Journal of Memory and Language 142 (2025) 104617

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Individual differences in prospective and retrospective memory offloading☆

Lauren L. Richmond[a,*], Lois K. Burnett[a], Julia Kearley[a,b], Sam J. Gilbert[c], Alexandra B. Morrison[d], B. Hunter Ball[e]

[a] Department of Psychology, Stony Brook University, Psychology B Building, Stony Brook, NY 11794-2500, USA
[b] Department of Psychology, McGill University, 2001 McGill College Ave., Montreal, Quebec H3A 1G1, Canada
[c] Institute of Cognitive Neuroscience, University College London, Alexandra House, 17-19 Queen Square, London WC1N 3AZ, UK
[d] Department of Psychology, California State University, Sacramento, 6000 J Street, Sacramento, CA 95819, USA
[e] Department of Psychology, University of Texas at Arlington, 701 S. Nedderman Drive, Arlington, TX 76019, USA

ARTICLE INFO

Keywords:
Working memory capacity
Prospective memory
Retrospective memory
Cognitive offloading

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 cognitive offloading 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 varying levels of 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, particularly under high memory load, 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 lower WMC estimates were associated with increased offloading frequency in either memory domain when participants were permitted to offload. Practical and theoretical implications of these findings are discussed.

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., 2023), akin to instructions for standard memory tasks administered in laboratory settings. 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 (e.g., poor working memory capacity [WMC]) 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

☆ This article is part of a special issue entitled: 'Individual differences in memory' published in Journal of Memory and Language.

* Corresponding author.
E-mail address: lauren.richmond@stonybrook.edu (L.L. Richmond).

https://doi.org/10.1016/j.jml.2025.104617
Received 2 April 2024; Received in revised form 11 October 2024; Accepted 8 January 2025
Available online 23 January 2025
0749-596X/© 2025 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

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 Fig. 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 other circles 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 special 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 Fig. 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 in this task has shown consistent benefit from having the opportunity to offload (Brown, 2021; Burnett & Richmond, 2023; Morrison & Richmond, 2020; Risko & Dunn, 2015).

Despite the consistency of findings across the literature around the benefits of cognitive offloading (Burnett & Richmond, under review; Gilbert et al., 2023), 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 information from long-term memory that has been displaced from focal awareness (Unsworth & Engle, 2007). These same attention and

