Individual Differences in Prospective and Retrospective Memory Offloading

Lauren L. Richmond[1], Lois K. Burnett[1], Julia Kearley[1,2], Samuel J. Gilbert[3], Alexandra B. Morrison[4], B. Hunter Ball[5]

[1]Department of Psychology, Stony Brook University
[2]Department of Psychology, McGill University
[3]Institute of Cognitive Neuroscience, University College London
[4]Department of Psychology, California State University, Sacramento
[5]Department of Psychology, University of Texas at Arlington

Author Note

Lauren L. Richmond https://orcid.org/0000-0001-6793-5773

Lois K. Burnett https://orcid.org/0000-0002-1679-4621

Julia Kearley https://orcid.org/0000-0002-6085-314X

Samuel J. Gilbert https://orcid.org/0000-0002-3839-7045

These data were presented at the 2023 Symposium for Individual Differences in Cognition in San Francisco, CA. Preregistration, data, and analysis scripts are available on OSF [This view-only link will be replaced with a publicly available link upon acceptance]. Correspondence concerning this article should be addressed to Lauren L. Richmond, Department of Psychology, Stony Brook University; Psychology B Building, Stony Brook, NY 11794-2500; email: lauren.richmond@stonybrook.edu

Abstract

Prior research focused on the relationship between cognitive offloading and working memory ability in the prospective and retrospective memory domains have produced conflicting results. Specifically, past work in the prospective memory domain has found that individuals with lower working memory capacity (WMC) choose to offload more often and benefit more from offloading than those with higher WMC (Ball, Peper, et al., 2022) while work in the retrospective memory domain has not found a relationship between WMC and the use of or benefit from offloading (Morrison & Richmond, 2020). However, task design across studies differed in several other respects aside from memory domain making it difficult to discern whether different mechanisms underlie performance across domains. The current study aimed to address these discrepancies by introducing similar procedures across offloading tasks. Results revealed that when offloading was required or permitted, participants with high and low WMC generally performed more similarly to one another than when the task had to be completed using internal memory alone. In addition, participants with lower WMC generally benefitted more from offloading compared to those with higher WMC when offloading was required and when participants had free choice about whether and when to engage in offloading. However, neither metacognitive underconfidence in internal memory capability nor low WMC estimates were associated with increased offloading frequency in either memory domain when participants were permitted to offload. Practical and theoretical implications of our findings are discussed.

Keywords: working memory capacity; prospective memory; retrospective memory; cognitive offloading

Individual Differences in Prospective and Retrospective Memory Offloading

Cognitive offloading, defined as the use of physical action to reduce internal cognitive demand (Risko & Gilbert, 2016), is ubiquitous in everyday life. Common examples of cognitive offloading include noting upcoming appointments in a calendar and jotting down notes during an important meeting. Cognitive offloading can be used to support both prospective (upcoming appointments) and retrospective (meeting notes) memory. Studies have consistently shown that people perform better when they have the opportunity to offload information compared to relying on internal memory alone (Burnett & Richmond, under review; Gilbert et al., 2022). Recent studies have also suggested that cognitive offloading can benefit internal memory by freeing up resources that can be reallocated to non-offloaded content (i.e., the saving enhanced memory effect; Storm & Stone, 2015). This suggests that individuals with poorer internal memory ability should benefit most from offloading. However, support for this idea is mixed. While differences in prospective memory performance due to poor internal working memory ability can be eliminated by offloading (Ball, Peper, et al., 2022), this has not been observed in the retrospective memory domain (Brown, 2021; Morrison & Richmond, 2020). The purpose of the current study was therefore to better understand the theoretical mechanisms that may underlie the benefits of offloading across different memory domains using an individual differences approach.

Prospective and Retrospective Memory Offloading

One laboratory paradigm that has been used extensively to study cognitive offloading for prospective memory intentions involves dragging numbered circles in numerical order to different locations on screen, referred to as the intention offloading task (Gilbert, 2015a). Most circles are to be dragged to the bottom of the screen, but some ‘special’ circles are indicated by appearing on screen in a color corresponding to the color appearing at the left, right, or top of the screen before fading to the same color as all other circles. See Figure 1 panel A for a depiction of this task. In the internal memory condition, participants must rely on their own internal memory to remember the number and intended location for each special circle while interacting with each circle in strict numerical order. In the cognitive offloading condition, however, participants can move the special circles closer to their goal locations out of order to serve as a reminder of where to move the circle when the time comes. Participants perform better when they have the opportunity to (Boldt & Gilbert, 2019; Chiu & Gilbert, 2023; Gilbert, 2015b, 2015a) or are forced to (Ball, Peper, et al., 2022; Engeler & Gilbert, 2020; Gilbert et al., 2020; Sachdeva & Gilbert, 2020; Scarampi & Gilbert, 2020) engage in cognitive offloading compared to relying on internal memory alone.

Another common task in the cognitive offloading literature that focuses on retrospective memoranda is the letter string offloading task (Risko & Dunn, 2015). As depicted in Figure 1, Panel C participants are presented with letter strings of varying lengths and asked to reproduce the letter string at the end of each trial. In the internal memory block, participants must report the letters they were presented in order from internal memory. In the cognitive offloading block, participants can create and use an external aid to support their ability to report the presented letter string. Performance is consistently benefitted by having the opportunity to offload (Brown, 2021; Burnett & Richmond, 2023; Morrison & Richmond, 2020; Risko & Dunn, 2015). See Figure 1 panel C for a task schematic.

Despite the consistency of findings across the literature around the benefits of cognitive offloading (Burnett & Richmond, under review; Gilbert et al., 2022), there are individual differences in the extent to which people may choose to engage in offloading and the benefits that they derive from doing so. One individual differences factor that may relate to offloading behavior and benefit is working memory. Working memory refers to the attention and memory processes needed to actively maintain goal-relevant information in the focus of attention and retrieve from long-term memory information that has been displaced from focal awareness (Unsworth & Engle, 2007).

In the prospective memory domain, Ball, Peper, et al. (2022) administered several versions of the intention offloading task that included blocks of forced internal (no reminder) trials, blocks of forced external (reminder) trials, and blocks in which participants were able to choose whether or not to offload on each trial (optional reminders). Participants with lower working memory capacity (WMC) performed worse on the intention offloading task when they had to rely on their own internal memory ability compared to participants with higher WMC, but this effect was eliminated when participants were forced to offload. Moreover, low WMC participants were more likely to offload when given the choice.

Turning to the retrospective memory domain, Morrison and Richmond (2020) had participants complete a version of the letter string offloading task that included forced internal (no reminder) and choice (optional reminders) blocks under varying memory load (e.g., 2-10 letters). As in the Ball, Peper, et al. (2022) study, participants with low WMC did worse on the retrospective memory task when they had to rely on their own internal memory, and the option to offload improved performance for participants across the board. However, in contrast to the findings of Ball, Peper, et al. (2022), low and high WMC participants exhibited similar benefits from offloading (i.e., offloading did not eliminate internal differences due to working memory) and WMC was unrelated to the frequency with which participants chose to offload. There are several possible reasons why these discrepant results emerged across these two studies, which we highlight below.

Theoretical Accounts of Cognitive Offloading

One influential account of people's decision to engage in cognitive offloading emphasizes the subjective metacognitive evaluations of one's own internal memory abilities. According to the metacognitive account, people use offloading as a memory aid when they are unconfident in the own internal memory abilities to increase the likelihood that the to-be-remembered information will successfully be retrieved. Indirect support for this idea comes from past work showing that participants offload more frequently in situations in which their internal memory performance is poor, such as under high memory load or when there are task interruptions. They also offload more when they receive negative (as opposed to positive) feedback about their task performance (Gilbert et al., 2020). More directly, research has found people offload more frequently when metacognitive predictions of internal memory are lower than actual performance (Gilbert, 2015b). These findings suggest that participants rely on metacognitive assessments of their own internal memory abilities to inform decisions to offload.

Notably, Ball, Peper, and colleagues (2022) found that low and high WMC participants were equally underconfident in their own internal memory abilities during the intention offloading task. It is possible, however, that low WMC participants in the study by Morrison and Richmond (2020) may have been overconfident in their ability, resulting in less effective offloading decisions that would otherwise compensate for their poorer internal memory ability. One notable difference between the two task types was that accuracy feedback was provided in the intention offloading task, whereby circles briefly turned red for incorrect responses and green for correct responses as they moved off screen, but not in the letter string offloading task. The provision of feedback in the intention offloading task may have helped participants better calibrate their metacognitive evaluations to their ability, resulting in more effective use of reminders. Another difference between these studies that might impact metacognition was that the study by Ball, Peper, and colleagues (2022) included forced external trials, which served to mitigate any differences in performance between low and high ability participants that may otherwise negatively influence metacognitive evaluations. The study by Morrison and Richmond (2020), in contrast, only included choice offloading trials.

