UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title
Effects of Offloading on Concurrent Prospective Memory Tasks

Permalink
https://escholarship.org/uc/item/2pq0p7m2

Author
Fellers, Craig

Publication Date
2024

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

EFFECTS OF OFFLOADING ON CONCURRENT PROSPECTIVE MEMORY TASKS

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

by

Craig Fellers

September 2024

The Dissertation of Craig Fellers is approved:

Professor Benjamin C. Storm, chair

Professor Hannah E. Hausman

Toshiya Miyatsu, Ph.D.

Peter Biehl
Vice Provost and Dean of Graduate Studies

Copyright © by
Craig Fellers
2024

Contents

List of Figures

Effects of Offloading on Concurrent Prospective Memory Tasks

Craig Fellers

Abstract

The current research adapted Einstein and McDaniel's (1990) prospective memory (PM) paradigm to explore how people manage multiple memory tasks while offloading some but not others. Participants engaged in an ongoing categorization task while monitoring for two specific PM cues. In an Offload condition, participants offloaded one of the two PM tasks, receiving reminders for that specific cue. In a Dual PM condition, participants remembered both PMs internally. This study aimed to address three specific hypotheses: (1) offloading one PM task would improve performance on a concurrent non-offloaded PM task, (2) offloading PMs is not entirely efficient and would lead to worse performance on non-offloaded PMs compared to a single PM condition, and (3) participants would show worse performance on a PM task when offloading was no longer available due to lack of practice. Performance was measured in terms of PM accuracy, reaction times, and ongoing task performance. We consistently failed to find evidence that offloading one PM led to improved performance on a concurrent non-offloaded PM. In some cases, offloading led to impaired performance on non-offloaded PMs. When participants were later required to perform previously offloaded tasks without reminders, their performance was consistently worse compared to those who had practiced these tasks without offloading. These findings challenge the assumption that offloading simply frees up cognitive resources for reallocation to other tasks and may stem from limited

flexibility in PM-related cognitive processes, hidden cognitive costs associated with offloading, or metacognitive illusions. These mechanisms, however, relied on the assumption that remembering 2 PMs requires more cognitive resources than remembering a single PM, and that offloading would reduce cognitive demands. In Experiment 1, however, participants did not perform better when they had a single PM compared to when they remembered two PMs. There may be benefits to remembering two PMs concurrently, possibly due to the increased frequency of cues with the additional PM; if so, offloading may hinder PM Performance. The current research suggests that while reminders improve immediate task completion, they may reduce performance on concurrent PM tasks and circumvent practice effects that strengthen PM learning.

Acknowledgment

I would like to thank Ben Storm for supporting the learning process and for his help in the research we have accomplished so far. I would also like to thank Hannah Hausman, Toshi Miyatsu, and Jeannie Fox Tree for their contributions to this project.

EFFECTS OF OFFLOADING ON CONCURRENT PROSPECTIVE MEMORY TASKS

Cognitive offloading can be defined as the use of physical action or external aids to alter the processing requirements of a task to reduce cognitive demand (Risko & Gilbert, 2016). It can encompass a wide range of actions aimed at transferring cognitive workload from the brain to external sources. Examples of cognitive offloading range from tilting one's head to read text at an angle (Jolicoeur, 1988) to writing down a shopping list to setting a digital calendar reminder. There are many benefits of cognitive offloading. In the context of remembering information for later retrieval, the primary benefit is that the offloaded information is saved externally and can thus remain available with higher fidelity than when remembered internally (Fellers, Miyatsu, & Storm, 2023).

Although cognitive offloading can effectively ensure future access to the saved information, it can also affect how people encode and remember internally. Sparrow, Liu, and Wegner (2011) tested participants' memory for trivia facts that were either saved (offloaded) or deleted (not offloaded). They found that participants remembered the saved facts less well than the deleted facts, a phenomenon often called digital amnesia. According to Sparrow et al., saving the facts onto the computer acts as a form of offloading within a transactive memory system (Wegner et al., 1991), freeing the participants from the need to retain those facts internally to the same degree as they would have if the facts were deleted. Subsequent research by Kelly and Risko (2019) confirmed that cognitive offloading allows participants to dismiss the offloaded information from internal memory once it is an external store. Murphy (2023) similarly observed that participants who offloaded items to an external store and then were given a test without access to those stores had worse memory for the offloaded items than participants who did not offload.

While offloading might impair memory for the offloaded content when it becomes unexpectedly inaccessible, it may improve memory in other ways. Storm and Stone (2015) found that, by offloading one list (List A), participants showed enhanced memory for a subsequently learned list (List B). The research investigated the premise that saving can serve as a mechanism for redirecting cognitive resources to other tasks. Participants studied 10 words from a single PDF (List A) and were informed that it would be saved or deleted. Subsequently, a second PDF (List B) with another set of 10 words was introduced for study. After a 20-second delay, a free recall test for List B items was given, followed by one for List A. If List A was saved, participants could review it again before test, as promised. Otherwise, no restudy opportunity was provided. Consistent with Storm and Stone's prediction, participants given a restudy opportunity for List A had a better memory for List B. This phenomenon was called the saving-enhanced memory effect. The proposed mechanism was that by offloading List A, there would be less interference for encoding and recalling items from List B.

In subsequent experiments, Storm and Stone (2015) explored the limitations of this effect. The second experiment introduced an unreliable saving mechanism, and they were only able to observe the saving enhanced effects when the save method was reliable. The third experiment varied the length of List A (2 words vs. 8 words). The saving enhanced memory effect was observed only in the 8-word condition. The researchers suggested if memory for List B were improved by reducing interference at encoding and/or retrieval, a shorter List A (2 words) would be less likely to induce such interference compared to a longer (8 word) List A.

Evidence also points to the saving enhanced memory effect being more general. Runge, Frings, and Temple (2019) not only replicated this effect using the paradigm from Storm and Stone (2015) but were also able to observe the benefits of saving within a new cognitive activity using math problems. After saving or deleting List A, participants answered several math problems before studying List B and then answered another set of math problems. Then they were tested on List B and finally List A (after a chance to restudy if List A was saved). Runge et al. confirmed the saving enhanced memory effect (with superior recall of List B items when List A was saved) and observed improved performance in solving math problems for the participants in the save condition compared to those in the delete condition. Interestingly, to our knowledge, research on the saving enhanced memory effect has not compared a save condition to a condition only requiring participants to remember a single list to test the efficiency of the offloading process in its ability to minimize interference from the offloaded items.

In recent theoretical work by Risko and colleagues (in press), it has been suggested that two independent dimensions can be used to characterize the ways in which people use external memory stores. The substitutive-duplicative dimension characterizes whether people use an external store as a replacement for internal memory or to duplicate items that are also stored in internal memory. The latter may occur particularly if the external store is unreliable. The redistribution dimension represents what people do with any processing capacity that is freed by cognitive offloading. According to Risko et al. (in press), individuals can either ‘conserve’ this capacity or redistribute it to other processing.

In a related concept of responsible remembering (Murphy & Castel, 2021), less critical information is intentionally either forgotten or offloaded to enhance the memory of more vital, goal-aligned data. Murphy (2021) revealed participants' tendency to remember items they deemed personally responsible for, a finding that aligns with selective rehearsal theories of item-method-directed forgetting (Benjamin, 2006; Woodward et al., 1973; MacLeod, 1975). Moreover, inhibitory mechanisms appear to complement selective rehearsal, with participants inhibiting certain information rather than completely forgetting it. Murphy observed that even when not responsible for remembering offloaded items, participants were still able to recall many of those items—demonstrating that under these conditions, the offloading process isn’t entirely efficient. However, this paradigm also has not been used to compare responsible remembering (intentionally forgetting less important items) to remembering a single set of items.

Offloading and Prospective Memory

One situation where cognitive offloading is often utilized is prospective memory, or remembering to perform intended actions in the future, a cornerstone of daily human functioning. PM tasks can be cognitively demanding, particularly when multiple intentions must be remembered concurrently (McDaniel & Einstein, 2000), and thus may particularly benefit from offloading. Cognitive offloading is an intuitive response to manage the increasing demands of multiple PMs. Indeed, Risko and Dunn (2015) and Gilbert (2015) observed that participants offloaded adaptively based on judgments of task difficulty.

Successful PM completion requires recognizing a cue and then remembering and performing the required action. A PM task involves this complete process. A PM task is done concurrently with an ongoing activity which distinguishes it from a vigilance task. For example, a PM in daily life could be remembering to take a new medication (PM task) when eating breakfast (cue) while going about the rest of the morning routine (ongoing task). One distinction within prospective memory is the type of target that will indicate the proper time to execute the future intention. Targets can either be time-based or event-based. In an event-based PM task, the to-be-performed action must be performed in response to an event (e.g., leaving the house or meeting a colleague). In a time-based task, one needs to remember an intended action at a particular time of day (e.g., pick up a child from daycare at 5:00) or after a specific duration (e.g., take the cookies out of the oven in 10 minutes).

Much of the laboratory research on prospective memory is based on a paradigm designed by Einstein and McDaniel (1990). In this paradigm, participants are first given an ongoing task to complete. This task typically consists of evaluating a string of letters to decide if they form a word (lexical decision-making task), a short-term remembering task, or a categorization task (does the exemplar fit within the category given). Participants are then told that they will have a secondary task to complete—they are to indicate when a specific cue is presented as a part of the ongoing task. For example, if they are performing a lexical decision-making task when they see the item "dance," they may need to press a specific key instead of performing the word/non-word judgment. The cue may also be an attribute of an item, such as when they see a word containing a double letter (e.g., ballot). Generally, the cue is presented between 3 and 6 times within the ongoing task, and PM performance is calculated as the proportion of PM cues that are accurately reported. Within this framework, experimenters have been able to control features of the ongoing task, the cue, and the relationship between them. While traditional laboratory research has often focused on the mechanisms underlying a singular PM task (e.g., Einstein & McDaniel, 1990; 1993), contemporary life necessitates juggling multiple PMs simultaneously. The implications of such multitasking with PMs are relatively unexplored, especially regarding how offloading one PM may affect the execution of another.