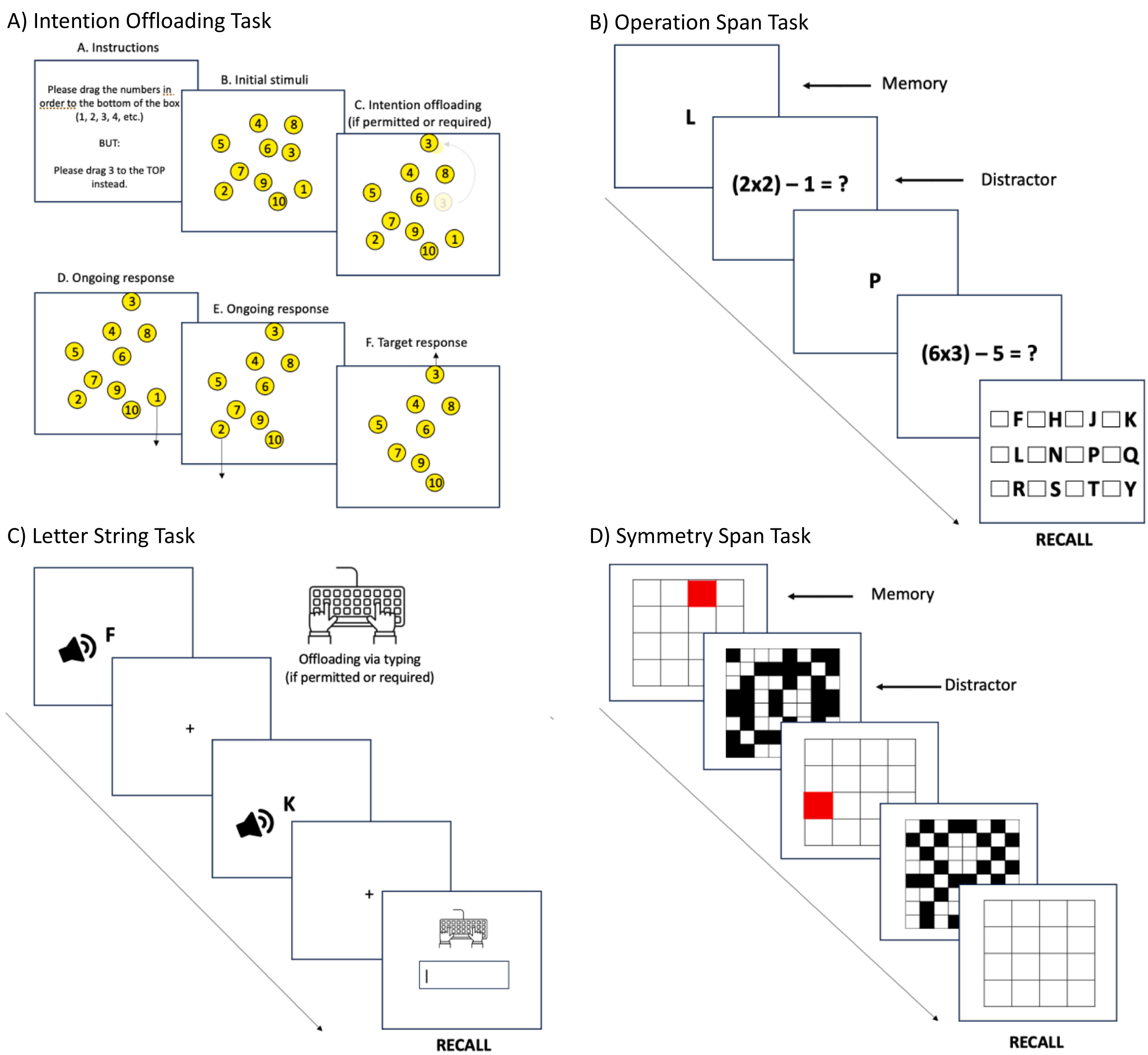

Fig. 1. Task Schematics.

memory processes are required to successfully complete both retrospective and prospective memory tasks. Therefore, WMC may well be related to the performance of these two different types of memory tasks when relying on internal memory alone. At the same time, offloading reduces the extent to which participants need to engage effortful memory maintenance processes by increasing the amount of available environmental support. In this context, a reasonable hypothesis is that offloading would be more beneficial for participants with lower WMC compared to those with higher WMC. However, differences among participants with varying levels of WMC around a host of other cognitive and non-cognitive factors, including for example metacognitive insights into their own memory capability, motivation, and interest in avoiding effort, could lead to patterns whereby participants with low WMC do not show the expected larger benefit of offloading than participants with higher WMC.

To test the relationship between WMC and memory task performance 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 to offload on each trial (optional reminders). Participants with lower 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, participants with low WMC 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 (i.e., 2–10 letters). As in the Ball, Peper, et al. (2022) study, participants with lower 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), participants with lower and higher WMC 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.

Interestingly, there are two key distinctions between the prospective intention offloading task and the retrospective letter string offloading task that may be relevant to consider in the context of these conflicting patterns. The first is that the intention offloading task necessarily requires participants to complete the ongoing task (dragging most numbered circles to the usual place) and the special/target circle task in which specific circles must be moved to specific locations that differ from those for the ongoing task circles in the same task context. That is, the intention offloading task is considered dual-task in nature (i.e., ongoing task + PM task). By contrast, in the letter string offloading task, participants need only to attend to and report back the presented letters. Second, the intention offloading task requires that participants themselves notice the target circles and self-initiate the target circle task component on their own. Again, this differs from the letter string offloading task in which participants are prompted to provide their responses at the close of each letter string presentation. There are several other possible theoretically-grounded accounts to consider that may underlie the discrepant results from 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 not confident in their 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, 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 (Chiu & Gilbert, 2023; Gilbert, 2024). 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.

Moreover, 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 an early 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 s 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 to 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.

The 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 therefore to address these discrepant findings by reducing the methodological differences that existed in past research (Ball, Peper, et al., 2022; Morrison & Richmond, 2020) while maintaining the key distinction from past studies: whether the offloading task was retrospective or prospective in nature. That is, the current study was designed to allow for better understanding of whether the benefits conferred by offloading are domain-general or if these effects are best considered to be more domain-specific. Moreover, examination of associations with WMC under both forced external and choice offloading conditions can help to address the potential for non-cognitive differences (e.g., motivation, effort avoidance) among participants with varying levels of WMC that could well impact choice offloading behavior but would be less influential in a forced offloading scenario.

Specifically, 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. Second, we hypothesized that the benefit conferred by offloading would be associated with WMC, and 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. Third, 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. Finally, we anticipated that lower WMC participants would offload more often than higher WMC participants when memory load was high. By the time that participants in the current study completed the trials in the choice block, they had already gained experience with both carrying out the task using internal memory alone (forced internal) and with completing the task when offloading was required (forced external). This experience 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. 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.

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 112 participants due to failure to meet minimum requirements for inclusion, leaving us with a final useable sample size of 251 participants.[1] Specifically, data were excluded for the following reasons: failing more than 50 % of the attention checks (n = 75), exhibiting lower performance in the forced external condition compared to the internal memory condition on one or both of the offloading tasks (n = 25), failing to offload on an adequate number of trials in the forced-external condition (n = 9), responding accurately on less than 50 % of trials at set size 2 in the choice block of the letter string offloading task (n = 1), and for failing to score above the guess rate for the processing portion of one of our complex span working memory tasks (n = 2; see Richmond et al., 2022). Demographic data were missing for two

[1] Our preregistration specified that we would also exclude participants who reported cheating on the task battery. The question read as follows: "For the integrity of the study, please answer the following truthfully. Note that your credit will not be affected by your answer. During the experiment, did you write anything down outside of the experimental program (i.e., on a piece of paper or an electronic note on your mobile device)?" We later considered the possibility that participants may have responded affirmatively to this question for writing down task instructions or completion codes and not task stimuli. Due to the unanticipated ambiguity in this question, we opted to include data from participants who responded affirmatively to this question but who otherwise met all inclusion criteria in our final sample.

participants, therefore sample characteristics reported below represent demographic data for 249 participants. Our final sample had a mean age of 20.93 years (SD = 6.11 years, Range: 17–57). With respect to race and ethnicity, 32.70 % of our sample identified as White, 28.10 % as Asian/Pacific Islander, 22.90 % as Hispanic/Latinx, 11.20 % as Black, 3.61 % of participants chose the response 'Other/Unknown,' and 2.01 % 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 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 a significant bivariate correlation between internal memory performance and offloading behavior as small as −.16 with 80 % power. This correlation is in line with the effect that we originally powered the study to detect.

