Journal of Experimental Psychology: General

The Metacognitive Optimization of Offloading Task (MOOT): Both Higher Costs to Offload and the Accuracy of Memory Predict Goodness of Offloading Performance

Dillon H. Murphy and Janet Metcalfe

Online First Publication, January 23, 2025. https://dx.doi.org/10.1037/xge0001726

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001726

The Metacognitive Optimization of Offloading Task (MOOT): Both Higher Costs to Offload and the Accuracy of Memory Predict Goodness of Offloading Performance

Dillon H. Murphy[1] and Janet Metcalfe[2]
[1] Department of Psychology, University of California, Riverside
[2] Department of Psychology, Columbia University

We developed a Metacognitive Offloading Optimization Task (MOOT) whereby participants were instructed to score as many points as possible by accessing words from a presented list either by remembering them (worth 10 points each) or by offloading them (worth less than 10 points each). Results indicated that participants were sensitive to the value of the offloaded items such that when offloaded items carried a high value (e.g., 8 points each), participants' scores were lower than if they had chosen to offload all items. Conversely, when offloaded items had a low value (e.g., 2 points each), participants' scores exceeded what they would have achieved had they offloaded all items. In Experiments 2 and 3, we investigated offloading optimality. Specifically, because each individual's maximum possible score depended on how much they could remember, each participant's memory ability was assessed in a pretest. The maximum score obtainable resulted from a strategy in which the participant opts to recall every item that they will be able to remember (obtaining 10 points for each) and offloads all other items (obtaining a value greater than 0 points for each), leaving no items unrecalled and not offloaded. To implement this strategy, the participant needs to have and use metaknowledge of exactly which items they will be able to recall. In each experiment, the MOOT scores—the ratio of participants' observed score to their maximum possible score—were closer to optimal for participants with better memory ability.

Public Significance Statement

The present study introduces a task called the Metacognitive Offloading Optimization Task (MOOT) which explores how individuals use external aids to enhance memory performance. The findings revealed that people often struggle to offload information optimally, especially when the costs associated with offloading are low. By understanding the conditions under which individuals are more or less effective at offloading information, this research can help improve educational and technological tools designed to aid memory. For the general public, the insights from this study highlight the importance of metacognitive awareness—our understanding of our own memory processes—in making effective use of technologies like smartphones and computers for remembering information. This work emphasizes the potential for tailored strategies that help individuals maximize their cognitive resources in both academic and everyday settings.

Keywords: offloading, metacognition, decision making, memory, value

The use of technology to offload (i.e., using an external store to remember information) information has become ubiquitous in everyday life (see Barr et al., 2015; Dror & Harnad, 2008; see also Storm et al., 2017). This can be beneficial since, compared with relying on memory alone, offloading information can reduce cognitive load and ensure access to a greater amount of information when it is needed (Cherkaoui & Gilbert, 2017; Gilbert et al., 2023; Henkel, 2014; Risko & Dunn, 2015; Risko & Gilbert, 2016;

Ashleigh Maxcey served as action editor.

Dillon H. Murphy https://orcid.org/0000-0002-5604-3494

The experiments reported in this article were not preregistered, but the stimuli and data have been made available on the Open Science Framework available at https://osf.io/ynhev/?view_only=1e6e1ecc0b15490b8d8cef70c4efb87c. The authors certify that they have no affiliations with or involvement in any organization or entity with any financial or nonfinancial interest in the subject matter or materials discussed in this article. The ideas and data appearing in the article have not been disseminated before (e.g., at a conference or meeting, posted on a listserv, shared on a website). The authors thank Drew Murphy for his assistance with coding the task.

Dillon H. Murphy played a lead role in data curation, formal analysis, writing–original draft, and writing–review and editing and an equal role in conceptualization. Janet Metcalfe played a lead role in supervision, a supporting role in writing–original draft and writing–review and editing, and an equal role in conceptualization.

Correspondence concerning this article should be addressed to Dillon H. Murphy, Department of Psychology, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, United States. Email: dillonm@ucr.edu

Sparrow et al., 2011; Storm & Stone, 2015; Weis & Wiese, 2019). Although offloading may sometimes come with distinct drawbacks—information that is offloaded tends to be more poorly remembered than information that is not offloaded (Eskritt & Ma, 2014; Lu et al., 2020; Marsh & Rajaram, 2019; Sparrow et al., 2011), and access to offloaded information may suffer should the external store unexpectedly become unavailable (Kelly & Risko, 2019a, 2019b; see also Kelly & Risko, 2022)—in this article, we focus on whether people can use (reliable) offloading effectively to optimize access to information that might otherwise be unavailable.

Given the drawbacks of offloading, it may be advantageous to use external stores selectively, as offloading all information could lead to inefficiencies and time-consuming searches through the external store to retrieve information. As such, effective offloading depends on the learner having accurate metacognitive knowledge of what they will remember without offloading. Ideally, if the person has metaknowledge of exactly which items they will be able to access and which they will not, they will offload only the latter items. Additionally, optimal offloaders need well-founded decision-making processes to allow for the evaluation and implementation of advantageous strategies that maximize performance given both the individual's memory constraints and any costs of offloading. Thus, the costs of offloading as compared to remembering, but also the value of the information, will come into play.

When presented with too much information to remember, learners typically better remember highly valued or important information relative to information of low value (Castel et al., 2002; Murphy, 2023a; see Knowlton & Castel, 2022; Madan, 2017, for a review). In attempting to maximize the utility of offloading, a similar reasonable strategy would be to prioritize valuable information (e.g., Murphy & Castel, 2023). Specifically, assuming the external store is reliable, whereas memory is fallible (e.g., Schacter, 1999, 2022), offloading information may provide a greater likelihood of information accessibility than attempting to remember it. For instance, if a person writes something down on their phone, they are probably more likely to have access to that information later than if they were only to try to remember it. However, if the external store is unreliable or at least less reliable than memory (e.g., if the phone is low on battery), it could be preferable to prioritize important information in memory to ensure the greatest likelihood of access to this information (see Murphy, 2023b). Despite the risks, prior work has demonstrated that learners do not selectively attempt to encode valuable information when they are told they can rely on an external store instead (Park et al., 2022). But should that external store become surprisingly unavailable, the result can be a selective and dysfunctional lack of accessibility of valuable information.

Maximizing the accessibility of needed information depends not only on an assessment of the reliability of the offloading device itself but also on an assessment of our memory. Metacognition, or the awareness of one's memory abilities, is typically considered in terms of monitoring and control (Nelson & Narens, 1990; see also Dunlosky et al., 2016; Nelson, 1996; see Rhodes, 2016, for a review). Prospective metacognitive monitoring involves evaluating the likelihood of later remembering some information, whereas metacognitive control involves the study decisions learners make based on their monitoring assessments. The latter also involves keeping the current situation, constraints, and goals—the agenda—in mind (Ariel et al., 2009, 2015; Bjork et al., 2013; DeLozier & Dunlosky, 2015). Prior work indicates that learners are metacognitively aware of their selective memory for more valuable or important information (Murphy et al., 2021; see also Murphy et al., 2022) and that metacognitive beliefs influence offloading behavior (Boldt & Gilbert, 2019; Dunn & Risko, 2016; Gilbert et al., 2020; Grinschgl et al., 2021; Murphy et al., 2024). Thus, metacognition is likely to play a crucial role in optimizing what is offloaded—the evaluation of one's memory in coordination with an appraisal of the value of the information is needed to allow effective offloading.

When making offloading decisions, learners need to monitor what information has been successfully encoded and what information will not be remembered later (Thiede, 1999). The information that has been successfully encoded and will be retrievable from memory presumably does not need to be offloaded, but offloading information that will not be remembered could afford later accessibility to otherwise lost information, provided that the external store is reliable. Thus, a person's knowledge of what will and will not be remembered should play a crucial role in offloading behavior, allowing strategic control of both cognitive and offloading resources. In sum, an effective offloader should be aware of what information will be remembered and offload information that will not be remembered (see also Tsai et al., 2023, for an examination of optimal offloading).

To maximize the utility of an external store when there is a cost associated with offloading as compared to remembering, individuals should offload the information that is most difficult to remember and use memory for the information that is easier to remember (assuming all of the information needs to be remembered; in real-world situations, the decision to offload information should be influenced not only by how difficult it is to remember but also by how likely it is that the information will be needed, i.e., how valuable it is). When studying lists of words, many stimulus characteristics contribute to a given item's memorability including concreteness, frequency, emotionality, word length, and animacy (see Murphy & Castel, 2022, for a recent demonstration of these effects)—these factors should play a role in people's offloading decisions. Narrowing in on animacy, animate items (e.g., animals) tend to be better remembered than inanimate items (e.g., objects), and this 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 selective offloading, using an external store to retain the most difficult items (in this case, the inanimate objects) and using memory to store the easiest items (in this case, the animate objects) would be a simple strategy that could increase the total amount of information that is accessible later.

Prior work on offloading conducted in the prospective memory domain (e.g., Ball et al., 2022; Fröscher et al., 2022; Gilbert et al., 2020; Sachdeva & Gilbert, 2020; Tsai et al., 2023) has shown that individuals' underconfidence in their memory capabilities may sometimes influence their decision to rely on external tools such as reminders. This previous work has also indicated that people may sometimes offload, even though it might not be strictly necessary, to avoid cognitive effort. It seems plausible that such effort avoidance might be especially likely if the value of the information or the cost of offloading were low. Other personality-related factors, such as risk aversion (Kimball, 1993) or the need for control (Leotti et al., 2010), may also contribute to people's offloading decisions.

In the present study, while not denying the importance of such motivational factors, our goal was to investigate whether people are capable of effectively balancing memory and offloading in an emotionally neutral and well-specified situation. To this end, we devised the Metacognitive Offloading Optimization Task (MOOT) in which participants "paid" to offload an item (i.e., offloaded items were worth less than remembered items). In our task, participants studied lists of words for a later test and were told that their job was to score as many points as possible by accessing these words on that test. They could do so in any combination of (a) remembering the words—in which case each was worth 10 points—or (b) deciding at study to offload particular words and access them in that manner at test. However, if the participant chose to offload an item, the accessed word would be worth less than 10 points (the amount less was specified in each case). Words that participants were unable to recall at test and had not offloaded were worth 0 points.

Participants were presented with a list of 20 words, one at a time, and told they would be awarded 10 points for every item recalled from memory but had to "pay" to offload an item (e.g., making offloaded items worth only, say, 5 points). If a participant remembered none of the items and offloaded none of them, they would get a score of 0. If they remembered none of the items but offloaded all of them, they would score $5 \times 20 = 100$ points. If they had a perfect memory and they offloaded none of the items, their score could potentially be the maximum of $10 \times 20 = 200$. Critically, if a participant knew exactly which items they would remember and which they would forget, their best strategy would be to offload only those items that they knew they would fail to remember and get 5 points (rather than 0 points) for each of them and remember the remainder, getting 10 points for each. This optimal strategy would result in exactly 20 items being either offloaded or remembered—no items would be scored as a 0 for not being accessible at test. This strategy would yield a score somewhere between 100 and 200, depending on how good the person's memory and metacognition were.

If the cost of offloading an item was lower than 5 points, say the offloaded items were worth 8 points per item instead of only 5, then participants—even those who could remember nothing—could score 160 points if they simply offloaded everything. Participants could beat 160, but only if they remembered and offloaded the right items (those that they knew they would not recall on the test). Beating the score they would attain by offloading everything would depend upon their metacognitive accuracy. If the price of offloading were higher, say offloaded items were worth only 2 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—there is a cost–benefit trade-off. Given the risk of not offloading a word and forgetting it on the test, as the price for offloading decreases (i.e., the value of the offloaded items increases), the criterion for offloading should shift toward offloading more items.