An alternative view proposes that people may offload to minimize the amount of effort they need to expend to complete the task. One way that people may reduce effort is by relying on an external store to reduce internal cognitive demand. This stems from the idea that internally representing information can be cognitively demanding, as committing multiple items to memory during encoding can be difficult and maintaining these representations can interfere with ongoing activities (Smith, 2003). Offloading provides a means to avoid engaging these effortful internal memory processes. Indirect support for this account comes from findings indicating that participants are more willing to rely on internal processes when motivated (financially) to do so or when the demands associated with checking reminders is high (Ball & Peper, 2022; Grinschgl et al., 2021; Sachdeva & Gilbert, 2020)**.** More directly, research shows that participants study to-be-remembered information for shorter/longer durations when they know reminders will/will not later be available during retrieval (Kelly & Risko, 2022; Peper & Ball, 2023). At the same time, the creation of the external store requires physical effort. In this context, if participants believe that it would be more effortful to create an external store than to retrieve information from memory, they may opt not to engage in cognitive offloading. In general, participants may weigh the relative demands of using internal versus external processes when deciding whether to offload, but it is possible that this computation differs for prospective and retrospective memory demands.

Finally, people may choose to offload because it allows for attention to be reallocated toward other information. For example, the saving-enhanced memory effect refers to the finding that offloading some previously learned information can improve memory for newly learned, but not offloaded, information (Storm & Stone, 2015). In the first study examining the saving-enhanced memory effect, participants were shown two lists of words, labeled List A and List B. Participants first studied List A, after which they were told to save the file or not to save the file. Next, all participants studied the words on List B and recalled the words on List B after a 20 second filled delay. Interestingly, Storm and Stone (2015) found that participants were able to recall a greater proportion of List B items on save versus non-save trials (see also Tsai et al., 2023). This effect replicated in a second experiment where the authors also demonstrated that the observed effect depended critically on the reliability of the saving process. These results suggest that when the saving process is reliable, the act of saving some information frees up internal memory resources to devote to the encoding and retrieval of other information. To the extent that the participants in the study by Ball, Peper, et al. (2022) experienced overall higher memory demand (i.e., six target circles per trial) than participants in the Morrison and Richmond (2020) study (i.e., trials ranged in length from 2-10 items, meaning that memory load was low on some trials), low WMC participants may have been more sensitive to the benefits of offloading for facilitating ongoing task completion in the intention offloading task.

Finally, there are shared and distinct processes that contribute to memory performance across the two domains. Prospective memory is comprised of two components: the *prospective component* involves noticing the target and becoming aware that an intended action should be initiated (i.e., remembering that there is *something* to do), while the *retrospective component* involves remembering the contents of the intention and retrieving the action from long-term memory (i.e., remembering what to remember; Einstein & McDaniel, 1990; Guynn et al., 1998). Critically, the retrospective component is common across both prospective and retrospective memory tasks, while the prospective component is unique to prospective memory tasks. Interestingly, Landsiedel and Gilbert (2015) measured brain activity during forced internal and external blocks and found that offloading reduced signal change in brain regions associated with remembering what to remember (i.e., retrospective component), but not in areas associated with remembering that something needed to be done (i.e., prospective component). Thus, it is possible that offloading is particularly effective for individuals with poor working memory ability in the intention offloading task (Ball, Peper, et al., 2022) because it frees up resources to allocate toward the prospective component. In contrast, it is possible that offloading the retrospective component, which is common to both prospective and retrospective memory tasks, may be equally beneficial for low and high WMC participants.

Current Study

The benefits from offloading as a function of WMC have been found to differ across prospective and retrospective memory tasks (Ball, Peper, et al., 2022; Morrison & Richmond, 2020; see also Meyerhoff et al., 2021). This could reflect important theoretical differences underlying decisions to offload and benefits from offloading across the two domains (e.g., metacognitive confidence, prospective vs. retrospective component, etc.), differences in methodology (e.g., feedback, load, choice, etc.) employed in the two task types, or some combination of both. The purpose of the current study was to hold constant the methodological differences across tasks to better understand whether the benefits conferred from offloading are domain-general or domain-specific. We detail our preregistered hypotheses below. In brief, we anticipated that when holding methodologies constant, low WMC participants would offload more frequently and benefit more from offloading than high WMC participants. We also anticipated that greater metacognitive underconfidence would be associated with increased offloading frequency.

Data Availability

Data, analysis scripts, and output resulting from our data analysis are available on OSF. [Note: This view-only link will be replaced with a publicly accessible link upon acceptance.]

Method

This study was preregistered: https://osf.io/42tb6

Participants

Full datasets were collected from a total of 363 participants from the Department of Psychology SONA participant pools at Stony Brook University and the University of Texas at Arlington. Of these, data were excluded from 155 participants due to failure to meet minimum requirements for inclusion as outlined in our preregistration, leaving us with a final usable sample size of 208 participants. Specifically, data were excluded for the following reasons: failing more than 50% of the attention checks (n = 60), reporting cheating on one or more of the tasks (n = 36), reporting that they believed that their data on one or more of the tasks should not be used (n = 25), exhibiting lower performance in the forced external condition compared to the internal memory condition on one or both of the offloading tasks (n = 23), failing to offload on an adequate number of trials in the forced offloading condition (n = 8), responding accurately on less than 50% of trials at set size 2 in the choice block of the letter string offloading task (n = 2), and for failing to score above the guess rate for processing portion of one of our complex span working memory tasks (n = 1) (n = 1; see Richmond et al., 2022). Our final sample had a mean age of 20.39 years (SD = 4.78 years, Range: 17-47). With respect to race and ethnicity, 32.70% of our sample identified as White, 27.90% as Asian/Pacific Islander, 21.60% as Hispanic/Latinx, 12.50% as Black, and 4.33% of participants chose the response 'Other/Unknown,' and 0.96% of participants did not provide a response to this question. All procedures were reviewed and approved by the Institutional Review Boards at Stony Brook University and University of Texas at Arlington.

Our *a priori* power analysis indicated a necessary sample size of 284 participants. However, the large number of data exclusions based on our preregistered exclusion criteria left us with a slightly smaller sample size than we powered for. Nonetheless, a post-hoc sensitivity analysis suggested that we were well powered to detect significant bivariate correlation between internal memory performance and offloading behavior as small as -.17 with 80% power. This correlation is in line with the effect that we originally powered the study to detect.

Procedure

Study procedures were completed online in the following fixed task order: intention offloading task, Operation Span, letter string offloading task, Symmetry Span. Each offloading task consisted of three blocks of trials presented in a fixed order: forced internal, forced external, cognitive offloading choice. Task order and block order within the offloading tasks were fixed in order to best capture individual differences (Robison et al., 2023). The intention offloading task was completed via a web-based interface; all other tasks were completed via EPrime GO (Psychology Software Tools, 2020). The detailed method for each task is described below. After participants completed all four tasks, they provided their demographic information, completed a post-study questionnaire, and were debriefed. The study took approximately 2 hours in total to complete, and participants were awarded partial course credit for their participation. Participants were permitted 24 hours to complete all study procedures. See Figure 1 for task schematics.

Intention Offloading Task

The materials and stimuli for the intention task were adapted from Gilbert et al. (2020; see Figure 1, panel A for a task schematic). At the start of each trial, six yellow circles were presented within a square on the computer. Each circle contained a number (1-6) and participants were to drag the circles in ascending numerical order to the bottom of the square. Each time a circle was dragged to the bottom of the square, a new circle appeared in its original location, continuing the numerical sequence until participants had been presented with all circles in a given trial. Occasionally, new circles (i.e., targets) initially appeared in blue, orange, or pink, rather than yellow, which corresponded with the left, top, and right side of the square, respectively. Two seconds after appearing on the screen, the color faded to yellow so that they matched the other circles. When a target appeared (e.g., in orange), this represented an instruction that it should eventually be dragged to its corresponding side of the square (e.g., top) when it was reached in the numerical sequence. For example, a participant first drags the circle labeled ‘1’ to the bottom of the screen where it disappears. Next, an orange circle labeled ‘7’ appears in its place, fading to yellow after two seconds. Meanwhile, the participant drags circles 2-6 to the bottom of the screen, before dragging the ‘7’ circle to the top. Importantly, targets can be remembered in two different ways. Participants can rely on their own internal representation of where it should eventually be dragged (i.e., no reminder) or can set an external reminder as soon as it appears by moving it near the location (e.g., top) where it eventually needs to be dragged.

The set size per trial was either 1, 3, 5, or 7 prospective memory targets, respectively, which corresponded to 7, 9, 11, or 13 ongoing task stimuli presented in ascending numerical order. The targets were randomly selected to appear for circles numbered 7-13 and were randomly allocated to the left, top, and right positions of the square. Feedback was provided by the circle changing color before disappearing if dragged to the correct location (green) or incorrect location (red). All circles correctly dragged to the bottom of the box turned purple before disappearing.