Monitoring for a PM cue is a cognitively demanding process. To establish the Preparatory Attentional and Memory Processes (PAM) theory, Smith (2003) explored the cognitive demands of event-based prospective memory by embedding a prospective memory task within an ongoing lexical decision task (identifying words from non-words) and measuring how doing so affected performance in terms of reaction time for the ongoing task. Non-cue reaction times specifically were evaluated to isolate the cognitive load imposed by monitoring for the PM cue. If the addition of a PM task leads to a longer reaction time on non-cue trials, that would suggest that participants allocate a proportion of their cognitive resources to monitoring for PM cues even when the cues themselves are not present. In contrast, the increased reaction times of cue items could be due to the additional demands of that particular trial. Participants given a PM exhibited longer reaction times on the lexical decision task for non-cue items than those who did not have the additional PM requirement, indicating that the prospective memory task competes for cognitive resources. Furthermore, amongst those given a PM to complete, those with better PM performance had longer reaction times for the lexical decision task, suggesting that the allocation of cognitive resources is balanced between these two tasks.

Adding additional PM tasks further impacts the performance of the ongoing task. Meier et al. (2015) observed that reaction times increased with the additional cognitive load required for monitoring multiple target events. The existence of multiple PM targets increases the cognitive effort required due to heightened monitoring requirements. Cantarella et al. (2023) also found that adding additional prospective memory tasks reduced the accuracy of the ongoing task.

Various attributes of the PM cue can affect PM performance. Cues that are perceptually distinct relative to items in the ongoing task result in improved PM performance (Brandimonte & Passolunghi, 1994; McDaniel & Einstein, 2000; McDaniel & Einstein, 1993). Brandimonte and Passolunghi made cues distinct by using upper-case text (when non-cue items were lower-case) and found that this led to large improvements in PM performance. McDaniel and Einstein (2000) had similar findings and suggested that when a perceptually distinct target is used, absolute levels of PM performance will be near perfect as these distinct targets engage an involuntary orienting process that leads to remembering the intention to perform that PM. They claimed that perceptually distinct targets should provide a reliable mechanism to improve PM performance. Furthermore, distinctive cues maintained high levels of performance even when participants were given a secondary task (Einstein et al., 1997) and for older adults who showed decreased PM performance when using non-distinct cues (McDaniel & Einstein, 2000). While not described in these terms, using cues with high perceptual distinctiveness may function as an effective reminder to complete the PM.

While the PAM theory emphasized the importance of strategic monitoring in PM tasks, the Multiprocess Framework proposed by McDaniel and Einstein (2000) offered a more flexible approach. This framework suggested that PM retrieval can occur through either strategic monitoring or spontaneous processes, depending on various factors such as whether the PM cue is integral to the ongoing task and cue salience. According to this theory, some PM tasks may be accomplished with minimal cognitive resources through spontaneous retrieval.

Building on this, the Dynamic Multiprocess Framework (Scullin et al., 2013) proposed that individuals can flexibly shift between strategic monitoring and spontaneous retrieval based on context and expectations. This framework suggests that people may engage in periods of strategic monitoring when they expect a PM cue to appear but may rely more on spontaneous retrieval at other times to conserve cognitive resources. This dynamic approach helps explain how individuals can maintain PM intentions over extended periods without constantly engaging in resource-demanding monitoring processes.

While less research has been done on offloading or setting reminders in the context of PM, Gilbert (2015) designed a paradigm for this line of inquiry that has been widely adopted. In this paradigm, participants are presented with 10 yellow circles, numbered 1-10, in a box on their computer screen and instructed to drag each circle in numerical order to the bottom of the box using their computer mouse. Once a circle is dragged to the bottom, it disappears. After the 10th circle disappears, the screen clears for the next trial. Participants are also given a PM task: they are instructed to drag either one circle or three circles to specific, alternative locations (left, right, or top). To offload PMs, participants can drag the target circles toward their intended locations at the beginning of each trial, serving as a perceptual cue and removing the need to internally remember the PM. Gilbert (2015) observed that offloading improved participants' performance for the PM. Furthermore, participants were more likely to offload PMs under conditions of higher memory load (three PMs compared to one PM).

Other more naturalistic study designs have also shown that offloading by setting reminders improves PM performance (Gurol-Urganci et al., 2013; Tam et al., 2021; Tran et al., 2014). Tam et al. (2021) ran a meta-analysis including 12 randomized controlled trials on the use of text messages to remind patients to take medications for chronic hypertension. Even though the cadence of text messages varied significantly by trial (from daily to every two weeks), Tam et al. found that objective clinical measures of blood pressure improved along with self-reported medication adherence with a text message intervention.

Most of the existing literature on offloading and PM is designed to demonstrate the improved performance of the PM being offloaded (Gilbert, 2015; Jones et al., 2021), with a limited exploration into how the effects of offloading could spill over to improve performance on other non-offloaded tasks. However, a recent study by Dupont et al. (2023), which used a modified experimental paradigm from Gilbert (2015), revealed a cognitive reallocation effect where cognitive resources were reallocated to remembering low-value PM tasks when high-value PMs were offloaded. These results suggested that, at least under these conditions, offloading one PM may allow participants to reallocate cognitive resources to enhance the non-offloaded PM performance. Dupont et al. proposed a mechanism similar to visual attention (Lavie 1995) where attention resources are utilized by relevant visual elements, and the remaining capacity automatically spills over to other features.

The delayed effects of offloading have also received little attention in the PM literature. However, in the context of time-based PM, research has revealed a practice effect (Roediger & Butler, 2011) where those who initially remember to perform a PM task without offloading perform better after a delay compared to those who had initially used a reminder (Guo & Huang, 2019; Hu & Feng, 2013). By repeatedly performing a time-based PM (without offloading), participants improve their ability to respond at the appropriate time. Using reminders improves PM performance for both time– and event-based PMs (Huang et al., 2014; Fellers et al., 2023). However, when reminders are used, less attention is allocated to monitoring for the time-based PM (Huang et al., 2014). Guo et al. (2023) observed the practice effect in the context of time-based PMs and explored possible mechanisms by measuring PM accuracy as well as reaction times in the ongoing task. In their experiment, participants experienced PM Phase 1 and a testing phase. For both, the ongoing task was a 1-back task, and the PM was to press the "1" key every minute. The PM response was successful if participants responded within 5 seconds of the minute, and PM Phase 1 consisted of 20 minutes. In the reminder condition, a timer would automatically display on the screen, starting at 55 seconds. Participants in the reminder condition did not see this reminder. Both groups, however, could check the time by pressing the spacebar, which would display the time for one second. In the testing phase, participants had 5 PM targets and were required to press the "1" key each minute with no reminders given to either group. Not surprisingly, in PM Phase 1, participants in the reminder condition had better PM performance and checked the time less frequently than participants in the no-reminder condition. Interestingly, no difference in ongoing task performance was observed between conditions. In the testing phase, participants in the reminder condition had worse PM performance than those in the no-reminder condition, although there was no significant difference in the number of time checks. Guo et al. speculated that the lack of difference in ongoing task performance in PM Phase 1 may have been due to the simple nature of the ongoing task. It was presumed that even if attentional resources were made available through the use of reminders, the limited cognitive demands of an easy ongoing task may have minimized any benefit from additional cognitive capacity yielded by using a reminder. While Guo (2023) was done in the context of time-based PM, it is reasonable to expect that using a reminder in an event-based PM may also compromise the learning effect of a repeated PM.

Cognitive offloading, which involves using external aids to reduce cognitive demand, improves memory for offloaded items. Offloading may also have a secondary benefit of enhancing memory for non-offloaded items through a reallocation of cognitive resources (Storm & Stone, 2015). While direct benefits of cognitive offloading have also been observed in the context of prospective memory (e.g., setting a reminder improves performance for the offloaded PM), less is known about how reminders affect other non-offloaded PMs, or how offloading may affect the ability to perform a PM when offloading is no longer available.

Aims of the Current Research

The current research adapted Einstein and McDaniel's (1990) prospective memory paradigm to explore how people manage multiple memory tasks while offloading some but not others. Participants engaged in an ongoing categorization task while simultaneously monitoring for two specific prospective memory cues (e.g., items that contain double letters or those beginning with a vowel). When participants were presented with a PM cue, they had to respond by pressing a designated key. In an Offload condition, participants offloaded one of the two PM tasks, and received reminders when a particular cue was presented. In a final catch trial, participants in the Offload condition received a final PM cue presented without the expected reminder to reveal to what extent they relied on the external reminders to remember the offloaded PM.

Within this experimental design, a few questions arose. First, how would performance on the non-offloaded PM compare to performance in another condition where the second PM needs to be remembered without offloading? The first hypothesis was that offloading one of two PM tasks would reduce cognitive demand required to monitor for the offloaded PM which would automatically be reallocated to monitoring for the non-offloaded PM. This hypothesis predicted that we would observe improved performance on the non-offloaded PM task compared to a condition where both PM tasks needed to be remembered without offloading. This hypothesis aligned with the Dynamic Multiprocess Framework (Scullin et al., 2013), which states that individuals adaptively shift between strategic monitoring and spontaneous PM retrieval based on context and expectations. It expanded this framework by suggesting that these adaptive shifts can occur not only within single PM tasks but also across multiple PM tasks, with offloading allowing for a spontaneous retrieval of the offloaded PM.