Materials

Intention offloading task

The materials and stimuli for the intention offloading task were adapted from Gilbert et al. (2020; see Fig. 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, representing the intended future action to be associated with that circle. This key manipulation serves to differentiate the intention offloading task from a standard short-term/working memory task. 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. There was no filler task between the appearance of a target circle and retrieval, other than completion of the ongoing task (dragging the circles numbered 2–6). In addition, the retention interval (i.e., the time from the appearance of a target circle until the target action was to be carried out) varied as a function of how quickly a given participant was able to complete each circle-dragging task. 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, 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 many 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 Fig. 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 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 Fig. 1 panel B for a task schematic. Symmetry Span involves making symmetry decisions regarding 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 Fig. 1 panel D for a task schematic.

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 that was programmed using Google Web Toolkit (https://gwtproject.org) and self-hosted by the researchers; all other tasks were completed via E-Prime GO (Psychology Software Tools, 2020). After participants completed all four tasks, they provided their demographic information, completed a post-study questionnaire, and were debriefed. The study took approximately 2 h in total to complete, and participants were awarded partial course credit for their participation. Participants were permitted 24 h to complete all study procedures. See Fig. 1 for task schematics.

Statistical approach

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.

Results

Results from our key research questions and corresponding analyses are presented below. False discovery rate corrections for multiple comparisons were applied for correlations conducted at each memory load level.

Descriptive statistics for all measures are presented in Table 1. Reliability for a few metrics in the intention offloading task and the letter string offloading task were below the acceptable range. Specifically, Cronbach's alphas for both the letter string offloading task and the intention offloading task were below the acceptable range (i.e., .70 or higher; Hair et al., 2010) in the forced internal and external blocks. Acceptable ranges for skew and kurtosis were based on recommendations from Kline (2011). In addition, skew was outside of the acceptable range (<= 3) for choice offloading blocks in both tasks and in the forced external block for the intention offloading task. Similarly, kurtosis was found to be outside of the acceptable range (<= 8) for the forced external and choice blocks in both offloading tasks. For analyses involving these measures, we conducted non-parametric rank-based

Table 1
Descriptive Statistics.

Measure	Mean	SD	Skew	Kurtosis	Reliability
LST – Forced Internal	0.70	0.37	−0.85	−0.79	0.46
LST – Forced External	0.94	0.14	−2.89	11.02	0.40
LST – Choice	0.94	0.15	−3.29	13.25	0.73
LST – Offloading Behavior	0.77	0.36	−1.23	−0.13	0.74
LST – Metacognitive Confidence	71.07	28.27	−0.76	−0.53	0.77
IO – Forced Internal	0.83	0.19	−0.98	0.03	0.62
IO – Forced External	0.98	0.05	−5.14	34.18	0.37
IO – Choice	0.97	0.07	−5.00	40.19	0.63
IO – Offloading Behavior	0.66	0.39	−0.77	−1.03	0.77
IO – Metacognitive Confidence	67.91	30.46	−0.66	−0.77	0.78
WMC Composite	0.46	0.10	−1.29	2.19	0.78

Note. The abbreviation "LST" refers to the retrospective letter string offloading task and the abbreviation "IO" refers to the prospective intention offloading task. Descriptives are collapsed across memory load for the offloading tasks and across the OSpan and SymSpan tasks for the WMC Composite. Reliability reflects Cronbach's alpha for all measures where there were multiple assessments of performance.

ANCOVAs (Quade, 1967) in addition to parametric statistical tests; we note below the instances where results differed across these analyses as footnotes. Full rank-based ANCOVA results can be found on OSF.

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?

Forced offloading

To test whether individuals with varying levels of WMC would perform more similarly to one another on the offloading tasks when offloading is required compared to when participants were required to rely on internal memory, we conducted a 2 (task: intention offloading vs. letter string) x 2 (load: high [levels 3 and 4] vs. low [levels 1 and 2]) x 2 (condition: forced internal vs. forced external) x WMC (continuous) ANCOVA.[2] Effects are reported in Table 2 for brevity. Of specific interest is the significant four-way interaction ($p = .001$), 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, condition, task, and WMC. See Fig. 2 panel A. Follow-up analyses were conducted separately for each task.

