload an item. If a participant offloads all the items, they would score 100 points but if they have perfect memory and recall all items (without offloading any), their score would be 200; if a participant offloads items they would not have remembered via memory, they can beat the score of 100 (the score one would achieve if all words were offloaded). For instance, imagine that a learner can recall 10 items on the test. If they offload the 10 items that they would not get correct, then they would score 150 points. Thus, with perfect metacognition and the ability to remember

10 items, the maximum score for this learner would be 150. However, if this learner's metacognition is not perfect, they would score less than 150 points. Therefore, participants should optimize their offloading by monitoring which items have been effectively encoded in memory and which items have not (and thus should be offloaded).

If the cost of offloading an item is lower, say 2 points per item, then participants would score 160 points if they offloaded everything but still have a maximum score of 200 if they offloaded nothing and have a perfect memory. Again, participants can beat 160 if they offload the right items but this depends on their memory and metacognitive accuracy. However, if the price of offloading is higher, say 8 points per item, then a participant's score would be 40 if they offloaded everything. Thus, the price of offloading is important in considering how much to offload and as the price for offloading goes down, the criterion for offloading should shift toward offloading more items.

In addition to the cost of offloading each item, participants should incorporate the effects of stimulus characteristics on the likelihood of remembering into their offloading decisions. For example, since animate items are better remembered than manmade objects (e.g., Popp & Serra, 2016, 2018), if the to-be-remembered list contains animate and manmade items, participants should use memory for the more memorable, animate items and offload the less memorable, inanimate items (i.e., if participants focused on inanimate items, they would likely be able to remember fewer total items). This strategy would maximize the number of items stored in memory and thus increase the participant's maximum score.

In this task, there should be an optimal solution for each participant offloading assuming the learner knows what the cost is for each offloaded item and assuming that they have perfect metacognition (i.e., for every item, participants know whether they will or will not remember that item). Since the reward for recalling any item from memory is the same (10 points), the optimal strategy is to recall as many items from memory as possible and offload all words that would not have been remembered. Thus, in the present task, the maximum score varies based on how good the person's memory is. Here, I first determined each participant's baseline memory ability using a separate set of lists where I tested their memory when no offloading was allowed. Specifically, by examining a given participant's recall ability without any offloading, I can compute what their maximal score would be on each list type if their metacognition were perfect. I also examined how good participants believed their memory was by collecting global predictions of performance which I used to compute participants' metacognitive calibration. I expected that participants would be closer to their maximal score when the cost of offloading is high (i.e., offloaded words are worth 2 points) than when the cost of offloading is lower (i.e., offloaded words are worth 8 points).

Method

Participants. After exclusions, participants were 136 undergraduate students (M_{age} = 20.41, SD_{age} = 3.65) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in two exclusions. I also excluded participants who did not recall a single word on the baseline recall tests which resulted in two exclusions. A sensitivity analysis based on the obtained data indicated that for a repeated-measures ANOVA with 3 levels (list type: worth 8, 5, 2), with an average correlation of .66 between repeated measures (score on each list type), assuming alpha = .05, power = .80, the smallest effect size the design could reliably detect is $\eta_p^2 = .01$.

Materials. Presented words were either animals (animate items) or manmade objects (inanimate items) taken from Popp and Serra (2016, 2018). Each list contained 10 animate items and 10 inanimate items; items in each list were presented in random order. Stimuli are available on OSF.

Procedure. Participants studied nine lists of words with each list containing 20 words presented one at a time for 6 seconds each. After the presentation of all 20 words in each list, participants were given a self-paced, immediate free recall test in which they had to recall as many words as they could from the list. However, on the first three lists, participants did not engage in any offloading. Specifically, participants were told that each item was worth 10 points if they recalled it on the test and that they should try to maximize their score. Prior to each of the first three lists, participants were asked to predict how many words (out of 20) that they would correctly recall on the test; participants were given as much time as they needed to make their predictions.

On the next six lists, participants were allowed to offload any word of their choice and could offload as many words as they liked. To offload a word, participants clicked a button labeled "Save current word" and offloaded words were provided to participants on the recall test. Critically, participants were told that if they saved a word, it would be worth fewer points—participants were given the specific values of offloaded words prior to studying each list (the value of offloaded words was either 8, 5, or 2 points; the value of words recalled from memory was always 10 points). Specifically, on two lists offloaded items were worth 8 points, on two lists offloaded items were worth 5 points, and on two lists offloaded items were worth 2 points; list order was counterbalanced but each list scoring type occurred in blocks (i.e., the two lists where offloaded words were worth 8 points occurred consecutively). At the end of each of the nine lists, participants were told their scores but were not given feedback about specific items.

Results

The number of items offloaded and recalled from memory as a function of list type is shown in Figure 5. To examine offloading behavior, I conducted a 3 (list type: worth 8, 5, 2) x 2 (animacy: animate, inanimate) repeated-measures ANOVA. Results revealed differences between list types [Mauchly's $W = .89$, $p < .001$; Huynh-Feldt corrected results: $F(1.82, 246.05) = 55.90$, $p < .001$, $\eta_p^2 = .29$] such that participants offloaded more items per list when offloaded items were worth 8 points ($M = 8.90$, $SD = 6.11$) than when they were worth 5 points ($M = 6.72$, $SD = 5.68$), [$p_{\text{holm}} < .001$, $d = .37$] and when they were worth 2 points ($M = 4.71$, $SD = 4.58$), [$p_{\text{holm}} < .001$, $d = .72$]; participants also offloaded more frequently when offloaded items were worth 5 points than when they were worth 2 points [$p_{\text{holm}} < .001$, $d = .35$]. There was a significant effect of animacy [$F(1, 135) = 5.30$, $p = .023$, $\eta_p^2 = .04$] such that inanimate items were offloaded more often per list ($M = 3.55$, $SD = 2.60$) than animate items ($M = 3.24$, $SD = 2.45$). However, list type did not interact with animacy [$F(2, 270) = .1.37$, $p = .257$, $\eta_p^2 < .01$].

To examine recall (only words recalled from memory (i.e., not including offloaded words), I conducted a 3 (list type: worth 8, 5, 2) x 2 (animacy: animate, inanimate) repeated-measures ANOVA. Results revealed differences between list types [$F(2, 270) = 21.21$, $p < .001$, $\eta_p^2 = .14$] such that participants recalled (from memory) fewer items per list when offloaded items were worth 8 points ($M = 6.33$, $SD = 4.67$) than when they were worth 5 points ($M = 7.21$, $SD = 4.58$), [$p_{\text{holm}} = .002$, $d = .18$] and when they were worth 2 points ($M = 8.06$, $SD = 4.55$), [$p_{\text{holm}} < .001$, $d = .35$]; additionally, participants recalled more items when they were worth 5 points than when they were worth 2 points [$p_{\text{holm}} = .002$, $d = .17$]. There was a significant effect of animacy [$F(1, 135) = 15.26$, $p < .001$, $\eta_p^2 = .10$] such that more animate items were recalled per list ($M = 3.85$, *SD* = 2.30) than inanimate items (*M* = 3.35, *SD* = 2.19), but list type did not interact with animacy [$F(2, 270) = 1.49$, $p = .227$, $\eta_p^2 = .01$].

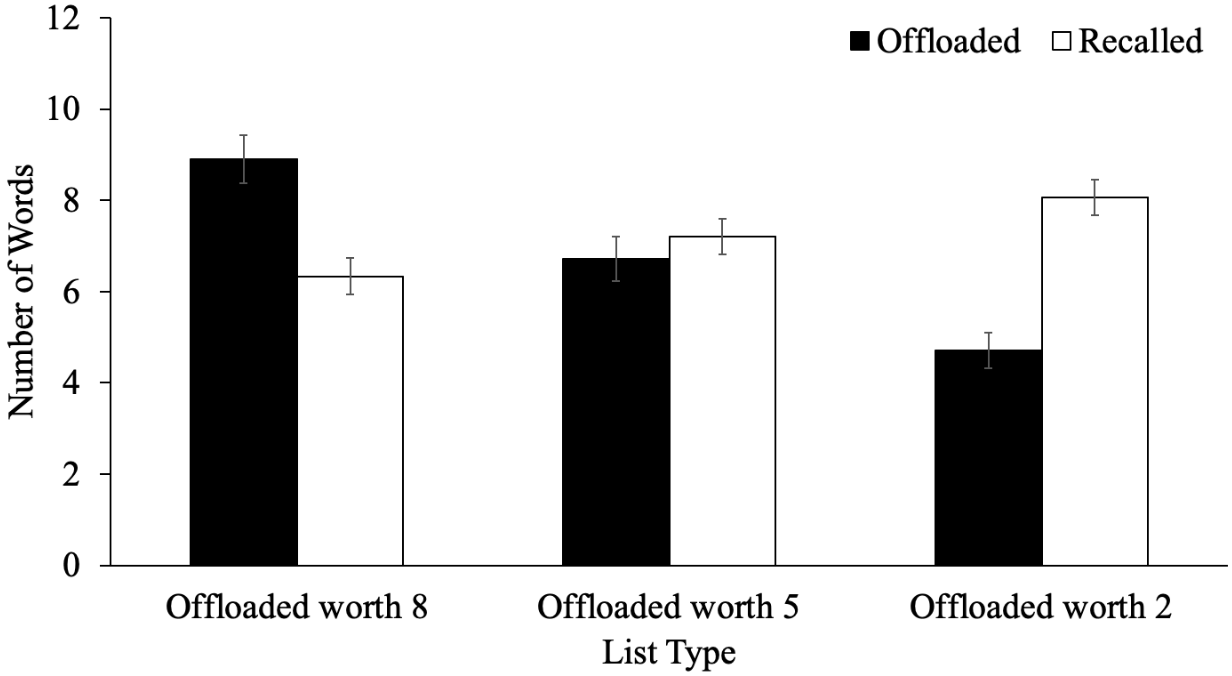


Figure 5. The average number of words offloaded or recalled from memory per list as a function of list type in Experiment 4. Error bars reflect the standard error of the mean.

To examine scores as a function of list type (see Figure 6), I conducted a repeated-measures ANOVA with 3 levels (list type: worth 8, 5, 2). Results revealed differences between list types [Mauchly's $W = .74$, $p < .001$; Huynh-Feldt corrected results: $F(1.60, 215.79) = 125.96$, $p < .001$, $\eta_p^2 = .48$] such that participants' scores were higher when offloaded words were worth 8 points ($M = 134.54$, $SD = 40.10$) than when they were worth 5 points ($M = 105.72$, $SD = 35.90$), [$p_{\text{holm}} < .001$, $d = .73$] and when they were worth 2 points ($M = 89.98$, $SD = 42.17$), [$p_{\text{holm}} < .001$, $d = 1.13$]; participants also scored higher when offloaded words were worth 5 points than when they were worth 2 points [$p_{\text{holm}} < .001$, $d = .40$].

I also examined how participants' scores compared with the scores they could have obtained if they offloaded every word. On lists where offloaded words were worth 8 points, a one-sample *t*-test indicated that participants' scores were lower than what they could have obtained if they offloaded every word (160), [$t(135) = -7.40$, $p < .001$, $d = -.64$]. On lists where offloaded words were worth 5 points, a one-sample *t*-test indicated that participants' scores were similar to what they could have obtained if they offloaded every word (100), [$t(135) = 1.86$, $p = .065$, $d = .16$]. On lists where offloaded words were worth 2 points, a one-sample *t*-test indicated that participants' scores were greater than what they could have obtained if they offloaded every word (40), [$t(135) = 13.82$, $p < .001$, $d = 1.19$].

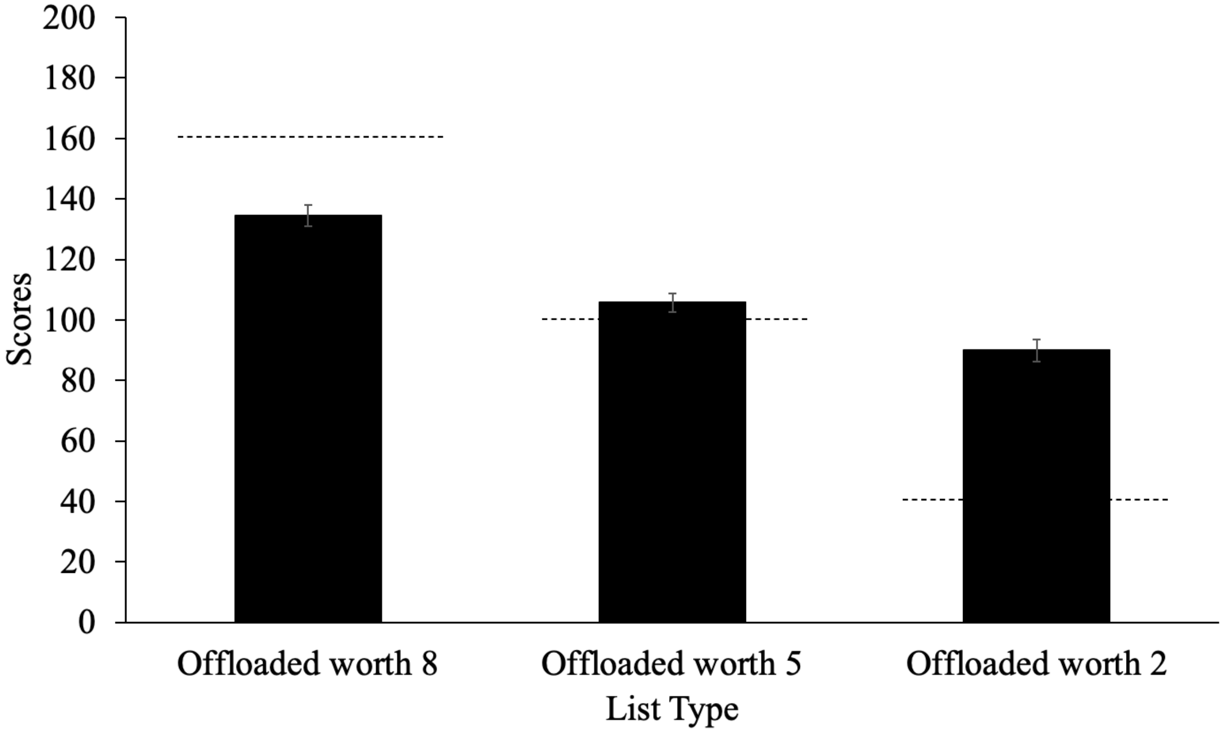


Figure 6. Average scores as a function of list type in Experiment 4. Dashed lines reflect the score participants would have obtained if they offloaded every word. Error bars reflect the standard error of the mean.

Table 1.

Correlations between baseline recall, the absolute value of participants' calibration scores (which reflect the accuracy of participants' metacognition whether they were over- or under-confident), the number of words offloaded per list for each list type, and participants' optimal score – their actual score for each list type in Experiment 4.

Measure	1	2	3	4	5	6	7	8
1) Baseline Recall	—							
2) Absolute Value of Calibration	.432***	—						
3) Offloaded Words Per List (Offloaded Worth 8)	-.182*	-.196*	—					
4) Offloaded Words Per List (Offloaded Worth 5)	-.203*	-.172*	.708***	—				
5) Offloaded Words Per List (Offloaded Worth 2)	-.170*	-.113	.542***	.721***	—			
6) Optimal Score – Actual (Offloaded Worth 8)	-.549***	-.098	-.468***	-.347***	-.205*	—		
7) Optimal Score – Actual (Offloaded Worth 5)	-.643***	-.254**	.039	-.006	.040	.683***	—	
8) Optimal Score – Actual (Offloaded Worth 2)	-.721***	-.313***	.286***	.220*	.266**	.500***	.761***	—

Note: * = $p < .05$, ** = $p < .01$, *** = $p < .001$

In the next section, I use each participant's baseline recall and metacognitive calibration (predicted recall – actual recall) to compute their optimal scores. Specifically, using the first three lists, I computed the mean number of words recalled for each participant as well as their calibration score. Correlations between baseline recall, the absolute value of participants' calibration scores (which reflect the accuracy of participants' metacognition whether they were over- or under-confident), the number of words offloaded per list for each list type, and participants' optimal score – their actual score for each list type are shown in Table 1. To narrow in on a few key findings, 1) participants' baseline recall was negatively correlated with how far off participants were from their max score on all three list types such that the greater a given participant's recall ability, the more they underperformed relative to their optimal score, 2) baseline recall was negatively related to the number of words offloaded on all three list types such that the better a given participant's baseline recall, the less they tended to offload, and 3) calibration predicts how far participants are from their optimal score such that the better a given participant's calibration, the closer they were to their optimal score on lists where offloaded words were worth 5 points and lists where offloaded words were worth 2 points.

Discussion

In Experiment 4, I examined the relationship between individuals' memory and optimal offloading. Specifically, I measured how good each participant's memory was and how good they thought it was. By looking at participants' performance on baseline lists, I computed how many items people should choose to offload and what their maximal score would be if they have perfect metacognition. Critically, I examined the relationship between participants' maximum scores based on their baseline recall ability, their metacognitive calibration, and performance on each of the list scoring types. Overall, the results indicated that 1) the greater a participant's memory, the farther they were from their optimal score, 2) participants with better memory offloaded less, and 3) the better a participant's metacognitive calibration, the closer they were to their optimal score.

I demonstrated that as the price for offloading goes down, the criterion for offloading shifts toward offloading more items. Additionally, when the cost of offloading was high, participants performed better than they would have if they simply offloaded every word. However, when the cost of offloading was low, participants performed worse than they would have if they offloaded every word. This indicates that participants did not effectively monitor their learning by identifying which words would be remembered and forgotten. This may be a form of overconfidence in memory performance where participants did not offload items that they should have. For example, one could beat the score of 160 (the score one would achieve if all words were offloaded) when offloaded words were worth 8 points (again, words recalled from memory were worth 10 points) by offloading 19 words and focusing on a single word to recall from memory (if successful, this would yield a score of 162).

When offloaded words are only worth 2 points, one only needs to recall five words to beat the score when all words are offloaded (40). In contrast, when the cost of offloading is low (offloaded words are worth 8 points), participants would need to recall 17 words to beat the score when all words are offloaded (160). Participants' performance below 160 indicates a sub-optimal use of the low-cost offloading, indicating overconfidence in memory performance and some degree of inaccurate metacognitive mentoring.

If a learner were to offload a word that would later be remembered, this would be inefficient (see Nelson & Leonesio, 1988 for a similar instance involving the labor-in-vein effect). For example, if a word could be recalled from memory it would be worth 10 points but if it is offloaded, it would be worth fewer points in the present task. Thus, fewer points will be scored and there would be an opportunity cost whereby memory for the offloaded word may reduce the ability to remember a different, not-offloaded word. As a result, maximizing performance in the present task requires metacognitive insight from the learner to monitor what will be remembered and make control decisions in the form of offloading words that will not be remembered.