Alternatively, there could be a benefit to remembering both PMs internally. If participants were required to remember two PMs, they would be exposed to a higher density of cues within the ongoing task compared to when they only had to remember one PM because the other was offloaded. In visual search, rare items are more often missed than more frequent stimuli (Wolfe, et al., 2005). In that case, reducing the monitoring requirements from one of the two PM cues by offloading would make the remaining cues a more rare target event, and would in turn reduce performance for the non-offloaded PM.

If improved PM performance is observed for the non-offloaded PM compared to a dual PM condition, that would provide evidence of the benefits of remembering fewer PMs and that cognitive resources made available by offloading are being reallocated to the non-offloaded PM. This would represent a benefit to offloading for PM similar to what is observed with saving enhanced memory (Storm & Stone, 2015). This result would bolster recommendations to offload PM tasks in the real world, especially for those with reduced cognitive capacity. However, if performance for the non-offloaded PM was reduced compared to a dual PM condition, this would provide evidence for a benefit of remembering multiple PMs concurrently. If this is observed, it would inform people to be more judicious with reminders, utilizing them only for the most important tasks to allow more PMs to be monitored concurrently to take advantage of the generalized benefits of monitoring multiple PMs together.

The second hypothesis was that offloading a PM is not entirely efficient. In this case we would observe lower PM performance on the non-offloaded PM task compared to a control condition where only a single PM task is required. This decrease in performance would be due to residual or redundant monitoring efforts for the offloaded PM task, which, despite the offloading, would pull cognitive resources away from monitoring for the non-offloaded PM. This hypothesis supported the

Preparatory Attentional and Memory Process theory proposed by Smith (2003) which suggests that attentional resources are required for all PM monitoring tasks, no matter their complexity.

Alternatively, there may be situations where performing the offloaded PM task is so simple, it can be performed spontaneously and requires not effortful monitoring. This hypothesis aligns with the Multiprocess Framework. An offloaded cue would be highly distinct and would be processed focally as a part of the ongoing task. Both of these cue attributes can mitigate costs to ongoing tasks and eliminate reductions in PM performance otherwise observed in older adult participants (McDaniel & Einstein, 2000). If offloading a PM cue led to spontaneous recognition of that cue, it may not require cognitive capacity to be reallocated from the non-offloaded PM task. In this case, no differences in PM performance would be expected between the non-offloaded PM and a condition with a single PM.

If performance of the non-offloaded PM is worse than PM performance for participants who only have to remember a single PM, it would indicate that offloading may require a non-zero amount of cognitive effort which would reduce the resources available to perform the non-offloaded PM. However, if participants have similar PM performance for the non-offloaded PM compared to a single PM, it would indicate that the offloaded PM may not be redundantly remembered internally. In other words, it would provide evidence that offloading one of two PMs functions like subtracting that part of the task, and that there are not additional cognitive demands involved with offloading (e.g., redundant monitoring for the offloaded PM cue or other management of an offloaded PM which would not be involved with internally-remembered PMs). It could even be observed that participants perform better on the non-offloaded PM compared to participants with a single PM if the additional cues for the offload condition (even if they are accompanied by a reminder) make seeing and responding to a PM cue less of a rare event, and in turn, improve PM performance.

The third hypothesis was that if a PM was initially offloaded, participants would exhibit worse performance on that task when offloading was no longer available due to the lack of practice in monitoring for the PM cue. This would predict that participants who initially offloaded a PM task (who would perform well on the offloaded PM while it was offloaded) would perform worse on that task later, when required to manage it without offloading, compared to those who never offloaded the task. This hypothesis aligned with the Testing Effect (Roediger & Karpicke, 2006) by extending its principles to prospective memory, suggesting that actively monitoring for PM cues (analogous to retrieval practice) would lead to better long-term performance than relying on external aids (analogous to passive restudying). This extension of the Testing Effect to PM tasks has important implications for understanding the potential drawbacks of reliance on cognitive offloading and for designing effective strategies to maintain PM skills in situations where external aids may not always be available.

However, monitoring for a PM cue may not function in the same way as retrieval practice. For retrieval practice, participants must attempt to reproduce the to-be-remembered item from memory. In many PM tasks, the limiting step is generally monitoring rather than recalling the to-be-performed items (Graf & Uttl, 2001). Monitoring for the proper PM cue may function more like recognition memory where participants must identify whether a particular item is old or new. Effects of retrieval practice in recognition memory are less conclusive than those with recall memory (Chan & McDermott, 2007). In this case, we may not expect to observe a reduction in practice effects after offloading.

In the case that worse PM performance is observed for PMs that were previously offloaded when offloading is no longer available, that would support the hypothesis that monitoring functions as a type of retrieval practice and that retrieval practice is reduced when offloading. However, if similar performance is observed between PMs that were previously offloaded and for those that were not, it could either indicate monitoring for a PM does not function in the same way as retrieval practice in retrospective memory, or that offloading does not mitigate any benefits that are imbued by this practice. An observation of a reduction in retrieval practice effects, especially if combined with other reductions in performance of the non-offloaded PM would support a recommendation to avoid unnecessary use of reminders as a tool.

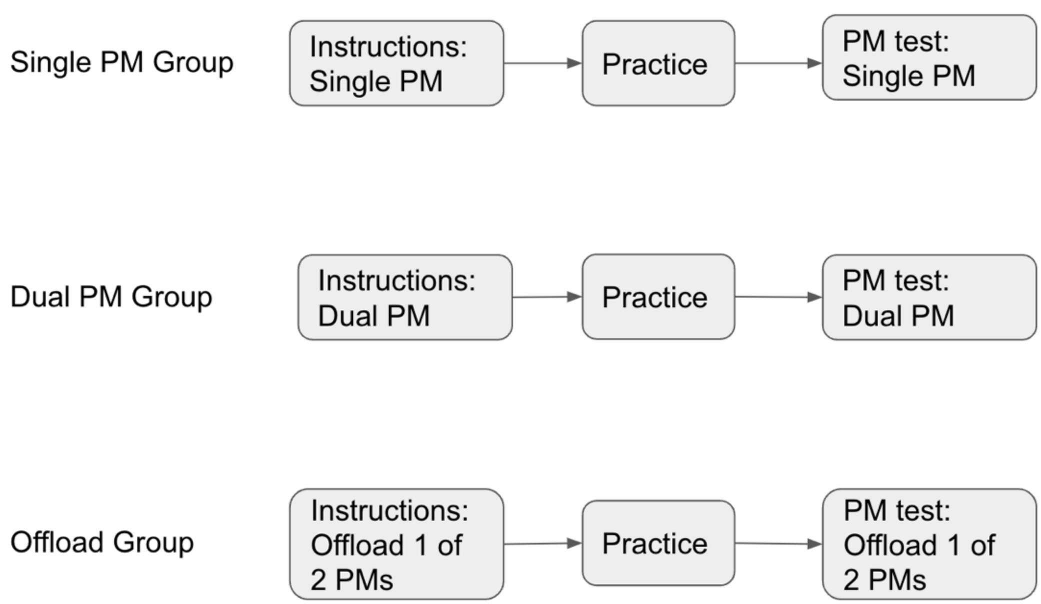
Experiment 1

Method

Participants. University of California, Santa Cruz undergraduates (N = 282) were recruited through the psychology subject pool and participated for partial class credit. We collected sufficient data to achieve .90 power to observe a Cohen's d of .5 for pairwise t-tests of PM performance across three groups.

Design. The critical independent variable was the offloading of one PM task. Participants were assigned to one of three groups. One group (Single) had a single PM, one group (Dual) had two PMs, and one group (Offload) had two PMs but was given reminders to help remember one of the PMs. The two dependent variables of interest were the proportion of the PMs correctly recalled as well as reaction times on the non-cue categorization trials. See Figure 1.

Figure 1. Experiment 1 Study Design.



Note. Between-subjects manipulation.

Experiment 1 examined the degree to which offloading one of two PMs might benefit the performance of the non-offloaded PM. The performance on the PM that was not offloaded was measured as a proportion of correct PM responses out of the 5 cues presented for each PM. For each group, the ongoing task was a categorization task adapted from Einstein and McDaniel (2005), where participants were asked to indicate whether an exemplar on one side of the screen fit within a category presented on the other side of the screen. Participants were instructed to respond by pressing the "1" key to indicate "yes" and the "0" key to indicate "no." The PMs were cued by specific word attributes within the ongoing task. One PM was cued by words containing a double letter (e.g., "gallon"). The other PM was cued when participants encountered words with three syllables (e.g., "elephant"); when either cue was observed, participants were instructed to press the spacebar. This easy-to-remember action was chosen to ensure that any differences in performance were due to monitoring for the cues rather than remembering the to-be-performed action.

For the Dual and Single PM groups, participants had to remember the PMs without reminders. In the Dual PM group, the order in which the PMs were presented was counterbalanced with half of the participants receiving instructions for either PM first. There were two versions of the Single PM group, one with each of the two PMs. In the Offload group, after being informed about both PMs, participants were told that for one of the two PMs (counterbalanced), they would receive a reminder to help them identify that cue within the ongoing categorization task. The cues with reminders were presented with the critical feature in bold and underlined. (e.g., a trial with a double letter cue in the offload condition would look like this: PROFESSION le**<u>tt</u>**er). For each PM task, participants were shown 5 cues spaced pseudo randomly across the 320 trials. All 320 trials were presented one after another without breaks.