In addition to the cost of offloading each item, we conjectured that offloading decisions would reflect the effects of stimulus characteristics on the likelihood of remembering to the extent that the participants realized that they would be more likely to remember particular classes of items. Since animate objects are usually better remembered than inanimate objects (e.g., Popp & Serra, 2016, 2018), if a to-be-remembered list contains animate and manmade items, participants should tend to opt to remember the more memorable, animate items and offload the less memorable, inanimate items. Such a strategy would maximize the number of items stored in memory and could increase the participant's score.

Experiment 1

In Experiment 1, learners were presented with lists of 20 words (10 animate, 10 inanimate) to access on a later test. Each participant's goal was to maximize their score on each list by accessing studied words on the test. In addition to using their memory, we allowed participants to offload words of their choosing but when they did so, those words became less valuable (each offloaded word was either worth 8 points, 5 points, or 2 points; participants completed multiple lists of each scoring type). We hypothesized that participants would offload items more frequently when the associated costs were lower (i.e., when offloaded words were worth higher point values) and that they would offload more inanimate items relative to animate items, informed by intrinsic cues that signal the relative difficulty of remembering specific types of items (see Koriat, 1997).

Method

Transparency and Openness

We report an analysis of our sample size and describe all data exclusions, manipulations, and measures in the study. All data and research materials are available on the Open Science Framework at https://osf.io/ynhev/?view_only=1e6e1ecc0b15490b8d8cef70c4efb87c (Murphy, 2024). Data were analyzed using JASP and Jamovi, and all information needed to reproduce the analyses is available. This study's design and its analysis were not preregistered. Informed consent was acquired, and the studies were completed in accordance with the University of California Los Angeles (UCLA) Institutional Review Board (Memory, Attention, Emotion and Aging: IRB No. 12-000617) and the Arizona State University Institutional Review Board (Online Cognitive Ability Testing: STUDY00018136).

Participants

Participants were 43 undergraduate students ($M_{\text{age}} = 22.42$, $SD_{\text{age}} = 5.56$) recruited from the UCLA Human Subjects Pool. In each experiment, participants were tested online, received course credit for their participation, and were excluded from analysis if they admitted to cheating (e.g., writing down answers) in a posttask questionnaire. To try to increase their forthrightness, participants were told they would still receive credit for the experiment even if they cheated. This exclusion process resulted in no exclusions.

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 the Open Science Framework at https://osf.io/ynhev/?view_only=1e6e1ecc0b15490b8d8cef70c4efb87c.

Procedure

After completing the informed consent process, participants completed a demographic survey, where they reported their gender (male, female, other), age, level of education (some high school, high school graduate, some college [no degree], associate's degree, bachelor's degree, graduate degree [master's, doctorate, etc.]), ethnicity (American Indian/Alaskan Native, Asian/Pacific Islander, Black, Hispanic, White, other/unknown), country of origin, and English language fluency. Participants were presented with nine lists of 20 unique to-be-remembered words. The stimulus words were presented for 6 s each. After the presentation of all 20 words in each list, participants were given a self-paced, immediate free recall/access test in which they had to recall or otherwise access as many words as they could from the list. Participants were told that each recalled word was worth 10 points, that they earned points by recalling/accessing the words on the test, and that they should try to maximize their scores.

During the presentation phase of 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 all offloaded words were provided to participants on the recall test (see Figure 1 for an example of the study and test phase). Critically, participants were told that if they saved a word, it would be worth fewer points at test, and participants were given the specific values of offloaded words before 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. Thus, on three lists, offloaded items were worth 8 points; on three lists, offloaded items were worth 5 points; and on three lists, offloaded items were worth 2 points. List order was counterbalanced, but each list scoring type occurred in blocks (i.e., the three lists where offloaded words were worth 8 points occurred consecutively). Immediately following the recall/access period, participants were told their score for that list but were not given feedback about specific items.

Results

Offloading and Recall

The number of items offloaded as compared to recalled from memory as a function of list scoring is shown in Figure 2. A 3 (list type: offloaded words worth 8, 5, 2) × 2 (animacy: animate, inanimate) repeated-measures analysis of variance (ANOVA) revealed differences between list types, $F(2, 84) = 12.73$, $p < .001$, $\eta_p^2 = .23$, such that participants offloaded more items per list when offloaded items were worth 8 points ($M = 9.20$, $SD = 7.44$) than when they were worth 5 points ($M = 7.12$, $SD = 6.23$), $p_{\text{holm}} = .020$, $d = .30$, and when they were worth 2 points ($M = 5.22$, $SD = 6.17$), $p_{\text{holm}} < .001$, $d = .58$; participants also offloaded more when offloaded items were worth 5 points than when they were worth 2 points ($p_{\text{holm}} = .020$, $d = .28$). Thus, as predicted, offloading decreased as the cost of offloading increased. The effect of animacy did not quite reach (two-tailed) significance, $F(1, 42) = 3.29$, $p = .077$, $\eta_p^2 = .07$: A similar number of animate items were offloaded per list ($M = 3.36$, $SD = 2.90$) as inanimate items ($M = 3.82$, $SD = 3.25$). This effect, though not statistically significant, is shown in Figure 3a to allow comparison to the parallel results in Experiment 2. List scoring did not interact with animacy, $F(2, 84) = .02$, $p = .983$, $\eta_p^2 < .01$.

To examine words recalled from memory alone (i.e., not including offloaded words), we conducted a 3 (list type: offloaded words worth 8, 5, 2) × 2 (animacy: animate, inanimate) repeated-measures ANOVA. As is shown in Figure 3b, results revealed differences between list types, $F(2, 84) = 8.10$, $p < .001$, $\eta_p^2 = .16$, such that participants recalled fewer items per list when offloaded items were worth 8 points ($M = 7.17$, $SD = 5.32$) than when they were worth 5 points ($M = 8.71$, $SD = 5.29$), $p_{\text{holm}} = .018$, $d = .27$, and when they were worth 2 points ($M = 9.43$, $SD = 5.53$), $p_{\text{holm}} < .001$, $d = .40$; the difference in recall when items were worth 5 points as compared to when they were worth 2 points was not significant ($p_{\text{holm}} = .208$, $d = .13$).[1] There was an effect of animacy, $F(1, 42) = 4.75$, $p = .035$, $\eta_p^2 = .10$, such that more animate items were recalled per list ($M = 4.44$, $SD = 2.47$) than inanimate items ($M = 4.00$, $SD = 2.62$). List scoring did not interact with animacy, $F(2, 84) = .08$, $p = .925$, $\eta_p^2 < .01$.

It is notable that the sum of offloaded and recalled words was less than 20 in all treatment combinations, indicating that people often failed to offload or remember some of the words in the list—a failure of metacognition. Specifically, the proportion of words that were neither recalled nor offloaded was .18 (.19), .21 (.18), and .27 (.22) for the offload worth 8, 5, and 2 conditions, respectively. The failure to either recall or offload a word is, of course, a violation of optimality.

Point Scores

The scores evaluated here were the actual number of points attained for each list as a function of list scoring (see Figure 4). A repeated-measures ANOVA with three levels (list scoring: offloaded words worth 8, 5, 2) revealed differences between list types, Mauchly's $W = .59$, $p < .001$; Huynh–Feldt corrected results: $F(1.45, 60.80) = 22.33$, $p < .001$, $\eta_p^2 = .35$, such that, as expected, participants' point scores were higher when offloaded words were worth 8 points ($M = 145.32$, $SD = 28.24$) than when they were worth 5 points ($M = 122.64$, $SD = 32.51$), $p_{\text{holm}} < .001$, $d = .62$), and when they were worth 2 points ($M = 104.78$, $SD = 47.12$), $p_{\text{holm}} < .001$, $d = 1.10$; participants also scored higher when offloaded words were worth 5 points than when they were worth 2 points ($p_{\text{holm}} = .004$, $d = .49$).

Finally, we examined how participants' point scores compared with the scores they would have obtained if they had simply offloaded every word. As is shown in Figure 4, on lists in which offloaded words were worth 8 points, a one-sample t test indicated that participants' mean scores were significantly *lower* than what they would have obtained if they had offloaded every word (160), $t(42) = -3.44$, $p = .001$, $d = -.52$. On lists in which offloaded words were worth 5 points, scores were *greater* than what they would have obtained if they had offloaded every word (100), $t(42) = 4.57$, $p < .001$, $d = .70$. On lists where offloaded words were worth 2 points, participants' scores were, again, *greater* than what they would have obtained if they had offloaded every word (40), $t(42) = 9.01$, $p < .001$, $d = 1.38$.

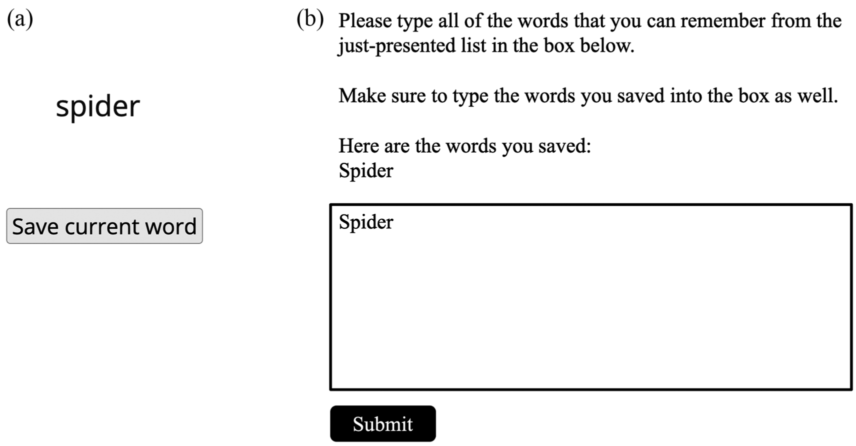
Discussion

In Experiment 1, results revealed that participants were more likely to offload items when the cost of offloading was lower (i.e., if offloaded items were worth 8 points). However, this difference in

[1] Since subjects could offload, however, many words they wanted, each subject had a different number of words available to be recalled (e.g., if one subject offloaded five words, they could recall a maximum of 15, but if another subject offloaded 10 words, they could only recall a maximum of 10 words). There were no differences in the proportion of not-offloaded words that were recalled between the three list types, $F(2, 70) = .34$, $p = .714$, $\eta_p^2 = .01$, these were .66 (.26), .65 (.27), and .63 (.25) for the offload worth 8, 5, and 2 conditions, respectively.