There were 3 separate blocks presented in a fixed order: forced internal (no reminder), forced external (reminder), and choice. Within each block, there were 16 trials, such that each set size was randomly selected to occur 4 times. In the forced internal block, circles were fixed in position on the screen (other than the current one that needed to be dragged in sequence) so that target circles could not be moved when they first appeared. In the forced external block, when a target circle appeared the task could only be continued after the participant moved it within the square. In the choice block, participants were given a free choice to use an internal strategy or set reminders in the upcoming trial.

After receiving the ongoing task instructions for the task, participants performed a practice trial with 7 circles where they moved the numbers, in order, to the bottom of the screen. Participants were then given the prospective memory instructions, followed by a 7-circle practice where the last circle was a target (set size 1). They had to drag the target circle to the correct location to proceed without the use of reminders. They then received practice with target set sizes of 3, 5, and then 7. Following this practice, participants were asked to predict what percentage of target circles (from 0-100%) they thought they would remember to drag to the appropriate side of the square during the actual task, separately for each set size. The comparison between predicted performance levels and actual performance levels formed the basis for our metacognitive calibration variable. Finally, participants were given instructions on how to set reminders and performed a full 13-circle practice phase (target set size 7) where reminder usage was required. They then completed the 3 separate blocks of 16 trials as described above.

Letter String Offloading Task

Participants were presented with letter strings ranging in length from two to eight letters in length by twos (2, 4, 6, or 8 letters; 4 cycles of each set size per block). Presentation was audiovisual, with letters being spoken aloud as well as displayed on the screen at a rate of 1 s each with a 500 ms ISI. Prior to task initiation, participants underwent a procedure to ensure that they were able to hear the audio portion of the task at a comfortable volume. Participants also completed internal memory practice trials (4 trials at set size 2) to get comfortable with the task and procedure for reporting the presented stimuli.

After the practice trials, participants completed one trial at each set size (2, 4, 6, and 8) using internal memory, and were then asked to predict what percentage of letters they would be able to recall in the correct order at each set size during the actual task. Comparison of predicted performance levels to actual performance levels formed the basis for our metacognitive calibration variable.

At the start of each trial, participants were shown a screen that told them how may letters they would be presented in the upcoming trial. Immediately after this screen, letter presentation began up to the length indicated on the initial screen. Immediately following presentation of the last letter in the trial, participants were asked to report the presented letters in serial order. In the internal memory block, participants had to rely on their own internal memory to reproduce the presented information. In the forced offloading block, participants were required to type the letter into a text box on screen as each letter was presented; their typed notes were made available to them at test. In the offloading choice block, participants had the opportunity to type letters during the encoding phase but were not required to do so. Typed letters were made available to them at test. After providing their responses for each trial, participants were provided feedback about their performance. Specifically, participants were told if their response was correct, incorrect, or if there was no response detected at the end of each trial. See Figure 1, panel C for a task schematic.

Complex Span Working Memory Tasks

Participants also completed shortened versions of the Operation Span and Symmetry Span tasks (Foster et al., 2015; Unsworth et al., 2005). Shortened versions of these tasks have been shown to have adequate reliability (Foster et al., 2015), and are less onerous and time consuming for participants to complete than the full (standard) task versions. Here, participants completed only two blocks of each task rather than three blocks as is the norm for the standard version of these tasks.

Both tasks consist of a processing component and a storage component, presented in an interleaved fashion. Participants are told that they should report the storage items in the order that they were presented. Participants were not told ahead of time how many items they would have to remember for each trial. After inputting their responses for the storage task, however, participants were provided with feedback about their performance level on both the processing and storage portions of the task before moving on to the next trial.

Briefly, Operation Span consisted of participants solving simple math problems (processing) and remembering letters (storage), presented in an interleaved fashion. Trials ranged in length from three to seven letters. See Figure 1, panel B for a task schematic. Symmetry Span involves making symmetry decisions about an 8x8 black-and-white grid about the vertical axis (processing) and remembering locations highlighted in red in a 4x4 grid (storage). Set sizes ranged from two to five items. See Figure 1, panel D for a task schematic.

Research Questions & Hypotheses

Our key research questions and hypotheses were pre-registered, but not all of our pre-registered analyses are reported in the manuscript. Pre-registered hypotheses and analyses not reported here can be found in the Supplemental Materials document. We also conducted some additional exploratory analyses to fully characterize our patterns across different offloading conditions (forced external vs. choice) and task (intention offloading vs. letter string offloading) that we report here. Such analyses are clearly marked as exploratory throughout the document.

Specifically, we first wanted to test whether individuals with high and low WMC would perform more similarly on the offloading tasks when offloading is required compared to when participants were required to rely on internal memory. We hypothesized that in the forced external block participants would perform similarly regardless of their WMC. However, we also considered the possibility that this effect would emerge only for the intention offloading task given that similarity in performance has not been observed in the letter string offloading task (albiet with only a choice offloading condition; see Morrison & Richmond, 2020). This was tested using a preregistered[1] mixed 2 (task: intention offloading vs letter string) x 4 (load: levels 1, 2, 3, 4) x 2 (condition: forced internal vs. forced external) x 2 (WMC: high vs. low) ANOVA. Participants were classified as high WMC and low WMC based on a tertile split of combined working memory capacity z-scores (Redick et al., 2012). If this omnibus test revealed a significant 4-way interaction, we planned to conduct separate follow-up tests for each offloading

---

[1] Our analysis plan on OSF did not clearly state that we intended for this analysis to focus specifically on high vs. low WMC participants.

task at each memory load level. Further, we conducted a series of exploratory analyses to determine whether similar patterns would emerge when participants were given the choice to offload rather than being required to do so, following the same analysis plan as described above to facilitate comparison across offloading conditions.

Next, we conducted a series of exploratory analyses to test whether WMC was associated with the magnitude of the benefit afforded by offloading. These analyses were motivated specifically by the opposing patterns reported by Ball, Peper, et al. (2022) where WMC was associated with the magnitude of the benefit afforded by offloading and the findings from Morrison and Richmond (2020) where this effect was not observed. Here, we hypothesized that the benefit conferred by offloading would be associated with WMC. We thought that this association would be stronger than has been reported by past research given that all participants in this study would have had prior experience completing the offloading tasks in an unaided fashion (forced internal) as well as experience with offloading (in the forced external block) prior to the offloading choice trials and participants received feedback about their performance. However, we considered that it was also possible that the expected pattern may not emerge given that this pattern did not emerge in the Morrison and Richmond (2020) study. We first tested this question using bivariate correlations between WMC with offloading benefit scores in each task at each load separately in both the forced-external and choice offloading conditions. Correlation strength across tasks was compared. Next, we used a simultaneous regression to predict offloading benefit in the choice offloading condition at each memory load with working memory and offloading frequency entered as predictors separately for each task.

In addition, we investigated the association between metacognition and offloading choice behavior. Given that metacognitive evaluations have been reported to be a better predictor of offloading behavior than overt memory ability in the intention offloading task (Boldt & Gilbert, 2019; Gilbert et al., 2020), we expected that, overall, participants would be underconfident in their own internal memory ability (Gilbert et al., 2020), and that the difference between actual and predicted performance may increase at higher memory loads. We further expected that the degree of underconfidence in one's internal memory would predict the frequency with which participants chose to offload. Specifically, we expected that greater underconfidence would be associated with higher rates of offloading in both tasks. We tested this pregistered hypothesis by correlating metacognitive calibration (predicted performance minus actual performance in the forced internal condition) with offloading choice behavior at each load separately for each task.

Finally, we tested whether WMC would be associated with the frequency of offloading in the choice offloading block. We anticipated that lower WMC participants would offload more often than higher WMC participants at higher memory loads. By the time that participants are completing the choice block, they have already had experience with both carrying out the task using internal memory alone (forced internal) and with completing the task when offloading was required, which we thought could allow them to both learn about their ability to perform the task in an unaided fashion and the benefits to performance conferred by offloading. Together, we expected that these experiences might drive participants to offload in ways that better corresponded to the limits of their internal memory abilities. However, there are some past data in both tasks that suggest that internal memory may not be a good predictor of offloading choice behavior (Morrison & Richmond, 2020; Scarampi & Gilbert, 2021), so we considered that a plausible alternative hypothesis would be that WMC would not predict offloading choice behavior in either task. To address this question, we conducted a preregistered correlation between offloading frequency in the choice block at each load with WMC separately for each task and compared correlation strengths to one another.

Results

Results from our key research questions and corresponding analyses are presented below. Results stemming from additional pre-registered research questions can be found in our Supplemental Materials document. Greenhouse-Geisser corrections were used when sphericity was violated, which is denoted with a “GG” subscript in the F statistic (F_{GG}). False discovery rate corrections for multiple comparisons were applied for correlations conducted at each memory load level.

Does the opportunity to offload (forced external, choice offloading) result in high and low WMC participants performing more similarly to one another than they do in the internal memory condition?