In addition to strategically offloading words that will not be remembered to maximize one's score, the optimal learner should prioritize the encoding of items that are easier to remember and offload items that are harder to remember. For example, of the 20 words presented on a given list, 10 were animals and 10 were manmade objects. Prior work has demonstrated that animate items are better remembered than inanimate items (e.g., Popp & Serra, 2016, 2018); thus, when using memory alone, participants' maximum number of items recalled should be higher if focusing on the animals compared with the manmade objects. As a result, if participants exclusively focus on the easier items to remember, this should increase the maximum number of items that can be recalled from memory which would subsequently increase their maximum potential score. In the present study, participants generally seemed to consider the relative difficulty of remembering each item such that the more difficult to remember, inanimate items were more likely to be offloaded. This indicates that learners used their metacognitive insight to inform offloading decision-making.

Since every learner has a different memory capacity and ability to monitor their learning, I included several lists at the beginning of the task where participants did not engage in any offloading to compute their baseline memory ability as well as metacognitive calibration. I then used these metrics for each participant to predict their performance in the present task. Some evidence suggests that people are more likely to offload when they believe they have a poor memory (Dunn & Risko, 2016; Hu et al., 2019). Accordingly, results revealed that participants who could recall more words tended to offload less, indicating that participants incorporated the knowledge of their own memory abilities in their propensity to offload. Similarly, participants who were more metacognitively calibrated generally achieved scores closer to their optimal score, indicating that the ability to monitor one's learning is critical in the successful implementation of strategic offloading. However, in the present paradigm, participants with greater memory capacities tended to be farther from their optimal scores and when the cost of offloading was low, participants tended to underperform relative to the scores that could have been obtained if participants relied solely on the external store. Thus, the present study demonstrates that learners may be overconfident in their memory abilities and underutilize external stores, particularly when the cost of offloading is low.

In sum, I demonstrated that participants use metacognitive insights regarding the difficulty of remembering information as well as their own memory abilities when making offloading decisions. Additionally, when the cost of offloading was relatively high, participants were able to use their metacognitive insight and memory to enhance task performance relative to relying solely on the external store. However, when the cost of offloading was low, participants underutilized the external store. Thus, I demonstrated that learners aimed to maximize the utility of the external store by 1) offloading items that would not be remembered and 2) offloading more difficult items while using memory for easier items, but when the cost of offloading is low, people may underutilize external stores.

Experiment 5

Metacognition and memory self-efficacy when offloading important information

Prior work suggests that metacognitive strategies influence offloading behavior (Boldt & Gilbert, 2019; Dunn & Risko, 2016; Gilbert et al., 2020; Grinschgl et al., 2021a). Additionally, there are individual differences in offloading tendencies (Meyerhoff et al., 2021; Morrison & Richmond, 2020) and learners are metacognitively aware of their memory selectivity (see Murphy et al., 2021a, 2021b). In Experiment 5, I was interested in how metacognitive monitoring abilities relate to a learner's ability to strategically use the external store and memory processes to maximize task performance.

Using an individual differences approach, I had participants complete a battery of cognitive tasks. Of particular interest, to assess metacognitive accuracy, I presented participants with lists of words followed by free recall tests for those words. After each word was presented, participants were asked to predict the likelihood of remembering it on the test (local JOLs). In another task, participants were presented with lists of words to remember for a later test, but each word was paired with a point value counting towards their score if recalled (and participants' job was to maximize their scores). Participants were allowed to offload some of the words on each list (they had access to these words on the test). I expected both absolute (average predictions – performance) and relative accuracy (i.e., Gamma correlations between local JOLs and recall) to predict selectivity scores such that participants with better metacognitive accuracy more effectively utilize the external store to maximize performance.

While metacognitive accuracy may be important for strategic offloading, potentially leading to better task performance, memory self-efficacy, or the self-assessment of one's memory abilities, may also impact offloading behavior. Memory self-efficacy generally positively affects memory performance (Bandura, 1989, 1997; Berry, 1999; see Beaudoin & Desrichard, 2011 for a review) and as a result, could influence how learners choose to offload information if the external store has a limited capacity. According to self-efficacy theory, there are four primary factors influencing memory self-efficacy: mastery (past memory successes and failures), modeling (the observation and implementation of the behaviors of other people), verbal persuasion or dissuasion (the overt feedback on memory capabilities from others), and level of physiological or psychological excitation or inhibition. According to these factors, researchers have suggested that heightened memory self-efficacy can relate to greater effort and motivation (Bandura, 1977), goal setting, and better task strategy (Locke et al., 1984). Since memory self-efficacy captures the belief in one's memory abilities, it may be a critical component of strategic offloading.

In addition to metacognitive accuracy and memory self-efficacy, I was also interested in other cognitive abilities which may be related to one's ability to effectively engage in offloading and selective memory such as working memory, fluid intelligence, the tendency to mind-wander, the ability to engage in directed forgetting, and retrieval tendencies.

Method

Participants. Participants were 348 undergraduate students (M_{age} = 20.08, SD_{age} = 2.36) recruited from the UCLA Human Subjects Pool and received course credit for their participation. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in six exclusions. I aimed to collect around 300 participants, consistent with other individual differences work (e.g., Unsworth et al., 2021).

Materials and Procedure. After signing informed consent, participants completed a set of questionnaires: demographics, the Memory Self-Efficacy Questionnaire (MSEQ) as well as the ADHD and the Dementia Worry scales. Next, all participants completed Raven's Progressive Matrices, the Remote Associates Task, operation span, a paired associate learning task with mind wandering probes, a free recall task with JOLs, a value-directed remembering betting task, a value-directed remembering task with offloading, offloading costs task, and a directed forgetting task. All tasks were administered in the order listed above.

Memory Self-Efficacy Questionnaire (MSEQ). Participants next completed the Memory Self-Efficacy Questionnaire for Items to assess participants' perception of their general memory abilities (MSEQ-I; Berry et al., 2013). The questionnaire consisted of 14 questions in which participants provided judgments of their ability to remember a list of items. They rated how confident they were that they could achieve a certain level of performance by selecting percentage responses ranging from 0 to 100% (in 10% increments). MSE scores were calculated by averaging participants' confidence ratings across all questions.

ADHD Worry Scale. Participants rate (on a 1-5 Likert Scale) how frequently 18 statements describe themselves over the past 6 months.

Dementia Worry Scale (DWS). Participants rate (on a 1-5 Likert Scale) 12 statements regarding Alzheimer's disease and related dementias in terms of how typical they are them.

Raven's Progressive Matrices (RPM). Participants completed the Raven's Progressive Matrices (Raven, 1938), a non-verbal test of abstract reasoning and fluid intelligence (e.g., Jarosz et al., 2019; Staff et al., 2014). The task is composed of 12 problems (of varying difficulty), each of which presents participants with a pattern that has a piece missing. Participants were instructed to select the option (out of eight choices) that correctly completes the pattern and then indicate their confidence in the accuracy of each response (from 0-100 with 0 being not confident and 100 being very confident). Participants were given a maximum of 2 minutes per problem and question difficulty was determined by Raven (1938) based on indices of participants' performance, such as mean response latency and accuracy rate. Participants completed the questions in random order, and fluid intelligence scores were calculated as the proportion correct.

Remote Associates Task (RAT). Participants were presented with three words along a theme. Their job was to generate a fourth word, which, when combined with each of the three presented words, results in word pairs that make up a common compound word or phrase (e.g., cottage/swiss/cake – cheese). Participants were given up to 15 seconds to answer each problem. Participants completed 25 problems taken from Bowden and Jung-Beeman (2003).

Operation Span. Participants solved arithmetic problems while trying to remember letters that appeared between those problems. All letters were from the pool (F,H,J,K,L,N,P,Q,R,S,T,Y), with 3 to 7 letters being presented in a given block. There were 10 blocks (two of each length, in random order) that were scored. In arithmetic problems, the first operation was always multiplication or division, and the second operation was always addition or subtraction (e.g., (8 x 8) – 1). Once participants had the answer, participants clicked the mouse/keypad to advance to a new screen displaying a number with "True" and "False" buttons below it—participants' task was to click "True" if the displayed number was the solution to the arithmetic and "False" if not. After responding, a letter appeared for 800 milliseconds; the task then advanced to the next arithmetic problem. If participants took too long on the math equation, the letter would appear, and the math problem would be counted as incorrect. Participants were instructed to keep their performance on the math problems above 85% while remembering as many letters as possible. Between blocks, a running average of math performance was shown (presented in green if at or above 85% and in red if below), and subjects were prompted to work more carefully on the equations if their performance dropped below 85%. For practice, participants began with 15 trials of only the math problems, with their performance in the practice used to set an upper limit on the time they could spend on each math problem in the rest of the task. They also had three short (two, two, and three letters) blocks of practice on the actual task before beginning the longer sets they would be evaluated on. Participants with math scores below 85% were not included in any analyses involving working memory.

Item-method directed forgetting. Participants were told that they would be presented with a list of words for a later test but that they only needed to remember some of them. Specifically, after each word was presented, a cue indicated whether the participant should try to remember the word (RRRR) or should try to forget the word (FFFF). Participants were presented with a total of 40 words (in random order) and for each participant, half of the words were randomly designated as to-be-remembered words, and half were designated as to-be-forgotten words. Each word, preceded by a 1-second fixation cross, was shown on the screen for 3 seconds, followed by a 1-

second inter-stimulus interval, and then the cue or value was displayed for 2 seconds. After the presentation of all 40 words, participants completed a 30-second distraction task requiring them to rearrange the digits of several three-digit numbers in descending order (e.g., 123 would be rearranged to 321). Participants were given 3 seconds to view each of the 10 three-digit numbers and subsequently rearrange the digits. Following the distractor task, participants completed a recognition test whereby they were shown the words from the studied list as well as 40 lures (in random order) and asked to indicate whether each item was on the list of studied items. Participants were given as much time as they needed for this portion of the task.

Paired Associate Learning. Participants were told that they would be presented with lists of word pairs (e.g., table : fan) to remember for a test and that each pair would be shown one at a time, for 5 seconds each. Following the study phase, participants were presented with the left word from each pair (in random order) and were asked to recall its associate. Each list contained 30 word pairs and there were two study-test cycles. Word pairs were unrelated nouns that were randomly paired together. Throughout the task, participants were periodically asked about their level of attentiveness. Specifically, they were asked to characterize their current conscious experience. Participants indicated whether they were either a) totally focused on the current task (i.e., paying attention to the word pairs and trying to learn them), b) thinking about their performance on the task or how long it is taking (i.e., whether they will remember the information but also wondering whether the task will end soon), c) distracted by information present in the room (e.g., sights and sounds), d) zoning out/mind wandering (i.e., their attention has shifted away from the task to self-generated thoughts unrelated to the task), or e) "Other". Participants were given as much time as they needed to report their conscious experience. I embedded four mind-wandering probes on each list (a mind-wandering probe occurred after every sixth word pair).

Free Recall. Participants were presented with words (unrelated) to remember for a later test. Participants studied each word for 5 seconds and then were asked to judge the likelihood of later remembering it (i.e., a JOL; 0 (not at all likely) – 100 (very likely)). After the presentation of all words (15 per list), participants completed a 30-second distraction task requiring them to rearrange the digits of several three-digit numbers in descending order (e.g., 123 would be rearranged to 321). Participants were given 3 seconds to view each of the 10 three-digit numbers and subsequently rearrange the digits. Following the distractor task, participants were asked to recall all the words they could remember from the just-studied list. Participants were given as much time to recall the words as they needed and were told how many words they were called correctly after each list. This was repeated for a total of three study-test cycles.

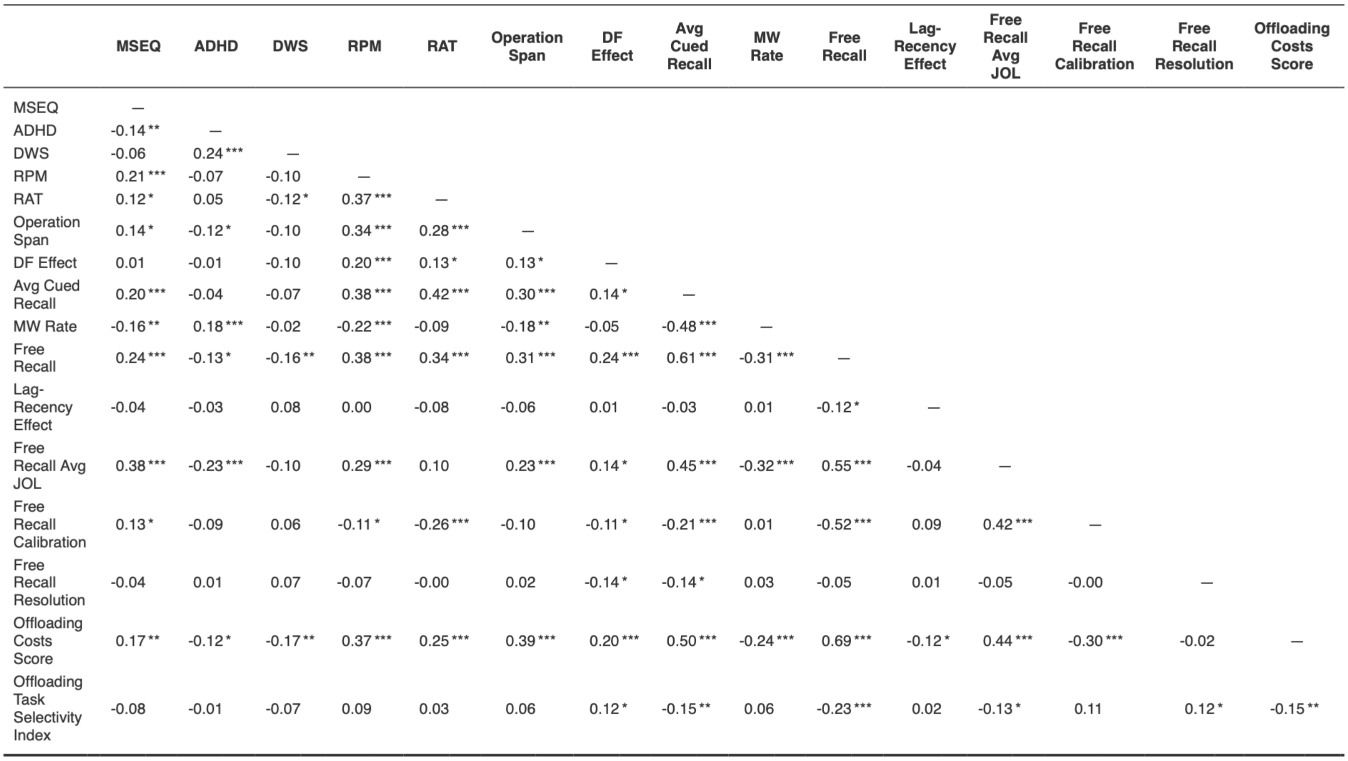
Offloading Costs. Participants were presented with three lists of unique to-be-remembered words. The stimulus words were presented for 5 seconds each. Participants were told that each word was worth 10 points, that they earned points by recalling the words on the test, and that they should try to maximize their scores. On each list, participants were allowed to offload any word of their choice and could offload as many words as they liked. To offload a word, participants clicked a button labeled "Save current word" and offloaded words were provided to participants on the recall test. Additionally, participants were told that if they saved a word, it would be worth fewer points—offloaded words were only worth 5 points. After the presentation of all 15 words in each list, participants completed a 30-second distraction task requiring them to rearrange the digits of several three-digit numbers in descending order (e.g., 123 would be rearranged to 321). Participants were given 3 seconds to view each of the 10 three-digit numbers and subsequently rearrange the digits. After the distractor task, participants were given a self-paced free recall test in which they had to recall as many words as they could from the list. Immediately following the recall period, participants were told their score for that list but were not given feedback about specific items.

Value-Directed Remembering—Offloading. Participants completed a similar value-directed remembering task where they were presented with lists of words paired with point values counting towards their score if recalled (but there was no betting). Each word was displayed for 5 seconds, and point values ranged from 1 to 15. Additionally, participants were allowed to offload five words of their choosing by clicking an on-screen button for words they wanted to save. Following the study phase, participants completed a 30-second distraction task requiring them to rearrange the digits of several three-digit numbers in descending order (e.g., 123 would be rearranged to 321). Participants were given 3 seconds to view each of the 10 three-digit numbers and subsequently rearrange the digits. After the distractor task, participants were asked to recall the words from the just-studied list and were provided with the words they saved during the study phase. Following the self-paced test phase, participants were told their scores but were not given feedback about specific items. This was repeated for three study test cycles.

Results and Discussion

Table 2 displays Pearson correlations between all collected measures. Regarding memory self-efficacy, consistent with prior work, memory self-efficacy was positively related to memory performance, and higher memory self-efficacy individuals reported less mind-wandering. People with high memory self-efficacy were more confident in their memory performance and subsequently better calibrated (but MSEQ scores did not significantly correlate with resolution). Memory self-efficacy was not related to selectivity in the value-directed remembering with offloading task but positively predicted scores on the costs of offloading task. This may arise due to higher memory self-efficacy individuals being able recall more words and better metacognitively calibrated, allowing them to minimize the costs of using an external store to remember information.

Table 2. Correlations between all collected measures in Experiment 5.