Materials and Procedure. The experiment was administered remotely using the online survey software Qualtrics. The study began by describing the ongoing task to participants, which was a word-categorization ongoing task adapted from Einstein and McDaniel (2005). All participants were told that a category would be presented in upper case on the left and an exemplar would be presented in lower case on the right. The task was to decide whether the exemplar on the right fit within the category presented on the left (e.g., SPORT—baseball; TREE—strawberry). Participants were instructed to press the "1" key to indicate "Yes" and the "0" key to indicate "No." Participants were told to respond as quickly and accurately as possible; however, the words stayed on the screen until an answer was given, in which case the survey auto-advanced to the next trial. After receiving instructions about the ongoing task, participants were then given a practice round with 18 category-exemplar pairs. Instructions were also provided for an attention check to encourage engagement and gauge attention. Ten simple math problems were embedded into the categorization task throughout the experiment. Participants were told to provide the answer when they saw a math problem (e.g., 3+5).

After practicing the ongoing task, participants were then told that while engaged in the word categorization task, they must also indicate when they identified specific cue(s) within that task.

Here, the three groups were given different instructions. In the Single PM group, instructions for one PM were given to participants. There were two versions of the Single PM group, so that half of the Single PM group received instructions for the first PM, and half the PM Single group received instructions for the second PM. Specifically, instructions for the two PMs were to press the spacebar when presented with an exemplar containing a double letter or three syllables. Participants were instructed to respond by pressing the spacebar even if their response was on the trial after the cue.

The Dual PM group was given instructions to perform both PMs in a counterbalanced order so that half of the participants received instructions for each PM task first.

In the Offload Group, participants received the same instructions for the two PMs as the Dual PM condition. However, after learning about the two PMs, they were informed that they would receive a reminder to help them remember to perform one of the two PMs. The Offload group was counterbalanced across participants such that half of the participants received reminders for one PM where the other half of the participants received reminders for the other PM. They were told that the critical aspect of the cue would be both bold and underlined. They were given examples of what that would look like (e.g., ba**ll**ot) for a double-letter cue.

Next, participants in all conditions were given another practice round consisting of 66 categorization trials with three PM cues (for the Single group) or six cues (for the Dual and Offload conditions) embedded pseudo randomly across the trials (with one PM cue presented every 11 trials, plus or minus 5, in the Dual PM and Offload conditions and one PM cue presented ever 22 trials, plus or minus 5, in the Single PM condition). In the Offload condition, participants also received the reminder for appropriate cues as expected. During the practice round, if participants did not respond to a cue by pressing the spacebar, they received feedback about the proper response. (e.g., “For the last trial, the item "caramel" has three syllables. For items with 3 syllables, you are supposed to press the SPACEBAR rather than "0" or "1."). Participants were then told that they had completed the practice round.

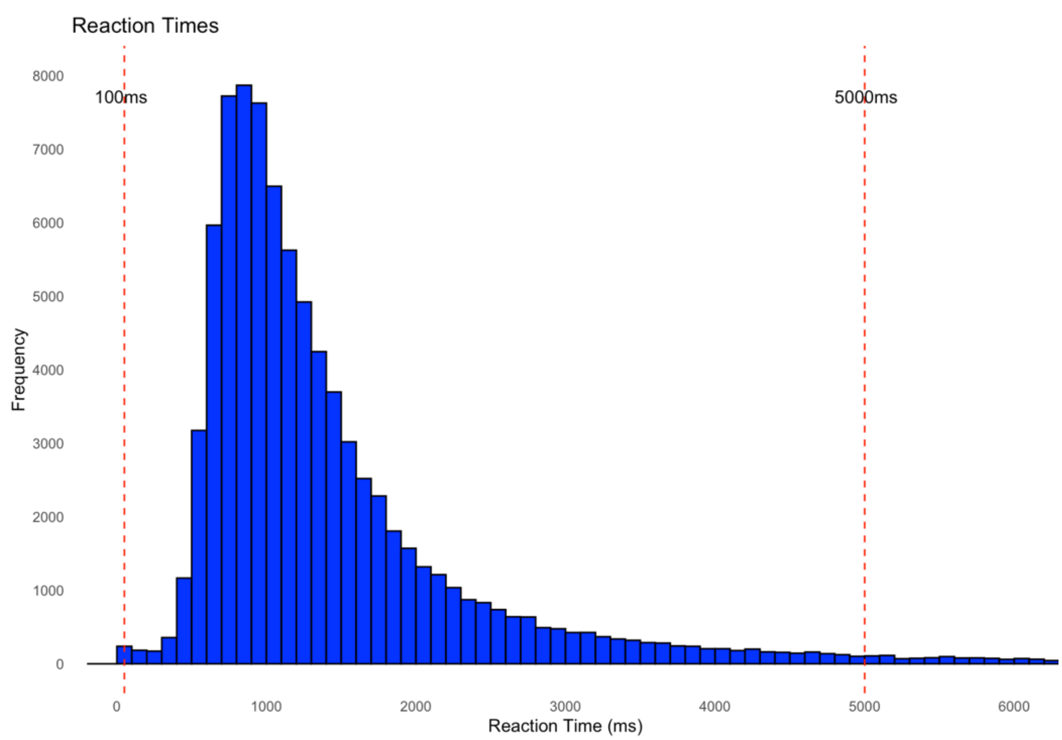
Participants then automatically advanced to the testing phase, where they were given the same instructions as they had received in the practice round. For all three groups, this phase consisted of 320 categorization trials using the same 8 categories and high-frequency exemplars. There were 5 instances of each cue (or 5 instances of a single cue for the Single PM Group) spaced pseudo randomly across 320 categorization trials (with one PM cue presented every 32 trials, plus or minus 7, in the Dual PM and Offload conditions, and one PM cue presented every 64 trials, plus or minus 7, in the Single condition). Lastly, participants in the Offload condition were given an additional 14 categorization trials followed by one additional PM cue—a catch trial for the cue that had been offloaded, although no reminder was given. The additional 14 categorization trials were only given to participants in the Offload condition. The testing phase lasted approximately 20 minutes. During the testing phase, a total of 10 math problems were presented as attention checks.

Participants were given credit for responding to a PM cue if they pressed the spacebar either on the appropriate or subsequent trial. No credit was given if they responded after that. Response times were measured using JavaScript code, which did not require additional software or plugin downloads. This code implemented a timing mechanism designed to enhance the accuracy of reaction time measurements by mitigating latency effects and minimizing technical differences across participants. The timer worked locally on the participants' devices to minimize the effects of internet connection speeds and the timer initiated only after the complete rendering of the stimulus was visible to the participant. The recording of reaction times was triggered by an initial keystroke. The survey advanced automatically after each response after a 250ms delay, allowing a small time buffer for loading the upcoming trials and helping to ensure that only deliberate responses were recorded.

Results

Data Cleaning. Only participants who finished the practice phase were considered for analysis. Data from 11 participants who did not correctly answer at least half of the attention check math problems were removed. Of the participants who were removed, the distribution was relatively even, with 5 from the Single PM condition, 3 from the Dual PM condition, and 3 from the Offload condition. This resulted in 91 participants in the Single PM condition, 90 in the Dual PM condition, and 90 in the Offload Condition. Next, outlier reaction times were removed. RTs were filtered based on task-specific expectations to include only RTs between 100 ms and 5000 ms, as shown in Figure 2. A 5000 ms upper cutoff was selected for the reaction time (RT) data for several reasons. There is no clearly agreed-upon cutoff time for reaction time data in the PM literature. Previous research has used a 2 to 3 standard deviation cutoff, which, when reported, typically removes 2-8% of the data (McBride et al., 2001; Hayes et al., 2018). In the current experiment, a 5000 ms cutoff eliminated 3.7% of the total RTs measured, aligning with this past work. The current RT data is left-skewed with a long tail. After eliminating clear outliers (RTs over 60 seconds, which are well outside reasonable expectations for completing a single trial of the categorization task), 5000 ms represents 2.25 standard deviations from the mean. This cutoff provides a consistent upper bound that can be applied equally across the experiments reported here. While log transformation of RT data is another potential method for reducing the effect of outliers, results are typically presented in raw format in the PM literature. We conformed with this approach to improve interpretability and comparability across the literature.

Figure 2. Experiment 1 Reaction Times.

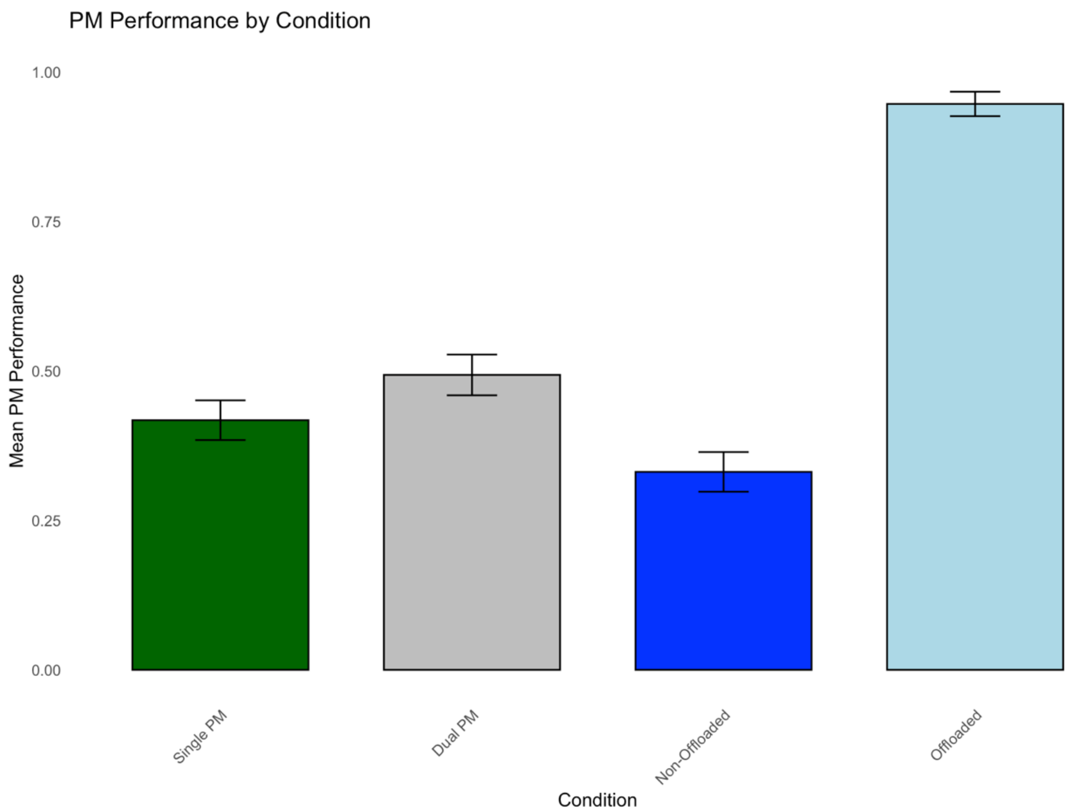


Note. Histogram represents reaction times pooled across all conditions. Outlier cutoffs indicated by red dashed lines.