Data were split into tertiles based on WMC estimates extracted from combined z-scores derived from performance on both Operation Span and Symmetry Span. This left us with 68 participants who had high WMC and 70 participants who had low WMC for the below analyses.

Forced Offloading

We conducted our planned 2 (task: intention offloading vs letter string) x 4 (load: levels 1, 2, 3, 4) x 2 (condition: forced internal vs. forced external) x 2 (WMC: high vs. low) ANOVA. Effects are reported in Table 1 for brevity. Of specific interest is the significant four-way interaction ($p = .047$), which motivated us to conduct the follow up tests reported below. The significant four-way interaction indicates that task performance differs as a function of load level, condition, task, and WMC. See Figure 2. Follow-up analyses were conducted separately for each task.

Prospective Memory: Intention offloading task. Our planned follow up 2 (span: low vs. high WMC) x 4 (load: levels 1, 2, 3, 4) x 2 (condition: forced internal vs. forced external) ANOVA revealed a significant main effect of span (F (1, 136) = 21.81, $p < .001$), a significant main effect of load (F_{GG} (2.85, 388.28) = 150.58, $p < .001$), and a significant main effect of condition (F (1, 136) = 299.38, $p < .001$). These main effects were qualified by significant span x condition (F (1, 136) = 18.89, $p < .001$), span x load (F_{GG} (2.85, 388.28) = 3.07, $p = .003$), and condition x load (F_{GG} (2.82, 383.92) = 122.66, $p < .001$) interactions. These effects were further qualified by a significant three-way span x load x condition interaction (F_{GG} (2.82, 383.92) = 3.14, $p = .028$). We followed up this significant three-way interaction be examining the interaction between span and condition at each memory load separately. At load level 1, there was a significant main effect of span (F (1, 136) = 4.84, $p = .030$) but no significant effect of condition (F (1, 136) = 1.11, $p = .294$) and no span x condition interaction (F (1, 136) = 2.55, $p = .112$). This indicates that participants with high WMC outperformed low WMC participants at load level 1 across both conditions. At load levels 2, 3, and 4, a different picture emerges. Specifically, there were significant main effects of span (largest $p < .001$) and condition (largest $p < .001$). These main effects were qualified by significant span x condition interactions (largest $p = .009$). Beyond load level 1, participants with low WMC exhibited a larger difference in performance when comparing performance in the forced internal and forced external blocks compared to participants with high WMC, suggesting that participants with low WMC benefitted more from being forced to offload than participants with high WMC. Further, these results support our hypothesis that high and low WMC participants would perform more similarly in the forced external condition compared to the forced internal condition. See Figure 2, panel A.

Retrospective Memory: Letter string offloading task. Our planned follow up 2 (span: low vs. high WMC) x 4 (load: levels 1, 2, 3, 4) x 2 (condition: forced internal vs. forced external) ANOVA revealed a significant main effect of span (F (1, 136) = 34.73, $p < .001$), a significant main effect of load (F_{GG} (2.45, 333.77) = 421.01, $p < .001$) and a significant main effect of condition (F (1, 136) = 488.18, $p < .001$). These effects were qualified by significant span x condition (F (1, 136) = 8.27, $p < .001$), span x load interaction (F_{GG} (2.45, 333.77) = 7.84, $p < .001$) and load x condition (F_{GG} (2.28, 310.65) = 303.26, $p < .001$) interactions. These effects were further qualified by a significant three-way span x load x condition interaction (F_{GG} (2.28, 310.65) = 6.49, $p = .001$). We followed up this significant three-way interaction be examining the interaction between span and condition at each memory load separately. At load levels 1, 2, and 3, there were significant main effects of span (largest $p = .044$) and condition (largest $p = .026$) that were qualified by significant span x condition interactions (largest $p = .018$). These patterns indicate that performance for low WMC participants benefitted more from being forced to offload than high WMC participants at load levels 1, 2, and 3. Further, participants with high and low WMC performed more similarly to one another in the forced external condition compared to the forced internal condition. At load levels 4, a significant main effect of span (F (1, 136) = 16.87, $p < .001$) and condition (F (1, 136) = 584.89, $p < .001$) emerged, but there was no interaction between span and condition (F (1, 136) = 2.47, $p = .119$). At the highest load level, participants with high WMC outperformed participants with low WMC, and performance was better in the forced external condition compared to the internal memory condition. See Figure 2, panel B.

Choice Offloading

To determine whether the patterns observed above are similar when people are given a choice to offload rather than being required to do so, we conducted a parallel exploratory analysis to the analysis described above focused on the comparison between choice offloading and forced internal conditions. As above, we conducted a 2 (task: intention offloading vs. letter string) x 4 (load: levels 1, 2, 3, 4) x 2 (condition: forced internal vs. choice offloading) x 2 (WMC: high vs. low) ANOVA. Effects are reported in Table 2 for brevity. Of note, in this four-way interaction did not reach significance ($p = .318$); other patterns observed here are qualitatively similar to what was observed for our four-way ANOVA involving the forced external condition (see Tables 1 and 2). Therefore, we moved forward with separate analyses of each task to facilitate comparisons to results from the forced offloading condition reported above.

Prospective Memory: Intention offloading task. Our follow up 2 (span: low vs. high WMC) x 4 (load: levels 1, 2, 3, 4) x 2 (condition: forced internal vs. choice offloading) ANOVA revealed a significant main effect of span (F (1, 136) = 24.10, $p < .001$), a significant main effect of load (F_{GG} (2.84, 385.68) = 135.95, $p < .001$), and a significant main effect of condition (F (1, 136) = 232.81, $p < .001$). These main effects were qualified by significant span x condition (F (1, 136) = 13.51, $p < .001$), span x load (F_{GG} (2.84, 385.68) = 3.44, $p = .019$), and condition x load (F_{GG} (2.87, 390.77) = 122.98, $p < .001$) interactions. The three-way span x load x condition interaction was not significant (F_{GG} (2.87, 390.77) = 2.50, $p = .062$). Nevertheless, to maintain consistency with the forced offloading analyses reported above, we followed up the three-way interaction be examining the interaction between span and condition at each memory load separately. Consistent with the forced offloading analyses, significant main effects of span (largest $p = .001$) and condition (largest $p < .001$) emerged; these were qualified by significant span x condition interactions (largest p = .011) at load levels 2, 3, and 4. At load level 1, only a significant main effect of span emerged (F (1, 136) = 6.31, p = .013; condition: F (1, 136) = 0.50, p = .479; span x condition: F (1, 136) = 1.45, p = .231). Beyond load level 1, low and high WMC participants performed more similarly to one another in the choice offloading condition compared to the forced internal condition, and performance for participants with low WMC benefitted more from the choice to offload compared to participants with high WMC.

Retrospective Memory: Letter string offloading task. Our follow up 2 (span: low vs. high WMC) x 4 (load: levels 1, 2, 3, 4) x 2 (condition: forced internal vs. choice offloading) ANOVA revealed a significant main effect of span (F (1, 136) = 39.50, p < .001), a significant main effect of load (F_{GG} (2.42, 328.99) = 449.16, p < .001) and a significant main effect of condition (F (1, 136) = 407.77, p < .001). These effects were qualified by significant span x condition (F (1, 136) = 5.83, p < .017), span x load interaction (F_{GG} (2.42, 328.99) = 13.12, p < .001) and load x condition (F_{GG} (2.48, 337.05) = 293.07, p < .001) interactions. These effects were further qualified by a significant three-way span x load x condition interaction (F_{GG} (2.48, 337.05) = 2.92, p = .044). We followed up this significant three-way interaction be examining the interaction between span and condition at each memory load separately. No significant main effects or interactions emerged at load level 1 (smallest p = .120). At load levels 2 and 4, significant main effects of span (largest p < .001) and condition (largest p = .004) emerged, but no significant interactions were observed (smallest p = .068). At load level 3, significant main effects of span (F (1, 136) = 36.16, p < .001), condition (F (1, 136) = 148.82, p < .001), and a span x condition interaction (F (1, 136) = 6.10, p < .015) were observed. We did not expect this interaction to emerge only at load level 3, and we are therefore reticent to offer a strong interpretation of this finding. Together, these data suggest that participants with high WMC outperformed participants with low WMC in both conditions, and that performance was better in the choice offloading condition compared to the forced internal condition. Participants with low WMC did not benefit more from having the opportunity to offload compared to high WMC participants (save for at load level 3) perhaps due to the underutilization of offloading. We will return to this point in the Discussion.

Is continuous WMC related to the magnitude of the benefit afforded by offloading when offloading is (a) required (i.e., forced external) or (b) when offloading is a choice (i.e., choice offloading)?

When offloading was required, we computed the benefit scores as performance in the forced offloading condition minus performance in the forced internal condition. When offloading was a choice, we computed the benefit scores as performance in the choice offloading condition minus performance in the forced internal condition. For these analyses, WMC was conceptualized as a continuous variable (rather than comparison performance from only low vs. high WMC participants as reported above).