Prospective Memory: Intention offloading task. Our planned follow up 2 (load: high [levels 3 and 4] vs. low [levels 1 and 2]) x 2 (condition: forced internal vs. forced external) x WMC (continuous) ANCOVA revealed a significant main effect of load ($F(1, 249) = 533.76$, $p < .001$), condition ($F(1, 249) = 585.99$, $p < .001$), and WMC ($F(1, 249) = 25.90$, $p < .001$). These main effects were qualified by significant load x condition ($F(1, 249) = 404.85$, $p < .001$), load x WMC ($F(1, 249) = 4.82$, $p = .029$), and condition x WMC ($F(1, 249) = 30.31$, $p < .001$) interactions. The three-way load x condition x WMC interaction ($F(1, 249) = 1.82$, $p = .179$) was not significant.

Despite not observing a significant three-way interaction, we followed up the significant condition x WMC interaction by examining the effects at high and low memory load separately in order to present consistent analyses across tasks. Under low memory load, there was a significant main effect of WMC ($F(1, 249) = 17.86$, $p < .001$), a

[2] Our analysis plan on OSF did not clearly state how WMC would be conceptualized for the current analysis. In addition, for brevity, all analyses reported below to test this question separates load into high (combining performance at load levels 3 and 4) vs. low (combining performance at load levels 1 and 2) categories rather than examining all four load levels, as visual inspection of the data suggested similar patterns of performance for high and low memory load trials.

Table 2
Results of 2x2x2xWMC ANCOVA for forced internal and forced external offloading task performance.

Effect	DF_B	DF_W	F	p-value	η_G^2
WMC	1	249	63.72	<.001	0.05
Task	1	249	240.34	<.001	0.12
Load Category	1	249	1509.70	<.001	0.40
Condition	1	249	1270.13	<.001	0.42
WMC*Task	1	249	8.71	.003	0.00
WMC*Load Category	1	249	19.90	<.001	0.01
WMC*Condition	1	249	45.09	<.001	0.03
Task*Load Category	1	249	385.13	<.001	0.13
Task*Condition	1	249	80.96	<.001	0.04
Load Category*Condition	1	249	1213.56	<.001	0.33
WMC*Task*Load Category	1	249	7.22	.008	0.00
WMC*Task*Condition	1	249	0.23	.629	0.00
WMC*Load Category*Condition	1	249	19.63	<.001	0.01
Task*Load Category*Condition	1	249	236.74	<.001	0.08
WMC*Task*Load Category*Condition	1	249	10.62	.001	0.00

Note. WMC was included as a z-score in the above analysis.

significant main effect of condition (F (1, 249) = 98.06, $p < .001$) and a significant WMC x condition interaction (F (1, 249) = 24.05, $p < .001$). A similar pattern emerged under high memory load (main effect of WMC: F (1, 249) = 21.77, $p < .001$; main effect of condition: F (1, 249) = 674.95, $p < .001$; WMC x condition interaction: F (1, 249) = 18.54, $p < .001$). Overall, these data suggest that participants with lower WMC benefitted more from being forced to offload than participants with higher WMC in the intention offloading task under both low and high memory load. Further, these results support our hypothesis that participants with varying levels of WMC would perform more similarly to one another in the forced external condition compared to the forced internal condition. See Fig. 2 panel A (left panel).

Retrospective Memory: Letter string offloading task. Our planned follow up 2 (load: high [levels 3 and 4] vs. low [levels 1 and 2]) x 2 (condition: forced internal vs. forced external) x WMC (continuous) ANCOVA revealed a significant main effect of load (F (1, 249) = 1103.91, $p < .001$), condition (F (1, 249) = 835.89, $p < .001$), and WMC (F (1, 249) = 50.55, $p < .001$). These effects were qualified by significant load x condition (F (1, 249) = 885.56, $p < .001$), load x WMC (F (1, 249) = 16.40, $p < .001$) and condition x WMC (F (1, 249) = 22.41, $p < .001$) interactions. These effects were further qualified by a significant three-way load x condition x WMC interaction (F (1, 249) = 20.47, $p < .001$). We followed up this significant three-way interaction by examining the interaction between condition and span separately at low and high memory loads. Under low load, only a significant main effect of WMC was revealed (F (1, 249) = 32.01, $p < .001$). The main effect of condition (F (1, 249) = 3.10, $p = .080$) and the condition x WMC interaction (F (1, 249) = 0.59, $p = .441$) were not significant. Under low memory load, participants with higher WMC outperformed participants with lower WMC overall. Under high memory load, though, a different pattern emerged. Here, a significant main effect of condition (F (1, 249) = 976.57, $p < .001$) and WMC (F (1, 249) = 36.94, $p < .001$) emerged. This effect was qualified by a significant condition x WMC interaction (F (1, 249) = 24.38, $p < .001$). These patterns indicate that participants with lower WMC benefitted more from being forced to offload than participants with higher WMC in the letter string offloading task, but only under high memory load. Moreover, these data again support our hypothesis that participants with varying levels of WMC would perform more similarly to one another in the forced external condition compared to the forced internal condition. See Fig. 2 panel A (right panel).