	MSEQ	ADHD	DWS	RPM	RAT	Operation Span	DF Effect	Avg Cued Recall	MW Rate	Free Recall	Lag-Recency Effect	Free Recall Avg JOL	Free Recall Calibration	Free Recall Resolution	Offloading Costs Score
MSEQ	—														
ADHD	-0.14 **	—													
DWS	-0.06	0.24 ***	—												
RPM	0.21 ***	-0.07	-0.10	—											
RAT	0.12 *	0.05	-0.12 *	0.37 ***	—										
Operation Span	0.14 *	-0.12 *	-0.10	0.34 ***	0.28 ***	—									
DF Effect	0.01	-0.01	-0.10	0.20 ***	0.13 *	0.13 *	—								
Avg Cued Recall	0.20 ***	-0.04	-0.07	0.38 ***	0.42 ***	0.30 ***	0.14 *	—							
MW Rate	-0.16 **	0.18 ***	-0.02	-0.22 ***	-0.09	-0.18 **	-0.05	-0.48 ***	—						
Free Recall	0.24 ***	-0.13 *	-0.16 **	0.38 ***	0.34 ***	0.31 ***	0.24 ***	0.61 ***	-0.31 ***	—					
Lag-Recency Effect	-0.04	-0.03	0.08	0.00	-0.08	-0.06	0.01	-0.03	0.01	-0.12 *	—				
Free Recall Avg JOL	0.38 ***	-0.23 ***	-0.10	0.29 ***	0.10	0.23 ***	0.14 *	0.45 ***	-0.32 ***	0.55 ***	-0.04	—			
Free Recall Calibration	0.13 *	-0.09	0.06	-0.11 *	-0.26 ***	-0.10	-0.11 *	-0.21 ***	0.01	-0.52 ***	0.09	0.42 ***	—		
Free Recall Resolution	-0.04	0.01	0.07	-0.07	-0.00	0.02	-0.14 *	-0.14 *	0.03	-0.05	0.01	-0.05	-0.00	—	
Offloading Costs Score	0.17 **	-0.12 *	-0.17 **	0.37 ***	0.25 ***	0.39 ***	0.20 ***	0.50 ***	-0.24 ***	0.69 ***	-0.12 *	0.44 ***	-0.30 ***	-0.02	—
Offloading Task Selectivity Index	-0.08	-0.01	-0.07	0.09	0.03	0.06	0.12 *	-0.15 **	0.06	-0.23 ***	0.02	-0.13 *	0.11	0.12 *	-0.15 **

Note. * p < .05, ** p < .01, *** p < .001

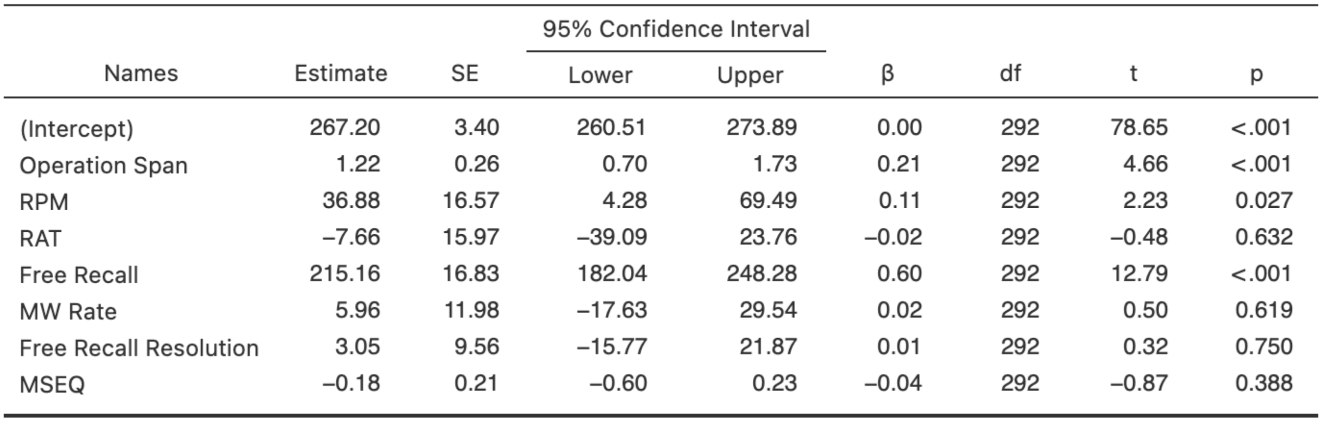
Next, I was interested in the role of fluid intelligence (as measured by the Raven's Progressive Matrices) in memory abilities. Fluid intelligence scores were highly related to creativity scores (RAT) as well as working memory (Operation Span). Participants with higher fluid intelligence were better able to engage in directed forgetting. Fluid intelligence was positively correlated with free and cued recall, perhaps due to reduced mind wandering. Although not predictive of selectivity in the value-directed offloading task, participants with higher fluid intelligence performed better on the costs of offloading task than lower fluid intelligence individuals.

I was also interested in the role of creativity (as measured by the Remote Associates Task) in memory abilities. Greater creativity was associated with better working memory capacity, the ability to engage in directed forgetting, associative memory, free recall, more under-confidence in memory, and greater scores on the costs of offloading task.

As measured by an Operation Span task, working memory capacity positively correlated with directed forgetting, cued and free recall, JOLs, and scores on the costs of offloading task. However, working memory capacity was negatively related to mind-wandering such that mind wandering was more frequent in high working memory people.

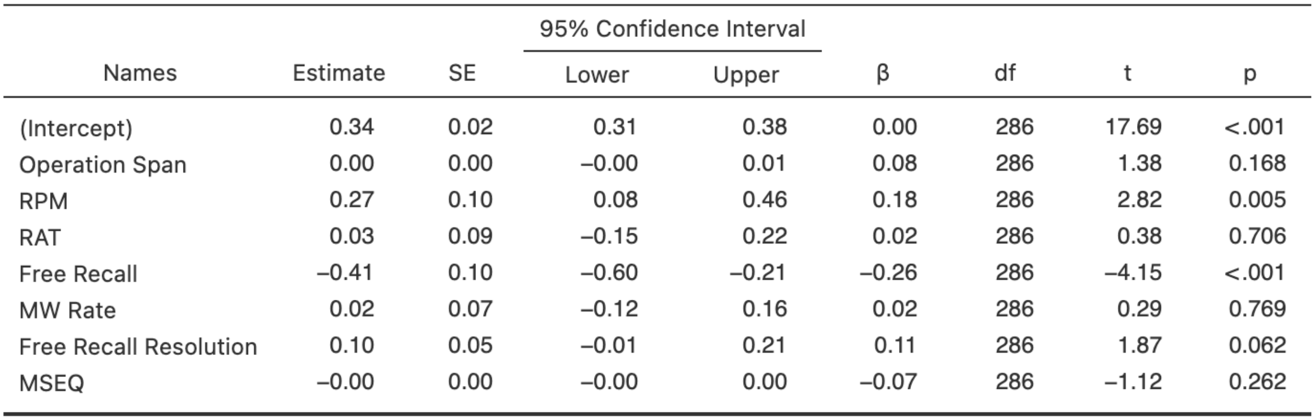
To isolate the effects of each variable, I conducted general linear models with working memory capacity (operation span), fluid intelligence (RPM), creativity (RAT), free recall ability, mind-wandering rate, metacognitive accuracy (as measured by resolution), and memory self-efficacy (MSE) predicting scores on the costs of offloading task and selectivity scores on the value-directed remembering with offloading task. The results of these models are shown in Tables 3 and 4. As can be seen, when controlling for other factors, working memory capacity, fluid intelligence, and free recall abilities positively predict performance. Specifically, learners with a greater working memory capacity and higher fluid intelligence who can remember more items are better able to maximize the recall of items from memory as well as score points for offloading items that are likely to be forgotten. In terms of selective offloading/memory, participants with greater fluid intelligence were better able to strategically use the external store plus their own memory to recall high-value words.

Table 3. The results of a general linear model with working memory capacity (operation span), fluid intelligence (RPM), creativity (RAT), free recall ability, mind-wandering rate, metacognitive accuracy (as measured by resolution), and memory self-efficacy (MSE) predicting scores on the costs of offloading task.



			95% Confidence Interval					
Names	Estimate	SE	Lower	Upper	β	df	t	p
(Intercept)	267.20	3.40	260.51	273.89	0.00	292	78.65	<.001
Operation Span	1.22	0.26	0.70	1.73	0.21	292	4.66	<.001
RPM	36.88	16.57	4.28	69.49	0.11	292	2.23	0.027
RAT	–7.66	15.97	–39.09	23.76	–0.02	292	–0.48	0.632
Free Recall	215.16	16.83	182.04	248.28	0.60	292	12.79	<.001
MW Rate	5.96	11.98	–17.63	29.54	0.02	292	0.50	0.619
Free Recall Resolution	3.05	9.56	–15.77	21.87	0.01	292	0.32	0.750
MSEQ	–0.18	0.21	–0.60	0.23	–0.04	292	–0.87	0.388

Table 4. The results of a general linear model with working memory capacity (operation span), fluid intelligence (RPM), creativity (RAT), free recall ability, mind-wandering rate, metacognitive accuracy (as measured by resolution), and memory self-efficacy (MSE) predicting scores on the value-directed remembering with offloading task.



			95% Confidence Interval					
Names	Estimate	SE	Lower	Upper	β	df	t	p
(Intercept)	0.34	0.02	0.31	0.38	0.00	286	17.69	<.001
Operation Span	0.00	0.00	–0.00	0.01	0.08	286	1.38	0.168
RPM	0.27	0.10	0.08	0.46	0.18	286	2.82	0.005
RAT	0.03	0.09	–0.15	0.22	0.02	286	0.38	0.706
Free Recall	–0.41	0.10	–0.60	–0.21	–0.26	286	–4.15	<.001
MW Rate	0.02	0.07	–0.12	0.16	0.02	286	0.29	0.769
Free Recall Resolution	0.10	0.05	–0.01	0.21	0.11	286	1.87	0.062
MSEQ	–0.00	0.00	–0.00	0.00	–0.07	286	–1.12	0.262

CHAPTER 3:

THE BENEFITS OF CHOICE

Predicting what we will remember and what we will forget is crucial for daily functioning. For example, when meeting someone, knowing that you will remember their face but not their name can help you focus on what will soon be forgotten and what needs to be remembered. For instance, if you know that you are likely to forget someone's name, you can prioritize memory for this information to avoid the potential consequences (e.g., social awkwardness) of forgetting. Likewise, if you know what is likely to be forgotten, you can prioritize memory for other information that is more likely to be remembered. Thus, evaluating what we know and what we do not know may be critical in optimizing memory, but it is unclear how these processes affect memory for information that is likely to be remembered and information that people feel will soon be forgotten.

Metamemory judgments often take the form of judgments of the likelihood of remembering (but see Finn, 2008; Li et al., 2021; for judgments of the likelihood of forgetting; Tauber & Rhodes, 2012, for estimates of the duration of retention), and making metamemory judgments can influence memory, an effect known as reactivity (cf. Arbuckle & Cuddy, 1969; Double & Birney, 2019; Double et al., 2018; Mitchum et al., 2016; Rivers et al., 2021; Soderstrom et al., 2015; Spellman & Bjork, 1992; Tekin & Roediger, 2020; Witherby & Tauber, 2017). Specifically, reactivity occurs when making metacognitive judgments while studying to-be-remembered information influences which or how much information is remembered (but this effect may be small and differ based on how memory is tested, see Myers et al., 2020). Thus, evaluating the memorability of information may change what is remembered, although it is unclear if learners are aware of how metacognitive judgments influence remembering and forgetting.

Experiment 6

Evaluating memorability can lead to remembering

I was interested in whether evaluating information as likely to be remembered or likely to be forgotten leads to enhanced memory for this information relative to information not subject to memory predictions. I presented participants with lists of words to remember for a later test and, on each list, participants identified words that they were confident that they would remember and words that they believed they were most likely to forget on a test. Since prior work illustrates that positive reactivity occurs for easier related pairs but negative reactivity occurs for more difficult pairs (but sometimes reversed; see Ericsson & Simon, 1980; Fox et al., 2011), in Experiment 6, if participants' predictions are accurate, then prior evidence suggests that the items predicted to be forgotten (the harder items, presumably) would show negative reactivity.

However, I expected words that participants indicated they were most likely to forget to be better recalled than words participants did not indicate as likely to be remembered or forgotten, potentially due to the benefits of selection (see Coverdale & Nairne, 2019; DuBrow et al., 2019; Gureckis & Markant, 2012; Markant et al., 2014; Markant & Gureckis, 2014; Rotem-Turchinski et al., 2019). This memory advantage for words strongly anticipated to be forgotten would suggest, paradoxically, that the act of judging information as likely to be forgotten may unexpectedly influence later memory for this information. Additionally, I expected participants to demonstrate elevated recall of words they indicated that they were most likely to remember relative to words they indicated that they were most likely to forget and words not given memory predictions.

Method

Participants. After exclusions, participants were 62 undergraduate students (M_{age} = 20.84, SD_{age} = 4.39) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in two exclusions.

Materials and Procedure. Participants were presented with lists of words to remember for a later test with each list containing 16 words. Words were presented simultaneously in two columns with eight words in each column. On each list, participants studied the words for as long as they liked and were asked to underline (by clicking on the word once) two words that they were confident that they would remember and circle (by clicking on the word twice) two words that they thought that they were most likely to forget on the test (with underlining and circling counterbalanced between-subjects). If participants clicked on a word a third time, it was no longer underlined or circled. Following the study phase, participants completed a 1-minute immediate free recall test whereby they typed all the words they could remember from the just-studied list into an on-screen text box. This was repeated for four study-test cycles.

Results

On average, participants spent 97.16 seconds studying each list ($SD = 80.13$). Recall as function of participants' predictions is shown in Figure 7. A repeated-measures ANOVA with 3 levels (predictions: forget, neither, remember) revealed a main effect of predictions [$F(2, 122) = 54.54$, $p < .001$, $\eta_p^2 = .47$] such that words participants said they would remember ($M = .90$, $SD = .19$) were better recalled than words predicted as to be forgotten ($M = .63$, $SD = .26$), [$p_{\text{holm}} < .001$, $d = .85$] and the words not predicted to be remembered or forgotten ($M = .49$, $SD = .28$), [$p_{\text{holm}} < .001$, $d = 1.31$]. Critically, recall was better for the words expected to be forgotten than words not selected as to be remembered or forgotten [$p_{\text{holm}} < .001$, $d = .45$]. Thus, predicting that a word would be forgotten made it more memorable.

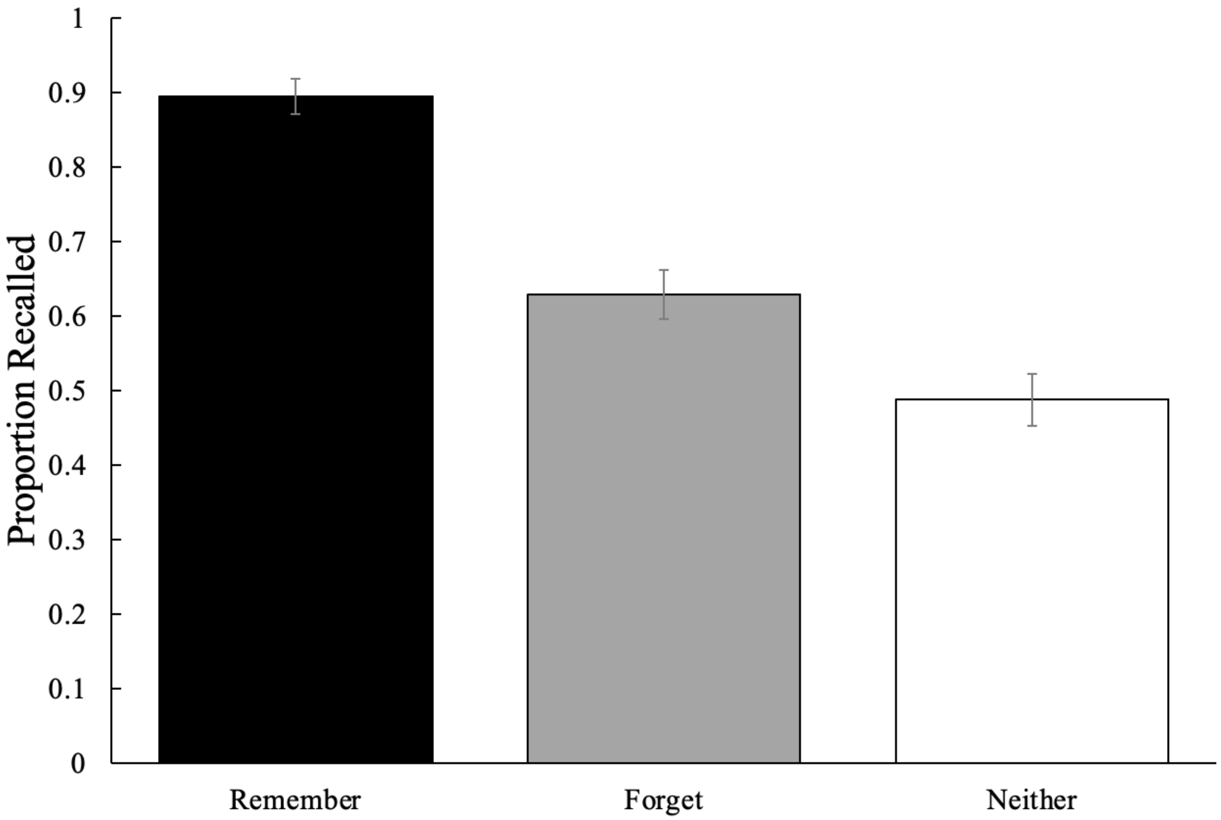


Figure 7. Average proportion recalled as a function of participants' predictions about remembering and forgetting in Experiment 6. Error bars reflect the standard error of the mean.

I also examined word frequency for remember words, forget words, and words not predicted as to be remembered or forgotten as an exploratory analysis. A repeated-measures ANOVA with 3 levels (predictions: forget, neither, remember) revealed a main effect of predictions [$F(2, 122) = 3.95$, $p = .022$, $\eta_p^2 = .06$] such that the words participants predicted that they would forget ($M = 8.52$, $SD = .75$) were less frequent than words not predicted to be remembered or forgotten ($M = 8.84$, $SD = .22$), [$p_{\text{holm}} = .025$, $d = -.34$], but not less frequent than the words predicted to be remembered ($M = 8.77$, $SD = .66$), [$p_{\text{holm}} = .082$, $d = -.26$]. Additionally, frequency for words predicted to be remembered was similar for words not given a prediction [$p_{\text{holm}} = .539$, $d = .08$]. Thus, to some extent, participants may have incorporated word characteristics that influence memorability into their decisions.

Finally, I examined participants' output by calculating the average output position of words participants indicated they would remember, forget, or did not select as to be remembered or forgotten. A repeated-measures ANOVA with 3 levels (predictions: forget, neither, remember) revealed a main effect of predictions [$F(2, 118) = 54.38$, $p < .001$, $\eta_p^2 = .48$] such that the words participants said they would remember ($M = 3.35$, $SD = 1.87$) were recalled earlier than words predicted as to be forgotten ($M = 6.01$, $SD = 2.39$), [$p_{\text{holm}} < .001$, $d = 1.22$] and words not predicted to be remembered or forgotten ($M = 5.76$, $SD = 1.73$), [$p_{\text{holm}} < .001$, $d = 1.11$]. However, the average output position for words predicted to be forgotten was similar for words not given a prediction [$p_{\text{holm}} = .388$, $d = .11$]. Thus, participants generally recalled to-be-remembered words before other words.

Discussion

Participants better recalled words they predicted would be remembered but also demonstrated enhanced recall for words they predicted would be forgotten. Guided by word frequency, participants may attend to words that they decide are both memorable and likely to be forgotten. However, these effects may be driven by the potential benefits of selecting a subset of words from a list, making selected words inherently more likely to be remembered.

Experiment 7

Selecting words can make them more memorable

Prior research has shown that when participants make choices about when and what to learn, memory for the chosen information is often enhanced, leading to a "choice effect" (Coverdale & Nairne, 2019; DuBrow et al., 2019; Gureckis & Markant, 2012; Markant et al., 2014; Markant & Gureckis, 2014; Rotem-Turchinski et al., 2019). For example, letting participants select cues or targets during paired-associate learning can improve cued recall (e.g., Monty & Permuter, 1975; Perlmuter et al., 1971; see also Watanabe & Soraci, 2004). Additionally, allowing participants to make decisions regarding aspects of learning such as presentation order or duration can benefit memory (Markant et al., 2014; Murty et al., 2015, 2019; Voss et al., 2011). Thus, these choice effects may be driving the enhanced memory for items selected as likely to be remembered or forgotten.

In Experiment 6, I again presented participants with lists of words to remember for later tests and asked participants to circle two of the words and underline two of the words (via mouse clicks). However, participants were not provided instructions regarding why or how to select which words to underline and which words to circle. It may be that circling and underlining words does not impact memory for those words since participants were not asked to consider memorability. However, if selecting words (i.e., the act of choosing words without instructions) elevates attention and enhances encoding, then these processes may impact memory. Thus, Experiment 7 permitted us to determine whether choice only conferred memory benefits in the absence of any direct instructions regarding predictions about what words would be later remembered or forgotten.