Bivariate Correlations

Below we report the results of our bivariate correlations[2] between the benefit of offloading and WMC at each memory load for each task separately. Correlation strength was then compared across tasks.

Forced Offloading. The benefit afforded by offloading (i.e., forced external performance minus forced internal performance) was correlated with WMC when considering this variable in a continuous fashion. In the intention offloading task, we observed significant correlations

---

[2] A hypothesis around the benefit of choice offloading compared to forced internal memory being more related to WMC than in past research was pre-registered; however, we did not outline a specific statistical test that corresponded to this hypothesis. Our analyses involving the benefit of forced external trials compared to forced internal trials as related to WMC are exploratory.

between offloading benefit and WMC at loads levels 2, 3, and 4 (load level 2: *r* (206) = -.32, adjusted *p* < .001; load level 3: *r* (206) = -.20, adjusted *p* = .006, load level 4: *r* (206) = -.27, adjusted *p* < .001), but this effect was not observed at load level 1 (*r* (206) = -.10, adjusted p = .169). In the letter string offloading task, we observed significant correlations between offloading benefit and WMC at all memory loads (load level 1: *r* (206) = .21, adjusted *p* = .005[3]; load level 2: *r* (206) = -.18, adjusted *p* = .014; load level 3: *r* (206) = -.34, adjusted *p* < .001; load level 4: *r* (206) = -.15, adjusted *p* = .033). Across tasks, correlations only differed statistically at the lowest load (i.e., load level 1; *p* = .009; see footnote 3). Correlations across tasks did not significantly differ at load levels 2, 3, or 4 (smallest *p* = .148) In general, these results suggests that those with lower WMC benefit more from cognitive offloading when offloading is required.

Choice Offloading. The benefit afforded by offloading (i.e., choice offloading performance minus forced internal performance) was correlated with WMC. In the intention offloading task, we observed significant correlations between offloading benefit and WMC at load levels 2, 3, and 4 (load level 1: r (206) = -.08, adjusted *p* = .240; load level 2: *r* (206) = -.29, adjusted *p* < .001; load level 3: *r* (206) = -.21, adjusted *p* = .004, load level 4: *r* (206) = -.22, adjusted p = .002). In the letter string offloading task, we only observed significant correlations between offloading benefit and WMC at load levels 3 and 4 (load level 1: *r* (206) = .08, adjusted *p* = .312; load level 2: *r* (206) = -.06, adjusted *p* = .369; load level 3: *r* (206) = -.23, adjusted *p* = .003; load level 4: r (206) = -.16, adjusted *p* = .044). Across tasks, the correlations at each memory load only differed statistically at load level 2 (*p* = .037). Correlations did not

---

[3] This surprising positive correlation is driven by a relatively small number of participants who did not achieve 100% accuracy on trials at load 2 in the internal memory condition. That is, if a participant achieved 100% accuracy at load 2 for the internal memory trials, their maximum possible benefit score would be 0 (if they also achieved 100% accuracy on offloading trials). Therefore, this finding seems to be driven by a relatively small number of participants who made one or more errors on load 2 trials in the internal memory condition for the letter string offloading task; because we do not think that this effect driven by a small number of incorrect trials exhibited by a small number of participants is a meaningful result, we do not further interpret this finding.

significantly differ at load levels 1, 3, or 4 (smallest *p* = .185) In general, these results suggests that those with lower WMC benefit more from cognitive offloading when offloading is a choice, particularly as memory load increased.

Simultaneous Regressions

Finally, we conducted additional exploratory analyses using simultaneous regressions to predict benefit scores in the choice offloading condition only (performance in the choice offloading block – performance in the internal memory block) from WMC and offloading frequency at each memory load. For the intention offloading task, the overall model was not significant at memory load level 1 (F (2, 205) = 0.72, R^2 = 0.01, adjusted R^2 < 0 .001; *p* = .490). At load levels 2, 3, and 4, the overall model was significant (load level 2: F (2, 205) = 16.47, R^2 = 0.14, adjusted R^2 =.13; *p* < 0.001; load level 3: F (2, 205) = 8.73, R^2 = 0.08, adjusted R^2 = .07; *p* < .001; load level 4: F (2, 205) = 7.39, R^2 = 0.07, adjusted R^2 = .06; *p* < .001) and both offloading frequency (all *p*s < .050) and WMC (all *p*s < .001) were significant predictors in these models. For the letter string offloading task, the overall model was not significant at memory load level 1 (F (2, 205) = 0.71, R^2 = 0.01, adjusted R^2 < .001; *p* = .494) or 2 (F (2, 205) = 2.23, R^2 = 0.02, adjusted R^2 = 0.01; *p* = .110). At loads level 3 and 4, the overall model was significant (load level 3: F (2, 205) = 16.64, R^2 = 0.14, adjusted R^2 =.13; *p* < 0.001; load level 4: F (2, 205) = 38.14, R^2 = 0.27, adjusted R^2 = .26; *p* < .001) and both offloading frequency (all *p*s < .001) and WMC (all *p*s < .001) were significant predictors in these models. In both tasks, as offloading frequency increased, so too did the benefit to performance conferred by offloading. Moreover, the benefit to performance afforded by offloading increased as WMC decreased. These findings suggest that participants who made the greatest use of the offloading strategy received the largest benefits to performance, and that participants with lower WMC benefitted from using offloading more than participants with high WMC.

Is metacognitive calibration related to offloading choice behavior?

We tested whether participants' underconfidence at each memory load was associated with offloading choice behavior (see Boldt & Gilbert, 2019; Gilbert et al., 2020; Sachdeva & Gilbert, 2020)[4]. Contrary to our hypothesis that greater underconfidence would be associated with offloading more often when participants had the opportunity to do so, we did not observe significant correlations between metacognitive calibration and offloading choice behavior at any loads for the intention offloading task (load level 1: r (206) = -0.05, adjusted $p = .503$; load level 2: r (206) = .03, adjusted $p = .844$; load level 3: r (206) = 0.07, adjusted $p = .844$; load level 4: r (206) = -.08, adjusted $p = .492$) or the letter string offloading task (load level 1: r (206) = .16, adjusted $p = .991$; load level 2: r (206) = 0.11, adjusted $p = .991$; load level 3: r (206) = 0.03, adjusted $p = .991$; load level 4: r (206) = 0.15, adjusted $p = .991$) . We return to this point in the Discussion.

Is WMC related to offloading choice behavior?

Finally, we tested whether offloading frequency (e.g., the proportion of trials on which participants chose to offload) was correlated with working memory capacity according to our preregistered analysis plan. In the intention offloading task, we did not observe significant correlations between offloading frequency and WMC at any load (load level 1: r (206) = -0.03, adjusted $p = .647$; load level 2: r (206) = -0.11, adjusted $p = .238$; load level 3: r (206) = 0.02, adjusted $p = .799$; load level 4: r (206) = 0.07, adjusted $p = .842$). In the letter string offloading task, we also did not observe any significant correlations between offloading frequency and

---

[4] While this hypothesis was pre-registered, we did not outline a specific statistical test that corresponded to this hypothesis.

working memory at any load (load level 1: r (206) < .001, adjusted p = .934; load level 2: r (206) = 0.02, adjusted p = .934; load level 3: r (206) = 0.09, adjusted p = .934; load level 4: r (206) = 0.10, adjusted p = .934). Unsurprisingly, the correlations at each memory load across tasks did not differ statistically (smallest p = .359). These results support an alternative hypothesis that we considered (outlined in our preregistration) that WMC would not be significantly associated with offloading choice behavior in either task.

Discussion

The current study attempted to address a discrepancy in the literature around the relationship between working memory ability and the use of and benefit afforded by cognitive offloading in the prospective and retrospective memory domains. The procedures for two commonly used offloading tasks – the intention offloading task and the letter string offloading task – were aligned with one another on several design dimensions while preserving the core key difference between the two tasks (i.e., memory domain). Overall, findings across tasks and offloading conditions were fairly consistent, though we note that stronger patterns were observed for forced-external compared to choice offloading analyses in some instances. First, many of the results of our omnibus four-way ANOVA for forced-external and choice offloading were qualitatively similar, though the four-way interaction reached significance in the forced-external analysis and did not for the choice offloading analysis. Further, participants with low and high WMC performed more similarly to one another under the forced-external offloading condition compared to relying on internal memory alone in both the intention offloading task and the letter string offloading task, particularly under high load. These findings supported our hypotheses, and the similarity in findings across tasks and offloading conditions is noteworthy. Results of these analyses suggest that offloading can be an effective strategy to 'close the gap' in performance that is typically exhibited for participants with low versus high WMC.