A Shapiro-Wilk test of the filtered data, which included a sampling of 5000 RT data points, indicated that those data may not be normally distributed, $W = .83$, $p < .001$. Next, participants were removed who did not correctly score above chance (at least .50) on the ongoing categorization task. This removed 1 participant from the Single PM condition, 4 from the Dual PM condition, and 3 from the Offload condition, leaving 90 from the Single, 86 from the Dual, and 87 from the Offload conditions in the filtered data to be analyzed.

Prospective Memory. Not surprisingly, participants in the Offload condition performed well on the offloaded PM task, responding to the PM cue accompanied by the reminder by pressing the appropriate key with a mean performance of .95 ($SE = .03$). Descriptive statistics were calculated for performance of the non-offloaded PM, as represented in Figure 3. Participants in the Single PM condition had a mean PM performance of .42 ($SE = .03$). In the Dual PM condition, PM performance is reported separately for each of the two PMs. The first PM is compared to the non-offloaded PM and the second is compared to the offloaded PM. With the order of the PMs counterbalanced, the Dual PM comparison to both the non-offloaded and offloaded PM both PMs contains both PM tasks (i.e., double letter and 3 syllable). Those in the Dual PM condition had a mean PM performance of .49 ($SE = .03$), and those in the Offload condition had a mean performance for the non-offloaded PM of .33 ($SE = .03$). A 1 x 3 (Condition: Single PM vs. Dual PM vs. Offload) ANOVA was conducted to examine the effect of condition on non-offloaded PM performance, and significant differences were observed, $F(2, 260) = 5.97$, $p = .003$, $\eta 2 = .04$. Follow up between subjects t-tests were then conducted. There was a significant difference in PM performance between participants in the Dual PM and Offload conditions, $t(171) = 3.45$, $p < .001$, $d = .53$, 95% CI [.22, .83]. However, a significant difference in PM performance between participants in the Single PM condition and participants in the Dual PM condition was not observed, $t(174) = 1.50$, $p = .136$, $d = .23$, 95% CI [-.07, .52]. There was a significant difference in PM performance between participants in the Single PM and participants in the Offload conditions, $t(175) = 1.98$, $p = .0496$, $d = .30$, 95% CI [.00, .59].

Figure 3. Experiment 1 Prospective Memory Performance, PM Phase 1.

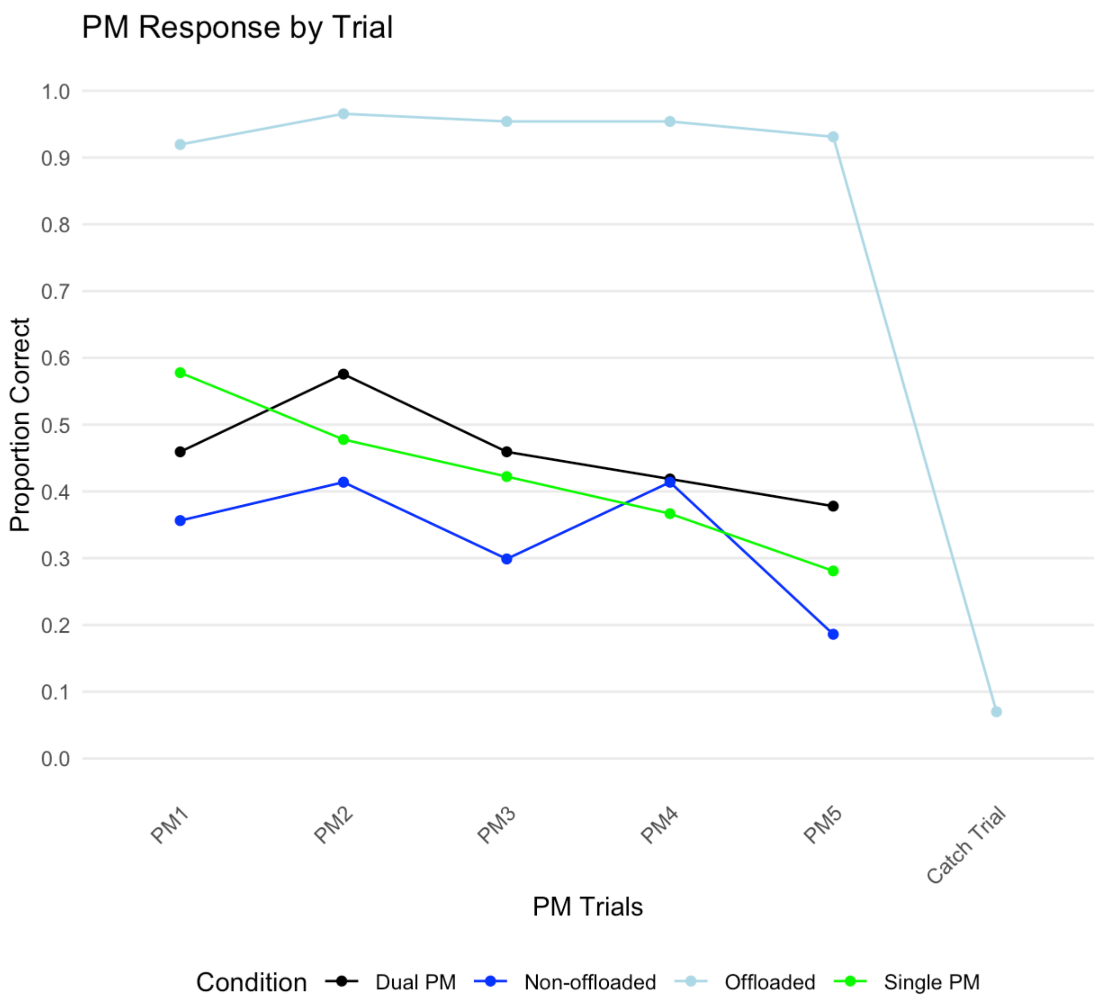


Note. In this figure, PM performance for the Dual PM condition was averaged across both counterbalance conditions. The Non-Offloaded and Offloaded PM performance was the performance from participants in the Offload condition.

Catch Trial. For the non-offloaded PM, performance generally decreased across trials: PM1 (M = .35, SE = .05), PM2 (M = .41, SE = .05), PM3 (M = .29, SE = .05), PM4 (M = .42, SE = .05), and PM5 (M = .19, SE = .04). However, participants had significantly higher PM performance in all non-offloaded PM trials than in the catch trial (M = .07, SE = .03). PM1: $t(85) = 4.74$, $p < .001$, $d = .51$, 95% CI [.29, .74], PM2: $t(85) = 5.53$, $p < .001$, $d = .60$, 95% CI [.37, .82], PM3: $t(85) = 3.79$, $p <$

.001, $d = .41$, 95% CI [.19, .63], PM4: $t(85) = 5.49$, $p < .001$, $d = .59$, 95% CI [.36, .82], PM5: $t(85) = 2.29$, $p = .024$, $d = .25$, 95% CI [.03, .46]. See Figure 4.

Figure 4. Experiment 1 Trial by Trial Analysis.



Note. In this figure, PM performance for the Dual PM condition was averaged across both counterbalance conditions. The Non-Offloaded and Offloaded PM performance was from participants in the Offload condition. Only those in the Offload condition were given a catch trial.

Reaction Times. Participants in the Single PM condition had a mean RT of 1419 ms (*SE* = 31), participants in the Dual PM condition had a mean RT of 1351 ms (*SE* = 32), and participants in the Offload condition had a mean RT of 1358 ms (*SE* = 31). A 1 x 3 (Condition: Single PM vs. Dual PM vs. Offload) ANOVA was conducted for the average RT for each participant, and significant differences were not observed, $F(2, 260) = 1.14$, $p = .239$, $\eta^2 = .01$.

A linear model was used to examine the impact of average RT and condition (Single PM, Dual PM, and Offload) on PM performance for the non-offloaded PM. The model was significant for predicting PM performance, $F(5, 257) = 4.90$, $p < .001$, (Adjusted $R^2 = .07$). Within the model, RT was a significant predictor of PM performance in the Dual PM condition (reference level) ($b = .00030$, $SE = .00011$, $p = .008$), indicating that longer RTs were associated with better PM performance in this condition. Controlling for RT, the main effect of condition was not significant for either the Single PM condition ($b = .3344$, $SE = .2240$, $p = .137$) or the Offload condition ($b = -.1012$, $SE = .2228$, p = .650) compared to the Dual PM condition. The interaction between RT and the Single condition did not reach significance ($b = -.00030$, $SE = .00016$, $p = .059$). The interaction between RT and the Offload condition was not significant ($b = -.00005$, $SE = .00016$, $p = .766$). The positive relationship between RT and PM performance observed in the Dual PM condition relative similar across conditions. The interaction between average RT and condition was not significant, $F(2, 257) = 2.08$, $p = .127$, $\eta^2 = .015$. The main effect of RT was significant, F(1, 257) = 7.97, p = .005, η2 = .028. The effect of condition was also significant, F(2, 257) = 6.22, p = .002, η^2 = .044.