Method

Participants. After exclusions, participants were 102 undergraduate students (M_{age} = 20.58, SD_{age} = 3.16) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in four exclusions. For participants self-pacing the study phase, participants were excluded if they circled more than two words or underlined more than two words. This exclusion process resulted in 10 exclusions.

Materials and Procedure. Participants were presented with lists of words to remember for a later test with each list containing 16 words. Words were presented simultaneously in two columns with eight words in each column. However, rather than circling and underlining words to indicate whether they would be remembered or forgotten, participants were asked to circle two words (by clicking on the word twice) and underline two words (by clicking on the word once); they were not given any instructions regarding the criteria for which words to circle or underline. If participants clicked on a word a third time, it was no longer underlined or circled. Study time on each list was either fixed (30 seconds; $n = 47$) or self-paced ($n = 55$). Following the study phase, participants completed a 1-minute immediate free recall test whereby they typed all the words they could remember from the just-studied list into an on-screen text box. This was repeated for four study-test cycles.

Results

On average, participants self-pacing their study time spent 85.18 seconds studying each list ($SD = 60.07$). Recall for words participants clicked on (either circled or underlined) as a function of whether study time was fixed or self-paced is shown in Figure 8. A 2 (action: clicked, not clicked) x 2 (study condition: fixed, self-paced) mixed-factor ANOVA revealed that words that were either circled or underlined ($M = .78$, $SD = .20$) were better recalled than words that were not clicked on ($M = .42$, $SD = .24$), [$F(1, 99) = 252.66$, $p < .001$, $\eta_p^2 = .72$]. Additionally, participants self-pacing their study time ($M = .57$, $SD = .24$) recalled more words than participants with fixed study time ($M = .41$, $SD = .11$), [$F(1, 99) = 13.13$, $p < .001$, $\eta_p^2 = .12$]. However, clicking did not interact with study condition [$F(1, 99) = 2.20$, $p = .142$, $\eta_p^2 = .02$]. A further analysis of memory for clicked words revealed that circled words ($M = .81$, $SD = .23$) were better recalled than underlined words ($M = .75$, $SD = .25$), [$t(97) = 2.16$, $p = .033$, $d = .22$], perhaps occurring because participants clicked an item twice to circle it but only once to underline it. Future work with a counterbalanced design would be necessary to test this conjuncture.

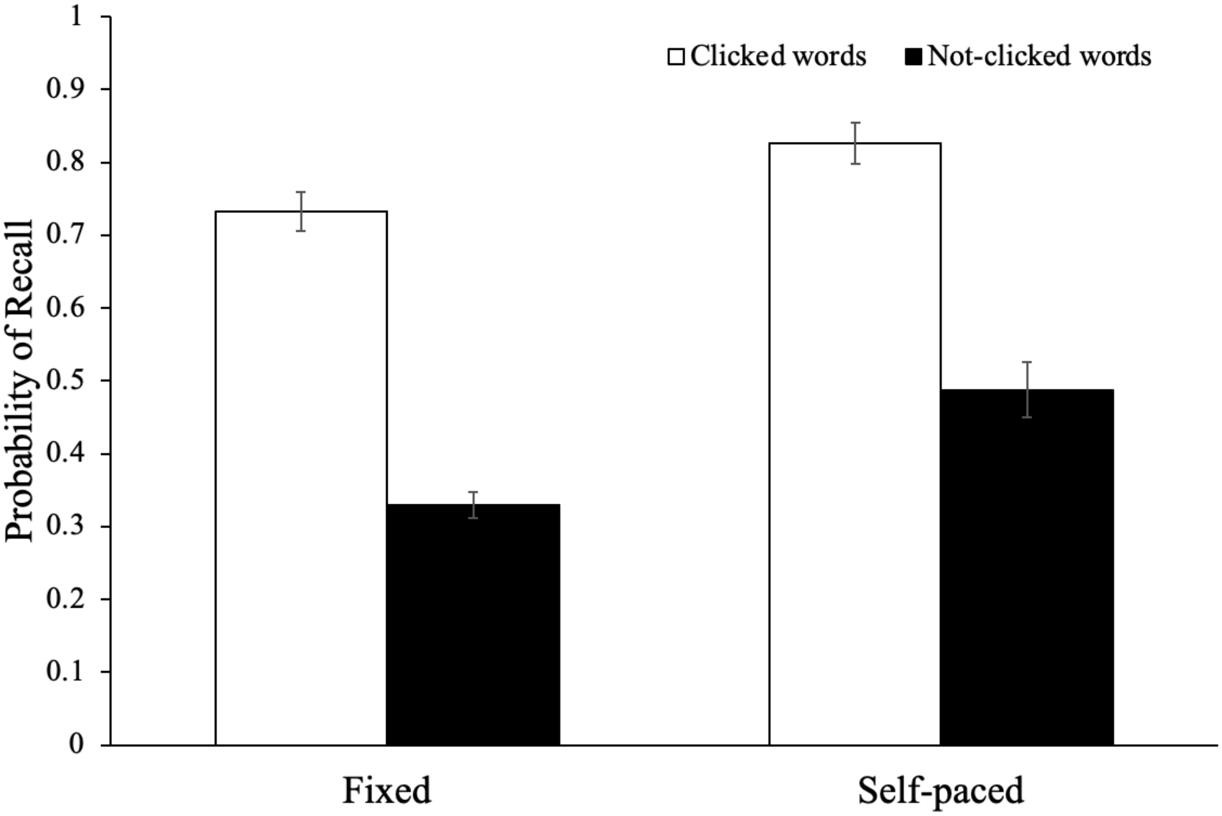


Figure 8. Average proportion recalled for words participants clicked on (either circled or underlined) as a function of whether study time was fixed or self-paced in Experiment 7. Error bars reflect the standard error of the mean.

I again examined word frequency for words participants either circled or underlined compared with words participants did not click on. A 2 (action: clicked, not clicked) x 2 (study condition: fixed, self-paced) mixed-factor ANOVA revealed that words that were either circled or underlined (M = 8.78, SD = .55) were similarly frequent as words that were not clicked on (M = 8.84, SD = .24), [$F(1, 99) = .62$, $p = .434$, $\eta_p^2 = .01$]. Additionally, word frequencies were similar for participants self-pacing their study time and participants with fixed study time, [$F(1, 99) =$

1.42, $p = .237$, $\eta_p^2 = .01$]. Moreover, clicking did not interact with study condition [$F(1, 99) = .44$, $p = .509$, $\eta_p^2 < .01$].

Finally, I examined participants' output by calculating the average output position of words participants either circled or underlined or did not click on. A 2 (action: clicked, not clicked) x 2 (study condition: fixed, self-paced) mixed-factor ANOVA revealed that words that were either circled or underlined ($M = 4.25$, $SD = 1.80$) were recalled earlier than words that were not clicked on ($M = 5.22$, $SD = 1.78$), [$F(1, 97) = 21.88$, $p < .001$, $\eta_p^2 = .18$]. Additionally, participants self-pacing their study time ($M = 5.37$, $SD = 1.79$) had a larger average output position (because they recalled more words) than participants with fixed study time ($M = 4.06$, $SD = .88$), [$F(1, 97) = 19.92$, $p < .001$, $\eta_p^2 = .17$]. However, clicking did not interact with study condition [$F(1, 97) = 3.33$, $p = .071$, $\eta_p^2 = .03$].

Discussion

In Experiment 6, participants better recalled words they selected (circled or underlined via mouse clicks) relative to the words that were not selected, consistent with the choice effect found in self-regulated learning contexts (Gureckis & Markant, 2012; Markant et al., 2014; Markant & Gureckis, 2014). Specifically, by selecting a word, participants' attention is likely drawn towards that word, leading to better memorability. Thus, drawing attention to some subset of words during encoding via the selection process can confer a memory benefit, as was observed in Experiment 6. Accordingly, the selection process may influence memory in the absence of any direct instructions regarding predictions about what words would be later remembered or forgotten.

Experiment 8

Does choosing what to highlight enhance memory?

When reading textbook chapters or other assigned readings for their courses, students often highlight portions of the text. Among college students, highlighting is a common strategy (Bell & Limber, 2010) but students likely differ in their strategy when highlighting. For example, some students may highlight the information they think is important to remember so they can later review this information (Hartwig & Dunlosky, 2012; Kornell & Bjork, 2007). In contrast, other students may use highlighting as a technique to remain engaged with the concepts and terminology without the intent of restudying the highlighted information. This approach may be more realistic as students may infrequently return to their course readings, raising the question as to whether highlighting (absent later reviewing the text) helps learning.

Compared with passive reading, prior work has suggested that highlighting or underlining can help learning (e.g., Annis & Davis, 1978; Fass & Schumacher, 1978; Fowler & Barker, 1974; Johnson, 1988; Leutner et al., 2007; Nist & Hogrebe, 1987; Nist & Simpson, 1988; Rickards & August, 1975; Yue et al., 2015). However, there is also evidence that has not shown a significant advantage for underlined or highlighted text (e.g., Dunlosky et al., 2013; Hoon, 1974; Idstein & Jenkins, 1972; Peterson, 1992; Stordahl & Christensen, 1956; Wade & Trathen, 1989). Additionally, pre-existing, inappropriate highlighting can impair learning (Silvers & Kreiner, 1997) and previous research suggests that highlighting is only beneficial for printed text, not digital text (Ben-Yehudah & Eshet-Alkalai, 2018). Thus, there is some disagreement about the impact of highlighting as well as the mechanisms that drive the potential benefits of highlighting.

Some have argued that the potential memory benefits from highlighting result from the von Restorff effect (see Johnson, 1988; Nist & Hogrebe, 1987; Wallace, 1965) whereby information that is more distinct or stands out compared with its neighbors (and perhaps the information likely to be highlighted) is better remembered (see Fowler & Barker, 1974; Silvers & Kreiner, 1997). However, rather than or perhaps in addition to the qualitative differences between what is and is not highlighted, choosing what information to highlight may contribute to any memory benefit from highlighting.

Active highlighting, rather than reading pre-highlighted text (not highlighted by the learner) can enhance memory (Fowler & Barker, 1974; Rickards & August, 1975; Rickards & Denner, 1979; but see Nist & Hogrebe, 1987), and when learners are trained in highlighting techniques, it may be more effective (see Leutner et al., 2007). Thus, the benefits of highlighting may arise as a result of deciding what to highlight, potentially engaging deeper levels of processing (Craik & Lockhart, 1972). Additionally, choosing what to highlight may require learners to think about the value or importance of information which can enhance memory for this information (see Murphy & Castel, 2021a, 2021b; Murphy et al., 2023). Furthermore, prior work has demonstrated that when learners must decide when and what to learn, memory for selected information is improved, illustrating a "choice effect" (Gureckis & Markant, 2012; Markant et al., 2014; Markant & Gureckis, 2014). Applied to highlighting, the benefits of highlighting may arise as a result of both the act of highlighting and choosing what information to highlight.

In Experiment 8, I was interested in the impact of highlighting on comprehension. Additionally, I wanted to examine whether there are ways highlighting can enhance memory. Specifically, I presented participants with passages to read in preparation for a subsequent test and either had participants read the passages, read and use a single highlighter, read and use three highlighters as they see fit, or read and use three highlighters with instructions regarding what to highlight (participants were told to highlight words, concepts, and ideas that they are likely to remember in green, words, concepts, and ideas they might remember in yellow, and words, concepts, and ideas they do not expect to remember in pink). I expect that highlighting will benefit test performance, particularly when using multiple colors. Additionally, I expected participants using three highlighters and instructions on how to use them to perform best as making decisions regarding what to highlight may benefit memory.

Method

Participants. Participants were 218 undergraduate students (M_{age} = 20.34, SD_{age} = 2.91) recruited from the UCLA Human Subjects Pool. Participants were tested in person and received course credit for their participation.

Materials and Procedure. Participants were told that they would be asked to read several passages and then answer questions based on those passages. Passages were taken from Little (2011), and each covered a single topic: the solar system or ferrets (topic order was counterbalanced). Passages were printed on standard printer paper. Participants were given 15 minutes to read the first passage. Participants then were asked to predict the number of questions (out of 32) that they expected to get correct and answered the 32 multiple-choice questions (each with three options) based on the just-read passage. Participants were then given the second passage (15 minutes), predicted the number of questions (out of 32) that they expected to get correct, and answered 32 multiple choice questions based on the just-read passage. If participants finished reading the passage early, they were told to continue reviewing it until the time expired. Participants were given as much time as they needed to answer the questions.

When reading the passages, participants were divided into four groups: no highlighting (n = 56), one highlighter (n = 54), three highlighters with no instructions (n = 56), or three highlighters with instructions (n = 52). All participants were instructed to read the passages and told that they would then be tested on their memory for information in the passages. Participants given no highlighters were not given any additional instructions. Participants given a single highlighter (the highlighter was yellow) were also told to use the highlighter as they normally would (and if they do not normally highlight, to highlight in the way they believe would best help their test performance).

Of the participants given three highlighters (the highlighter colors were yellow, green, and pink), some were told to use the highlighters as they normally would (and if they do not normally highlight, to highlight in the way they believe would best help their test performance) but to make sure to use all three colors. However, other participants were told to highlight certain aspects of the passages in different colors. Specifically, participants were told to highlight words, concepts, and ideas that they are likely to remember in green, words, concepts, and ideas they might remember in yellow, and words, concepts, and ideas they do not expect to remember in pink.

Results

Predictions of performance and test performance as a function of passage topic for each of the four groups are shown in Figure 9. To examine differences in participants' predictions, I conducted a 2 (topic: ferrets, solar system) x 4 (group: no highlighting, one highlighter, three highlighters with no instructions, or three highlighters with instructions) mixed ANOVA. Results revealed a main effect of topic such that learners expected to answer a greater proportion of questions correct for the ferrets passage (M = .61, SD = .20) than the solar system passage (M = .56, SD = .21), [$F(1, 214) = 31.43$, $p < .001$, $\eta_p^2 = .13$]. However, there were no group differences in predictions [$F(3, 214) = .78$, $p = .509$, $\eta_p^2 = .01$] such that participants not highlighting (M = .60, SD = .18), with one highlighter (M = .61, SD = .19), with three highlighters and no instructions

(*M* = .57, *SD* = .21), and with three highlighters and instructions (*M* = .56, *SD* = .19) predicted similar performance. Topic did not interact with group [$F(3, 214) = .90$, $p = .443$, $\eta_p^2 = .01$].

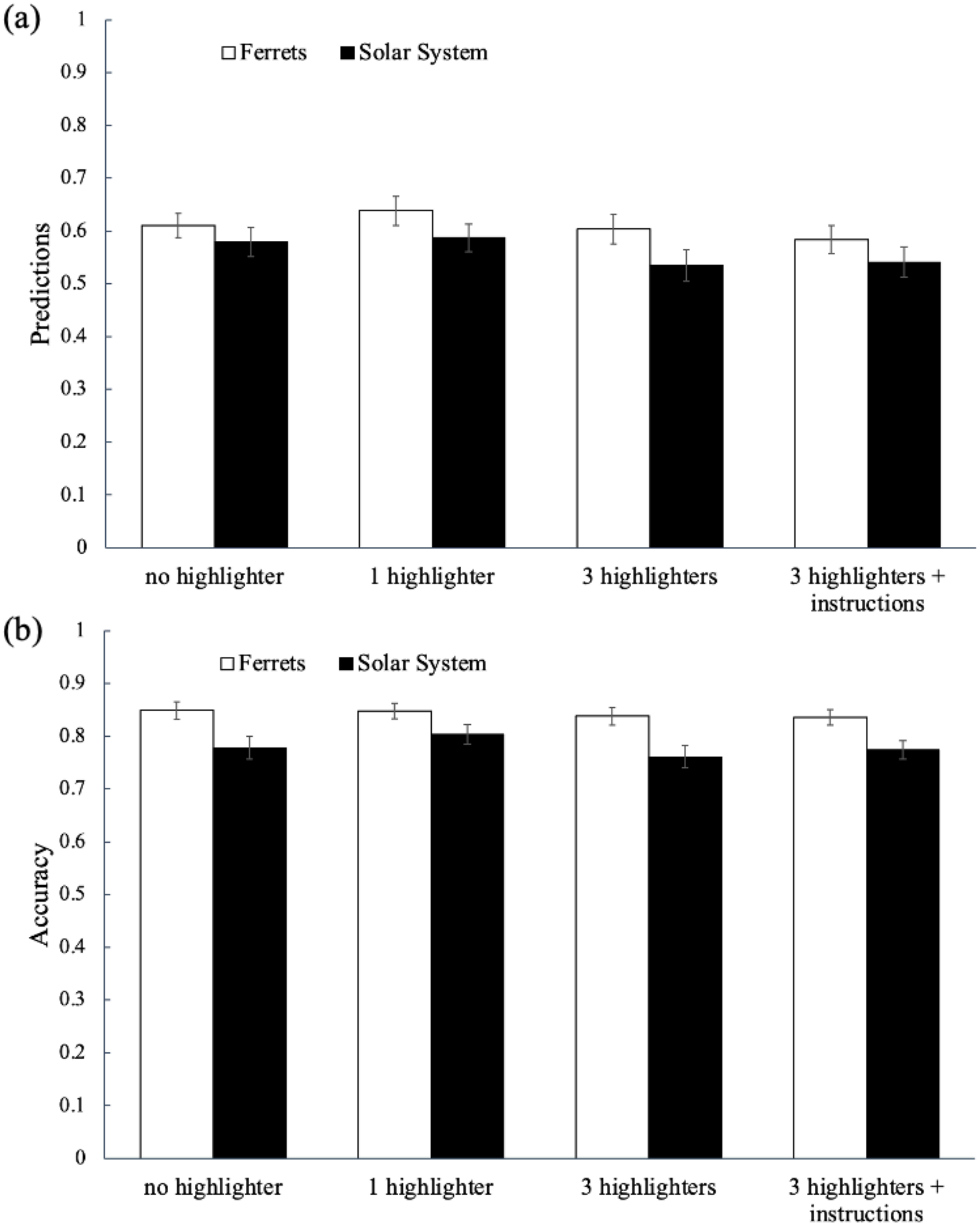


Figure 9. Predictions of performance (a) and performance on the comprehension test (b) as a function of passage topic for each group in Experiment 8. Error bars reflect the standard error of the mean.

To examine differences in performance on the comprehension tests, I conducted a 2 (topic: ferrets, solar system) x 4 (group: no highlighting, one highlighter, three highlighters with no instructions, or three highlighters with instructions) mixed ANOVA. Results revealed a main effect of topic such that a greater proportion of questions were answered correctly for the ferrets passage (M = .84, SD = .12) than the solar system passage (M = .78, SD = .14), [$F(1, 214) = 75.16$, $p < .001$, $\eta_p^2 = .26$]. However, there were no group differences in performance [$F(3, 214) = .68$, $p = .682$, $\eta_p^2 = .01$] such that participants not highlighting (M = .81, SD = .13), with one highlighter (M = .83, SD = .11), with three highlighters and no instructions (M = .80, SD = .13), and with three highlighters and instructions (M = .81, SD = .11) predicted similar performance. Topic did not interact with group [$F(3, 214) = .96$, $p = .411$, $\eta_p^2 = .01$].