In addition, participants with low WMC tended to benefit more from offloading both when they were forced to offload and when they had the choice to offload, and the magnitude of this benefit was similar at high memory loads (load levels 3 and 4) across tasks. Moreover, the expected pattern did emerge for the benefit conferred by offloading in the choice block compared to the internal memory block. In both tasks, the frequency with which participants chose to offload was associated with larger benefits to performance. Critically, performance benefits from offloading were larger overall for individuals with low WMC compared to those with high WMC.

The finding that WMC was associated with offloading benefit in the choice offloading condition is consistent with the pattern reported by Ball and colleagues (2022) but stands in contrast to the findings reported by Morrison and Richmond (2020). Importantly, in the current study, participants had the opportunity to complete a forced external block, and block order was fixed such that the choice condition always came after both forced internal and forced external trials. Additionally, participants were given feedback on their performance after each trial, which was only done previously in the intention offloading task. It is possible that this prior task experience and/or feedback allowed participants to make use of offloading more optimally than the sample reported by Morrison and Richmond (2020). At the same time, however, in light of the nonsignificant correlations between WMC and offloading frequency, these findings suggest that participants with low WMC may not have used the strategy as frequently as they should have in order to obtain maximal performance benefit (see Scarampi & Gilbert, 2021 for similar results with older adults). This interpretation is bolstered by patterns observed in the letter string offloading task comparing performance in the choice offloading and forced internal blocks where significant span x condition interactions were not observed at most memory load levels (1, 2, and 4). Why low WMC participants may not make use of the offloading strategy as much as they should given their more limited internal memory capability is an interesting avenue to be explored in future studies. An important future study would be to directly examine whether the benefit of offloading for participants with varying levels of WMC differs depending on whether feedback is provided.

Surprisingly, neither metacognitive calibration nor WMC were found to be related to the frequency with which participants chose to offload, which diverges from past findings with the intention offloading task (Ball, Peper, et al., 2022; Boldt & Gilbert, 2019; Gilbert et al., 2020; Sachdeva & Gilbert, 2020). Moreover, the role of metacognition in offloading behavior in the letter string offloading task was tested here for the first time, so additional research is needed to replicate the patterns reported here.

With respect to why the expected relationship between WMC and metacognitive calibration was not observed in the current study, it is possible that our measure of metacognition was too coarse. To avoid inflating the time and demands of the task battery beyond what participants were already being required to do in the context of the current study, we opted to use a relatively simple approach similar to what has been used in prior research for testing metacognitive associations with offloading choice behavior (see for example Ball, Peper, et al., 2022; Boldt & Gilbert, 2019). We also opted to use this relatively simple approach to measure metacognition given the asynchronous and unsupervised nature of data collection for this study. Nonetheless, our approach may not have captured the relationship between metacognition and offloading behavior in the choice offloading block given that the choice block in the current study was always presented as the final block after participants had extensive task experience. Future studies designed to test questions around metacognitive contributions to offloading behavior may also choose to include metacognitive measures that could index item-level metacognitive judgements instead of or in addition to global-level measures of metacognition. For example, future studies may choose to sample metacognitive beliefs multiple times throughout the task, use a larger number of probes embedded within the task, and/or include retrospective confidence judgements or other such procedures.

The lack of association found between WMC and offloading choice behavior in the current study stands in contrast to what was observed by Ball, Peper, et al. (2022), but is consistent with the pattern observed by Morrison and Richmond (2020). Specifically, Ball, Peper, and colleagues (2022) found that low WMC participants were more likely to offload when given the choice. However, a critical difference in that study was that choosing to offload was associated with different point values (ranging from 2-8) while choosing to complete the task internally was always associated with a higher fixed value (i.e., 10 points). This allowed for the examination of whether working memory was associated with *optimal* offloading decisions based on one's own forced internal and external memory ability (e.g., if a participants can remember on average 5 of 10 targets on internally and 10 of 10 targets externally, then they should choose to offload when given a value of 6 points but not when given a value of 4 points). Working memory was not associated with the optimality of offloading, but low WMC participants did offload more overall (independent of value). Based on these findings, coupled with the fact that metacognitive calibration was not associated with working memory, it was suggested by Ball, Peper, and colleagues (2022) that low WMC participants may choose to offload to avoid effort. The results of the current study are not consistent with this interpretation, as working memory capacity was also not associated with offloading frequency. However, it is unclear whether the two procedures are directly comparable, as the current study did not include point values that might otherwise change decisions to offload. Future studies may explore whether the use of point values specifically encourages the expected patterns to emerge.

An alternative explanation as to why offloading choice behavior was not associated with WMC is referred to as "strategy perseveration." Scarampi and Gilbert (2020) had participants either use their own internal memory ability or offload in the first block, while in the second block participants were given the choice to offload. Results showed that participants chose to offload more frequently in the second block if they had previously been forced to offload than if they had to use their own memory, suggesting that choice decisions were made based on prior strategies. In the current study, participants completed forced internal and external trials prior to the choice trials, but external trials always occurred immediately prior. The finding that working memory was not associated with offloading frequency suggests that low WMC participants are not more likely to perseverate on an immediately prior strategy than high WMC participants.

Together, the results of the current study are inconsistent with the idea the prospective and retrospective memory task performance and offloading behavior relies on fundamentally different mechanisms. The dual component model of working memory (Unsworth & Engle, 2007) argues that working memory underlies the ability to flexibly control both attention and memory. Ball and colleagues (Ball, Wiemers, et al., 2022) have argued that the relation between working memory and prospective memory occurs because attention is needed to remember *that* something needs to be done (i.e., the prospective component) and memory processes are needed to remember *what* needs to be done (i.e., the retrospective component). In retrospective memory tasks, participants only need to remember *what* needs to be done. However, attentional processes are still needed at encoding to focus attention on to-be-remembered information (Unsworth, 2019) and are likely re-engaged at retrieval to reinstate the temporal context needed to retrieve information in serial order (Spillers & Unsworth, 2011). Thus, attention and memory processes that differ as a function of WMC are likely operating in both tasks, and offloading can reduce the demands placed on both processes.

Overall, our findings do not provide overwhelming support for any extant theories of cognitive offloading. Perhaps most surprisingly, our current findings did not support a metacognitive account for offloading behavior in either task context. This finding stands in contrast to prior work using variants of the intention offloading task reported here (Ball, Peper, et al., 2022; Boldt & Gilbert, 2019; Gilbert et al., 2020; Sachdeva & Gilbert, 2020; Scarampi & Gilbert, 2021). It is possible that the methodological changes made for the purposes of this study to better align with the letter string offloading task may have impacted the use of metacognition to guide offloading behavior in unexpected ways. Patterns from our comparisons involving forced external and choice offloading conditions were similar, though in some instances our findings were stronger in the forced external condition compared to the choice offloading condition. This may suggest some subtle support for the effort avoidance account in that participants may not have made best use of the opportunity to offload when they had the choice to do so rather than being forced to offload due to the effort associated with creating the external reminder. This was particularly true for low WMC participants who stood to benefit from the performance boost afforded by offloading but didn’t always choose to offload as much as they should have given their more limited internal memory capability. However, we note that manipulations that serve to increase the effort associated with making and/or using offloaded notes would be better suited to test this account (e.g., Grinschgl et al., 2021). Our findings do not seem to provide strong support for the saving enhanced memory account, though participants were not limited in the amount of information that they were allowed to offload in the choice offloading condition in the current study. Such a manipulation may encourage participants to offload to-be-remembered information more selectively than was observed here and may better demonstrate the potential for offloading to produce the saving enhanced memory effect (see for example Fellers & Storm, 2024).

Despite not finding clear support for any extant accounts of cognitive offloading, we believe that the overall consistency of our findings across tasks in the prospective and retrospective memory domains suggests that the theoretical basis for offloading benefits may be similar across tasks. Thus, our results support the development of domain-general theories of cognitive offloading. However, different theoretical accounts may be needed to explain the mechanism(s) by which cognitive offloading confers benefits to performance and the behavior exhibited by participants when they have the choice to engage in offloading, including why some sub groups (inlcuding participants with low WMC and older adults; see also Scarampi & Gilbert, 2021) do not make optimal use of the opportunity to offload. These are important future directions for the field to address.

Limitations

There are a few limitations of the current study that should be noted. First, data for the current study were collected online, and participants completed the tasks in an unsupervised manner. While our stringent exclusion criteria make us confident that we retained data only from participants who completed the tasks as instructed and exhibited performance within a reasonable range, it is possible that in-person studies would produce different patterns of results given the more stringent testing conditions that can be enforced in the lab. At the same time, the use of online data collection in this study increases the potential for these findings to translate to real-world behaviors outside of the lab. Although a recent meta-analysis suggests that patterns do not differ significantly based on data collection site (i.e., in person vs. online; see Burnett & Richmond, under review), future research could examine whether the patterns reported here replicate for lab-based data collection.