Ongoing Categorization Performance. The ongoing categorization task was scored for the three groups. The Single PM group had a mean of .88 (*SE* = .01), the Dual PM group had a mean of .87 (*SE* = .01), and the Offload group had a mean of .87 (*SE* = .01). A 1 x 3 condition (Single PM vs. Dual PM vs. Offload) ANOVA was conducted to examine the effect of condition on the performance of the ongoing categorization task, $F(2, 260) = .18$, $p = .834$, $\eta^2 = .001$, and significant differences were not observed.

Discussion

In Experiment 1, we observed that offloading one of two PMs by using a reminder reduced the performance of the non-offloaded PM compared to that of participants in a condition where both PMs were remembered internally. Performance on the non-offloaded PM was also worse than that of participants in a condition with only a single PM. However, participants in the Single PM condition were only presented with 5 total PM cues, while the Dual PM condition had 5 PM cues for each PM for a total of 10 cues. The Offload condition also had 10 total PM cues, although 5 were accompanied by a reminder. In visual search, rare items are more often missed than more frequent stimuli (Wolfe, et al., 2005). Indeed, in this experiment, participants in the Dual PM condition did not perform worse than those in the Single PM condition as was initially expected. This finding compromises our ability to convincingly test the first two hypotheses.

Still, these results were surprising and contrary to the hypothesis that offloading would reduce the monitoring demands for the offloaded PM. This hypothesis predicted that resulting cognitive capacity would be reallocated to monitoring for the non-offloaded PM, an effect that was recently observed by Dupont et al. (2023). If a cognitive reallocation effect was not observed in terms of PM performance for the non-offloaded PM task, it was possible that cognitive resources would be reallocated to the ongoing task. This would manifest as increased speed or accuracy in the ongoing categorization task. However, no significant differences were found between conditions for either measure.

One possible explanation for why we did not observe better PM performance for the non-offloaded PM for participants in the Offload condition compared to participants in the Dual PM condition, and why we did observe significantly lower PM performance for the non-offloaded PM compared to PM performance in the Single PM condition is that participants in the Offload condition did not completely offload the to-be-offloaded PM. If participants did not trust that a reminder would be provided as instructed, they might have continued to redundantly monitor for the offloaded PM, which could have consumed cognitive capacity. However, it is unclear whether the non-offloaded PM performance was worse than the Single PM performance because of inefficient offloading. In the Dual PM condition, PM performance was numerically higher than in the Single PM condition; not fully offloading the offloaded task could have even been beneficial for the performance on the non-offloaded task if there is some benefit to remembering multiple PMs internally. In terms of the offloaded PM, participants had near-ceiling performance (.95) and performed poorly in the catch trial (.07), where one additional instance of the offloaded cue was provided without the reminder. This indicates that participants were indeed offloading the PM task to some degree and relying on the reminder.

Experiment 2a

Experiment 2a was designed to address two primary questions. First, it aimed to replicate the unexpected result from Experiment 1 that participants in the Offload condition had worse PM performance for the non-offloaded PM than those in the Dual PM condition. Second, it sought to investigate the delayed effects of offloading when learning multiple PM tasks. While previous research has shown both immediate costs (Sparrow et al., 2011; Murphy, 2023) and benefits (Storm & Stone, 2015; Dupont et al., 2023) of cognitive offloading, less is known about its impact after a delay when offloading is no longer available.

Delayed effects of offloading can be thought of in terms of the practice effect (Roediger & Karpicke, 2006; Roediger & Butler, 2011; Bjork & Bjork, 2011). In traditional memory studies, actively retrieving information during learning leads to better long-term retention compared to passive restudy. In the context of prospective memory, monitoring for a cue without a reminder could be analogous to retrieval practice, while relying on reminders might be more similar to restudy. By comparing performance between those who initially monitored for PM cues without a reminder versus those who relied on reminders, Experiment 2a aimed to explore whether a similar practice effect could be observed.

Experiment 2a was structured to address these two questions. Specifically, this experiment included both a PM Phase 1 and PM Phase 2 and utilized a within-subjects design for additional statistical power. PM Phase 1 was similar to Experiment 1 where participants in the Dual PM condition had to remember two PM tasks without reminders and participants in the Offload condition received reminders for one of the two PM tasks. This provided an opportunity to replicate the result that those in the Offload condition had worse PM performance for the non-offloaded PM compared to those in the Dual PM condition. After PM Phase 1, participants proceeded to PM Phase 2 where participants remembered both PMs without reminders. Unlike the catch trial in Experiment 1, in PM Phase 2, participants were given instructions that no reminders will be provided. While the catch trial helped to confirm a theoretical prediction that participants would have worse memory for offloaded items when the reminder was unexpectedly removed, PM Phase 2 allowed us to observe whether practice effects would be reduced in the Offload condition.

The hypothesis that could then be tested is that if a PM was initially offloaded, participants would exhibit worse performance on that task when offloading was no longer available, due to the lack of practice in monitoring for the PM cue. That hypothesis predicted that participants who initially offloaded a PM task would perform worse on that task in PM Phase 2 compared to those who never offloaded the task.

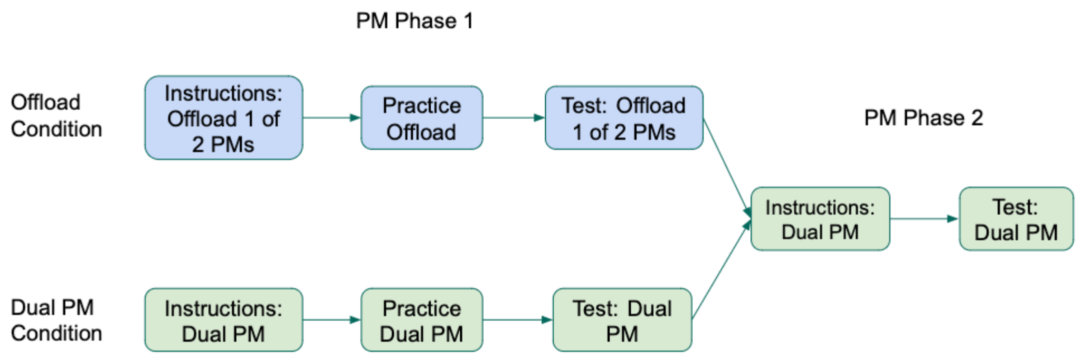
Method

Participants. University of California, Santa Cruz undergraduates (N = 130) were recruited through the psychology subject pool and participated for partial class credit. Although Experiment 1 revealed a Cohen's *d* of .53 for the between-subjects difference in PM performance for the non-offloaded PM, the effect observed in Experiment 1 was unexpected and novel, and replication attempts often yield smaller effect sizes than original studies. For that reason, the a priori sample size for .90 power to observe a Cohen's *d* of .30 with an alpha level of .05 (two-tailed) for the within-subjects difference in the PM performance utilized. Furthermore, using a smaller effect size for power analysis ensured adequate power to detect potentially subtler effects in the new PM Phase 2, where delayed effects of offloading were being explored for the first time in this context.

Design. Experiment 2a was divided into PM Phases 1 and 2, as depicted in Figure 5. There were several within-subjects variables. First was the PM performance for the non-offloaded PM between the Dual PM and Offload in PM Phase 1. Next was the PM performance in PM Phase 2 for the PM that had previously been offloaded compared to PM performance in PM Phase 2 in the Dual PM condition.

The procedure for the Dual PM and Offload conditions in PM Phase 1 was similar to the PM test in Experiment 1. After completing PM Phase 1, participants in both conditions followed the same instructions for PM Phase 2s, which required them to perform the same two PM tasks from PM Phase 1 without offloading. Non-cue reaction times were also compared between the Dual PM and Offload conditions for both the PM Phase 1 and 2.

Figure 5. Experiment 2a Study Design.



Note. Within-subjects manipulation.

Materials. Two sets of materials were implemented to accommodate a within-subjects manipulation of condition (Dual PM vs. Offload). All participants completed both the Dual PM and Offload conditions. The order of conditions was counterbalanced across materials so half of participants were placed into the Dual PM condition first while half of participants were placed into the Offload condition first. Within the Dual PM condition, the order of the PMs was counterbalanced as in Experiment 1 and within the Offload condition, the offloaded PM was counterbalanced as in Experiment 1. The first set of materials was similar to those used in Experiment 1. However, Experiment 2a required PM Phase 2 that was not utilized in Experiment 1. Due to the additional length of this experimental design, the Practice Phase was shortened to 20 categorization trials and contained 1 instance of each PM. PM Phase 1 contained 160 categorization trials, and PM Phase 2 contained

320 ongoing categorization trials, both were divided into the same 8 categories but used different exemplars. Both PM Phases 1 and 2 had 10 different PM cues each (5 for each PM) embedded pseudo randomly into the categorization trials (with 1 PM cue presented every 16 trials, plus or minus 5 trials, for PM Phase 1 and one PM cue presented every 32, trials plus or minus 7, in PM Phase 2). The two PM cues were items with double letters and items with three syllables, to which participants were to respond by pressing the spacebar. In the Offload condition, offloaded cues were presented in bold and underlined. Ten attention-check math problems were given in both PM Phases 1 and 2.