Figure 1

Example of the Study (a) and Test (b) Phase in Each Experiment

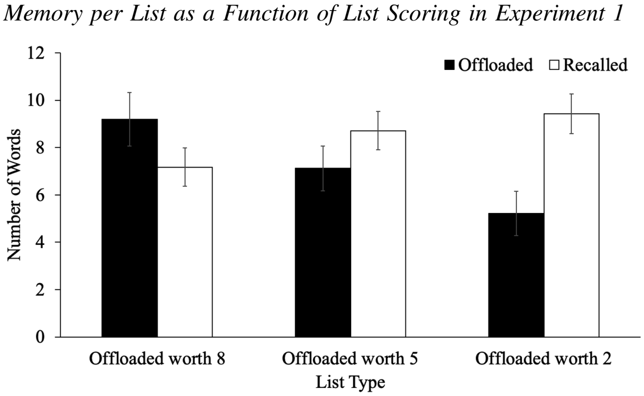


Note. Example of the study (a) and test (b) phase in each experiment (note that in Experiment 3; see the Open Science Framework at https://osf.io/ynhev/?view_only=1e6e1ecc0b15490b8d8cef70c4efb87c, offloaded words did not appear during the test, and participants did not need to type them into the box). Participants pressed the "Save current word" button to add the currently presented word to the external store. During the recall/access test, the words that were saved appeared above the text box that participants used to recall words (in Experiments 1 and 2 only).

offloading as a function of value was not significantly impacted by whether words were animate or inanimate, although animate items were better recalled than inanimate items. Conversely, participants recalled more words when the cost of offloading was higher (i.e., if offloaded items were worth 2 points). This may have resulted, in part, because they offloaded less often, leaving a larger pool of words that counted toward being recalled from memory. Critically, only when the cost of offloading was relatively high (i.e., when offloaded words were worth only 5 or 2 points) did participants achieve a point score that was higher than they would have obtained had they simply offloaded all words. When the cost of offloading was low (i.e., the offloaded words were worth 8 points), participants achieved a point score that was worse than they would have attained if they had simply offloaded all words and not used their memory at all. These findings indicate that participants took value into account in their offloading behavior.

Figure 2

The Average Number of Words Offloaded or Recalled From Memory per List as a Function of List Scoring in Experiment 1



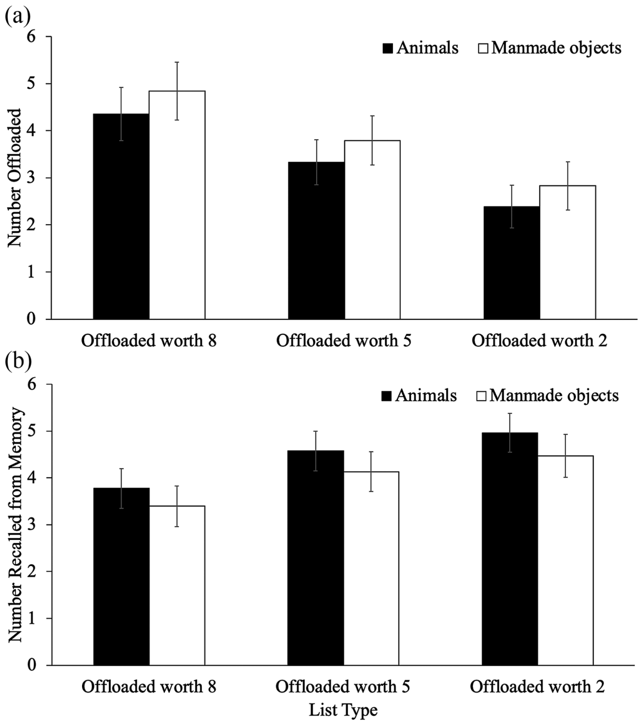
Note. Error bars reflect the standard error of the mean.

In Experiment 1, we were unable to determine the optimality of participants' offloading behavior since defining optimality requires considering each individual's memory capabilities. If an individual had no memory at all, then the optimal strategy—that which would produce that person's own best score—would be to offload everything. Their maximum point score would be 20 times the value of the offloaded items (40 when the items were worth 2, 100 when the items were worth 5, and 160 when the items were worth 8). In contrast, if an individual were capable of remembering everything, that person's optimal strategy (and the strategy that would award them with the maximum point score of 200, regardless of what offloaded items were worth) would be to offload nothing and remember everything. People with memory capabilities between these two extremes could optimize their own strategy by choosing to offload only and exactly those items that they would not remember. However, their maximum scores would differ depending on how many items they could remember. Choosing to remember (rather than offload) those exact items that they would later be able to remember depends on knowing what they are; it depends upon the person's accurate prospective metacognition about what they will remember and what they will not remember.

The MOOT task, in conjunction with an estimate of each individual's memory capacity, provides the possibility of determining what each person's own maximum possible score is if they were to coordinate remembering and offloading decisions as effectively as possible for their own memory capacity. By contrasting that maximum possible score with their actual score, we could potentially get a measure of the goodness of their offloading strategy, but doing so necessitates our knowing how much they could remember without offloading. Because of the lack of an independent estimate of each participant's memory in Experiment 1, we were unable to determine

Figure 3

The Average Number of Words Offloaded (a) and Recalled From Memory (b) per List as a Function of List Scoring and Animacy in Experiment 1

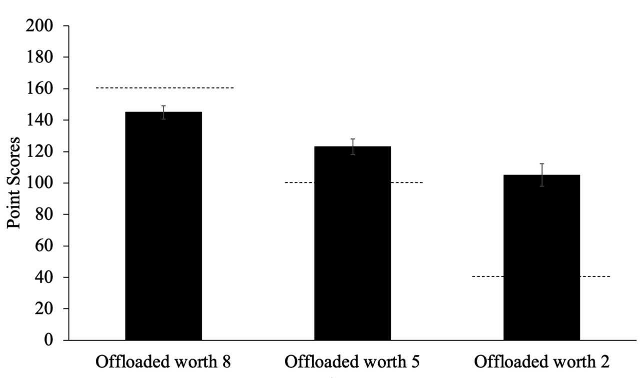


Note. Error bars reflect the standard error of the mean.

the maximum point score for each participant. Experiment 2 was designed to overcome this limitation by attaining an estimate of each participant's memory ability. This estimate allowed us to compute what their maximum score could be, given their own memory constraints, and to assess how close they came to attaining this individualized maximum score. In short, it allowed us to compute a MOOT score—a measure of the optimality of their memory offloading decisions.

Figure 4

Average Point Scores as a Function of List Scoring in Experiment 1



Note. Dashed lines represent the score participants would have obtained if they had offloaded every word. Error bars reflect the standard error of the mean.

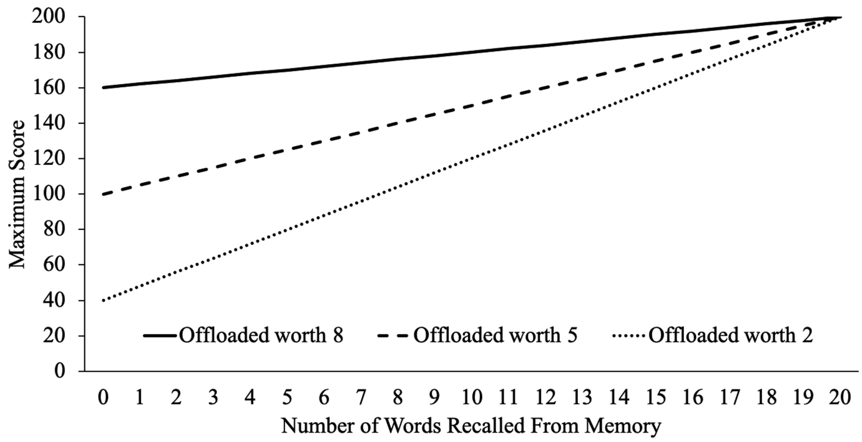
Experiment 2

As noted above, there is an optimal solution for each participant's offloading decisions if they are metacognitively perfect. Unless a given participant can recall all 20 words, their maximum is not 200. Instead, the maximum point score attainable depends on how many items the participant can remember, but it also depends on how much the offloaded items are worth. The participant should recall all the items they can since they are always worth more than offloaded items, and they should offload all the remaining items—leaving no items unrecalled and not offloaded. If they can recall an item, they should do so rather than offloading it since recalled items are always worth more than offloaded items. The maximum score (S) for any particular list is simply $10\ R + V\ (20 - R)$, where R is the number of items the participant can recall, and V is the value of each offloaded item assigned in a given condition. To obtain this value, R needs tobe estimated (which is the reason for the baseline measures of memory included in this experiment). Figure 5 illustrates the maximum point score a participant could obtain with perfect metacognition and offloading decisions as a function of the number of words they can recall from memory on the test. As can be seen from the figure, the maximum point score is greater for people with better memory as compared to people with poor memory. Except in the case in which an individual can and does remember everything, the maximum point scores are also greater when offloaded items are in higher valued conditions than lower valued conditions.

The optimal strategy that will allow the participant to obtain the score illustrated in Figure 5 at their own level of recall is to recall as many items from memory as possible and offload all of the words that would not have been remembered. To be able to do this, the participant needs to know exactly which words they will recall and which they will not recall and assign the latter, and only the latter, to be offloaded—if a word that could be remembered is offloaded, the participant will score fewer points than is possible. Similarly, if the participant makes a mistake and thinks they can recall a word and does not offload it, and then it turns out they cannot recall it, they get a score of 0 for that word. This will also result in a lower score than shown in the Figure 5. Thus, participants need to not only offload the right number of words but also make no mistakes about *which* particular words will be recallable and opt to recall just those words, and they need to offload all of the other words to achieve their optimal point score.

Pragmatically, it is probably impossible that people can perfectly know with much certainty which particular items they will be able to later remember. Even with excellent metaknowledge of their own capabilities, such as whether they have attentively and deeply encoded particular items, and the ability to correctly infer whether particular items are easy or difficult to remember, achieving perfect metaknowledge would still require additional insights. Specifically, participants would also have to know what the impact of trying to remember earlier items in the list will have on their ability to remember the present items. Even worse, they would need to know what the impact of items that are going to occur later in the list will be. They would also need to correctly anticipate conditions at recall. It is worth pointing out that it is

Figure 5
Maximum Possible Scores Based on Memory Ability



Note. The maximum point scores a participant could attain, assuming the perfect offloading of all the words that they cannot recall, as a function of the number of words they can recall from memory for each list scoring type in the present study.

not just MOOT scores that require prescience for perfection but all prospective metacognitive judgments insofar as they all are constrained by the individual's inherent inability to know the future. Thus, although it is possible to approximate the maximum score, through insight and experience, a person's predictions may improve so that they possibly approach it. It is, nevertheless, in our universe, at least not fully realizable.

In Experiment 1, we had no independent assessment of participants' memory. As such, we could not compute an approximation of what individuals' optimal scores would be. To overcome this problem, in Experiment 2, we used a separate set of lists, structured like those in the offloading task, in which we tested participants' memory when no offloading was permitted. This allowed us to compute an approximation (based on the average number of items recalled on these lists) to what their maximal point score would be. Knowing the maximum possible point score for each participant, we then computed the ratio of their obtained point score to their estimated maximum possible point score to obtain a measure of the goodness of their metacognitively based offloading—the MOOT score. The only other change from Experiment 1 to Experiment 2 was the inclusion of a question asking participants to evaluate their memory by making a global prediction about their upcoming performance before viewing each list.

In Experiment 2, we also had the opportunity to investigate whether participants with different levels of baseline memory performance might also exhibit different patterns of optimality in their offloading decisions (see also Kruger & Dunning, 1999; as well as Burson et al., 2006; Hacker & Bol, 2019; Händel & Fritzsche, 2016; Jansen et al., 2021; Krajc & Ortmann, 2008; Krueger & Mueller, 2002; Miller & Geraci, 2011; Serra & DeMarree, 2016). The ability to effectively offload may be especially important for individuals who have memory limitations since offloading directly impacts their ability to access information when needed. Thus, offloading could serve as a strategy to offset deficiencies in memory. While there is some research on the relation between goodness of performance and metacognition, there is little on the relation between memory capability and offloading optimization. Understanding how well participants can use their metacognition about what they will and will not be able to remember in a context in which they may strategically opt out of using their memory for access has widespread implications.