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 focused on the comparison between choice offloading and forced internal conditions. As above, we conducted a 2 (task: intention offloading vs. letter string) x 2 (load: high [levels 3 and 4] vs. low [levels 1 and 2]) x 2 (condition: forced internal vs. choice offloading) x WMC (continuous) ANCOVA. Effects are reported in Table 3 for brevity. Of note, the four-way interaction again reached significance ($p = .002$), and all other patterns observed here were qualitatively similar to what was observed for our four-way ANCOVA involving the forced external condition (see Tables 2 and 3).[3] Therefore, we again conducted follow-up analyses for each task separately.

Prospective Memory: Intention offloading task. Our follow up 2 (load: high vs. low) x 2 (condition: forced internal vs. choice offloading) x WMC (continuous) ANCOVA revealed a significant main effect of load (F (1, 249) = 433.31, $p < .001$), a significant main effect of condition (F (1, 249) = 438.94, $p < .001$), and a significant main effect of WMC (F (1, 249) = 37.18, $p < .001$). These main effects were qualified by significant load x condition (F (1, 249) = 369.75, $p < .001$), load x WMC (F (1, 249) = 7.41, $p = .007$), and condition x WMC (F (1, 249) = 13.03, $p < .001$) interactions. The three-way load x condition x WMC interaction was not significant (F (1, 249) = 0.22, $p = .643$). Nevertheless, to maintain consistency with the forced offloading analyses reported above, we followed up this analysis by examining the interaction between condition and WMC separately under low and high memory load. Consistent with the forced offloading analyses, significant main effects of condition (low: F (1, 249) = 67.62, $p < .001$; high: F (1, 249) = 551.63, $p < .001$) and WMC (low: F (1, 249) = 26.36, $p < .001$; high: F (1, 249) = 29.38, $p < .001$) emerged; these were qualified by significant condition x WMC interactions (low: F (1, 249) = 12.96, $p < .001$; high: F (1, 249) = 7.22, $p = .008$) under both low and high memory load. Participants with lower WMC benefitted more from having the choice to offload than participants with higher WMC in the intention offloading task regardless of memory load. Again, participants with varying levels of WMC performed more similarly to one another in the choice offloading condition compared to the forced internal condition in accordance with our predictions. See Fig. 2 panel B (left panel).

Retrospective Memory: Letter string offloading task. Our follow up 2 (load: high vs. low) x 2 (condition: forced internal vs. choice offloading) x WMC (continuous) ANCOVA revealed significant main effects of load (F (1, 249) = 1283.47, $p < .001$), condition (F (1, 249) = 562.58, $p < .001$) and WMC (F (1, 249) = 64.23, $p < .001$). These effects were qualified by significant load x condition (F (1, 249) = 769.67, $p < .001$), load x WMC (F (1, 249) = 23.21, $p < .001$) and condition x WMC (F (1, 249) = 6.59, $p = .001$) interactions. These effects were further qualified by a significant three-way load x condition x WMC interaction (F (1, 249) = 14.40, $p < .001$). We followed up this significant three-way interaction by examining the interaction between condition and WMC under low and high memory load separately. Under low load, only the main effect of WMC was significant (F (1, 249) = 34.81, $p < .001$; main effect of condition: F (1, 249) = 1.15, $p = .284$[4]; condition x WMC interaction: F (1, 249) = 0.70, $p = .405$) Under high memory load, significant main effects of both condition (F (1, 249) = 752.82, $p < .001$ and WMC (F (1, 249) = 52.90, $p < .001$) emerged, as well as a significant condition x WMC interaction (F (1, 249) = 11.26, $p < .001$). Participants with lower WMC benefitted more from having the choice to offload than participants with higher WMC in the letter string offloading task only under high memory load. In addition, participants with varying levels of WMC performed more similarly to one another in the choice offloading condition compared to the forced internal condition under high memory load, somewhat in line with our predictions. See Fig. 2 panel B (right panel).

[3] In our non-parametric analysis, the task x condition interaction was not significant (p =.364).

[4] In our non-parametric analysis, the main effect of condition under low memory load was significant (p =.032).