Discussion

Results did not reveal advantages to using highlighters—each group performed similarly on the tests (and predictions mapped onto this trend). However, performance on the tests may have been too high to reveal potential differences between groups. Additionally, multiple-choice questions may be a less sensitive measure of learning and memory compared with more open-ended/free-recall questions. Although Experiment 8 suggests no benefit to highlighting, future work should further explore whether there are ways highlighters can be used to enhance memory.

CHAPTER 4:

DESIRABLE DIFFICULTIES AND VALUE

Portions of the following introductory comments, description of Experiment 9, and conclusion are taken directly from Murphy et al. (2022)

To achieve successful learning, students need to embrace, rather than avoid, certain *desirable difficulties* (Bjork & Bjork, 2014; Bjork, 1994) such as testing and generating rather than restudying (DeWinstanley & Bjork, 2004; Halamish & Bjork, 2011; Roediger & Karpicke, 2006), varying the environmental context when studying (Smith et al., 1978), interleaving rather than blocking practice (e.g., Kornell & Bjork, 2008), and spacing rather than massing study sessions (i.e., the spacing effect; Bjork & Allen, 1970; Cepeda et al., 2006; Greene, 2008; Karpicke & Bauernschmidt, 2011), which have all been shown to enhance learning outcomes. In the present research, I focused on whether the benefits of spacing might be realizable even when a student might be either unaware of the benefits of spacing or unwilling/unable to restudy after a substantial delay.

Across the long history of research on the effects of spaced versus massed opportunities to study various types of to-be-learned material, a variety of theoretical mechanisms have been proposed to account for the benefits of spacing (see Delaney et al., 2010; Hintzman, 1974, 1976; Maddox, 2016; Toppino & Gerbier, 2014 for reviews). Early research provided evidence for the attenuation of attention whereby attention declines more during massed presentations compared with spaced presentations (Melton, 1970; Shaughnessy et al., 1972; Underwood, 1969; but see Zimmerman, 1975). Other work proposed a consolidation account whereby long-term recall depends on a to-be-learned item's representation in memory being consolidated and that massed repetition of an item does not provide enough time for the effects of a first study trial to be consolidated before a second study trial is presented (see Bjork & Allen, 1970, for a test of the consolidation idea).

Experiment 9

Distributing practice in a massed encoding session

Some prior work has shown the benefits of repetition and spacing within a single encoding session. For example, in a continuous paired-associate task, Glenberg (1976) demonstrated that distributing study time across multiple presentations within a list can enhance memory (see also Delaney et al., 2010; Maddox, 2016; see Maddox & Balota, 2015; Peterson et al., 1963 for examinations of the spacing effect utilizing different retention intervals). In Experiment 9, I was interested in the potential benefits of distributed practice within a single encoding session (using unrelated word lists and free recall tests) and how this effect differs as a function of the difficulty of to-be-learned words.

I presented participants with lists of words to remember for a later test and each word received the same amount of total study time, but I manipulated (within-subjects) how that study time was distributed within each list (i.e., fewer, but longer, study opportunities or more, but shorter, study opportunities). I expected that increased (but shorter) encoding opportunities within a given study session would lead to better subsequent free recall of the studied words, but that any such benefits might be reduced or eliminated for more difficult (i.e., more abstract/less concrete) words.

In part, this expectation follows from Paivio's dual-coding theory (1971, 2013), which assumes that concrete words activate perceptual as well as verbal memory and are easier to remember compared to more abstract words (see also Schwanenflugel et al., 1988). Another consideration is that effective encoding strategies—such as interactive imagery, sentence generation, and grouping—lead to enhanced recall compared to less effective strategies, such as passive reading and simple repetition (Hertzog et al., 1998; Richardson, 1998; Unsworth, 2016), and participants studying more concrete (easier) words may be better able to utilize effective encoding strategies resulting in better performance on a later memory test than participants studying more abstract words. In addition, participants may only have time to engage in imagery, sentence generation, and other recall-enhancing activities when the to-be-remembered words are presented for longer study durations versus shorter, but more frequent, study opportunities (i.e., presentations for only 1 or 2 seconds at a time). As a result, repeated studying may only be beneficial for easier-to-remember words or when learners have enough study time to utilize elaborative encoding strategies for each item.

Method

Participants. After exclusions, participants were 86 undergraduate students (M_{age} = 19.85, SD_{age} = 1.43) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in one exclusion. A sensitivity analysis based on the obtained sample indicated that for a 2 (word difficulty: easy, hard) x 3 (study schedule: 1 second x 4, 2 seconds x 2, 4 seconds x 1) mixed ANOVA, assuming alpha = .05, power = .80, and a high correlation (r = .63) between repeated measures, the smallest effect (recall as a function of study schedule) the design could reliably detect is $\eta_p^2 = .02$ which is larger than most effects reported in memory research (see Morris & Fritz, 2013).

Materials. All studied words were nouns that contained four letters and participants either studied lists containing more concrete words (i.e., easier to remember; n = 44) or more abstract words (i.e., more difficult to remember; *n* = 42). Words were classified according to the English Lexicon Project website (Balota et al., 2007) and word lists were formed by randomly sampling unique sets of 20 words from a pool of 303 (177 easy words and 126 hard words). Thus, each participant received different lists of words with a different combination of words in each list, and each word could appear on a list with any of the different study schedules.

For participants presented with easier words to remember, on the log-transformed Hyperspace Analogue to Language (HAL) frequency scale (with lower values indicating lower frequency in the English language and higher values indicating higher frequency), words ranged from 5.48-12.88 and averaged a score of 9.63 (*SD* = 1.44). In terms of concreteness (with lower values indicating lower concreteness and higher values indicating higher concreteness), these words ranged from 4.26-5.00 and averaged a score of 4.74 (*SD* = .20). For participants presented with harder-to-remember words, frequency levels ranged from 7.43-14.35 and averaged a score of 10.70 (*SD* = 1.16), and their concreteness levels ranged from 1.25-4.24 and averaged a score of 3.26 (*SD* = .71). Words I classified as "hard" were significantly more concrete than words I classified as "easy" [$t(301) = 26.34$, $p < .001$, $d = 3.07$, $BF_{10} > 100$] but were also significantly less frequent than "easy" words [$t(301) = 6.89$, $p < .001$, $d = .80$, $BF_{10} > 100$][1].

Procedure. Participants were told that they would be presented with lists of words with each list containing 20 words and that their task was to remember the words for a later test. Participants were presented with six lists in total and on each list, participants either viewed each word once for 4 seconds (two lists), twice for 2 seconds (two lists), or four times for 1 second (two lists). List order was counterbalanced but study conditions occurred in blocks (i.e., the two lists

---

[1] Since we failed to control for frequency, word frequency effects (see Popov & Reder, in press for a review of the effects of word frequency on memory) may have reduced the strong effects of concreteness commonly seen in the literature.

where words were studied once for 4 seconds occurred consecutively). On lists where participants viewed the words more than once, the order of words was the same across cycles throughout the list (i.e., words 1-20 were presented once, then again in the same order). After the presentation of all 20 words, participants were given a 1-minute immediate free recall test in which—in an on-screen text box—they recalled as many words as they could from the just-studied list in any order they wished. Immediately following the recall period, participants were informed of the number of correctly recalled words for that list but were not given feedback about specific words.

Following the test of the final to-be-remembered list, participants reported what encoding strategies (if any) they had used to remember the words using a check-off list of possible strategies. Specifically, participants indicated whether they simply read each word as it appeared, repeated the words as much as possible, developed rhymes for the words, used sentences to link the words together, developed mental images of the words, grouped the words in a meaningful way, or did not use any strategies. Participants could select some, all, or none of the suggestions to indicate which strategies they used).

Results

Recall performance for each study schedule as a function of word difficulty is shown in Figure 10 and to analyze potential differences, I conducted a 2 (word difficulty: easy, hard) x 3 (study schedule: 1 second x 4, 2 seconds x 2, 4 seconds x 1) mixed ANOVA. Results did not reveal a significant effect of word difficulty [$F(1, 84) = 3.64$, $p = .060$, $\eta_p^2 = .04$, $BF_{01} = .82$], with participants recalling a similar proportion of easy words ($M = .55$, $SD = .14$) as hard words ($M = .49$, $SD = .14$). However, there was a significant main effect of study schedule [$F(2, 168) = 9.91$, $p < .001$, $\eta_p^2 = .11$, $BF_{10} > 100$] such that words studied four times for 1 second ($M = .54$, $SD = .17$) were recalled better than the words studied once for 4 seconds ($M = .48$, $SD = .16$), [$p_{\text{bonf}} =$

.003, $d = .37$] but not better than the words studied twice for 2 seconds ($M = .54$, $SD = .16$), [$p_{\text{bonf}} > .999$, $d = .04$]; additionally, recall for the words studied twice for 2 seconds was greater than that for the words studied once for 4 seconds [$p_{\text{bonf}} < .001$, $d = .46$]. Word difficulty did not interact with study schedule [$F(2, 168) = .52$, $p = .596$, $\eta_p^2 = .01$, $BF_{01} = 9.67$].

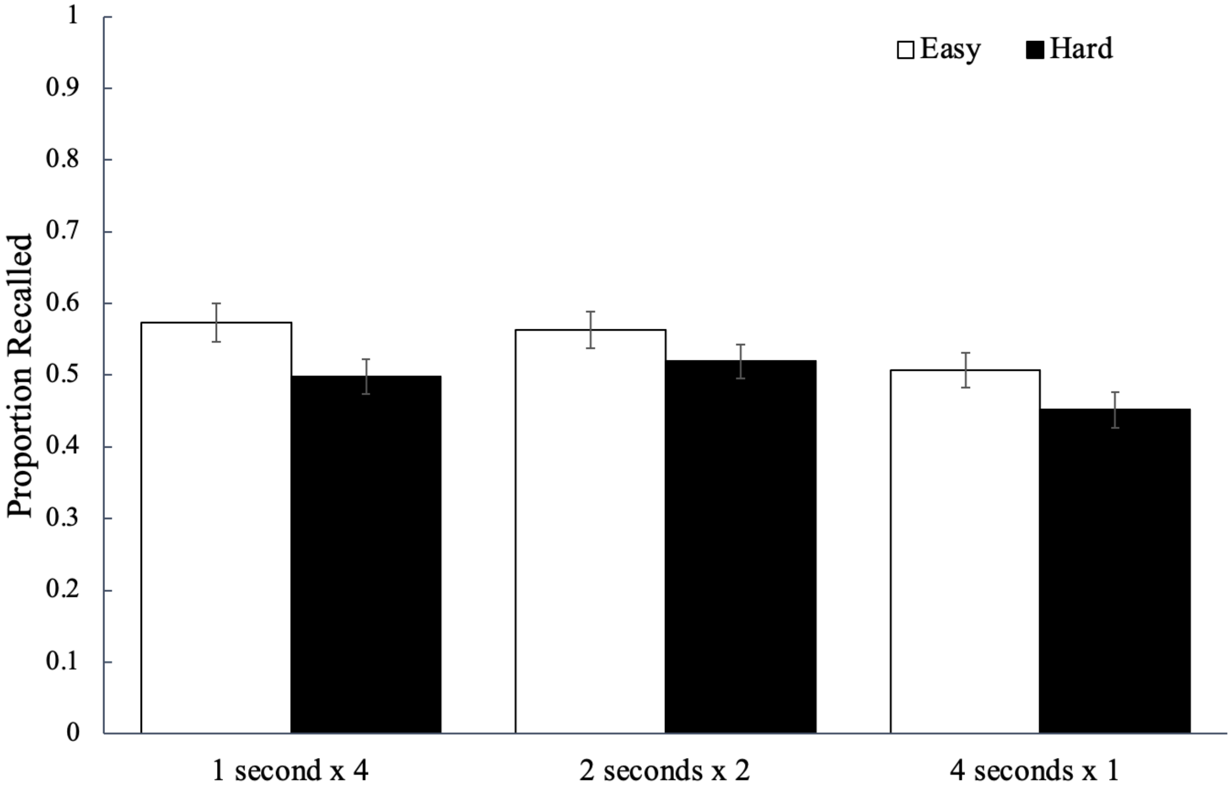


Figure 10. The average proportion of words correctly recalled as a function of word difficulty and how a fixed study time was distributed across presentations of a given word in Experiment 9. Error bars reflect the standard error of the mean.

Although results indicate that studying a word multiple times but for a shorter duration can enhance memory, the retention interval differs between these study schedules. For example, when studying each word once for 4 seconds, the time until the recall test is 76 seconds for the first word whereas when the first word is studied for the final time when studying each word four times for 1 second each, the time until the recall test is 19 seconds. To examine the potential benefits of distributing study time while controlling for retention interval, I conducted a logistic multilevel model (MLM) where I treated the data as hierarchical or clustered (i.e., multilevel) and items nested within individual participants. In this analysis, the regression coefficients are given as logit units (i.e., the log odds of correct recall). I report exponential betas (e^B), and their 95% confidence intervals, which give the coefficient as an odds ratio (i.e., the odds of correctly recalling a word divided by the odds of not recalling a word). Thus, e^B can be interpreted as the extent to which the odds of correctly recalling a word changed. Specifically, values greater than 1 represent an increased likelihood of recall while values less than 1 represent a decreased likelihood of recall.

To examine recall as a function of the number of word presentations while controlling for retention interval (for the final presentation of each word, I calculated the time remaining until the beginning of the recall test), I conducted a logistic MLM with item-level recall modeled as a function of number of word presentations and retention interval. Results revealed that, when controlling for retention intervals, the number of word presentations significantly predicted recall [$e^B = 1.11$, CI: 1.07 – 1.16, $z = 5.25$, $p < .001$] such that distributed practice enhanced memory.

Discussion

To summarize, the pattern of results obtained in Experiment 9 revealed a recall advantage for words studied multiple times compared to once (despite no differences in total study time) but this finding did not differ according to item difficulty.

Experiment 10

Is the spacing effect greater for high-value information?

To better recall valuable information, often at the expense of low-value information, people sometimes use more effective encoding strategies (e.g., mental imagery and sentence generation) when studying high-value items but less effective strategies (e.g., passive reading and simple repetition) when studying low-value items (Hennessee et al., 2019). Specifically, when studying high-value words, learners selectively use elaborative encoding strategies that arise from deeper semantic and associative processing, leading to a stronger memory trace and greater memory for this information (Craik & Lockhart, 1972; Richardson, 1998). Thus, other techniques that benefit memory performance (i.e., desirable difficulties) may also be selectively used to enhance memory for valuable information.

In addition to enhancing memory overall, desirable difficulties may offer a greater memory benefit for valuable information. Specifically, high point values may lead to an additive effect of spacing, generation, and testing as a result of the benefits of desirable difficulties as well as strategic processing. For example, based on the dual-process accounts of value-directed remembering (automatic versus controlled/strategic processing; see Knowlton & Castel, 2022), high-value words may see more of a memory benefit due to additional strategic processing when given later study opportunities (i.e., when study time is distributed) as opposed to the massed encoding of each word. Particularly, the use of strategic processing when given later (but shorter) study opportunities may be enhanced as the learner gains task experience (see Cohen et al., 2017 for a demonstration of strategy-driven effects of value on memory). Similar memory benefits for high-value words may arise when generating or testing the information rather than restudying as learners may exert strategic processing of high-value words that, when combined with desirable difficulties, enhance memory selectivity.

In Chapter 4, I was interested in whether the memory benefits of desirable difficulties such as the spacing effect, the generation effect, and the testing effect differentially benefit memory as a function of value. Specifically, if desirable difficulties offer a larger memory benefit for valuable information, this could further illustrate their effectiveness in the classroom. In contrast, if desirable difficulties benefit memory for information regardless of value, this suggests that learners should be strategic with their employment of desirable difficulties as a study strategy and selectively utilize desirable difficulties when studying valuable or important information.

In Experiment 10, I investigated whether the spacing effect differentially enhances memory for high-value information (see Cohen et al., 2013 for work on value and spacing). Specifically, I presented learners with words paired with point values and either allowed participants to study each word for 4 seconds (massed) or four times for 1 second each. Experiment 9 demonstrated that the benefits of spacing can occur in a single study such that studying a word four times for 1 second each resulted in better memory than studying a word once for 4 seconds. Applied here, I expected spacing to benefit memory more for high-value compared with low-value words, particularly after learners gain task experience.

Method

Participants. Participants were 108 undergraduate students (M_{age} = 20.63, SD_{age} = 2.53) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in one exclusion.

Materials and Procedure. Participants were presented with 20 to-be-remembered words with each word paired with point values indicating how much the word is “worth” with participants’ goal being to maximize their score. On three lists participants viewed each word once for 4 seconds and on three lists viewed each word four times for 1 second each. On lists where participants viewed the words four times for 1 second each, the order of words was the same across cycles throughout the list (i.e., words 1-20 were presented once, then again in the same order three more times). Whether participants studied each word once for 4 seconds on the first three lists and studied each word four times for 1 second each on the last 3 lists (or vice versa) was counterbalanced. Following the first presentation of all 20 word pairs, participants completed a 30-second distraction task requiring them to rearrange the digits of several three-digit numbers in descending order (e.g., 123 would be rearranged to 321). Finally, participants were given a 1-minute free recall test in which they had to type as many words as they could from the list (they did not need to recall the point values) into an on-screen text box. Immediately following the recall period, participants were informed of their score for that list but were not given feedback about specific items.

Results and Discussion

Recall as a function of study schedule and value is shown in Figure 11. To examine potential differences in recall, I conducted a logistic MLM with item-level accuracy modeled as a function of value and whether study time was distributed. Results revealed that value significantly predicted recall [$e^B = 1.09$, $CI_{95\%} = 1.08 – 1.09$, $z = 24.87$, $p < .001$] such that high-value words were better recalled than low-value words. Additionally, participants recalled a greater proportion of words when study time was distributed ($M = .54$, $SD = .17$) than when study time was massed ($M = .47$, $SD = .17$), [$e^B = 1.35$, $CI_{95\%} = 1.26 – 1.46$, $z = 7.99$, $p < .001$]. However, value did not interact with study schedule [$e^B = 1.01$, $CI_{95\%} = 1.00 – 1.02$, $z = 1.26$, $p = .209$]. Thus, spacing benefits memory but not especially for items of value.