Method

Participants

After exclusions, participants were 136 undergraduate students ($M_{\text{age}} = 20.41$, $SD_{\text{age}} = 3.65$) recruited from the UCLA Human Subjects Pool. Two participants were excluded for cheating. We also excluded two participants who did not recall a single word on the baseline recall tests, resulting in a total of four exclusions.

Materials and Procedure

The materials used in Experiment 2 were the same as the materials used in Experiment 1. Participants again studied nine lists of words, with each list containing 20 words presented one at a time for 6 s each. However, on the first three lists, participants did not engage in any offloading. Before these first three lists, participants were told that each item was worth 10 points if they recalled it on the self-paced, immediate free recall test and that they should try to maximize their score. Before each of the first three lists, participants were asked to predict how many words (out of 20) 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 engaged in the same offloading procedure used in Experiment 1. 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. At the end of each of the nine lists, participants were told their score for the list they just completed but were not given feedback about specific items.

Results

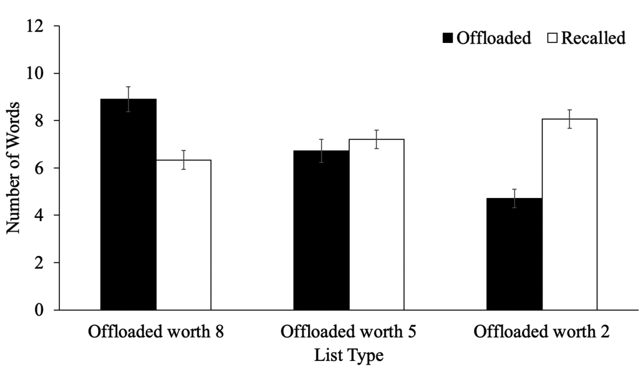
Offloading and Recall

The number of items offloaded and recalled from memory on the six experimental lists as a function of list scoring is shown in Figure 6. A 3 (list type: offloaded words worth 8, 5, 2) × 2 (animacy: animate, inanimate) repeated-measures ANOVA 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 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 ($M = 3.55$, $SD = 2.60$) than animate items ($M = 3.24$, $SD = 2.45$); see Figure 7a. List scoring did not interact with animacy, $F(2, 270) = 1.37$, $p = .257$, $\eta_p^2 < .01$. These results replicate the main findings of Experiment 1.

Next, we examined the frequency of offloading relative to baseline recall. Paired-samples *t* tests revealed that on lists where offloaded words were worth 8 points, participants offloaded a similar number of items ($M = 8.90$, $SD = 6.11$) that their initial three lists had indicated they would not be able to recall ($M_{\text{20-baseline}} = 9.60$, $SD = 3.78$), $t(135) = 1.24$, $p = .217$, $d = .11$. On lists where offloaded words were only worth 5 points, participants offloaded fewer items ($M = 6.72$, $SD = 5.68$) than would be ideal, $t(135) = 5.47$, $p < .001$, $d = .47$. On lists where offloaded words were worth 2 points, participants also offloaded fewer items ($M = 4.71$, $SD = 4.58$) than 20-baseline, $t(135) = 10.56$, $p < .001$, $d = .91$. This result indicates that only in the condition in which offloaded items were worth a lot of points were people offloading close to the right number of items. When offloaded words were worth fewer points, people did not offload enough items. Importantly, this measure does not address

Figure 6

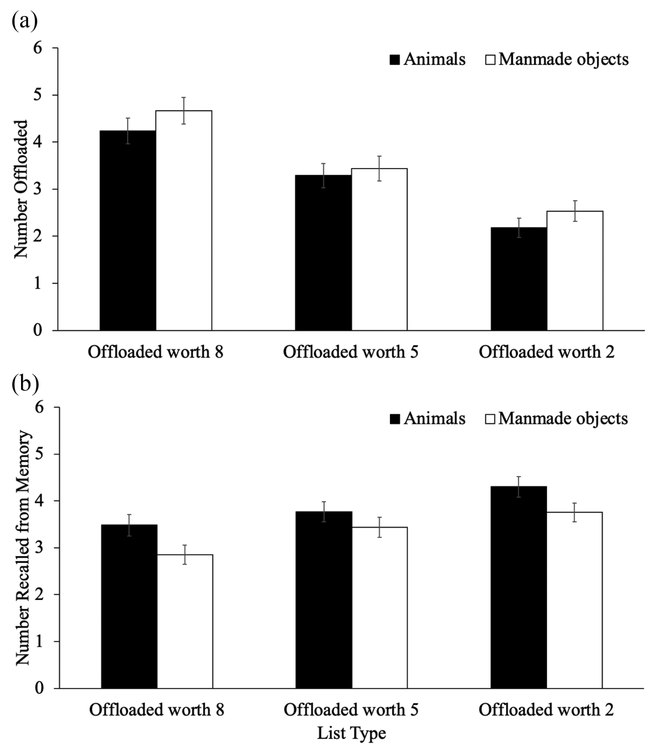
The Average Number of Words Offloaded or Recalled From Memory per List as a Function of List Scoring in Experiment 2



Note. Error bars reflect the standard error of the mean.

Figure 7

The Average Number of Words Offloaded (a) and Recalled From Memory (b) per List as a Function of List Scoring and Animacy in Experiment 2



Note. Error bars reflect the standard error of the mean.

which items were offloaded and whether those items that were offloaded were, indeed, the items that could not be remembered.

To examine recall (excluding offloaded words), we conducted a 3 (list type: offloaded words worth 8, 5, 2) × 2 (animacy: animate, inanimate) repeated-measures ANOVA. As is shown in Figure 7b, there were differences between list types, $F(2, 270) = 21.21$, $p < .001$, $\eta_p^2 = .14$, such that participants recalled fewer items per list when offloaded items were worth 8 points ($M = 6.33$, $SD = 4.67$) than when offloaded items were worth 5 points ($M = 7.21$, $SD = 4.58$), $p_{\text{holm}} = .002$, $d = .18$, and when offloaded items were worth 2 points ($M = 8.06$, $SD = 4.55$), $p_{\text{holm}} < .001$, $d = .35$; additionally, participants recalled more items when offloaded items were worth 5 points than when they were worth 2 points ($p_{\text{holm}} = .002$, $d = .17$).[2] More animate items were recalled per list ($M = 3.85$, $SD = 2.30$) than inanimate items ($M = 3.35$, $SD = 2.19$), $F(1, 135) = 15.26$, $p < .001$, $\eta_p^2 = .10$, but list scoring did not interact with animacy, $F(2, 270) = 1.49$, $p = .227$, $\eta_p^2 = .01$.

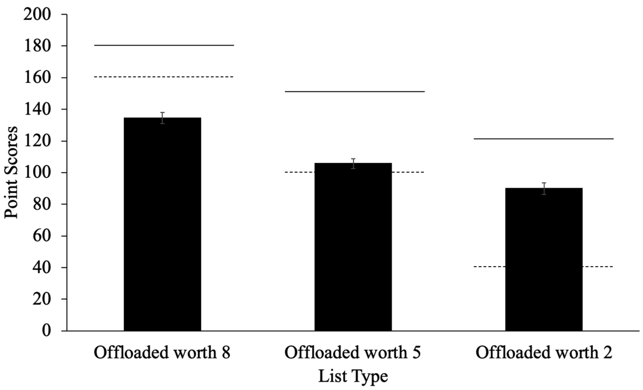
Point Scores

To examine point scores as a function of list scoring (see Figure 8), we conducted a repeated-measures ANOVA with three levels (list scoring: offloaded words worth 8, 5, 2). Results revealed differences

[2] There were no differences in the proportion of not-offloaded words that were not recalled between the three list types, Mauchly's $W = .94$, $p = .023$; Huynh–Feldt corrected results: $F(1.92, 235.55) = 3.03$, $p = .052$, $\eta_p^2 = .02$.

Figure 8

Average Point Scores as a Function of List Scoring in Experiment 2



Note. Dashed lines represent the score participants would have obtained if they had offloaded every word. Solid lines represent the average maximum scores attainable given perfect metacognition and offloading across all participants (as given by the first three lists during which offloading was not allowed). Error bars reflect the standard error of the mean.

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 offloaded words 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).

Next, we examined how participants' scores compared with the scores they would have obtained if they had 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 had they offloaded every word (160), $t(135) = -7.40$, $p < .001$, $d = -.64$. On lists where offloaded words were worth 5 points, 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, 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.

MOOT Scores and Optimality

We examined how participants' point scores compared with the scores they could have obtained if they had had perfect metacognition regarding which words they would recall and had offloaded all of the words that they would not recall. To calculate each participant's MOOT score for each of the three list scoring conditions, we first estimated their optimal scores based on their recall ability, which we assessed from the initial three lists where offloading was not permitted. To obtain the participant's maximum possible score for each treatment combination, participants' average baseline recall (R) was multiplied by 10 (the value given to recalled items) and added to $(20 - R)$ times the value of the offloaded items in the treatment combination in question (either 2, 5, or 8). The MOOT score—the ratio of their actual score (we averaged scores across the two lists for each list scoring condition) to this possible maximum score—was then computed for each list scoring condition:

$$\text{MOOT Score} = \frac{\text{points scored}}{\text{maximum possible score}|\text{memory level}}. \quad (1)$$

It should be noted that, theoretically, MOOT scores should never exceed 1.0, but because we had to *estimate* recall from the pretest, and the estimated recall obtained from the first three lists could be different from participants' recall during the offloading task, the observed ratios of actual score and *estimated* maximum score were sometimes greater than 1.[3] One obvious reason for this was that sometimes participants learned how to do the task and remembered better than their baseline would indicate as the session proceeded (e.g., participants may have learned how to perform better during the task, used mnemonic techniques or deeper encoding, or focused their attention better during the task). Such learning to learn would make their baseline an underestimate of their actual recall ability on the task. Aside from discrepancies induced by such error in the baseline estimates, a value of 1, as indicated by the solid horizontal line in Figure 9, indicates optimality.

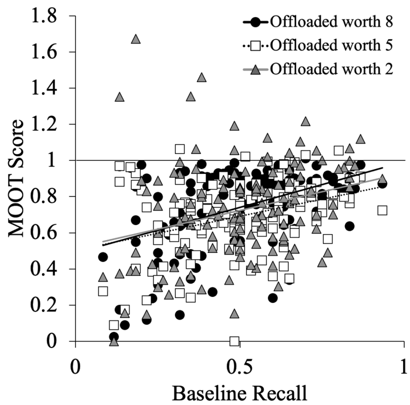
To examine MOOT scores as a function of list scoring (see Figure 10), we conducted a repeated-measures ANOVA with three levels (list scoring: offloaded words worth 8, 5, 2). Results did not reveal differences between list types, Mauchly's W = .70, p < .001; Huynh–Feldt corrected results: $F(1.55, 208.78) = 3.00$, p = .065, η_p^2 = .02, such that MOOT scores were similar when offloaded words were worth 8 points (M = .74, SD = .21), 5 points (M = .69, SD = .20), and 2 points (M = .73, SD = .27).

Next, we examined MOOT scores as a function of individuals' own baseline memory performance. As shown in Figure 9, participants whose baseline recall was low showed less optimal offloading than participants who had higher baseline recall. To analyze this effect statistically, we used each participant's baseline recall as a predictor of the proportion of each participant's MOOT score for each list scoring condition. The lists in which offloaded items were worth 2 points were chosen as the reference group. A multilevel model (items nested within participants) predicting MOOT scores as a function of participants' baseline recall and list scoring revealed that baseline recall predicted MOOT scores, $t(134) = 5.58$, $p < .001$, estimate: .02, such that the greater a given participant's baseline recall, the closer they were to optimal. There were no differences between lists where offloaded words were worth 2 points and offloaded words were worth 5 or 8 points (both ps > .117), and the comparisons between lists where offloaded words were worth 2 points and offloaded words were worth 5 or 8 points did not interact with baseline recall (both ps > .408).