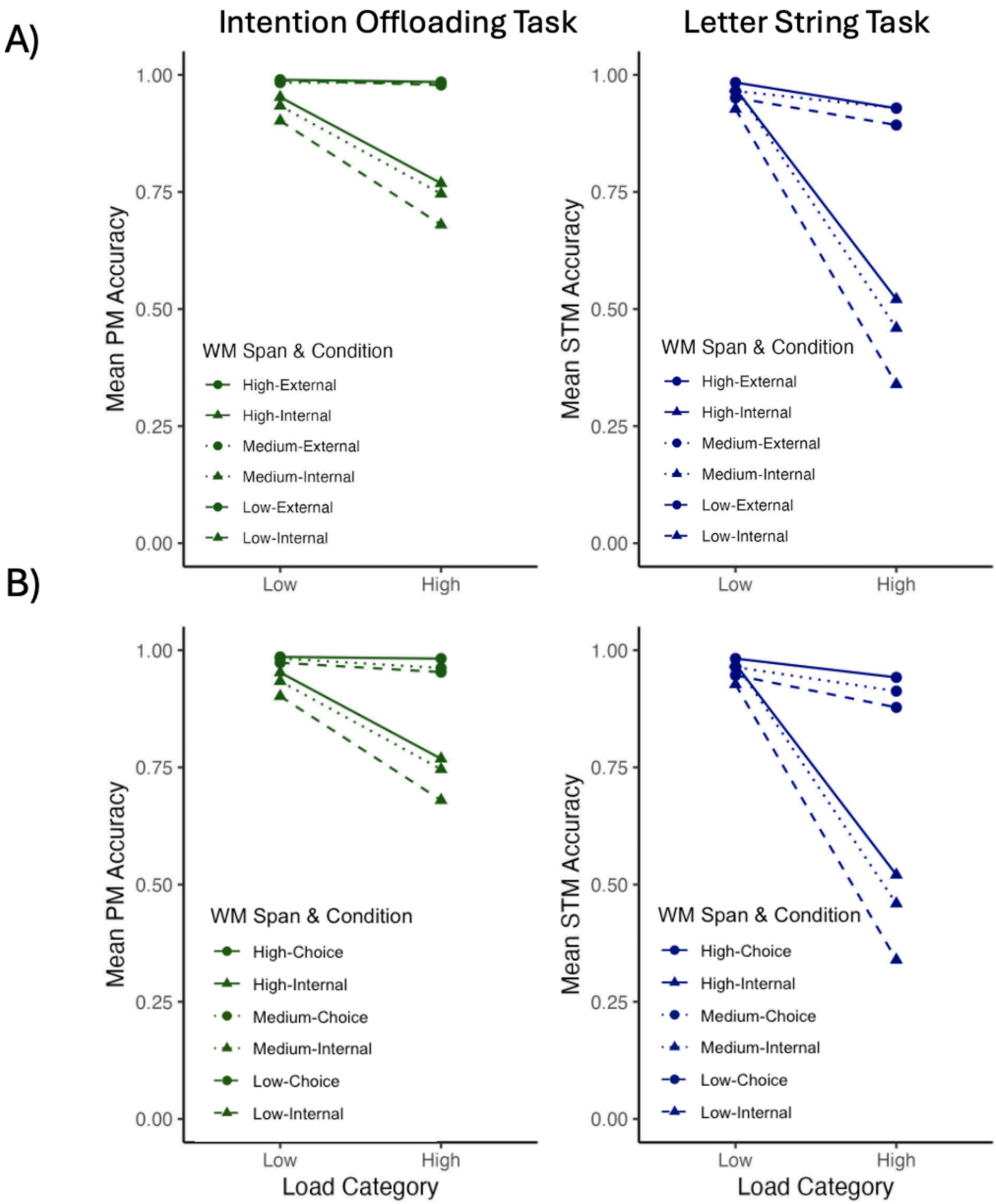

Fig. 2. ***Task performance across all blocks as a function of WMC*** *Note.* Panel A depicts performance in the forced internal and forced external blocks. Panel B depicts performance in the forced internal and choice offloading blocks. A tertiary split was applied to WMC as shown here solely for the purpose of visualization. In all analyses, WMC was treated as a continuous variable.

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)?

In this series of exploratory analyses, we aimed to test whether WMC was associated with the magnitude of the benefit afforded by offloading. This analysis was 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.

Table 3
Results of 2x2x2xWMC ANCOVA for forced internal and choice offloading task performance.

Effect	DF_B	DF_W	F	*p*-value	η^2_G
WMC	1	249	77.90	<.001	0.07
Task	1	249	225.12	<.001	0.10
Load Category	1	249	1553.82	<.001	0.38
Condition	1	249	869.08	<.001	0.36
WMC*Task	1	249	10.84	.001	0.00
WMC*Load Category	1	249	27.68	<.001	0.01
WMC*Condition	1	249	15.19	<.001	0.01
Task*Load Category	1	249	398.59	<.001	0.11
Task*Condition	1	249	72.61	<.001	0.04
Load Category*Condition	1	249	1085.70	<.001	0.29
WMC*Task*Load Category	1	249	7.45	.007	0.00
WMC*Task*Condition	1	249	0.01	.908	0.00
WMC*Load Category*Condition	1	249	11.94	.001	0.00
Task*Load Category*Condition	1	249	221.76	<.001	0.07
WMC*Task*Load Category*Condition	1	249	9.81	.002	0.00

Note. WMC was included as a z-score in the above analysis.

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.

Bivariate correlations

Below we report the results of our bivariate correlations[5] between the benefit of offloading and WMC at low (levels 1 and 2) and high (levels 3 and 4) memory load for each task separately. Correlation strength was then compared across tasks.

Forced Offloading. In general, the benefit afforded by offloading (i.e., forced external performance minus forced internal performance) under high memory load was negatively correlated with WMC. In the intention offloading task, we observed significant correlations between offloading benefit and WMC at both high (r (249) = -.26, adjusted $p <$ .001) and low memory loads (r (249) = -.30, adjusted $p < .001$). In the letter string offloading task, we observed significant correlations between offloading benefit and WMC at high (r (249) = -.30, adjusted $p <$ .001), but not low (r (249) = -.05, adjusted p = .441) memory load. Correlations across tasks differed significantly at low (adjusted p = .007), but not high (adjusted p = .640), memory load. In general, these results suggests that those with lower WMC benefit more from cognitive offloading when offloading is required, particularly under high memory load.