The second set of materials consisted of a pleasantness rating ongoing task. Participants were asked to rate how conventionally attractive they find common items (e.g., lake, raccoon, club, kelp, etc.) on a 1–7 scale with the instructions, “For the next set of trials, please indicate how conventionally attractive each item is on a 1-7 scale. Please be as quick and accurate as possible.” The two PMs in the second set of materials were items beginning with a vowel and items containing ‘ou.’ Instructions for performing the PM task were, “While performing the conventional attractiveness task that you just practiced, you will have some additional tasks to remember. Whenever you see an item that starts with a vowel (such as the word apple), please press the spacebar INSTEAD of rating the items. Please press the spacebar even if it is on the subsequent trial. In addition, whenever you see an item that contains an 'ou' (such as the word house), please press the spacebar INSTEAD of rating the item. Please press the spacebar even if it is on the subsequent trial. Now we'll practice incorporating these additional tasks." The order of the PMs being presented was counterbalanced so half of the participants had each PM presented first. Again, for participants in the Offload condition, offloaded cues were presented in bold and underlined and explained with the additional instruction, "In order to make it easier to recognize the items with an 'ou' [starting with a vowel], those letters will be written in bold with the 'ou' underlined such as: h**ou**se [**a**pple]". The PM that was offloaded was counterbalanced so half of the participants offloaded each PM.

Procedure. The experiment was administered remotely using the online survey software Qualtrics. All participants completed both the Dual PM and Offloaded conditions in counterbalanced order with half of the participants starting with the Dual PM condition, and the other half starting with the Offload condition. Participants in the first condition used the categorization ongoing tasks and the 'double letter' and 'three syllable' PMs. Participants in the second condition used the pleasantness rating task and the 'starts with a vowel' and 'contains ou' PMs.

Participants were first introduced to the ongoing task in both the Dual and Offload conditions. For the categorization task, they were told to indicate whether the item on the right fit within the category on the left for each trial. On the keyboard, they were told press "1" for "yes" and "0" for "no" and be as quick and accurate as possible. For the pleasantness rating task, they were asked to indicate how conventionally attractive each item was on a 1-7 scale and to be as quick and accurate as possible. As with Experiment 1, participants were instructed about math problems that were interspersed within the ongoing task as attention checks. The survey auto-advanced to the next trial after each response. After performing 20 trials of the ongoing task, participants were then introduced to the two PMs. In the first set of materials, instructions indicated, "While performing the ongoing categorization task that you just practiced, you will have some additional tasks to remember. Whenever you see an item with 3 syllables (such as the word elephant), please press the spacebar INSTEAD of categorizing the items. Please press the spacebar even if it is on the subsequent trial. In addition, whenever you see an item with a double letter (such as the word ballot), please press the spacebar INSTEAD of categorizing the item. Please press the spacebar even if it is on the subsequent trial. Now, we'll practice incorporating these additional tasks. However, it is important to continue to categorize items correctly." Following those instructions, participants were given 20 more trials of the ongoing task with a single instance of each PM embedded. For the Offload condition, the following additional instruction was provided, "In order to make it easier to recognize the items with a double letter [3 syllables], they will be written in bold with a double letter underlined such as: ba**<u>ll</u>**ot [**<u>el</u>** **<u>e</u>** **<u>phant</u>**]. Now we'll practice incorporating these additional tasks." Analogous instructions were given for the pleasantness rating materials. During the practice phase, if participants did not correctly respond to the PM cue, they were provided with feedback reminding them what to do when they saw a PM cue. After completing these trials, participants were told that they were done with the practice round and would advance directly to PM Phase 1.

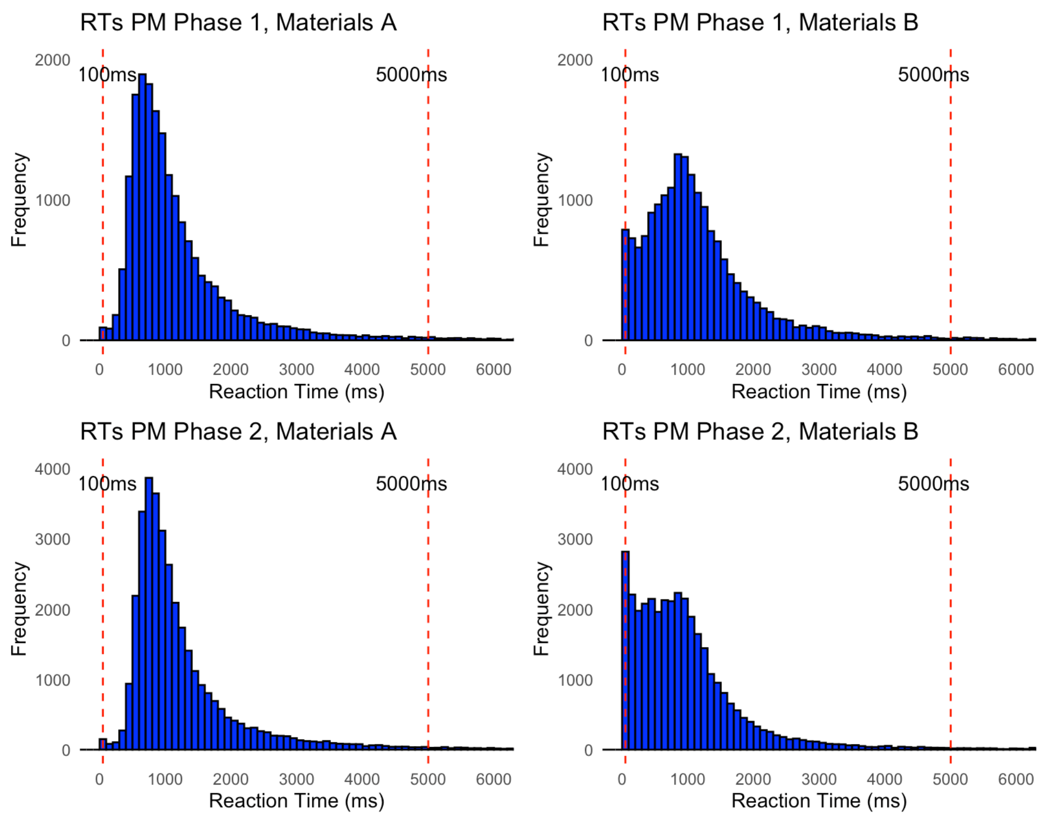
To begin PM Phase 1, instructions for completing the PM task were repeated to participants. They then proceeded through the 160 trials of PM Phase 1 and were presented with 10 PM cues in total (5 for each of the two PMs) interspersed pseudo-randomly across the trials as described in Materials. Offloaded PM cues were written in bold and underlined as expected. Ten attention-check math problems were embedded into PM Phase 1.

After completing PM Phase 1, participants transitioned directly into PM Phase 2. Participants from both the Dual and Offload conditions were given the same instructions, "We have one more block of categorization trials. Please read the following instructions. You will continue to categorize each item. While doing so, you will still need to press the spacebar rather than rating that item if it obeys either of the two rules from before. No reminders will be provided to help you remember to press the spacebar. So, whenever you see an item with 3 syllables (such as the word elephant), please press the spacebar INSTEAD of categorizing the items. Please press the spacebar even if it is on the subsequent trial. Also, whenever you see an item with a double letter (such as the word ballot), please press the spacebar INSTEAD of categorizing the item. Please press the spacebar even if it is on the subsequent trial." PM Phase 2 consisted of a new set of 320 trials with 10 new PM cues (5 for each of the two PMs) spaced pseudo randomly across the trials. One PM cue was presented every 32 items, plus or minus 7. Different cue items were used in PM Phase 2 than were used in PM Phase 1. Ten attention-check math problems were also be embedded into PM Phase 2.

Results

Data Cleaning. 130 participants who finished the practice phase were considered for analysis and 20 were removed; 11 were from the Dual PM first condition, and 9 were from the Offload first condition. Of those, 7 did not complete the survey, 7 did not respond correctly to at least half of the attention checks, and 6 did not follow instructions for the ongoing task. Of the 6 who did not follow instructions, one participant typed letters as a response (rather than a 1 or 0 to categorize), and five pressed the advance button without providing any response for most of the categorization task. Next, outlier reaction times were removed. RTs were filtered based on task-specific expectations to include only RTs between 100 ms and 5000 ms, as shown in Figure 6.

Figure 6. Experiment 2a Reaction Times.



Note. Histogram represents reaction times pooled across all conditions. Outlier cutoffs indicated by red dashed lines.

This eliminated 5.6 percent of the total RTs that were measured. A Shapiro-Wilk test of the filtered data, which included a sampling of 5000 RT data points, indicated that those data may not be normally distributed. For PM Phase 1 Materials A, $W = .81$, $p < .001$; for PM Phase 1 Materials B, $W = .88$, $p < .001$; for PM Phase 2 Materials A, $W = .79$, $p < .001$; for PM Phase 2 Materials B, $W = .85$, $p < .001$.