[3] As Figures 10 and 15 show, this happened more frequently, when the value of the offloaded items was 2 than when it was 5 or 8. The reason it happened more frequently in that particular condition is that a greater proportion of the obtained total score is attributable to the memory portion of the score, when the offloaded items are worth less, making the variation (and, in particular, the underestimation) in the estimates of the memory score more important in that condition.

Figure 9

The Relation Between Memory Ability and Optimal Offloading for Each List Scoring Condition in Experiment 2



Note. The horizontal solid black line indicates "optimal." The optimal score represents the highest possible score for a given participant considering the participant's memory ability and strategic use of metacognition to determine which information to offload and which to remember. Observations greater than 1.0 are due to errors in the estimation of memory. MOOT = Metacognitive Offloading Optimization Task.

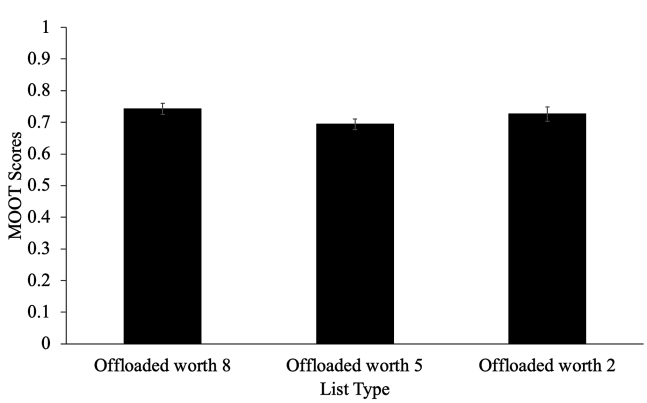
Correlations

A summary of the correlations between baseline recall, the number of words offloaded per list for each list type, point scores for each list type, and MOOT scores for each list type are shown in Table 1. As is shown in the table, baseline recall was negatively related to the number of words offloaded on all three list types such that the better a given participant's memory ability, the less they tended to offload. Baseline recall was positively related to point scores on all three list types such that the better a given participant's memory ability, the more points they scored (which makes sense given that words recalled from memory were worth more). Moreover, as noted above, participants' baseline recall was positively correlated with MOOT scores such that the greater a participant's memory ability, the better they were at optimizing the combination of memory and offloading.[4] Finally, calibration (baseline predictions—baseline recall; negative values would indicate underconfidence, whereas positive values would indicate overconfidence; $M = -.06$, $SD = .12$), a global measure of metacognitive accuracy was a significant predictor of MOOT scores such that the more overconfident the learner, the lower their MOOT score.

We also examined whether the propensity to offload inanimate items over animate items correlated with MOOT scores. We calculated an offloading difference score which was the number of animate words offloaded minus the number of manmade words offloaded. On lists where offloaded words were worth 8 points, there was a significant correlation ($r = -.22$, $p = .010$), but on lists where offloaded words were worth 5 points, the correlation was not significant ($r = -.06$, $p = .486$); the correlation was also not significant on lists where offloaded words were worth 2 points ($r = -.07$, $p = .437$).

Figure 10

Average MOOT Scores (Point Scored/Maximum Possible Score | Memory Level) as a Function of List Scoring in Experiment 2



Note. Error bars reflect the standard error of the mean. MOOT = Metacognitive Offloading Optimization Task.

Discussion

In Experiment 2, we replicated the key findings observed in Experiment 1. People were sensitive to the cost of offloading—doing so more when the cost was low than when it was high. Additionally, Experiment 2 afforded an estimate of how close each participant came to having optimal metacognitively guided offloading in the form of their MOOT score—participants were less than optimal in their offloading decisions on all list types. Moreover, high memory performers were closer to their own optimal scores than were poor memory performers. This new finding is similar to the widely cited "unskilled and unaware" findings of Kruger and Dunning (1999; see also Burson et al., 2006; Händel & Fritzsche, 2016; Jansen et al., 2021; Krajc & Ortmann, 2008; Krueger & Mueller, 2002; Miller & Geraci, 2011; Serra & DeMarree, 2016; see

[4] MOOT scores, which assess the optimality of offloading decisions based on participants' actual performance relative to their theoretical maximum, are derived using baseline recall measures. However, our approach aims to mitigate any circularity by using these baseline measures primarily to establish a benchmark for memory capacity that informs the theoretical maximum score a participant could achieve if they offloaded optimally. The key here is that the MOOT score calculation adjusts for each participant's baseline memory capacity, essentially normalizing their performance across varying levels of memory ability. This method does not inflate the association between baseline memory and MOOT scores but instead controls for individual differences in memory capacity, allowing us to assess how well participants utilize their memory and offloading strategies relative to their personal best potential outcome. Furthermore, when we analyze the relationship between baseline recall and MOOT scores, we are effectively examining whether participants who generally perform well on memory tasks are also better at optimizing their offloading decisions—not merely reconfirming their memory ability. This analysis seeks to uncover the extent to which good rememberers are also good at metacognitively evaluating when to rely on memory versus offloading. However, because the method used to assess the correlation between memory ability and MOOT violated the assumption of independence, we also examined the correlation between MOOT scores and memory in a different way—by splitting the trials used to derive each measure. Specifically, we collapsed the data across the three baseline lists. We then used average recall for even trials as baseline memory ability and average recall for odd trials to compute MOOT scores. Results revealed similar correlations between baseline recall and MOOT scores: There were strong correlations between memory and MOOTs on all list types: lists where offloaded words were worth 8 points ($r = .43$, $p < .001$), lists where offloaded words were worth 5 points ($r = .35$, $p < .001$), and lists where offloaded words were worth 2 points ($r = .29$, $p < .001$).

Table 1
Correlations Between Variables of Interest in Experiment 2

Measure	1	2	3	4	5	6	7	8	9	10	11
1. Baseline recall	—										
2. Calibration	−.539***	—									
3. Offloaded words per list—Offloaded worth 8	−.182*	.027	—								
4. Offloaded words per list—Offloaded worth 5	−.203*	.028	.708***	—							
5. Offloaded words per list—Offloaded worth 2	−.170*	.052	.542***	.721***	—						
6. Point scores—Offloaded worth 8	.555***	−.372***	.464***	.344***	.202*	—					
7. Point scores—Offloaded worth 5	.658***	−.468***	−.043	.000	−.044	.688***	—				
8. Point scores—Offloaded worth 2	.738***	−.459***	−.285***	−.221**	−.265**	.511***	.771***	—			
9. MOOT scores—Offloaded worth 8	.444***	−.316***	.533***	.410***	.250**	.991***	.637***	.432***	—		
10. MOOT scores—Offloaded worth 5	.350***	−.334***	.081	.148	.060	.624***	.927***	.595***	.618***	—	
11. MOOT scores—Offloaded worth 2	.286***	−.253**	−.207*	−.106	−.180*	.317***	.579***	.823***	.295***	.593***	—

Note. Correlations between baseline recall, calibration (baseline predictions − baseline recall), the number of words offloaded per list for each list scoring condition, point scores for each list scoring condition, and MOOT scores (the proportion of participants' optimal score obtained) for each list scoring condition in Experiment 2. MOOT = Metacognitive Offloading Optimization Task.
* $p < .05$. ** $p < .01$. *** $p < .001$.

also Hacker & Bol, 2019) except that, rather than involving metacognitive calibration, our findings indicate that the strategic use of offloading to later access information was worse in people with poor than good memory. In sum, participants with better memory offloaded less, scored higher, and, most importantly, also achieved higher MOOT scores (i.e., closer to optimal offloading efficiency and superior metacognitive awareness) than participants with poorer memory ability.

Experiment 3

In Experiment 2, the display of participants' offloaded words during recall may have inadvertently introduced a part-list cuing effect (Bäuml & Aslan, 2004; Roediger, 1973; Slamecka, 1968), potentially influencing participants' ability to remember nonoffloaded items. Part-list cuing is a phenomenon, whereby providing some items from a to-be-remembered list of words can impair the recall of the remaining items. When some of the items are given "for free" at the beginning of recall, their presentation can disrupt retrieval strategies and inhibit the recall of other items that are not so provided. Unlike a traditional part-list cuing experiment in which cues are preselected by the experimenter, participants in our previous experiments determined their own cues by choosing which words to offload. Furthermore, Experiment 2 showed that participants with poor memory offloaded more often than participants with good memory. Thus, the presence of the offloaded words at the time of retrieval may have hurt memory performance This was particularly worrisome because if part-list cuing had a deleterious effect on MOOT scores, that effect would have been larger for people with lower baseline memory because they had offloaded more items than did people with higher memory baselines. Accordingly, we conducted Experiment 3, which mirrored the setup of Experiment 2 but with one alteration: Offloaded words were not displayed during recall. This repeat of the experiment with a small change in procedure also provided the desirable opportunity to replicate the results of Experiment 2.

Method

Participants

After exclusions, participants were 185 undergraduate students ($M_{\text{age}} = 19.07$, $SD_{\text{age}} = 1.66$) recruited from the Arizona State University Human Subjects Pool. Eight participants were excluded for cheating. No participants were excluded for not recalling a single word on the baseline recall tests.

Materials and Procedure

The materials and procedure in Experiment 3 were the same as in Experiment 2. However, offloaded words were not visible to participants during recall and did not need to be typed into the text box.

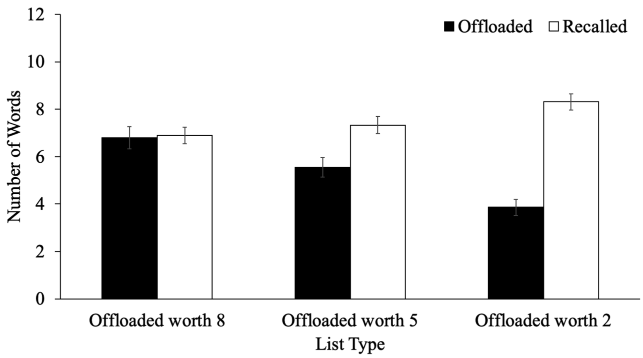
Results

Offloading and Recall

The number of items offloaded and recalled from memory on the six experimental lists as a function of list scoring is shown in Figure 11. A 3 (list type: offloaded words worth 8, 5, 2) × 2 (animacy: animate, inanimate) repeated-measures ANOVA revealed differences between list types, Mauchly's $W = .84$, $p < .001$; Huynh–Feldt corrected results: $F(1.74, 320.52) = 40.69$, $p < .001$, $\eta_p^2 = .18$, such that participants offloaded more items per list when offloaded items were worth 8 points ($M = 6.80$, $SD = 6.25$) than when they were worth 5 points ($M = 5.55$, $SD = 5.53$), $p_{\text{holm}} < .001$, $d = .22$, and when they were worth 2 points ($M = 3.87$, $SD = 4.64$), $p_{\text{holm}} < .001$, $d = .51$; participants also offloaded more when offloaded items were worth 5 points than when they were worth 2 points ($p_{\text{holm}} < .001$, $d = .29$). There was no effect of animacy, $F(1, 184) = 1.08$, $p = .300$, $\eta_p^2 = .01$, inanimate items were offloaded similarly frequently ($M = 2.75$, $SD = 2.58$) as animate items ($M = 2.66$, $SD = 2.44$); see Figure 12a. List scoring did not interact with animacy, $F(2, 368) = .20$, $p = .818$, $\eta_p^2 < .01$.