Choice Offloading. The benefit afforded by offloading (i.e., choice offloading performance minus forced internal performance) was negatively correlated with WMC under high memory load in both tasks. Specifically, in the intention offloading task, we observed significant correlations between offloading benefit and WMC at both high (r (249) = -.17, adjusted p = .008) and low (r (249) = -.22, adjusted p = .001) memory loads. In the letter string offloading task, we only observed significant correlations between offloading benefit and WMC under high load (r (249) = -.21, adjusted p = .002) but not low load (r (249) = .05, adjusted p = .405). Across tasks, the correlations did not differ under high load (adjusted p = .618), but task differences were observed at low load (adjusted p = .002). 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 under high and low memory load. In the intention offloading task, the overall model was significant for both high and low memory load (high: F (2, 248) = 16.17, R^2 = 0.12, adjusted R^2 = 0.11; $p < .001$; low: F (2, 248) = 13.58, R^2 = 0.10, adjusted R^2 = 0.09; $p < .001$). Offloading frequency (all *p*s < .001) and WMC (all *p*s < .001) were significant predictors in both the high and low load models. In the letter string offloading task, the overall model was significant for both high and low memory load (high: F (2, 248) = 45.02, R^2 = 0.27, adjusted R^2 = 0.26; $p < .001$; low: F (2, 248) = 3.75, R^2 = 0.03, adjusted R^2 = 0.02; p = .020). Offloading frequency was a significant predictor in both the high and low load models (largest p = .010) but WMC was only a significant predictor in the high load model (high load $p < .001$; low load p = .390). 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 in the intention offloading task overall and in the letter string task at high memory loads. 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 tended to benefit from using offloading more than participants with higher WMC.

Is metacognition related to offloading choice behavior?

To investigate the association between metacognition and offloading choice behavior, we followed the plan outlined in our preregistration to correlate metacognitive calibration (predicted performance minus actual performance in the forced internal condition) with offloading choice behavior. We tested whether participants' underconfidence was associated with offloading choice behavior (see Boldt & Gilbert, 2019; Gilbert et al., 2020; Sachdeva & Gilbert, 2020).[6] 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 in the intention offloading task for low (r (249) = -0.04, adjusted p = .544) or high (r (249) = 0.01, adjusted p = .544) memory load. Similar patterns were observed in the letter string offloading task for both low (r (249) = 0.14, adjusted p = .986) and high (r (249) = 0.10, adjusted p = .986) memory load. We note, however, that metacognitive calibration is based on a difference score. Difference scores can have low reliability (e.g., Lord, 1956). Indeed, we found that reliability for our metacognitive calibration measure was outside of the typical acceptable range for both the intention offloading task (Cronbach's α = .67) and the letter string offloading task (Cronbach's α = .67). However, reliability was within the acceptable range for global metacognitive confidence (see Table 1).

Therefore, to further test whether overall metacognitive judgments were associated with offloading choice behavior with a more reliable measure, we conducted an exploratory analysis repeating the approach described above that referenced participant's metacognitive responses directly rather than metacognitive calibration. Here, we observed a significant negative association only in the intention offloading task under low memory load (r (249) = -0.14, adjusted p = .023). At high memory load in this task (r (249) = 0.02, adjusted p = .624) and for both low (r (249) = 0.06, adjusted p = .844) and high (r (249) = -0.01,

[5] A hypothesis around the benefit of choice offloading compared to forced internal memory being more related to WMC than in past research was preregistered; 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.

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

adjusted $p = .844$) memory load in the letter string offloading task no significant associations were observed.

Together, the general lack of association between metacognitive assessments and offloading choice behavior, particularly under high memory load in the intention offloading task where this relationship has been reported in the past, was surprising. 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 WMC according to our preregistered analysis plan. In the intention offloading task, we did not observe significant correlations between offloading frequency and WMC for either low (r (249) = -0.08, adjusted $p = .212$) or high (r (249) = 0.06, adjusted $p = .811$) memory load. Again, similar patterns emerged in the letter string offloading task (low load: r (249) = -0.01, adjusted $p = .907$; high load: r (249) = 0.14, adjusted $p = .985$). Unsurprisingly, the correlations at low and high memory load across tasks did not differ statistically (adjusted $p = .310$ for both low and high memory load comparisons). 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 addressed 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, including significant four-way interactions in both analyses. Further, participants with lower and higher levels of 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 lower 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 under high memory load in both tasks were negatively associated with WMC, suggesting that the benefits of offloading were larger for individuals with lower WMC.

The finding that WMC was associated with offloading benefit at high memory load in the choice offloading condition is consistent with the pattern reported by Ball, Peper, 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 lower 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 the pattern observed in the letter string offloading task comparing performance in the choice offloading and forced internal blocks where a significant WMC x condition interaction was not observed under low memory load. Why lower 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 reported pattern for this task. However, we note that the finding that WMC was not related to offloading frequency does replicate prior work in the letter string offloading task (Brown, 2021; Morrison & Richmond, 2020).