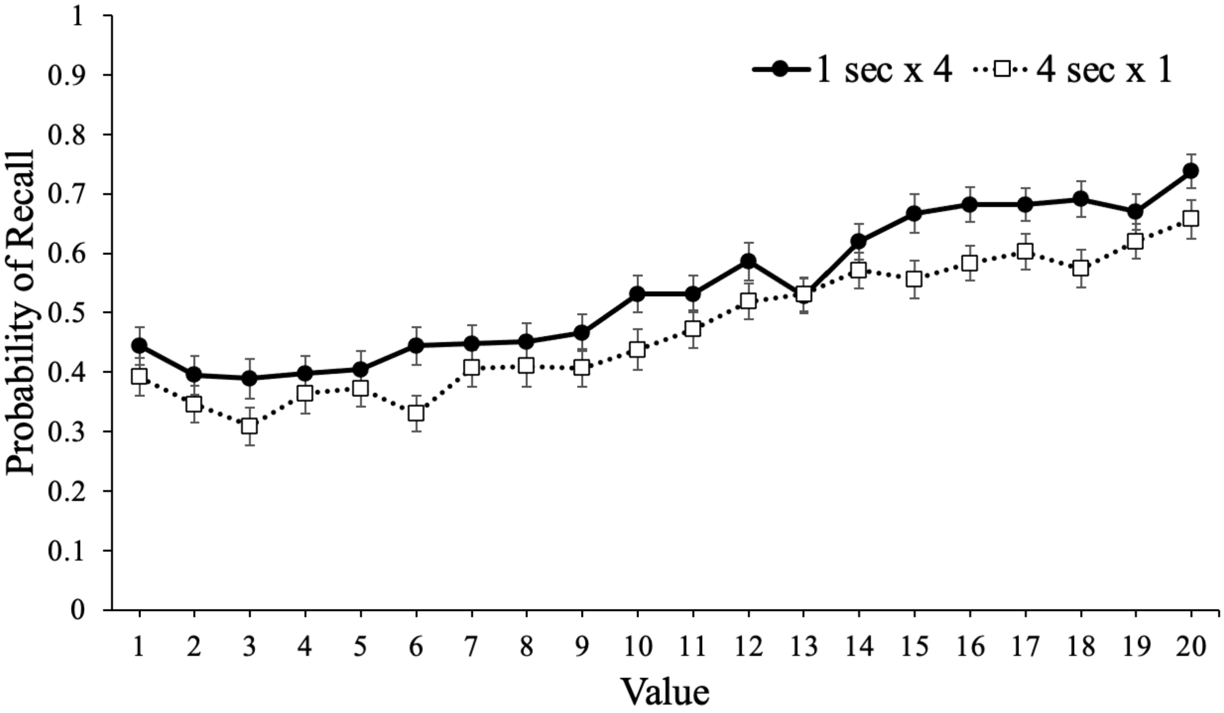


Figure 11. The average proportion of words recalled as a function of value and how a fixed study time was distributed across presentations of a given word in Experiment 10. Error bars reflect the standard error of the mean.

Experiment 11

Is the generation effect greater for high-value information?

Experiment 11a

In Experiment 11, I examined whether the generation effect is greater for high-value relative to low-value information. In Experiment 11a, I presented participants with related word pairs (e.g., airplane : travel) to remember for a later test. On the recall test, participants were presented with the left word (cue) and asked to recall the right word (target). Each word pair was paired with a point value counting towards participants' score if later remembered, with participants' goal being to maximize their score. On some lists, after an initial study opportunity, participants restudied the complete word pairs. However, on other lists, after an initial study opportunity, participants studied the word pairs a second time with the target word fragmented (e.g., airplane : trav _ _). Thus, on these lists participants must generate the target word rather than simply restudying it. Prior work indicates that generation effects are stronger when manipulated within-subjects than between-subjects (see Hertel, 1989; Schmidt, 1992 for a reduced generation effect in a between-subjects design) so I had participants complete both lists where they restudied the to-be-remembered information and lists where they needed to generate it. I expected the generation effect to benefit memory more for valuable pairs than low-value word pairs and for this effect to increase with on later lists.

Method

Participants. Participants were 104 undergraduate students (M_{age} = 20.63, SD_{age} = 2.84) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in one exclusion.

Materials and Procedure. Participants were presented with a series of to-be-remembered word pairs (e.g., airplane : travel) with each word pair paired with a unique, randomly assigned value between 1 and 20 indicating how much the word was “worth” (point values appeared beneath each word pair). Each point value was used only once within each list and the order of the point values within lists was randomized. Word pairs were related and taken from the 1998 Florida State University word norms. Each word pair was initially shown one at a time, for 5 seconds each, in random order. Following the study phase, participants completed a 30-second distraction task.

After the distractor task, participants either restudied the word pairs a second time for 5 seconds each (3 lists) or restudie the word pairs for 5 seconds each with the target word fragmented (e.g., airplane : trav _ _; 3 lists), with each word presented in a random order. Whether participants restudid the full word pairs on the first three lists and restudy the word fragments on the last 3 lists (or vice versa) was counterbalanced. To create word fragments, I removed the last two characters of the target word and inserted two blanks. Following this second study opportunity, participants again completed the 30-second distractor task. Next, in the test phase, participants were shown the left word from each pair one at a time, in random order, and asked to recall the associated word. Participants were given as much time as they needed to recall each pair and were given any general or specific feedback at the end of each list, only a reminder to maximize their scores.

Results and Discussion

Cued recall as a function of study technique and value is shown in Figure 12. To examine potential differences in cued recall, I conducted a logistic MLM with item-level accuracy modeled as a function of value and study technique. Results revealed that value significantly predicted recall [$e^B = 1.02$, $CI_{95\%} = 1.01 – 1.03$, $z = 3.75$, $p < .001$] such that high-value word pairs were better recalled than low-value word pairs. Additionally, participants recalled a greater proportion of words when the target word was fragmented ($M = .81$, $SD = .22$) than when participants simply restudied the full word pairs ($M = .75$, $SD = .25$), [$e^B = .62$, $CI_{95\%} = .56 – .69$, $z = -9.40$, $p < .001$]. However, value did not interact with study technique [$e^B = 1.01$, $CI_{95\%} = .99 – 1.03$, $z = .85$, $p = .397$]. Thus, the generation effect does not uniquely benefit valuable information.

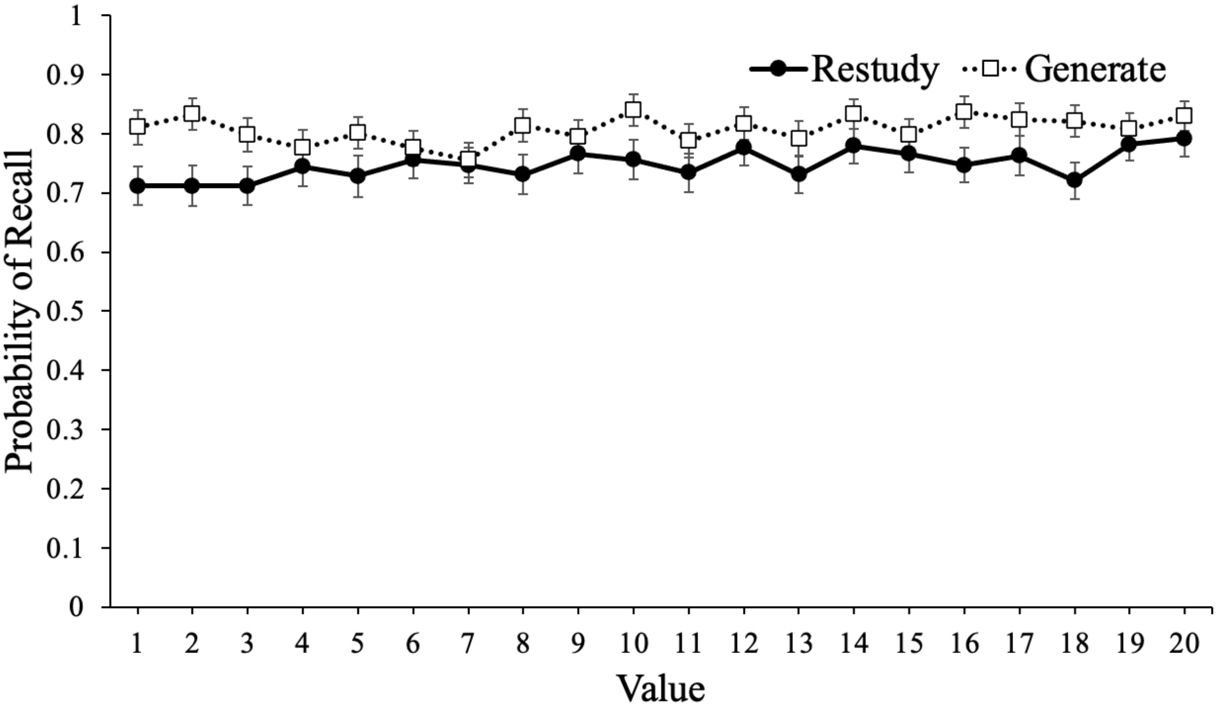


Figure 12. The average proportion of words recalled as a function of value and study technique in Experiment 11a. Error bars reflect the standard error of the mean.

Experiment 11b

In Experiment 11b, I used a similar design as Experiment 11a but used a free recall paradigm for unassociated words rather than cued recall.

Method

Participants. Participants were 105 undergraduate students (M_{age} = 21.22, SD_{age} = 4.61) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in three exclusions.

Materials and Procedure. Participants were presented with a series of to-be-remembered words with each word pair paired with a unique, randomly assigned value between 1 and 20 indicating how much the word is “worth” (point values will appear beneath each word pair). Each point value was used only once within each list and the order of the point values within lists was randomized. Each word was initially shown one at a time, for 4 seconds each, in random order. The words on each list were unrelated words and each word was 4 letters in length. Following the study phase, participants completed a 30-second distraction task.

After the distractor task, participants either restudied the words a second time for 4 seconds each (3 lists) or restudied the word for 4 seconds each with the word fragmented (the middle two letters removed), with each word presented in a random order. Whether participants restudied the full word on the first three lists and restudy the word fragments on the last 3 lists (or vice versa) was counterbalanced. Following this second study opportunity, participants again completed the 30-second distractor task. Next, in the test phase, participants were given 1 minute to recall all the words from the just-studied list that they can remember. Participants were told their score at the end of each list.

Results and Discussion

Recall as a function of study technique and value is shown in Figure 13. To examine potential differences in recall, I conducted a logistic MLM with item-level accuracy modeled as a function of value and study technique. Results revealed that value significantly predicted recall [$e^B = 1.08$, $CI_{95\%} = 1.07 – 1.08$, $z = 21.10$, $p < .001$] such that high-value words were better recalled than low-value words. Additionally, participants recalled a greater proportion of words when participants restudied the full words ($M = .59$, $SD = .21$) than when the word was fragmented on the second study opportunity ($M = .49$, $SD = .20$), [$e^B = 1.62$, $CI_{95\%} = 1.50 – 1.75$, $z = 12.14$, $p < .001$]. However, value did not interact with study technique [$e^B = 1.00$, $CI_{95\%} = .99 – 1.02$, $z = .25$, *p* = .806]. Thus, I provide further evidence that techniques that enhance memory overall may not particularly benefit high-value items.

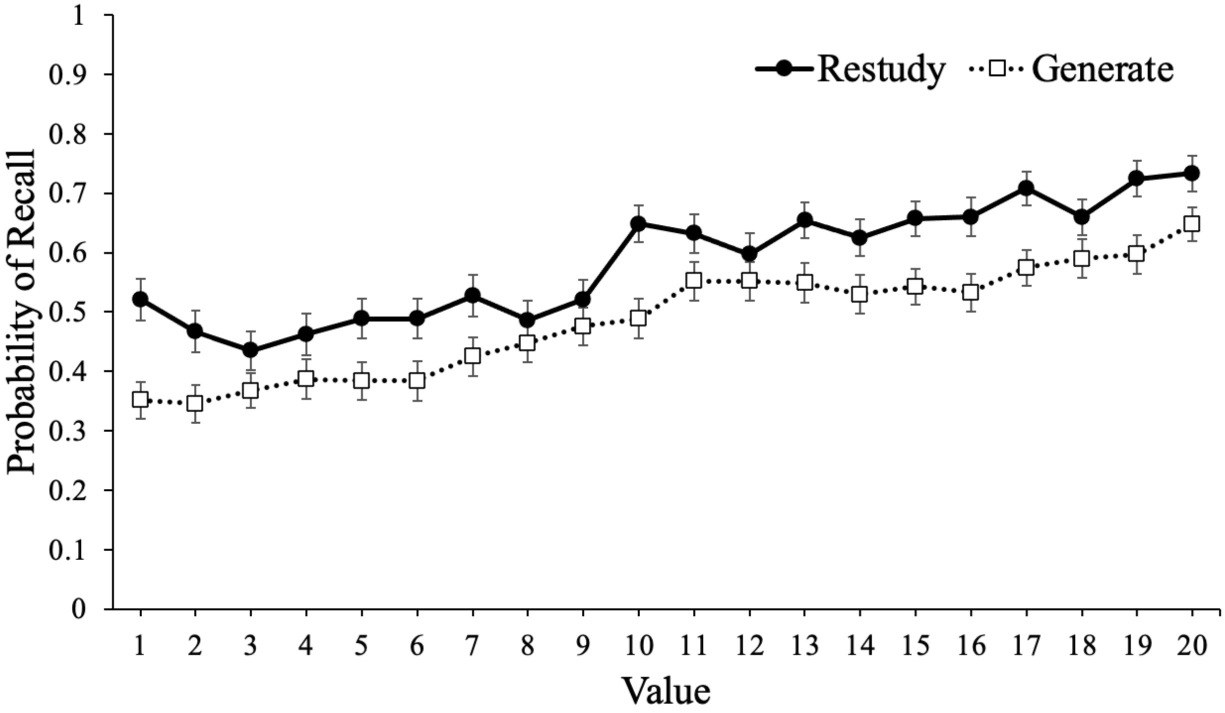


Figure 13. The average proportion of words recalled as a function of value and study technique in Experiment 11b. Error bars reflect the standard error of the mean.

Experiment 12

Is the testing effect greater for high-value information?

Experiment 12a

In Experiment 12, I examined how the benefits of testing interact with information value. Specifically, in Experiment 12a, I presented participants with words paired with point values and either had participants restudy the words or take a practice recall test for the words prior to a final free recall test on each list. I expected the practice recall test to enhance memory for high-value words at the expense of low-value words, especially on later lists.

Method

Participants. Participants were 110 undergraduate students (M_{age} = 20.35, SD_{age} = 2.41) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in seven exclusions.

Materials and Procedure. The task in Experiment 12 was similar to Experiment 10. Participants were presented with 20 to-be-remembered words with each word paired with a point value indicating how much the word is "worth" (e.g., twig : 5) with participants' goal being to maximize their score. Words and point values were simultaneously displayed for 4 seconds each. After the presentation of all 20 word-number pairs in each list, participants completed a 30-second distraction task.

Next, participants either restudied the words a second time for 4 seconds each, in random order (3 lists), or completed a practice free recall test (80 seconds, the same duration as the restudy phase). On lists with the practice free recall test, participants were told "This is a PRACTICE test. The actual test will occur shortly. Please use this time to type all of the words that you can remember from the just-presented list in the box below." Participants were not given any feedback following the practice test. Whether participants restudied the full words on the first three lists and completed a practice test on the last 3 lists (or vice versa) was counterbalanced. Participants then completed the 30-second distractor task prior to the official test phase whereby participants were given a 1-minute free recall test in which they had to type as many words as they can from the list (they did not need to recall the point values) into an on-screen text box. Immediately following the recall period, participants were informed of their score for that list but not given feedback about specific items.

Results and Discussion

Recall as a function of study technique and value is shown in Figure 14. To examine potential differences in recall, I conducted a logistic MLM with item-level accuracy modeled as a function of value and study technique. Results revealed that value significantly predicted recall [e^B = 1.07, $CI_{95\%}$ = 1.07 – 1.08, z = 21.07, p < .001] such that high-value words were better recalled than low-value words. Additionally, participants recalled a greater proportion of words when participants restudied the words (M = .55, SD = .19) than when taking a practice test (M = .42, SD = .18), [e^B = .54, $CI_{95\%}$ = .51 – .59, z = -15.97, p < .001]. However, value did not interact with study technique [e^B = .99, $CI_{95\%}$ = .98 – 1.00, z = -1.90, p = .058]. Although restudying benefitted memory (not testing), the memory benefits were not specific to valuable items.

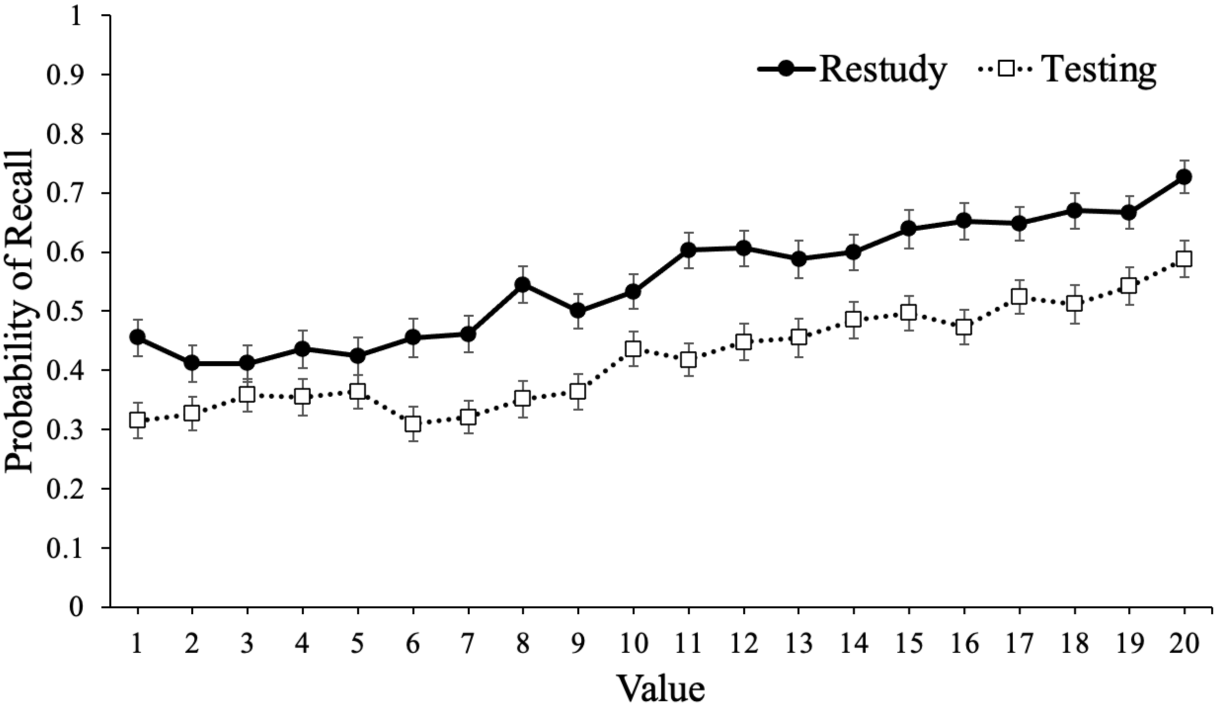


Figure 14. The average proportion of words recalled as a function of value and study technique in Experiment 12a. Error bars reflect the standard error of the mean.

Experiment 12b

In Experiment 12b, participants studied a list of words, either did a practice recall test or a distractor task, then studied the words again before completing the recall test.

Method

Participants. Participants were 106 undergraduate students (M_{age} = 21.19, SD_{age} = 3.56) recruited from the UCLA Human Subjects Pool. Participants were tested online and received course credit for their participation. Participants were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a post-task questionnaire (they were told they would still receive credit if they cheated). This exclusion process resulted in five exclusions.