Figure 11

The Average Number of Words Offloaded or Recalled From Memory per List as a Function of List Scoring in Experiment 3

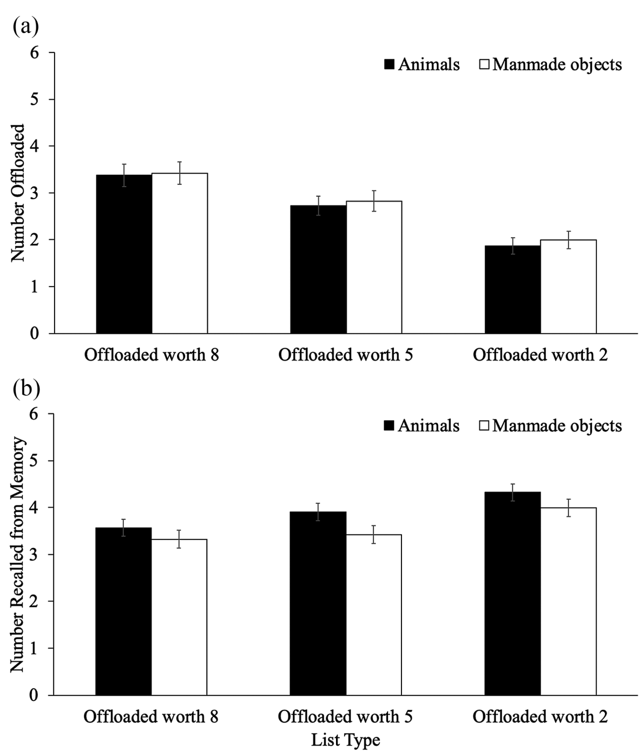


Note. Error bars reflect the standard error of the mean.

To examine participants' frequency of offloading relative to the number of items each participant should have offloaded ($M_{20\text{-baseline}}$ = 10.02, *SD* = 3.44), we conducted paired-samples *t* tests. On lists where offloaded words were worth 8 points, participants offloaded fewer than 20-baseline items, *M* = 6.80, *SD* = 6.25), *t*(184) = 6.70, $p < .001$, *d* = .49, as was the case on lists where offloaded words were worth 5 points, *M* = 5.55, *SD* = 5.53, *t*(184) = 10.39, $p < .001$, *d* = .76, and lists where offloaded words were worth 2 points, *M* = 3.87, *SD* = 4.64, *t*(184) = 17.07, $p < .001$, *d* = 1.26.

Figure 12

The Average Number of Words Offloaded (a) and Recalled From Memory (b) per List as a Function of List Scoring and Animacy in Experiment 3



Note. Error bars reflect the standard error of the mean.

As is shown in the recall data (which excludes offloaded items) in Figure 12b, a 3 (list type: offloaded words worth 8, 5, 2) × 2 (animacy: animate, inanimate) repeated-measures ANOVA revealed differences between list types, Mauchly's *W* = .93, $p < .001$; Huynh–Feldt corrected results: *F*(1.88, 346.12) = 15.68, $p < .001$, $\eta_p^2 = .08$, such that participants recalled a similar number of items per list when offloaded items were worth 8 points (*M* = 6.89, *SD* = 4.82) as when they were worth 5 points (*M* = 7.33, *SD* = 4.84), p_{holm} = .093, *d* = .09, but recalled fewer items when offloaded items were worth 8 points than when they were worth 2 points (*M* = 8.31, *SD* = 4.66), $p_{\text{holm}} < .001$, *d* = .28; additionally, participants recalled more items when offloaded items were worth 5 points compared with 2 points ($p_{\text{holm}} < .001$, *d* = .20).[5] There was a significant effect of animacy, *F*(1, 184) = 16.35, $p < .001$, $\eta_p^2 = .08$, such that more animate items were recalled per list (*M* = 3.93, *SD* = 2.18) than inanimate items (*M* = 3.58, *SD* = 2.29), but list scoring did not interact with animacy, *F*(2, 368) = 1.40, *p* = .249, $\eta_p^2 = .01$.

Point Scores

As shown in Figure 13, a repeated-measures ANOVA with three levels (list scoring: offloaded words worth 8, 5, 2) revealed point score differences between list types, Mauchly's *W* = .74, $p < .001$; Huynh–Feldt corrected results: *F*(1.60, 294.65) = 74.00, $p < .001$, $\eta_p^2 = .29$, such that scores were higher when offloaded words were worth 8 points (*M* = 123.28, *SD* = 43.71) than when they were worth 5 points (*M* = 101.05, *SD* = 40.43), $p_{\text{holm}} < .001$, *d* = .53, or 2 points (*M* = 90.87, *SD* = 45.23), $p_{\text{holm}} < .001$, *d* = .77; participants also scored higher when offloaded words were worth 5 points than when they were worth 2 points ($p_{\text{holm}} < .001$, *d* = .24).

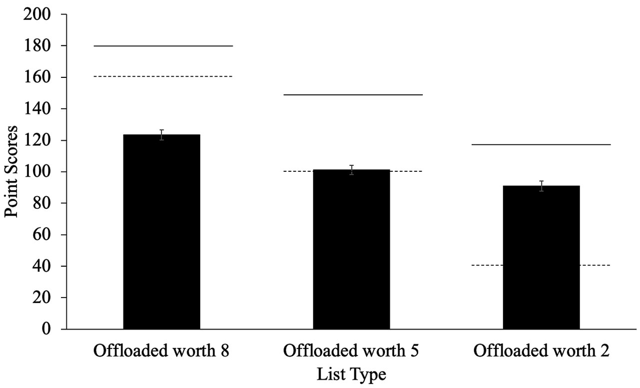
On lists where offloaded words were worth 8 points, a one-sample *t* test indicated that participants' scores were *lower* than what they could have obtained had they offloaded every word (160), *t*(184) = −11.43, $p < .001$, *d* = −.84. On lists where offloaded words were worth 5 points, participants' scores were *similar* to what they could have obtained if they offloaded every word (100), *t*(184) = .36, *p* = .723, *d* = .03. On lists where offloaded words were worth 2 points, participants' scores were *greater* than what they could have obtained if they offloaded every word (40), *t*(184) = 16.27, $p < .001$, *d* = 1.20.

MOOT Scores and Optimality

As shown in Figure 14, a repeated-measures ANOVA on MOOT scores with three levels (list scoring: offloaded words worth 8, 5, 2) revealed small differences between list types, Mauchly's *W* = .69, $p < .001$; Huynh–Feldt corrected results: *F*(1.54, 283.29) = 11.73, $p < .001$, $\eta_p^2 = .06$, such that MOOT scores were higher when offloaded words were worth 2 points (*M* = .75, *SD* = .27) than 5 points (*M* = .67, *SD* = .23), $p_{\text{holm}} < .001$, *d* = .34, and 8 points (*M* = .69, *SD* = .24), $p_{\text{holm}} < .001$, *d* = .28, but MOOT scores were similar when offloaded words were worth 8 and 5 points (p_{holm} = .445, *d* = .06).

[5] Again, there were no differences in the proportion of not-offloaded words recalled between the three list types, Mauchly's *W* = .96, *p* = .018; Huynh–Feldt corrected results: *F*(1.93, 334.60) = 1.19, *p* = .304, $\eta_p^2 = .01$.

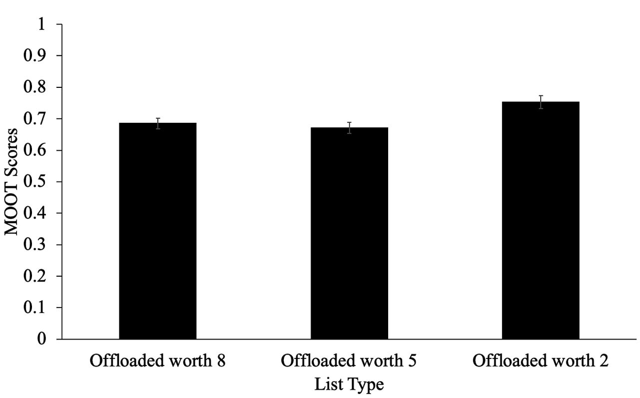
Figure 13
Average Point Scores as a Function of List Scoring in Experiment 3



Note. Dashed lines represent the score participants would have obtained if they had offloaded every word. Solid lines represent the average maximum scores attainable given perfect metacognition and offloading across all participants (as given by the first three lists during which offloading was not allowed). Error bars reflect the standard error of the mean.

The relation between participants' MOOT scores and their baseline memory performance is shown in Figure 15. As before, we used each participant's baseline recall as a predictor of MOOT scores for each list type. The lists in which offloaded items were worth 2 points were chosen as the reference group. A multilevel model (items nested within participants) predicting MOOT scores as a function of participants' baseline recall, and list scoring replicated the finding that baseline recall predicted MOOT scores, $t(183) = 7.32$, $p < .001$, estimate = .03, such that the greater a given participant's baseline recall, the closer they were to their optimal score. There were differences between lists where offloaded words were worth 2 points and offloaded words were worth 5 or 8 points (both *p*s < .001). However, an analysis of the simple effects indicates similar trends when offloaded words were worth 2 points, $t(419) = 5.97$, $p < .001$, estimate = .03; 5 points, $t(419) = 6.00$, $p < .001$, estimate = .03; and 8 points, $t(419) = 4.97$, $p < .001$, estimate = .02.

Figure 14
Average MOOT Scores (Point Scored/Maximum Possible Score | Memory Level) as a Function of List Scoring in Experiment 3



Note. Error bars reflect the standard error of the mean. MOOT = Metacognitive Offloading Optimization Task.

The comparisons between lists where offloaded words were worth 2 points and offloaded words were worth 5 or 8 points did not interact with baseline recall (both *p*s > .365).

Correlations

A summary of the correlations between the variables of interest is shown in Table 2. As is shown in the table, the better a given participant's baseline recall, the less they tended to offload. Additionally, the better a given participant's baseline recall, the higher their point scores. Moreover, the greater a participant's memory, the better their ability to optimize the combination of memory and offloading, as given by their MOOT scores.[6] Finally, calibration ($M = .00$, $SD = .16$) was a significant predictor of MOOT scores such that the more overconfident the learner, the lower their MOOT score.

We again examined the correlation between offloading difference scores (the number of animate words offloaded minus the number of manmade words offloaded) and MOOT scores. There was no significant correlation on lists where offloaded words were worth 8 points ($r = -.03$, $p = .692$), 5 points ($r = -.10$, $p = .188$), or 2 points ($r = -.11$, $p = .129$).

Discussion

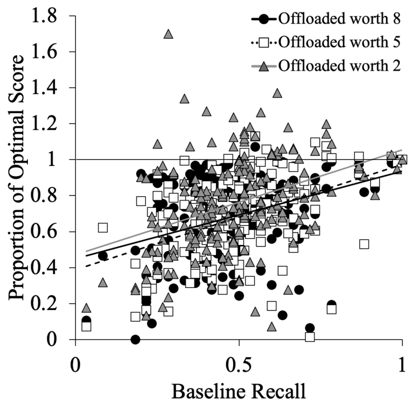
In Experiment 3, the tendency to offload more frequently when offloaded words were worth more, and the differential recall of animate over inanimate items was replicated. However, a few differences between Experiments 2 and 3 emerged: Experiment 3 saw a shift toward a more uniform offloading rate between animate and inanimate items, diverging slightly from Experiments 1 and 2, where more inanimate items were offloaded (though this effect was marginal in Experiment 1). In Experiment 3, participants offloaded fewer items than 20-baseline recall levels across all valuation levels of offloaded words. Participants' MOOT scores showed that they were slightly closer to optimal when offloaded words were less valuable in this experiment than was shown in Experiment 2. However, despite these nuances, the overarching patterns remained the same in Experiment 3, as was observed in the earlier experiments, including the influence of baseline recall on MOOT scores. Once again, people with better memory ability also offloaded in a manner that was closer to optimal than did people with lower memory abilities.