Another limitation is that our load manipulation was not directly comparable across tasks. That is, although both tasks contained four load levels, the intention offloading task contained memory loads of 1, 3, 5, and 7 while the letter string offloading task contained memory loads of 2, 4, 6, and 8. However, we note that in both tasks load increased by two items at each load level from the lowest memory load. The lowest memory load was not consistent across tasks because we wanted demands associated with each memory load to be roughly similar across tasks. That is, based on previous research we expected that remembering 1 intention and 2 letters would be relatively low demand for all participants, and that remembering 7 intentions and 8 letters would be a much higher demand. In examining the means across tasks, performance levels were generally comparable across the first three loads, although performance dropped more considerably at the highest load in the retrospective letter string offloading task compared to the prospective intention offloading task. Future research comparing prospective and retrospective memory offloading may wish to directly equate memory loads across tasks in order to test whether patterns reported here persist under those conditions.

Conclusion

The use of similar procedures for offloading tasks in the prospective and retrospective memory domains brought findings across these tasks into better alignment with one another. Participants had experience with both forced internal memory and forced external trials in each task before moving on to offloading choice trials, which may have helped participants make better choices about when and how to use cognitive offloading. However, low WMC participants did not seem to use the cognitive offloading strategy as much as they should have, suggesting that extended practice and/or explicit instruction around the use of cognitive offloading may be beneficial to these individuals. Finally, the findings here suggest that more research is needed around the association between metacognition and offloading behavior, particularly in the retrospective memory domain.

References

Ball, B. H., & Peper, P. (2022). *Cost avoidance underlies decisions to use prospective memory reminders* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/sqxme

Ball, B. H., Peper, P., Alakbarova, D., Brewer, G., & Gilbert, S. J. (2022). Individual differences in working memory capacity predict benefits to memory from intention offloading. *Memory*, *30*(2), 77–91. https://doi.org/10.1080/09658211.2021.1991380

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

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

Brown, M. (2021). *Enhancing short-term memory storage through cognitive offloading*. California State University, Sacramento.

Burnett, L. K., & Richmond, L. L. (2023). Just write it down: Similarity in the benefit from cognitive offloading in young and older adults. *Memory & Cognition*. https://doi.org/10.3758/s13421-023-01413-7

Burnett, L. K., & Richmond, L. L. (under review). *Cognitive offloading benefits performance and reduces interindividual variability: A meta-analysis*.

Chiu, G., & Gilbert, S. J. (2023). Influence of the physical effort of reminder-setting on strategic offloading of delayed intentions. *Quarterly Journal of Experimental Psychology (2006)*, 17470218231199977. https://doi.org/10.1177/17470218231199977

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 717–726. https://doi.org/10.1037/0278-7393.16.4.717

Engeler, N. C., & Gilbert, S. J. (2020). The effect of metacognitive training on confidence and strategic reminder setting. *PLOS ONE*, *15*(10), e0240858. https://doi.org/10.1371/journal.pone.0240858

Fellers, C., & Storm, B. C. (2024). The saving enhanced memory effect can be observed when only a subset of items are saved. *Memory & Cognition*. https://doi.org/10.3758/s13421-024-01545-4

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*(2), 226–236. https://doi.org/10.3758/s13421-014-0461-7

Gilbert, S. J. (2015a). Strategic offloading of delayed intentions into the external environment. *The Quarterly Journal of Experimental Psychology*, *68*(5), 971–992. https://doi.org/10.1080/17470218.2014.972963

Gilbert, S. J. (2015b). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. *Consciousness and Cognition*, *33*, 245–260. https://doi.org/10.1016/j.concog.2015.01.006

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. (2022). *Outsourcing memory to external tools: A review of 'intention offloading'* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/ahqtz

Grinschgl, S., Papenmeier, F., & Meyerhoff, H. S. (2021). Consequences of cognitive offloading: Boosting performance but diminishing memory. *Quarterly Journal of Experimental Psychology*, *74*(9), 1477–1496. https://doi.org/10.1177/17470218211008060

Guynn, M. J., Mcdaniel, M. A., & Einstein, G. O. (1998). Prospective memory: When reminders fail. *Memory & Cognition*, *26*(2), 287–298. https://doi.org/10.3758/BF03201140

Kelly, M. O., & Risko, E. F. (2022). Study effort and the memory cost of external store availability. *Cognition*, *228*, 105228. https://doi.org/10.1016/j.cognition.2022.105228

Landsiedel, J., & Gilbert, S. J. (2015). Creating external reminders for delayed intentions: Dissociable influence on "task-positive" and "task-negative" brain networks. *NeuroImage*, *104*, 231–240. https://doi.org/10.1016/j.neuroimage.2014.10.021

Meyerhoff, H. S., Grinschgl, S., Papenmeier, F., & Gilbert, S. J. (2021). Individual differences in cognitive offloading: A comparison of intention offloading, pattern copy, and short-term memory capacity. *Cognitive Research: Principles and Implications*, *6*(1), 34. https://doi.org/10.1186/s41235-021-00298-x

Morrison, A. B., & Richmond, L. L. (2020). Offloading items from memory: Individual differences in cognitive offloading in a short-term memory task. *Cognitive Research: Principles and Implications*, *5*(1). https://doi.org/10.1186/s41235-019-0201-4

Peper, P., & Ball, B. H. (2023). *Great expectations: Anticipating a reminder influences prospective memory encoding and unaided retrieval* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/hcn5a

Psychology Software Tools. (2020). *E-prime (E-prime Go)* [Computer software]. Psychology Software Tools Pittsburgh, PA.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164–171.

Richmond, L. L., Burnett, L. K., Morrison, A. B., & Ball, B. H. (2022). Performance on the processing portion of complex working memory span tasks is related to working memory capacity estimates. *Behavior Research Methods*, *54*(2), 780–794. https://doi.org/10.3758/s13428-021-01645-y

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

Robison, M. K., Celaya, X., Ball, B. H., & Brewer, G. A. (2023). Task sequencing does not systematically affect the factor structure of cognitive abilities. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-023-02369-0

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

Scarampi, C., & Gilbert, S. J. (2020). The effect of recent reminder setting on subsequent strategy and performance in a prospective memory task. *Memory*, *28*(5), 677–691. https://doi.org/10.1080/09658211.2020.1764974

Scarampi, C., & Gilbert, S. J. (2021). Age differences in strategic reminder setting and the compensatory role of metacognition. *Psychology and Aging*, *36*(2), 172–185. https://doi.org/10.1037/pag0000590

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 347–361. https://doi.org/10.1037/0278-7393.29.3.347

Spillers, G. J., & Unsworth, N. (2011). Variation in working memory capacity and temporal–contextual retrieval from episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(6), 1532–1539. https://doi.org/10.1037/a0024852

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

Tsai, P., Sachdeva, C., Gilbert, S. J., & Scarampi, C. (2023). An investigation of the saving□ enhanced memory effect: The role of test order and list saving. *Applied Cognitive Psychology*, *37*(4), 736–748. https://doi.org/10.1002/acp.4067

Unsworth, N. (2019). Individual differences in long-term memory. *Psychological Bulletin*, *145*(1), 79–139. https://doi.org/10.1037/bul0000176

Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038–1066. https://doi.org/10.1037/0033-2909.133.6.1038

Unsworth, N., Heitz, R., Schrock, J., & Engle, R. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505.

Table 1

Results of 2x4x2x2 ANOVA for offloading task performance when offloading is required

Effect	DF_b	DF_w	F	p-value	η^2_G
WMC	1.00	136.00	50.04	< .001	0.04
Task	1.00	136.00	138.70	< .001	0.09
Load Level	2.59	352.45	559.94	< .001	0.38
Condition	1.00	136.00	763.83	< .001	0.34
WMC*Task	1.00	136.00	4.08	.045	0.00
WMC*Load Level	2.59	352.45	9.78	< .001	0.01
WMC*Condition	1.00	136.00	23.72	< .001	0.02
Task*Load Level	2.69	365.26	144.70	< .001	0.14
Task*Condition	1.00	136.00	46.17	< .001	0.03
Load Level*Condition	2.52	342.47	402.16	< .001	0.31
WMC*Task*Load Level	2.69	365.26	3.48	.020	0.00
WMC*Task*Condition	1.00	136.00	0.18	.668	0.00
WMC*Load Level*Condition	2.52	342.47	8.04	< .001	0.01
Task*Load Level*Condition	2.59	352.74	88.85	< .001	0.08
WMC*Task*Load Level*Condition	2.59	352.74	2.81	.047	0.00

Note. Greenhouse-Geisser correction applied to degrees of freedom as appropriate.