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 or was plagued by issues of low reliability. 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. Importantly, however, we did observe low Cronbach's alphas for our metacognitive calibration measure in both tasks, suggesting that low reliability for this measure could explain our unexpected patterns. Analyses focused on the responses that participants provided to the metacognitive questions themselves, rather than the calibration measure, only fared slightly better. That is, the expected association between metacognition and offloading behavior only emerged under low memory load in the intention offloading task.

Despite using an approach to measuring metacognition that has been used in prior research, it is possible that we failed to observe strong support for the expected relationship between metacognition and offloading behavior in the choice offloading block because the choice block in the current study was always presented as the final block after participants had extensive task experience. In addition, the large task battery used in the current study differs from past literature on this topic; it is possible that participants' awareness of the number of tasks and the length of the study overall may have impacted their metacognitive assessments. Future studies designed to test questions around metacognitive contributions to offloading behavior may choose to include fewer tasks and a larger number of measurements of metacognition. Future studies may also focus on item-level metacognitive judgements instead of or in addition to global-level measures of metacognition. For example, researchers 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 to better characterize the relationship between metacognition and offloading choice behavior. Another possibility is that the influence of metacognitive beliefs on cognitive offloading may be stronger when participant rewards are determined by task performance. In designs such as the present study, where participants receive a fixed reward for participation, non-metacognitive factors may have a proportionately greater influence on offloading strategies (Ngai & Gilbert, 2024).

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 to 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 participant can remember on average 5 of 10 targets 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 WMC 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 provide strong support for 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 the lack of significant correlations between WMC and offloading choice behavior may suggest some subtle support for the effort avoidance account. That is, 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 lower 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 (including 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 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. In addition, participants in the current study were permitted to complete the task battery over a 24-hour period, rather than in one or two in-laboratory sessions as would be the typical approach for in person studies. Participants also completed these tasks on their own devices rather than using a standardized laboratory setup. Therefore, it is possible that these methodological differences necessitated by the online data collection method had an impact on the current pattern of results. 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.

It should be noted that participants from across the WMC spectrum exhibited near-ceiling performance in the forced-external condition. This may not be desirable from the standpoint of assessing individual differences. However, we note that WMC was correlated with benefits of offloading at high memory load in both offloading tasks and in both offloading blocks (forced-external, choice offloading) and at low memory load for the intention offloading task in both offloading blocks. These data suggest that participants with higher WMC still stood to benefit from offloading when they were required or could choose to do so relative to relying on their own internal memory. Moreover, the general consistency of findings that emerged from comparisons involving the forced-external block and choice offloading block serve to clarify that any effects associated with WMC are not simply due to suboptimal choice behavior in lower WMC participants. However, future research may choose to include higher memory loads than were used in the current study to avoid participants performing near ceiling in offloading trials.

Finally, we note 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. 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. Despite our efforts to equate these tasks in terms of difficulty, it is possible that participants did not experience these tasks as comparably difficult. Future research comparing prospective and retrospective memory offloading may wish to directly equate memory loads across tasks to test whether patterns reported here persist under those conditions. It may also be useful to collect subjective difficulty ratings across tasks in future studies in order to bolster understanding of participants' experiences of completing these tasks.

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.

CRediT authorship contribution statement

Lauren L. Richmond: Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Lois K. Burnett:** Writing – review & editing, Writing – original draft, Visualization, Formal analysis, Data curation. **Julia Kearley:** Writing – review & editing, Software, Project administration, Investigation. **Sam J. Gilbert:** Writing – review & editing, Software, Methodology, Conceptualization. **Alexandra B. Morrison:** Writing – review & editing, Methodology, Conceptualization. **B. Hunter Ball:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

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 (pregistration: https://osf.io/42tb6; data and analysis scripts: https://osf.io/tsnp6/). 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.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2025.104617.

Data availability

Data, analysis scripts and output resulting from our data analysis are available on OSF: https://osf.io/tsnp6/

References

Ball, B. H., & Peper, P. (2022). *Cost avoidance underlies decisions to use prospective memory reminders* [Preprint]. PsyArXiv. doi: 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.* Sacramento: California State University.

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*, Article 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), Article 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. (2024). Cognitive offloading is value-based decision making: Modelling cognitive effort and the expected value of memory. *Cognition, 247*, Article 105783. https://doi.org/10.1016/j.cognition.2024.105783

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., 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

Hair, J. F., Black, W. C., & Babin, B. J. (2010). Multivariate data analysis: A global perspective. *Pearson Education.* https://books.google.com/books?id=SLRPLgAACAAJ.

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

Kline, R. B. (2011). *Principles and practice of structural equation modeling, 3rd ed.* (pp. xvi, 427). Guilford Press.

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

Lord, F. M. (1956). The measurement of growth. *ETS Research Bulletin Series, 1956*(1). https://doi.org/10.1002/j.2333-8504.1956.tb00058.x

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

Ngai, C., & Gilbert, S. (2024). *Metacognitive training facilitates optimal cognitive offloading* [Preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/dq85e.

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.

Quade, D. (1967). Rank analysis of covariance. *Journal of the American Statistical Association, 62*(320), 1187–1200. https://doi.org/10.1080/01621459.1967.10500925

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*, Article 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.