Materials and Procedure. The task in Experiment 12b was similar to Experiment 12a. Participants were presented with the 20 to-be-remembered words paired with point values. After the presentation of all 20 word-number pairs in each list, participants either completed a 60-second distraction task (3 lists) or a 60-second practice recall test (3 lists). Next, participants either restudied the words a second time for 4 seconds each, in random order. Whether the first three lists included the distractor task or the practice recall test was counterbalanced. Participants then completed the 30-second distractor task again prior to the official test phase whereby participants were given a 1-minute free recall test in which they had to type as many words as they could from the list (they did not need to recall the point values) into an on-screen text box. Immediately following the recall period, participants were informed of their score for that list but not given feedback about specific items.

Results

Recall as a function of study technique and value is shown in Figure 15. To examine potential differences in recall, I conducted a logistic MLM with item-level accuracy modeled as a function of value and study technique. Results revealed that value significantly predicted recall [e^B = 1.06, $CI_{95\%}$ = 1.05 – 1.07, z = 16.92, $p < .001$] such that high-value words were better recalled than low-value words. Additionally, participants recalled a greater proportion of words when taking a practice test (M = .56, SD = .21) compared with a distractor task (M = .52, SD = .20), [e^B = 1.18, $CI_{95\%}$ = 1.09 – 1.27, z = 4.16, $p < .001$]. However, value did not interact with study technique [e^B = 1.00, $CI_{95\%}$ = .98 – 1.01, z = -.55, p = .584].

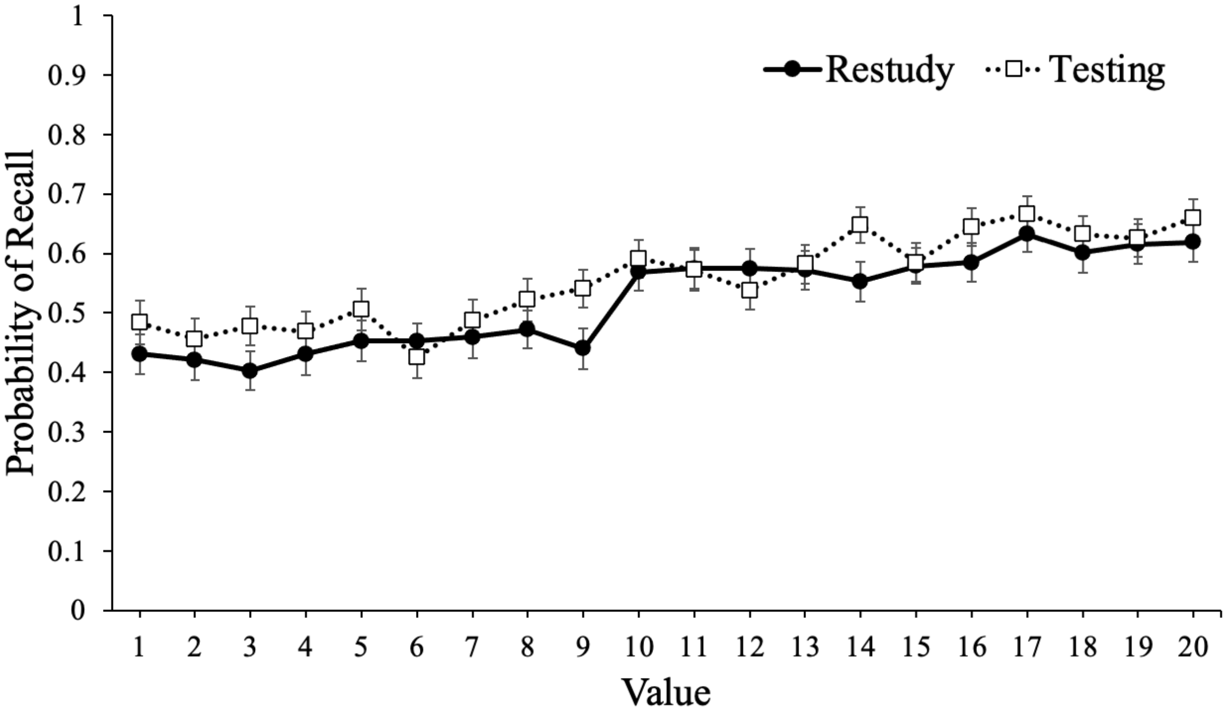


Figure 15. The average proportion of words recalled as a function of value and study technique in Experiment 12b. Error bars reflect the standard error of the mean.

Discussion

In Chapter 4, I aimed to investigate whether desirable difficulties, such as the spacing effect, the testing effect, and the generation effect selectively enhance memory for high-value information. Contrary to our expectations, desirable difficulties did not lead to a greater memory benefit for high-value words compared to low-value words, even after learners gained task experience. Thus, while both value-directed remembering and desirable difficulties can independently improve memory performance, they may not always interact to selectively benefit valuable information.

The mechanisms underlying value-directed remembering and desirable difficulties may operate differently and can sometimes even conflict with each other. For example, desirable difficulties involve creating challenges or obstacles during learning which creates cognitive effort and increases the difficulty of the learning task, which can initially impair performance or reduce the immediate recall of information. However, these challenges also promote deeper processing, elaboration, and engagement of cognitive resources, which can enhance long-term retention and recall of the learned information. On the other hand, value-directed remembering involves prioritizing valuable information, which may involve the use of more shallow or passive processing strategies when the information is less valuable, which may not promote the kind of deep processing and elaboration that desirable difficulties aim to enhance.

As a result, when desirable difficulties are introduced during learning, they may create challenges or obstacles that conflict with the prioritization and processing strategies associated with value-directed remembering. Specifically, the challenges posed by desirable difficulties may initially impair the processing and encoding of valuable information, leading to reduced immediate recall or performance compared to low-value information. However, these challenges may promote long-term retention and recall of the valuable information by enhancing deeper processing and elaboration. Future research should examine the long-term effects of desirable difficulties on memory for valuable information.

Another possible explanation for the lack of differential effects of spacing on memory for high-value versus low-value words is that other factors, such as the nature of the encoding strategies used by learners, may have influenced memory performance. Previous research has shown that learners tend to use more effective encoding strategies, such as mental imagery and sentence generation, when studying high-value items, compared to less effective strategies, such as passive reading and simple repetition, when studying low-value items (Hennessee et al., 2019). It is possible that these differential encoding strategies may have contributed to similar memory performance for high-value and low-value words, regardless of the desirable difficulties manipulation.

Overall, the findings from Chapter 4 did not find evidence for a differential effect of desirable difficulties on memory for high-value versus low-value information. Thus, while both value-directed remembering and desirable difficulties can independently enhance memory, these processes may not lead to selective benefits for valuable information.

CHAPTER 5:

CONCLUSIONS

In everyday life, we encounter more information than we can remember. However, if we are strategic with how we use our memory, we can maximize our ability to remember important information. For example, forgetting unimportant or outdated information can benefit memory for important information we want to remember but if we have little faith in our memory abilities, we can strategically use cognitive offloading to remember valuable information. Specifically, if our memory is less reliable than an external store such as a cell phone, it may be beneficial to offload information we need to remember to avoid the negative consequences of forgetting. Despite the obvious benefits of offloading, there may also be circumstances when we should use our memory to remember important information (i.e., if the external store is unreliable).

This strategic offloading likely includes a metacognitive component and how we evaluate our own memory can also influence what we remember. For example, simply evaluating whether something is likely to be remembered or forgotten, or even just selecting information in some way, can enhance the likelihood that it is remembered. Thus, an awareness of what we might forget can reduce the susceptibility to forget that information and this may have important implications for classroom learning.

Finally, techniques known to enhance memory performance do not differentially enhance memory for important information. Specifically, strategies like distributing study time, generating information rather than restudying, and testing yourself rather than re-reviewing material benefit memory regardless of the value of a given item. However, we often engage in the strategic processing of valuable information (see Knowlton & Castel, 2022) and, if we selectively employ desirable difficulties when studying valuable items, learners may be able to enhance memory for high-value information. Taken together, strategic forgetting and offloading, an awareness of our own memory, and the strategic use of effective learning techniques may help us maximize memory for important information.

References

Aguirre, C., Gómez-Ariza, C. J., Andrés, P., Mazzoni, G., & Bajo, M. T. (2017). Exploring Mechanisms of Selective Directed Forgetting. *Frontiers in Psychology, 8*, 316.

Annis, L., & Davis, J. K. (1978). Study techniques and cognitive style: Their effect on recall and recognition, *The Journal of Educational Research*, *71*, 175-178.

Anselmi, N., Montaldo, S., Pomilla, A., & Maraone, A. (2020). Forgotten baby syndrome: Dimensions of the phenomenon and new research perspectives. *Rivista di psichiatra, 55*, 112-118.

Arbuckle, T. Y., & Cuddy, L. L. (1969). Discrimination of item strength at time of presentation. *Journal of Experimental Psychology, 81*, 126-131.

Ariel, R., Dunlosky, J., & Bailey, H. (2009). Agenda-based regulation of study-time allocation: when agendas override item-based monitoring. *Journal of Experimental Psychology: General*, *138*, 432-447.

Armitage, K. L., Bulley, A., & Redshaw, J. (2020). Developmental origins of cognitive offloading. *Proceedings of the Royal Society B*, *287*, 20192927.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39*, 445-459.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84,* 191-215.

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental Psychology, 25*, 729-735.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York: W. H. Freeman.

Barr, N., Pennycook, G., Stolz, J. A., & Fugelsang, J. A. (2015). The brain in your pocket: Evidence that Smartphones are used to supplant thinking. *Computers in Human Behavior, 48*, 473-480.

Basden, B. H., & Basden, D. R. (1998). Directed forgetting: A contrast of methods and interpretations. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 139-172). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 603-613.

Beaudoin, M., & Desrichard, O. (2011). Are memory self-efficacy and memory performance related? A meta-analysis. *Psychological Bulletin, 137*, 211-241.

Bell, K. E., & Limber, J. E. (2010). Reading skill, textbook marking, and course performance. *Literacy Research and Instruction*, *49*, 56-67.

Ben-Yehudah, G., & Eshet-Alkalai, Y. (2018). The contribution of text-highlighting to comprehension: A comparison of print and digital reading. *Journal of Educational Multimedia and Hypermedia*, *27*, 153-178.

Berry, J. M. (1999). Memory self-efficacy in its social cognitive context. In T.M. Hess & F. Blanchard-Fields (Eds.), *Social cognition and aging* (pp. 69-96). San Diego, CA: Academic Press.

Berry, J. M., Williams, H. L., Usubalieva, A., & Kilb, A. (2013). Metacognitive awareness of the associative deficit for words and names. *Aging, Neuropsychology and Cognition*, *20*, 592-619.

Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition, 5*, 176-196.

Bjork, E. L., & Bjork, R. A. (2014). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher and J. Pomerantz (Eds.), *Psychology and the real world: Essays illustrating fundamental contributions to society (2nd edition)*. (pp. 59-68).

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), Coding processes in human memory (pp. 217-235). Washington, DC: Winston.

Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 309-330). Hillsdale, NJ: Erlbaum.

Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). MIT Press.

Bjork, R. A. (1998). Intentional forgetting in perspective: Comments, conjectures, and some directed remembering. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 453-481). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior, 9*, 567-572.

Bjork, R. A., LaBerge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science, 10*, 55-56.

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, *4*, 1-16.

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: Further evidence from word and picture stimuli. *Memory & Cognition*, *42*, 370–382.

Bonin, P., Gelin, M., Laroche, B., Méot, A., & Bugaiska, A. (2015). The 'how' of animacy effects in episodic memory. *Experimental Psychology, 62*, 371-384.

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers*, *35*, 634-639.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717-726.

Carter, J. A. (2018). Autonomy, cognitive offloading, and education. *Educational Theory*, *68*, 657-673.

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging, 20,* 718- 721.

Castel, A. D. (2008). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A. S. Benjamin & B. H. Ross (Eds.). *The psychology of learning and motivation* (Vol. 48, pp. 225-270). London, UK: Academic Press.

Castel, A. D., Benjamin, A. S., Craik, F. I. M., & Watkins, M. J. (2002). The effects of aging on selectivity and control in short-term recall. *Memory & Cognition, 30*, 1078-1085.

Castel, A. D., Farb, N. A. S., & Craik, F. I. M. (2007). Memory for general and specific value information in younger and older adults: Measuring the limits of strategic control. *Memory & Cognition, 35*, 689-700.

Castel, A. D., Murayama, K., Friedman, M. C., McGillivray, S., & Link, I. (2013). Selecting valuable information to remember: Age-related differences and similarities in self-regulated learning. *Psychology and Aging, 28*, 232-242.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354-380.

Chiew, K., & Braver, T. S. (2017). Context processing and control in the human brain: From gating models to dual mechanisms. In T. Egner (Ed.), The Wiley handbook of cognitive control. Chichester, UK: John Wiley & Sons.

Cohen, M. S., Rissman, J., Hovhannisyan, M., Castel, A. D., & Knowlton, B. J. (2017). Free recall test experience potentiates strategy-driven effects of value on memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 1581-1601.

Cohen, M. S., Yan, V. X., Halamish, V., & Bjork, R. A. (2013). Do students think that difficult or valuable materials should be restudied sooner rather than later? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1682-1696.

Coverdale, M. E., & Nairne, J. S. (2019). The mnemonic effect of choice. *Psychonomic Bulletin & Review*, *26*, 1310-1316.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

Craik, F. I. M. (2002). Human memory and aging. In L. Bäckman & C. von Hofsten (Eds.), *Psychology at the turn of the millennium* (pp. 261-280). Hove, UK: Psychology Press.

Craik, F. I. M., & Bosman, B. A. (1992). Age-related changes in memory and learning. In H. Bouma & J. A. M. Graafmans (Eds.), *Gerontechnology* (pp. 79-92). Amsterdam, The Netherlands: IOS Press.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *Journal of Verbal Learning & Verbal Behavior*, *11*, 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268-294.

Dawson, P. (2020). Cognitive offloading and assessment. In *Re-imagining University Assessment in a Digital World* (pp. 37-48). Springer, Cham.

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation, 53*, 63-147.

DeWinstanley, P. A., & Bjork, E. L. (2004). Processing strategies and the generation effect: Implications for making a better reader. *Memory & Cognition, 32*, 945-955.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135-168.

Double, K. S., & Birney, D. P. (2019). Reactivity to measures of metacognition. *Frontiers in Psychology, 10*, 2755.

Double, K. S., Birney, D. P., & Walker, S. A. (2018). A meta-analysis and systematic review of reactivity to judgements of learning. *Memory, 26*, 741-750.

Dror, I. E., & Harnad, S. (2008). Offloading cognition onto cognitive technology. In I. E. Dror & S. Harnad (Eds.), *Cognition distributed: How cognitive technology extends our minds* (pp. 1-23). Amsterdam, The Netherlands: John Benjamins.

DuBrow, S., Eberts, E. A., & Murty, V. P. (2019). A common mechanism underlying choice's influence on preference and memory. *Psychonomic Bulletin & Review*, *26*, 1958-1966.

Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2016). Methodology for investigating human metamemory: Problems and pitfalls. In J. Dunlosky & S. K. Tauber (Eds.), *Oxford library of psychology. The Oxford handbook of metamemory* (p. 23-37). Oxford University Press.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning and comprehension: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest, 14*, 4-58.

Dunn, T. L., & Risko, E. F. (2016). Toward a metacognitive account of cognitive offloading. *Cognitive Science*, *40*, 1080-1127.

Egner, T. (2017). *The Wiley handbook of cognitive control*. Hoboken, NJ: Wiley Blackwell.

Einstein, G. O., & Hunt, R. R. (1980). Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 588-598.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-251.

Eskritt, M., & Ma, S. (2014). Intentional forgetting: Note-taking as a naturalistic example. *Memory & Cognition*, *42*, 237-246.

Fass, W., & Schumacher, G. M. (1978). Effects of motivation, subject activity, and readability on the retention of prose materials. *Journal of Educational Psychology*, *70*, 803-807.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Finn, B. (2008). Framing effects on metacognitive monitoring and control. *Memory & Cognition*, *36*, 813-821.

Fowler, R. L., & Barker, A. S. (1974). Effectiveness of highlighting for retention of text material. *Journal of Applied Psychology*, *59*, 358-364.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*, 316-344.

Friedman, M. C., & Castel, A. D. (2011). Are we aware of our ability to forget? Metacognitive predictions of directed forgetting. *Memory & Cognition, 39*, 1448-1456.

Geiselman, R. E., & Bagheri, B. (1985). Repetition effects in directed forgetting: Evidence for retrieval inhibition. *Memory & Cognition, 13*, 57-62.

Geiselman, R. E., Bjork, R. A., & Fishman, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General, 112*, 58-72.

Gilbert, S. J., Bird, A., Carpenter, J. M., Fleming, S. M., Sachdeva, C., & Tsai, P.-C. (2020). Optimal use of reminders: Metacognition, effort, and cognitive offloading. *Journal of Experimental Psychology: General, 149*, 501-517.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15*, 1-16.

Golding, J. M., Long, D. L., & MacLeod, C. M. (1994). You can′t always forget what you want: Directed forgetting of related words. *Journal of Memory and Language, 33*, 493-510.

Greene, R. L. (2008). Repetition and spacing effects. In H. L. Roediger (Ed.), *Learning and memory: A comprehensive reference. Vol. 2: Cognitive psychology of memory* (pp. 65-78). Elsevier.

Grinschgl, S., Meyerhoff, H. S., & Papenmeier, F. (2020). Interface and interaction design: How mobile touch devices foster cognitive offloading. *Computers in Human Behavior*, *108*, 106317.

Grinschgl, S., Meyerhoff, H. S., Schwan, S., & Papenmeier, F. (2021a). From metacognitive beliefs to strategy selection: does fake performance feedback influence cognitive offloading? *Psychological Research*, *85*, 2654-2666.

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021b). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, *74*, 1477-1496.

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, *7*, 464-481.

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

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin Review, 19*, 126-134.

Henkel, L. A. (2014). Point-and-shoot memories: The influence of taking photos on memory for a museum tour. *Psychological Science*, *25*, 396-402.

Hennessee, J. P., Patterson, T. K., Castel, A. D., & Knowlton, B. J. (2019). Forget me not: Encoding processes in value-directed remembering. *Journal of Memory and Language, 106*, 29-39.

Hertel, P. T. (1989). The generation effect: A reflection of cognitive effort? *Bulletin of the Psychonomic Society*, 27, 541-544.

Hertzog, C., McGuire, C. L., & Lineweaver, T. T. (1998). Aging, attributions, perceived control, and strategy use in a free recall task. *Aging, Neuropsychology and Cognition, 5*, 85-106.

Hintzman, D. L. (1974). *Theoretical implications of the spacing effect.* In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp.77-99*).* Lawrence Erlbaum.

Hintzman, D. L. (1976). Repetition and memory. *Psychology of Learning and Motivation, 10*, 47-91.

Hollingshead, A. B. (2001). Cognitive interdependence and convergent expectations in transactive memory. *Journal of Personality and Social Psychology, 81*, 1080-1089.