Table 2

Results of 2x4x2x2 ANOVA for offloading task performance when offloading is a choice

Effect	DF_b	DF_w	F	***p*-value**	η^2_G
WMC	1.00	136.00	54.13	< .001	0.05
Task	1.00	136.00	131.76	< .001	0.08
Load Level	2.66	361.14	555.02	< .001	0.37
Condition	1.00	136.00	601.94	< .001	0.32
WMC*Task	1.00	136.00	3.98	.048	0.00
WMC*Load Level	2.66	361.14	14.54	< .001	0.01
WMC*Condition	1.00	136.00	16.16	< .001	0.01
Task*Load Level	2.63	357.20	147.69	< .001	0.13
Task*Condition	1.00	136.00	49.84	< .001	0.03
Load Level*Condition	2.66	362.29	404.72	< .001	0.30
WMC*Task*Load Level	2.63	357.20	5.74	.001	0.01
WMC*Task*Condition	1.00	136.00	0.10	.753	0.00
WMC*Load Level*Condition	2.66	362.29	4.40	.007	0.00
Task*Load Level*Condition	2.65	359.89	86.73	< .001	0.08
WMC*Task*Load Level*Condition	2.65	359.89	1.17	.318	0.00

Note. Greenhouse-Geisser correction applied to degrees of freedom as appropriate.

Figure 1

Task Schematics

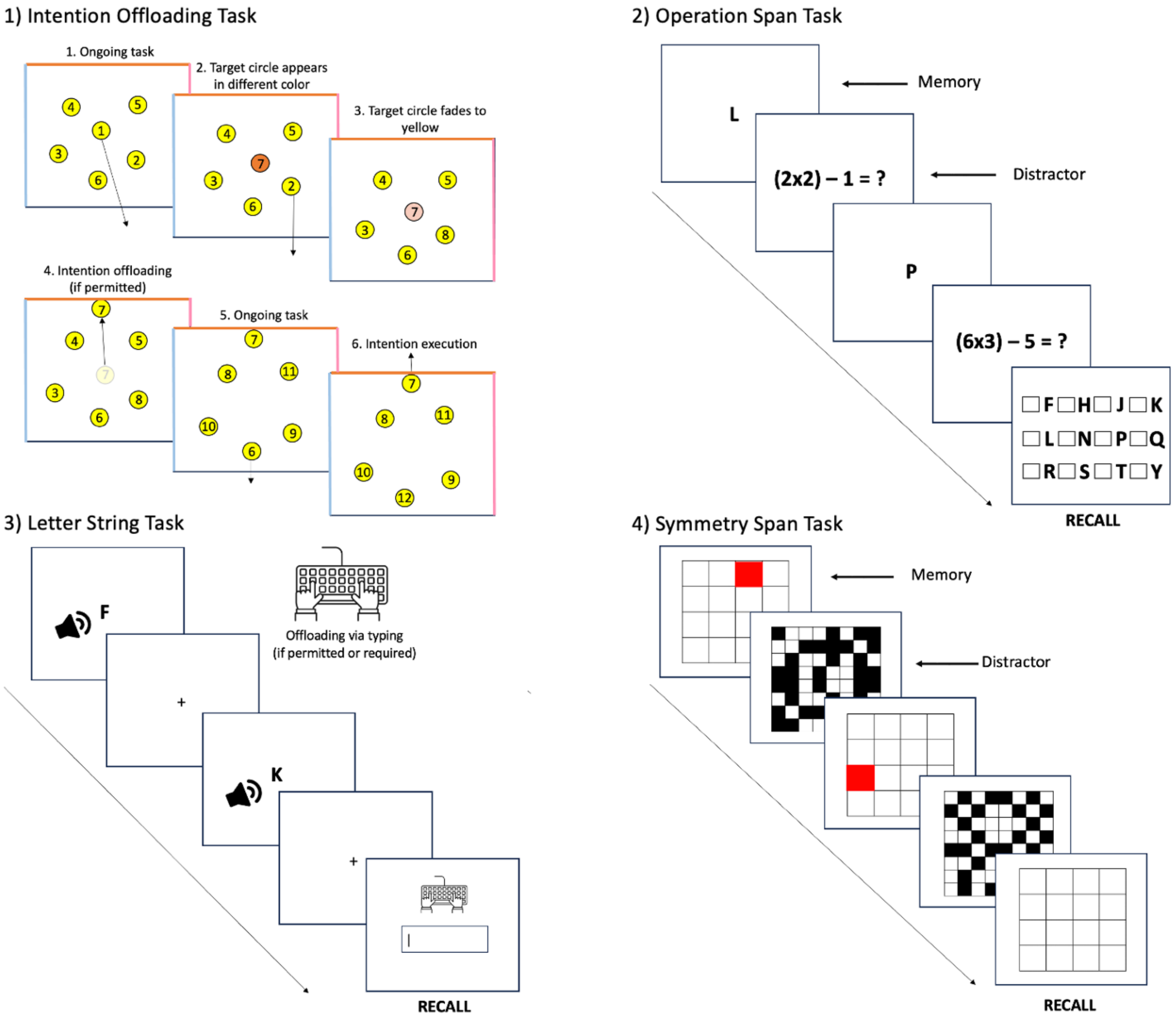
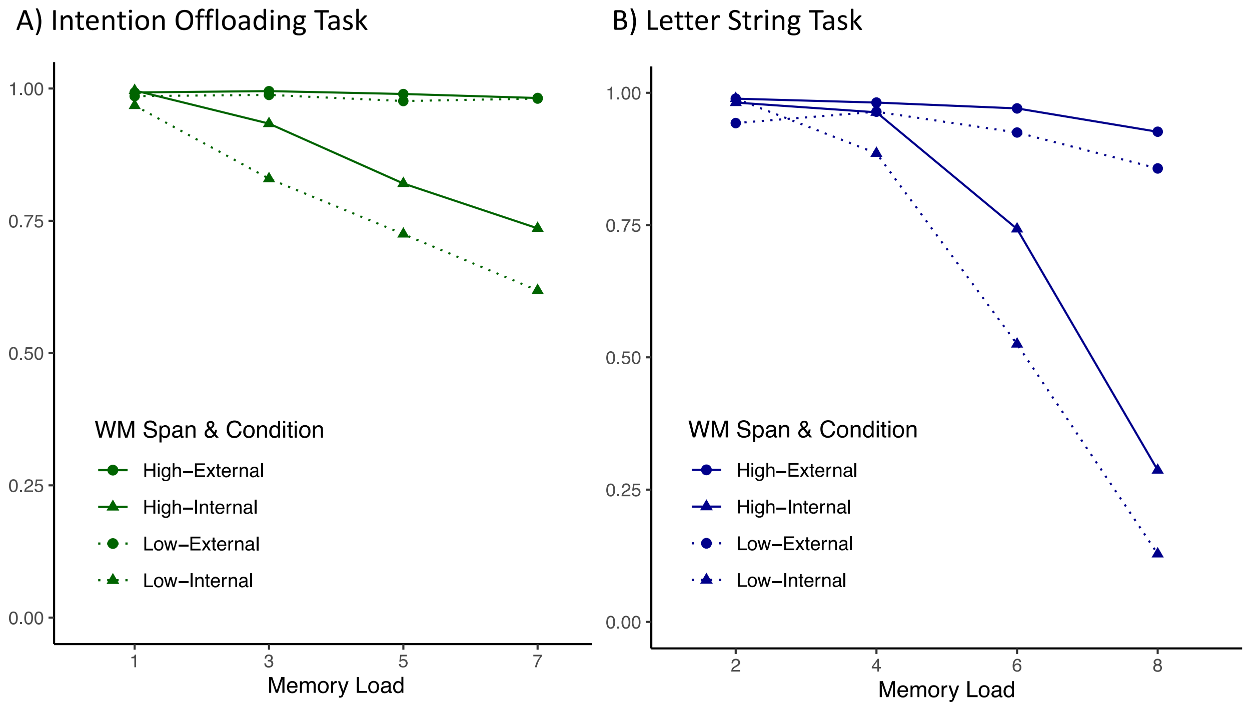


Figure 2

Forced Internal and Forced External block performance for low vs. high WMC participants



Individual Differences in Prospective and Retrospective Memory Offloading

Lauren L. Richmond[1], Lois K. Burnett[1], Julia Kearley[1,2], Samuel J. Gilbert[3], Alexandra B. Morrison[4], B. Hunter Ball[5]

[1]Department of Psychology, Stony Brook University
[2]Department of Psychology, McGill University
[3]Institute of Cognitive Neuroscience, University College London
[4]Department of Psychology, California State University, Sacramento
[5]Department of Psychology, University of Texas at Arlington

Author Note

Lauren L. Richmond https://orcid.org/0000-0001-6793-5773

Lois K. Burnett https://orcid.org/0000-0002-1679-4621

Julia Kearley https://orcid.org/0000-0002-6085-314X

Samuel J. Gilbert https://orcid.org/0000-0002-3839-7045

These data were presented at the 2023 Symposium for Individual Differences in Cognition in San Francisco, CA. Preregistration, data, and analysis scripts are available on OSF [This view-only link will be replaced with a publicly available link upon acceptance]. Correspondence concerning this article should be addressed to Lauren L. Richmond, Department of Psychology, Stony Brook University; Psychology B Building, Stony Brook, NY 11794-2500; email: lauren.richmond@stonybrook.edu