General Discussion

Our primary objective was to explore people's ability to effectively engage in memory offloading, developing the MOOT task to do so. The results revealed that as the price for offloading decreased, the criterion for offloading shifted toward offloading more items, indicating that participants were sensitive to the cost–benefit trade-off associated with offloading. This criterion shift highlighted the importance people place on the value and cost

[6] Similar to Experiment 2, we examined this correlation using independent data. Results revealed similar correlations between baseline recall and MOOT scores: There were strong correlations on lists where offloaded words were worth 8 points ($r = .34$, $p < .001$), lists where offloaded words were worth 5 points ($r = .42$, $p < .001$), and lists where offloaded words were worth 2 points ($r = .38$, $p < .001$).

Figure 15
The Relation Between Memory Ability and MOOT Scores for Each List Scoring in Experiment 3



Note. The horizontal solid black line indicates "optimal." The optimal score represents the highest possible score for a given participant considering the participant's memory ability and strategic use of metacognition to determine which information to offload and which to remember. Observations greater than 1.0 are due to errors in the estimation of memory. MOOT = Metacognitive Offloading Optimization Task.

associated with offloading and their sensitivity to the trade-off between the potential gains from offloading and the potential losses from relying solely on memory. As a result of this strategy, when the cost of offloading was high (i.e., when offloaded words were worth fewer points), participants performed better than they would have if they offloaded every word, whereas when the cost of offloading was medium or low, they performed worse than they would have if they simply offloaded every word.

People's MOOT scores—a new measure of the optimality of their metacognitively based strategies for offloading—were systematically lower than optimal. A perfect MOOT score would entail that the individual not only knew how many items they would recall but also which items they would be able to recall and chose exactly the right items—those that they correctly knew they would not recall—to offload. People who exhibited better baseline memory demonstrated MOOT scores that were closer to optimal.

In all three experiments and list scoring types, a significant proportion of items were neither recalled nor offloaded, leading to 0 points for those items (across experiments, people got 0 points for around 31% of the items on each list). Thinking one will remember something and then *not being able to* remember it is the essence of overconfidence, and this is indicated by those items that were neither offloaded nor recalled—a phenomenon that occurred repeatedly and consistently in these experiments. The consistent occurrence of this outcome across different experimental conditions implies that participants often overestimate their memory capabilities, choosing not to secure at least some points through offloading, because they expect their recall to be more successful than it is.

This pervasive overestimation of memory capabilities underscores a fundamental challenge in the effective management of cognitive resources. Moving beyond the empirical results, the insightful work of Gilbert et al. (2020; see also Sachdeva & Gilbert, 2020) sheds light on additional dynamics at play, such as risk tolerance and effort avoidance, which might influence offloading decisions in contexts like the MOOT paradigm. It is currently unknown how such factors, or others such as the need for control, might play in the MOOT paradigm—further nuanced modeling taking them into account is needed. While the MOOT score is an idealized measure, it is also important to recognize that individual differences, such as these, are important factors and could also significantly impact offloading decisions. Additionally, not all information may warrant remembering in real-world scenarios, highlighting the selective nature of offloading.

Prior work has suggested that people with poor memory may have a metacognitive deficit (i.e., unskilled and unaware; Kruger & Dunning, 1999, 2002) such that they are more overconfident than people who have higher levels of expertise. However, this conclusion is controversial (see Ehrlinger et al., 2008). Furthermore, the role of

Table 2
Correlations Between Variables of Interest in Experiment 3

Measure	1	2	3	4	5	6	7	8	9	10	11
1. Baseline recall	—										
2. Calibration	−.494***	—									
3. Offloaded words per list—Offloaded worth 8	−.185*	.032	—								
4. Offloaded words per list—Offloaded worth 5	−.206**	−.008	.811***	—							
5. Offloaded words per list—Offloaded worth 2	−.282***	.026	.581***	.656***	—						
6. Point scores—Offloaded worth 8	.445***	−.222**	.476***	.367***	.198**	—					
7. Point scores—Offloaded worth 5	.649***	−.263***	.062	.025	−.079	.743***	—				
8. Point scores—Offloaded worth 2	.754***	−.320***	−.219**	−.224**	−.348***	.447***	.670***	—			
9. MOOT scores—Offloaded worth 8	.352***	−.188*	.529***	.417***	.247***	.994***	.697***	.378***	—		
10. MOOT scores—Offloaded worth 5	.436***	−.184*	.178*	.144	.027	.743***	.959***	.533***	.726***	—	
11. MOOT scores—Offloaded worth 2	.373***	−.174*	−.149*	−.132	−.250***	.332***	.481***	.868***	.306***	.457***	—

Note. Correlations between baseline recall, calibration (baseline predictions − baseline recall), the number of words offloaded per list for each list scoring condition, point scores for each list scoring condition, and MOOT scores (the proportion of participants' optimal score obtained) for each list scoring condition in Experiment 3. MOOT = Metacognitive Offloading Optimization Task.
* $p < .05$. ** $p < .01$. *** $p < .001$.

metacognition in bootstrapping learning and memory is not straightforward, so although it is tempting to interpret this finding as implying that their learning is bound to be dysfunctional, such an inference is not necessarily true. In some situations, overconfidence may be adaptive in bolstering what might otherwise be flagging motivation. In the present experiments, the observation that participants with poorer memory also demonstrated less effective offloading presents a novel finding but the reasons behind this phenomenon, its broader implications, and how widely it can be generalized remain unclear at this stage. Participants with better memory may be more proficient in assessing which items they are likely to remember and which ones they are more likely to forget, enabling them to strategically offload the items they anticipate forgetting and rely on their memory for the ones they are confident they can recall. However, there may be other reasons for these observed differences in optimality such as individual differences in risk tolerance, the perceived value of the information, or variations in cognitive effort.

In real-world scenarios, individuals frequently rely on external sources such as technology to store and retrieve information. Gaining a greater understanding of how these external resources can be more effectively coordinated with our own cognitive and memory capabilities is critical in today's increasingly information-rich society. This enhanced understanding could lead to the development of more efficient strategies for managing information overload, which in turn can improve productivity and reduce the cognitive load associated with trying to remember vast amounts of information.

Limitations

Many factors influence memory and metacognition that should and probably do affect offloading that we did not consider. For instance, we did not investigate the effects of different retention intervals or consider the dynamics of forgetting over time. We also did not investigate the effects of attention. Moreover, we used a very limited range of to-be-remembered/offloaded materials—undoubtedly, memory and offloading will be affected if more meaningful materials are used (see Monin & Oppenheimer, 2014). Furthermore, our variations of reward structure were very limited and we did not consider motivational or personality factors—people may try to remember valuable information in ways that are different from what the simple point value assignments used in these experiments could capture. Additionally, the (lack of) learning consequences that accrue to offloaded items is a topic deserving intensive study in its own right—and one that we did not touch on at all here. Finally, offloading may have profound consequences for people's long-term acquisition of knowledge. Clearly, other goals deserve intensive research beyond the overly simple point values used here.

We also note that the measure of optimality employed in the present study, while useful, is not perfect, and there were some inherent difficulties in the baseline measure of memory, which was always conducted first and was undoubtedly noisy. People may have either learned to learn over the session and thereby attained higher scores than this baseline indicated or become fatigued over time and done worse. Possibly, interspersing baseline measures throughout the session might produce more reliable results. Discrepancies in this baseline from people's actual memory capability pose a serious problem for the practical problem of accurately assessing offloading optimality.

Another potential direction for future studies would be to incorporate metacognitive assessments where participants predict and later reflect on their performance. Specifically, asking participants to estimate their scores before and after the task, and to evaluate how they think they performed relative to the strategy of offloading all information, could provide insights into the role of metacognition in optimizing offloading behavior.

We also acknowledge that the assumption of perfect metacognition required for achieving the maximum score in our task is not realistic. Future work would benefit from a more nuanced understanding of the dynamic probabilities involved in memory and metacognitive judgments, and how these probabilities contribute to the offloading strategies tailored to different payoff structures. While the current optimization calculations assume an ideal scenario of perfect metacognition, understanding actual participant behavior will undoubtedly require a more sophisticated model that includes dynamic probabilistic memory assessments and cost/benefit trade-offs—a complex interplay that deserves deeper investigation.

Summary

The present study investigated the optimality of people's offloading strategies. In a gamelike paradigm, people had to decide whether to try to remember individual items (for a high-point reward, but with the risk that they would forget it and get no reward) or to offload items (for a lower reward but with no risk of failure) to have access to the items later. This was called the MOOT. Higher scores on this task indicated success, and we were interested in factors that influence participants' ability to achieve these high scores. We proposed a new measure for evaluating the optimality of offloading—the MOOT score. This score takes into account (a) the value of remembered items, (b) the value of offloaded items, (c) the (lack of) value that eventuates from forgetting and not having offloaded, and (d) the limitations imposed by the individual's own memory abilities. The individual's MOOT score indicates how close they came to achieving their own maximum possible score, given their own memory capability (i.e., how close to optimal their offloading strategy was). The present research showed that people were sensitive to the value of the offloaded items and that their offloading was far from optimal. Most interestingly, the research revealed a new finding: People who exhibited poorer baseline memory ability also showed less optimal offloading—their MOOT scores were systematically lower than those of those people who exhibited better memory.

Constraints on Generality

The present study's findings are primarily applicable to adult populations with typical cognitive abilities, as the experiments were conducted using university students. While these results provide valuable insights into memory offloading and metacognitive strategies within this demographic, they may not directly generalize to younger individuals, older adults, or those with significant cognitive impairments or developmental disorders. Additionally, the experimental tasks involved structured and controlled settings that may not perfectly replicate everyday memory offloading scenarios, such as those involving more complex or interactive technologies. Future research should explore the applicability of these findings across different age groups, varied cognitive profiles, and more naturalistic settings to fully understand the dynamics and limitations of offloading in a broader context.

References

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*(3), 432–447. https://doi.org/10.1037/a0015928

Ariel, R., Price, J., & Hertzog, C. (2015). Age-related associative memory deficits in value-based remembering: The contribution of agenda-based regulation and strategy use. *Psychology and Aging*, *30*(4), 795–808. https://doi.org/10.1037/a0039818

Ball, B. H., Wiemers, E. A., & Brewer, G. A. (2022). Individual differences in memory and attention processes in prospective remembering. *Psychonomic Bulletin & Review*, *29*(3), 922–933. https://doi.org/10.3758/s13423-022-02059-3

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. https://doi.org/10.1016/j.chb.2015.02.029

Bäuml, K. H., & Aslan, A. (2004). Part-list cuing as instructed retrieval inhibition. *Memory & Cognition*, *32*(4), 610–617. https://doi.org/10.3758/BF03195852

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, *4*(1), Article 45. https://doi.org/10.1186/s41235-019-0195-y

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: Further evidence from word and picture stimuli. *Memory & Cognition*, *42*(3), 370–382. https://doi.org/10.3758/s13421-013-0368-8

Bonin, P., Gelin, M., Laroche, B., Méot, A., & Bugaiska, A. (2015). The 'how' of animacy effects in episodic memory. *Experimental Psychology*, *62*(6), 371–384. https://doi.org/10.1027/1618-3169/a000308

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, *90*(1), 60–77. https://doi.org/10.1037/0022-3514.90.1.60

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*(7), 1078–1085. https://doi.org/10.3758/BF03194325

Cherkaoui, M., & Gilbert, S. J. (2017). Strategic use of reminders in an 'intention offloading' task: Do individuals with autism spectrum conditions compensate for memory difficulties? *Neuropsychologia*, *97*, 140–151. https://doi.org/10.1016/j.neuropsychologia.2017.02.008

DeLozier, S., & Dunlosky, J. (2015). How do students improve their value-based learning with task experience? *Memory*, *23*(6), 928–942. https://doi.org/10.1080/09658211.2014.938083

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). John Benjamins. https://doi.org/10.1075/bct.16.02dro

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* (pp. 23–37). Oxford University Press.