Hoon, P. W. (1974). Efficacy of three common study methods. *Psychological Reports, 35*, 1057-1058.

Hu, X., Luo, L., & Fleming, S. M. (2019). A role for metamemory in cognitive offloading. *Cognition*, *193*, 104012.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 497-514.

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distinctiveness. *Journal of Memory and Language, 32*, 421-445.

Idstein, P., & Jenkins, J. R. (1972). Underlining versus repetitive reading. *Journal of Educational Research, 65*, 321-323.

Jarosz, A. F., Raden, M. J., & Wiley, J. (2019). Working memory capacity and strategy use on the RAPM. *Intelligence, 77*, 101387.

Johnson, L. L. (1988). Effects of underlining textbook sentences on passage and sentence retention. *Reading Research and Instruction*, *28*, 18-32.

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1250-1257.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773-795.

Kelly, M. O., & Risko, E. F. (2019a). The isolation effect when offloading memory. *Journal of Applied Research in Memory and Cognition*, *8*, 471-480.

Kelly, M. O., & Risko, E. F. (2019b). Offloading memory: Serial position effects. *Psychonomic Bulletin & Review*, *26*, 1347-1353.

Kelly, M. O., & Risko, E. F. (2021). Revisiting the influence of offloading memory on free recall. *Memory & Cognition*, 1-12.

Knowlton, B. J., & Castel, A. D. (2022). Memory and reward-based learning: A value-directed remembering perspective. *Annual Review of Psychology*, *73,* 25-52.

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General, 133*, 643-656.

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*, 36-69.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated study. *Psychonomic Bulletin & Review, 14*, 219-224.

Kornell, N., & Bjork, R. A. (2008). Learning Concepts and Categories: Is Spacing the “Enemy of Induction”? *Psychological Science*, *19*, 585-592.

Kornell, N., Rhodes, M. G., Castel, A. D., & Tauber, S. K. (2011). The ease of processing heuristic and the stability bias: Dissociating memory, memory beliefs, and memory judgments. *Psychological Science, 22*, 787-794.

Leding, J. K. (2019). Adaptive memory: Animacy, threat, and attention in free recall. *Memory & Cognition*, *47*, 383-394.

Leutner, D., Leopold, C., & den Elzen-Rump, V. (2007). Self-regulated learning with a text-highlighting strategy: A training experiment. *Zeitschrift für Psychologie*, *215*, 174-182.

Li, B., Zhao, W., Zheng, J., Hu, X., Su, N., Fan, T., Yin, Y., Liu, M., Yang, C., & Luo, L. (2021). Soliciting judgments of forgetting reactively enhances memory as well as making judgments of learning: Empirical and meta-analytic tests. *Memory & Cognition.*

Little, J. L. (2011). *Optimizing multiple-choice tests as learning events*. University of California, Los Angeles.

Locke, E. A., Frederick, E., Lee, C., & Bobko, P. (1984). Effect of self-efficacy, goals, and task strategies on task performance. *Journal of Applied Psychology, 69,* 241-251.

Lu, X., Kelly, M. O., & Risko, E. F. (2020). Offloading information to an external store increases false recall. *Cognition*, *205*, 104428.

MacLeod, C. (1998). Directed forgetting. In J. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 139-172). Mahwah, NJ: Erlbaum.

MacLeod, C. M. (1975). Long-term recognition and recall following directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory, 1*, 271-279.

Madan, C. R. (2017). Motivated cognition: Effects of reward, emotion, and other motivational factors across a variety of cognitive domains. *Collabra: Psychology, 3*, 24.

Maddox, G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: A case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology, 28*, 684-706.

Maddox, G. B., & Balota, D. A. (2015). Retrieval practice and spacing effects in young and older adults: An examination of the benefits of desirable difficulty. *Memory & Cognition*, *43*, 760-774.

Mandler, J. M. (1984). Representation and recall in infancy. In M. Moscovitch (Ed.), *Infant memory: Its relation to normal and pathological memory in humans and other animals* (pp. 75- 101). New York: Plenum.

Markant, D., DuBrow, S., Davachi, L., & Gureckis, T. M. (2014). Deconstructing the effect of self-directed study on episodic memory. *Memory & Cognition*, *42*, 1211-1224.

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? Learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General, 143*, 94-122.

Marsh, E. J., & Rajaram, S. (2019). The digital expansion of the mind: Implications of internet usage for memory and cognition. *Journal of Applied Research in Memory and Cognition*, *8*, 1-14.

McGillivray, S., & Castel, A. D. (2011). Betting on memory leads to metacognitive improvement by younger and older adults. *Psychology and Aging, 26*, 137-142.

McGillivray, S., & Castel, A. D. (2017). Older and younger adults' strategic control of metacognitive monitoring: The role of consequences, task experience and prior knowledge. *Experimental Aging Research, 43*, 362-374.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 596-606.

Meyerhoff, H. S., Grinschgl, S., Papenmeier, F., & Gilbert, S. J. (2021). Individual differences in cognitive offloading: a comparison of intention offloading, pattern copy, and short-term memory capacity. *Cognitive Research: Principles and Implications*, *6*, 1-18.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167-202.

Mitchum, A. L., Kelley, C. M., & Fox, M. C. (2016). When asking the question changes the ultimate answer: Metamemory judgments change memory. *Journal of Experimental Psychology: General, 145*, 200-219.

Monty, R. A., & Permuter, L. C. (1975). Persistence of the effects of choice on paired-associate learning. *Memory & Cognition, 3*, 183-187.

Morris, P. E., & Fritz, C. O. (2013). Effect sizes in memory research. *Memory*, *21*, 832-842.

Morrison, A. B., & Richmond, L. L. (2020). Offloading items from memory: individual differences in cognitive offloading in a short-term memory task. *Cognitive Research: Principles and Implications*, *5*, 1-13.

Murphy, D. H. (2023). Strategic offloading: How the value of to-be-remembered information influences offloading decision-making. *Applied Cognitive Psychology*.

Murphy, D. H., Agadzhanyan, K., Whatley, M. C., & Castel, A. D. (2021a). Metacognition and fluid intelligence in value-directed remembering. *Metacognition and Learning, 16*, 685-709.

Murphy, D. H., Bjork, R. A., & Bjork, E. L. (2022). Going beyond the spacing effect: Does it matter how time on a task is distributed? *Quarterly Journal of Experimental Psychology*.

Murphy, D. H., & Castel, A. D. (2020). Responsible remembering: How metacognition impacts adaptive selective memory. *Zeitschrift für Psychologie, 228*, 301-303.

Murphy, D. H., & Castel, A. D. (2021a). Metamemory that matters: Judgments of importance can engage responsible remembering. *Memory, 29*, 271-283.

Murphy, D. H., & Castel, A. D. (2021b). Responsible remembering and forgetting as contributors to memory for important information. *Memory & Cognition, 49*, 895-911.

Murphy, D. H., & Castel, A. D. (2022). Selective remembering and directed forgetting are influenced by similar stimulus properties. *Memory, 30*, 1130-1147.

Murphy, D. H., Hoover, K. M., & Castel, A. D. (2023). The role of metacognition when remembering important information: Strategic metacognitive control when engaging in responsible remembering. *Memory & Cognition, 51*, 234-251.

Murphy, D. H., Huckins, S. C., Rhodes, M. G., & Castel, A. D. (2021b). The effect of perceptual processing fluency and value on metacognition and remembering. *Psychonomic Bulletin & Review*.

Murty, V. P., DuBrow, S., & Davachi, L. (2015). The simple act of choosing influences declarative memory. *Journal of Neuroscience, 35*, 6255-6264.

Murty, V. P., DuBrow, S., & Davachi, L. (2019). Decision-making increases episodic memory via postencoding consolidation. *Journal of Cognitive Neuroscience*, *31*, 1308-1317.

Myers, S. J., Rhodes, M. G., & Hausman, H. E. (2020). Judgments of learning (JOLs) selectively improve memory depending on the type of test. *Memory & Cognition, 48*, 745-758.

Nairne, J. S. (2010). Adaptive memory: Evolutionary constraints on remembering. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 53, pp. 1-32). Burlington: Academic Press.

Nairne, J. S. (2013). Adaptive memory: Controversies and future directions. In B. L. Schwartz, M. L. Howe, M. P. Toglia, & H. Otgaar (Eds.), *What is adaptive about adaptive memory?* (pp 308-321). New York: Oxford University Press.

Nairne, J. S. (2015). Adaptive memory: Novel findings acquired through forward engineering. In D. S. Lindsay, C. M. Kelley, A. P. Yonelinas, & H. L. Roediger (Eds.), *Remembering: Attributions, processes, and control in human memory*. New York: Psychology Press.

Nairne, J. S. (2016). Adaptive memory: Fitness-relevant "tunings" help drive learning and remembering. In D. C. Geary & D. B. Berch (Eds.), *Evolutionary perspectives on child development and education* (pp. 251–269). Basel, Switzerland: Springer International Publishing.

Nairne, J. S., & Pandeirada, J. N. S. (2008). Adaptive memory: Remembering with a stone-age brain. *Current Directions in Psychological Science, 17*, 239-243.

Nairne, J. S., & Pandeirada, J. N. S. (2010). Adaptive memory: Ancestral priorities and the mnemonic value of survival processing. *Cognitive Psychology, 61*, 1-22.

Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 263-273.

Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science, 26*, 22-27.

Nairne, J. S., VanArsdall, J. E., Pandeirada, J. N. S., Cogdill, M., & LeBreton, J. M. (2013). Adaptive memory: The mnemonic value of animacy. *Psychological Science, 24*, 2099-2105.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*, 102-116.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of Experimental Psychology. Learning, Memory, and Cognition, 14*, 676-686.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), (p. 80) *The psychology of learning and motivation* (pp. 125-173). New York, NY: Academic Press.

Nist, S. L., & Hogrebe, M. C. (1987). The role of underlining and annotating in remembering textual information. *Reading Research and Instruction*, *27*, 12-25.

Nist, S. L., & Simpson, M. L. (1988). The effectiveness and efficiency of training college students to annotate and underline text. *National Reading Conference Yearbook*, *37*, 251-257.

Paivio, A. (1971). Imagery and language. In *Imagery* (pp. 7-32). Academic Press.

Paivio, A. (2013). Dual coding theory, word abstractness, and emotion: A critical review of Kousta et al. (2011). *Journal of Experimental Psychology: General, 142*, 282-287.

Park, J. S., Kelly, M. O., Hargis, M. B., & Risko, E. F. (2022). The effect of external store reliance on actual and predicted value-directed remembering. *Psychonomic Bulletin & Review, 29*, 1367-1376.

Perlmuter, L., Monty, R. A., & Kimble, G. A. (1971). Effect of choice on paired-associate learning. *Journal of Experimental Psychology, 91*, 47-53.

Peterson, L. R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. *Journal of Experimental Psychology, 66*, 206-209.

Peterson, S. E. (1992). The cognitive functions of underlining as a study technique. *Reading Research and Instruction*, *31*, 49-56.

Popov, V., & Reder, L. (in press). Frequency effects in recognition and recall. In E., Tulving, & F. I. M. Craik, (Eds.). *The Oxford Handbook of Memory.* Oxford University Press.

Popp, E. Y., & Serra, M. J. (2016). Adaptive memory: Animacy enhances free recall but impairs cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 186-201.

Popp, E. Y., & Serra, M. J. (2018). The animacy advantage for free recall performance is not attributable to greater mental arousal. *Memory, 26*, 89-95.

Racsmány, M., & Conway, M. A. (2006). Episodic inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 44- 57.

Rajaram, S., & Pereira-Pasarin, L. P. (2010). Collaborative memory: Cognitive research and theory. *Perspectives on Psychological Science*, *5*, 649-663.

Raven, J. C. (1938). *Progressive matrices: A perceptual test of intelligence*. London, England: H. K. Lewis.

Rhodes, M. G. (2016). Judgments of learning. In J. Dunlosky and S. K. Tauber (Eds.), *The Oxford handbook of metamemory* (pp. 65-80). New York: Oxford University Press.

Richardson, J. T. E. (1998). The availability and effectiveness of reported mediators in associative learning: A historical review and an experimental investigation. *Psychonomic Bulletin & Review, 5*, 597-614.

Rickards, J., & August, G. (1975). Generative underlining strategies in prose recall. *Journal of Educational Psychology*, *67*, 860-865.

Rickards, J., & Denner, P. R. (1979). Depressive effects of underlining and adjunct questions on children's recall of text. *Instructional Science*, *8*, 81-90.

Risko, E. F., & Dunn, T. L. (2015). Storing information in-the-world: Metacognition and cognitive offloading in a short-term memory task. *Consciousness and Cognition, 36,* 61-74.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences, 20*, 676-688.

Risko, E. F., Kelly, M. O., Patel, P., & Gaspar, C. (2019). Offloading memory leaves us vulnerable to memory manipulation. *Cognition*, *191*, 103954.

Rivers, M. L., Janes, J. L., & Dunlosky, J. (2021). Investigating memory reactivity with a within-participant manipulation of judgments of learning: support for the cue-strengthening hypothesis. *Memory*, 1-12.

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

Rotem-Turchinski, N., Ramaty, A., & Mendelsohn, A. (2019). The opportunity to choose enhances long-term episodic memory. *Memory*, *27*, 431-440.

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1064-1072.

Schmidt, S. R. (1992). Evaluating the role of distinctiveness in the generation effect. *Quarterly Journal of Experimental Psychology*, 44A, 237-260.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27*, 499-520.

Serra, M. J. (2021). Animate and inanimate words demonstrate equivalent retrieval dynamics despite the occurrence of the animacy advantage. *Frontiers in Psychology*, *12*, 2027.

Serra, M. J., & England, B. D. (2012). Magnitude and accuracy differences between judgments of remembering and forgetting. *Quarterly Journal of Experimental Psychology, 65*, 2231-2257.

Shaughnessy, J. J., Zimmerman, J., & Underwood, B. J. (1972). Further evidence on the MP-DP effect in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 11*, 1-12.

Sheard, E. D., & MacLeod, C. M. (2005). List method directed forgetting: Return of the selective rehearsal account. In N. Ohta, C. M. Mac- Leod, & B. Uttl (Eds.), *Dynamic cognitive processes* (pp. 219-248). Tokyo: Springer.

Silvers, V. L., & Kreiner, D. S. (1997). The effects of pre-existing inappropriate highlighting on reading comprehension. *Reading Research and Instruction*, *36*, 217-223.

Smith, S. M., Glenberg, A. M., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6*, 342-353.

Soderstrom, N. C., Clark, C. T., Halamish, V., & Bjork, E. L. (2015). Judgments of learning as memory modifiers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 553-558.

Soderstrom, N. C., & McCabe, D. P. (2011). The interplay between value and relatedness as bases for metacognitive monitoring and control: Evidence for agenda-based monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1236-1242.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, *333*, 776-778.

Spellman, B. A., & Bjork, R. A. (1992). When predictions create reality: Judgments of learning may alter what they are intended to assess. *Psychological Science, 5*, 315-316.

Staff, R. T., Hogan, M. J., & Whalley, L. J. (2014). Aging trajectories of fluid intelligence in late life: The influence of age, practice and childhood IQ on Raven's Progressive Matrices. *Intelligence, 47,* 194-201.

Stordahl, K. E., & Christensen, C. M. (1956). The effect of study techniques on comprehension and retention. *Journal of Educational Research, 49*, 561-570.

Storm, B. C., & Stone, S. M. (2015). Saving-enhanced memory: The benefits of saving on the learning and remembering of new information. *Psychological Science*, *26*, 182-188.

Storm, B. C., Stone, S. M., & Benjamin, A. S. (2017). Using the Internet to access information inflates future use of the Internet to access other information. *Memory*, *25*, 717-723.

Tan, P., Ensor, T. M., Hockley, W. E., Harrison, G. W., & Wilson, D. E. (2020). In support of selective rehearsal: Double-item presentation in item-method directed forgetting. *Psychonomic Bulletin & Review, 27*, 529-535.

Tauber, S. K., & Rhodes, M. G. (2012). Measuring memory monitoring with judgments of retention interval (JOR). *Quarterly Journal of Experimental Psychology, 65*, 1376-1396.

Tekin, E., & Roediger, H. L. (2020). Reactivity of judgments of learning in a levels-of-processing paradigm. *Zeitschrift für Psychologie*.

Toppino, T. C., & Gerbier, E. (2014). About practice: Repetition, spacing, and abstraction. In *Psychology of learning and motivation* (Vol. 60, pp. 113-189). Academic Press.

Underwood, B. J. (1969). Some correlates of item repetition in free recall learning. *Journal of Verbal Learning and Verbal Behavior, 9*, 573-580.

Unsworth, N. (2016). Working memory capacity and recall from LTM: Examining the influences of encoding strategies, study time allocation, search efficiency, and monitoring abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 50-61.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*, 104-132.

Unsworth, N., Robison, M. K., & Miller, A. L. (2021). Individual differences in lapses of attention: A latent variable analysis. *Journal of Experimental Psychology: General*, *150*, 1303-1331.

Verde, M. F. (2012). *Retrieval-induced forgetting and inhibition: A critical review.* In B. H. Ross (Ed.), *The psychology of learning and motivation: Vol. 56. The psychology of learning and motivation* (p. 47-80). Elsevier Academic Press.

Voss, J. L., Gonsalves, B. D., Federmeier, K. D., Tranel, D., & Cohen, N. J. (2011). Hippocampal brain-network coordination during volitional exploratory behavior enhances learning. *Nature Neuroscience, 14*, 115-120.

Wade, S. E., & Trathen, W. (1989). Effect of self-selected study methods on learning. *Journal of Educational Psychology, 81*, 40-47.

Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, *63*, 410-424.

Watanabe, T., & Soraci, S. A. (2004). The self-choice effect from a multiple-cue perspective. *Psychonomic Bulletin & Review, 11*, 168-172.

Wegner, D. M. (1986). Transactive memory: A contemporary analysis of the group mind. In B. Mullen & G. R. Goethals (Eds.), *Theory of group behavior (pp.* 185-208). New York: Springer-Verlag.

Weiner, B., & Reed, H. (1969). Effects of the instructional sets to remember and to forget on short-term retention: Studies of rehearsal control and retrieval inhibition (suppression). *Journal of Experimental Psychology, 79*, 226-232.

Weis, P. P., & Wiese, E. (2019). Problem solvers adjust cognitive offloading based on performance goals. *Cognitive Science*, *43*, e12802.

Witherby, A. E., & Tauber, S. K. (2017). The influence of judgments of learning on long-term learning and short-term performance. *Journal of Applied Research in Memory and Cognition*, *6*, 496-503.

Woodward, A. E., & Bjork, R. A. (1971). Forgetting and remembering in free recall: Intentional and unintentional. *Journal of Experimental Psychology, 89*, 109-116.

Yue, C. L., Storm, B. C., Kornell, N., & Bjork, E. L. (2015). Highlighting and its relation to distributed study and students' metacognitive beliefs. *Educational Psychological Review*, *27*, 69-78.

Zimmerman, J. (1975). Free recall after self-paced study: A test of the attention explanation of the spacing effect. *The American Journal of Psychology, 88*, 277-291.