Dunn, T. L., & Risko, E. F. (2016). Toward a metacognitive account of cognitive offloading. *Cognitive Science*, *40*(5), 1080–1127. https://doi.org/10.1111/cogs.12273

Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., & Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, *105*(1), 98–121. https://doi.org/10.1016/j.obhdp.2007.05.002

Eskritt, M., & Ma, S. (2014). Intentional forgetting: Note-taking as a naturalistic example. *Memory & Cognition*, *42*(2), 237–246. https://doi.org/10.3758/s13421-013-0362-1

Fröscher, L., Friedrich, A. K., Berentelg, M., Widmer, C., Gilbert, S. J., & Papenmeier, F. (2022). Framing cognitive offloading in terms of gains or losses: Achieving a more optimal use of reminders. *Cognitive Research: Principles and Implications*, *7*(1), Article 61. https://doi.org/10.1186/s41235-022-00416-3

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*(3), 501–517. https://doi.org/10.1037/xge0000652

Gilbert, S. J., Boldt, A., Sachdeva, C., Scarampi, C., & Tsai, P. C. (2023). Outsourcing memory to external tools: A review of 'intention offloading'. *Psychonomic Bulletin & Review*, *30*(1), 60–76. https://doi.org/10.3758/s13423-022-02139-4

Grinschgl, S., Meyerhoff, H. S., Schwan, S., & Papenmeier, F. (2021). From metacognitive beliefs to strategy selection: Does fake performance feedback influence cognitive offloading? *Psychological Research*, *85*(7), 2654–2666. https://doi.org/10.1007/s00426-020-01435-9

Hacker, D. J., & Bol, L. (2019). Calibration and self-regulated learning: Making the connections. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 647–677). Cambridge University Press. https://doi.org/10.1017/9781108235631.026

Händel, M., & Fritzsche, E. S. (2016). Unskilled but subjectively aware: Metacognitive monitoring ability and respective awareness in low-performing students. *Memory & Cognition*, *44*(2), 229–241. https://doi.org/10.3758/s13421-015-0552-0

Henkel, L. A. (2014). Point-and-shoot memories: The influence of taking photos on memory for a museum tour. *Psychological Science*, *25*(2), 396–402. https://doi.org/10.1177/0956797613504438

Jansen, R. A., Rafferty, A. N., & Griffiths, T. L. (2021). A rational model of the Dunning–Kruger effect supports insensitivity to evidence in low performers. *Nature Human Behaviour*, *5*(6), 756–763. https://doi.org/10.1038/s41562-021-01057-0

Kelly, M. O., & Risko, E. F. (2019a). Offloading memory: Serial position effects. *Psychonomic Bulletin & Review*, *26*(4), 1347–1353. https://doi.org/10.3758/s13423-019-01615-8

Kelly, M. O., & Risko, E. F. (2019b). The isolation effect when offloading memory. *Journal of Applied Research in Memory and Cognition*, *8*(4), 471–480. https://doi.org/10.1037/h0101842

Kelly, M. O., & Risko, E. F. (2022). Revisiting the influence of offloading memory on free recall. *Memory & Cognition*, *50*(4), 710–721. https://doi.org/10.3758/s13421-021-01237-3

Kimball, M. S. (1993). Standard risk aversion. *Econometrica*, *61*(3), 589–611. https://doi.org/10.2307/2951719

Knowlton, B. J., & Castel, A. D. (2022). Memory and reward-based learning: A value-directed remembering perspective. *Annual Review of Psychology*, *73*(1), 25–52. https://doi.org/10.1146/annurev-psych-032921-050951

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349

Krajc, M., & Ortmann, A. (2008). Are the unskilled really that unaware? An alternative explanation. *Journal of Economic Psychology*, *29*(5), 724–738. https://doi.org/10.1016/j.joep.2007.12.006

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology*, *82*(2), 180–188. https://doi.org/10.1037/0022-3514.82.2.180

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Kruger, J., & Dunning, D. (2002). Unskilled and unaware—But why? A reply to Krueger and Mueller (2002). *Journal of Personality and Social Psychology*, *82*(2), 189–192. https://doi.org/10.1037/0022-3514.82.2.189

Leding, J. K. (2019). Adaptive memory: Animacy, threat, and attention in free recall. *Memory & Cognition*, *47*(3), 383–394. https://doi.org/10.3758/s13421-018-0873-x

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control. *Trends in Cognitive Sciences*, *14*(10), 457–463. https://doi.org/10.1016/j.tics.2010.08.001

Lu, X., Kelly, M. O., & Risko, E. F. (2020). Offloading information to an external store increases false recall. *Cognition*, *205*, Article 104428. https://doi.org/10.1016/j.cognition.2020.104428

Madan, C. R. (2017). Motivated cognition: Effects of reward, emotion, and other motivational factors across a variety of cognitive domains. *Collabra. Psychology*, *3*(1), Article 24. https://doi.org/10.1525/collabra.111

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), 1–14. https://doi.org/10.1016/j.jarmac.2018.11.001

Miller, T. M., & Geraci, L. (2011). Unskilled but aware: Reinterpreting overconfidence in low-performing students. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 502–506. https://doi.org/10.1037/a0021802

Monin, B., & Oppenheimer, D. M. (2014). The limits of direct replications and the virtues of stimulus sampling. *Social Psychology*, *45*, 299–300. https://doi.org/10.1027/1864-9335/a000202

Murphy, D. H. (2023a). Does point value structure influence measures of memory selectivity? *Memory*, *31*(8), 1074–1088. https://doi.org/10.1080/09658211.2023.2221006

Murphy, D. H. (2023b). Strategic offloading: How the value of to-be-remembered information influences offloading decision-making. *Applied Cognitive Psychology*, *37*(4), 749–767. https://doi.org/10.1002/acp.4051

Murphy, D. H. (2024). *The metacognitive optimization of offloading task (MOOT): Both higher costs to offload and the accuracy of memory predict goodness of offloading performance*. https://osf.io/ynhev/?view_only=1e6e1ecc0b15490b8d8cef70c4efb87c

Murphy, D. H., Agadzhanyan, K., Whatley, M. C., & Castel, A. D. (2021). Metacognition and fluid intelligence in value-directed remembering. *Metacognition and Learning*, *16*(3), 685–709. https://doi.org/10.1007/s11409-021-09265-9

Murphy, D. H., & Castel, A. D. (2022). Selective remembering and directed forgetting are influenced by similar stimulus properties. *Memory*, *30*(9), 1130–1147. https://doi.org/10.1080/09658211.2022.2092152

Murphy, D. H., & Castel, A. D. (2023). Age-related differences in memory when offloading important information. *Psychology and Aging*, *38*(5), 415–427. https://doi.org/10.1037/pag0000750

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*(3), 910–921. https://doi.org/10.3758/s13423-021-02030-8

Murphy, D. H., Rhodes, M. G., & Castel, A. D. (2024). The perceived importance of words in large font guides learning and selective memory. *Memory & Cognition*, *52*(7), 1463–1476. https://doi.org/10.3758/s13421-024-01555-2

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). Springer International Publishing. https://doi.org/10.1007/978-3-319-29986-0_10

Nairne, J. S., VanArsdall, J. E., & Cogdill, M. (2017). Remembering the living: Episodic memory is tuned to animacy. *Current Directions in Psychological Science*, *26*(1), 22–27. https://doi.org/10.1177/0963721416667711

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*(10), 2099–2105. https://doi.org/10.1177/0956797613480803

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist*, *51*(2), 102–116. https://doi.org/10.1037/0003-066X.51.2.102

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 125–173). Academic Press. https://doi.org/10.1016/S0079-7421(08)60053-5

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*(4), 1367–1376. https://doi.org/10.3758/s13423-022-02064-6

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*(2), 186–201. https://doi.org/10.1037/xlm0000174

Popp, E. Y., & Serra, M. J. (2018). The animacy advantage for free-recall performance is not attributable to greater mental arousal. *Memory*, *26*(1), 89–95. https://doi.org/10.1080/09658211.2017.1326507

Rhodes, M. G. (2016). Judgments of learning. In J. Dunlosky & S. K. Tauber (Eds.), *The oxford handbook of metamemory* (pp. 65–80). Oxford University Press.

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. https://doi.org/10.1016/j.concog.2015.05.014

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Roediger, H. L., III. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning & Verbal Behavior*, *12*(6), 644–657. https://doi.org/10.1016/S0022-5371(73)80044-1

Sachdeva, C., & Gilbert, S. J. (2020). Excessive use of reminders: Metacognition and effort-minimisation in cognitive offloading. *Consciousness and Cognition*, *85*, Article 103024. https://doi.org/10.1016/j.concog.2020.103024

Schacter, D. L. (1999). The seven sins of memory. Insights from psychology and cognitive neuroscience. *American Psychologist*, *54*(3), 182–203. https://doi.org/10.1037/0003-066X.54.3.182

Schacter, D. L. (2022). The seven sins of memory: An update. *Memory*, *30*(1), 37–42. https://doi.org/10.1080/09658211.2021.1873391

Serra, M. J. (2021). Animate and inanimate words demonstrate equivalent retrieval dynamics despite the occurrence of the animacy advantage. *Frontiers in Psychology*, *12*, Article 661451. https://doi.org/10.3389/fpsyg.2021.661451

Serra, M. J., & DeMarree, K. G. (2016). Unskilled and unaware in the classroom: College students' desired grades predict their biased grade predictions. *Memory & Cognition*, *44*(7), 1127–1137. https://doi.org/10.3758/s13421-016-0624-9

Slamecka, N. J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology*, *76*(4, Pt. 1), 504–513. https://doi.org/10.1037/h0025695

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, *333*(6043), 776–778. https://doi.org/10.1126/science.1207745

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*(2), 182–188. https://doi.org/10.1177/0956797614559285

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*(6), 717–723. https://doi.org/10.1080/09658211.2016.1210171

Thiede, K. W. (1999). The importance of monitoring and self-regulation during multitrial learning. *Psychonomic Bulletin & Review*, *6*(4), 662–667. https://doi.org/10.3758/BF03212976

Tsai, P.-C., Scarampi, C., Kliegel, M., & Gilbert, S. J. (2023). Optimal cognitive offloading: Increased reminder usage but reduced proreminder bias in older adults. *Psychology and Aging*, *38*(7), 684–695. https://doi.org/10.1037/pag0000751

Weis, P. P., & Wiese, E. (2019). Problem solvers adjust cognitive offloading based on performance goals. *Cognitive Science*, *43*(12), Article e12802. https://doi.org/10.1111/cogs.12802

Received May 2, 2024
Revision received November 21, 2024
Accepted December 8, 2024 ■