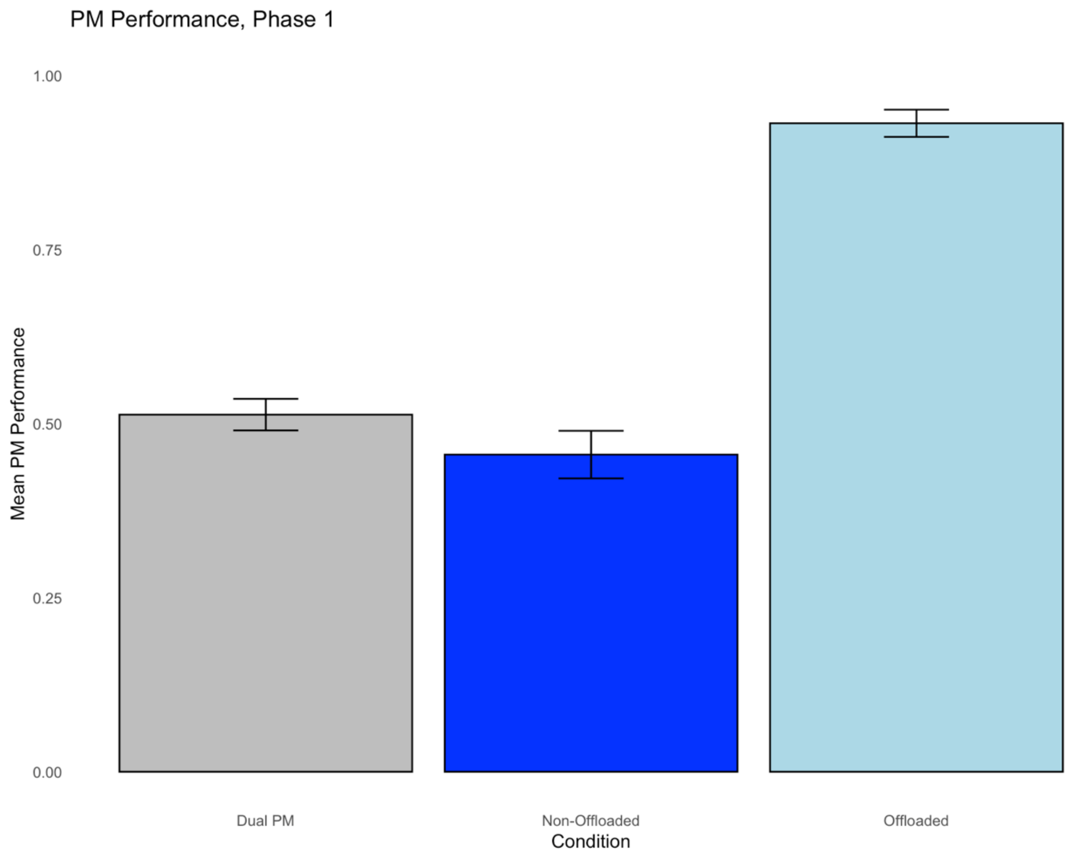
Next, participants who did not correctly score at least .50 on the ongoing categorization task were removed. This removed 2 additional participants from the

Dual PM condition, leaving 50 from the Dual PM first counterbalance condition and 58 from the Offload PM first in the filtered data to be analyzed.

PM Phase 1.

Prospective Memory. A within-subjects t-test was performed on performance of the non-offloaded PM between conditions. Here and in Phase 2, in the Dual PM condition, PM performance is reported separately for each of the two PMs. The first PM is compared to the non-offloaded PM and the second is compared to the offloaded PM. With the order of the PMs counterbalanced, the Dual PM comparison to both the non-offloaded and offloaded PM both PMs contains both PM tasks (i.e., double letter and 3 syllable). The mean PM performance for the non-offloaded PM in the Dual PM condition (M = .51, SE = .03) was not significantly different from that in the Offload condition (M = .46, SE = .03), $t(107) = 1.25$, $p = .214$, $d = .12$, 95% CI [-.07, .31]. The mean PM performance of the Offloaded PM was .93 (SE = .02). See Figure 7. To avoid any potential effects of participants receiving the catch trial in the Offload condition, non-offloaded PM performance was compared using only the first set of materials between-subjects. The mean PM performance for the non-offloaded PM in the Dual PM condition (M = .58, SE = .05) was not significantly different from that in the Offload condition (M = .51, SE = .04), $t(106) = 1.25$, $p = .306$, $d = .20$, 95% CI [-.18, .58].

Figure 7. Experiment 2a Prospective Memory Performance, PM Phase 1



Note. In this figure, PM performance for the Dual PM condition was averaged across both counterbalance conditions. The Non-Offloaded and Offloaded PM performance was from participants in the Offload condition.

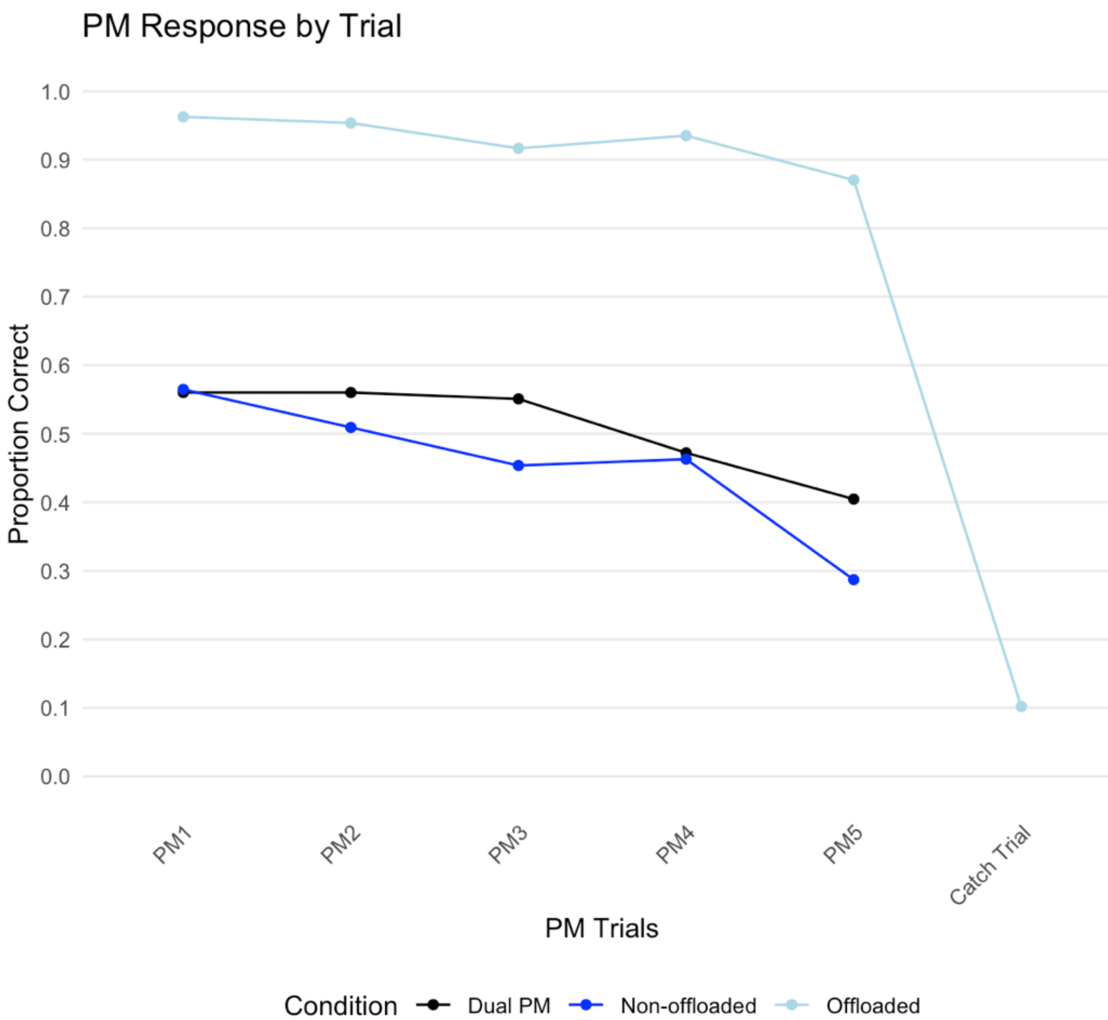
Reaction Times. Average participant RT was compared for non-cue items, and in the Dual PM condition participants showed longer RTs (M = 1201, SE = 41) compared to the Offload condition (M = 1121, SE = 35), $t(107) = 2.09$, $p = .039$, d = .20, 95% CI [-.01, .39].

A linear regression analysis was conducted to assess the impact of RT and condition (Dual PM vs. Offload) on PM performance for the non-offloaded PM. The model was not significant for predicting PM performance, $F(3, 212) = 2.61$, $p = .052$, (Adjusted $R^2 = .02$). However, with the model's p-value of .052, the model was still reported. Within the model, controlling for condition, RT was a significant predictor of PM performance ($b = .00019$, $SE = .00008$, $p = .015$), indicating that longer RTs were associated with better PM performance. Controlling for RT, the main effect of condition was not significant ($b = .1054$, $SE = .1453$, $p = .469$). There was no significant interaction between RT and condition ($b = -.00013$, $SE = .00012$, $p = .276$). The positive relationship between RT and PM performance was similar across the Offload and Dual PM conditions.

Ongoing Categorization Performance. Given the nature of the ongoing tasks, ongoing task scores could only be calculated for Materials A, which consisted of the categorization task. Participants performed better on the ongoing categorization task in the Dual condition ($M = .88$, $SE = .01$) than in the Offload condition ($M = .84$, $SE = .01$), $t(106) = 3.53$, $p < .001$, $d = .68$, 95% CI [.29, 1.07].

Catch Trial. For the non-offloaded PM, performance declined across trials: PM1 ($M = .56$, $SE = .05$), PM2 ($M = .51$, $SE = .05$), PM3 ($M = .45$, $SE = .05$), PM4 ($M = .46$, $SE = .05$), and PM5 ($M = .29$, $SE = .04$). However, in all trials, participants had significantly higher PM performance for the non-offloaded PM they did than for the catch trial ($M = .10$, $SE = .03$). PM1: $t(107) = 8.43$, $p < .001$, $d = .81$, 95% CI [.59, 1.03], PM2: $t(107) = 7.29$, $p < .001$, $d = .70$, 95% CI [.49, .91], PM3: $t(107) = 7.33$, $p < .001$, $d = .71$, 95% CI [.49, .92], PM4: $t(107) = 7.22$, $p < .001$, $d = .70$, 95% CI [.48, .90], PM5: $t(107) = 3.74$, $p < .001$, $d = .36$, 95% CI [.17, .55]. See Figure 8.

Figure 8. Experiment 2a Trial by Trial Analysis.

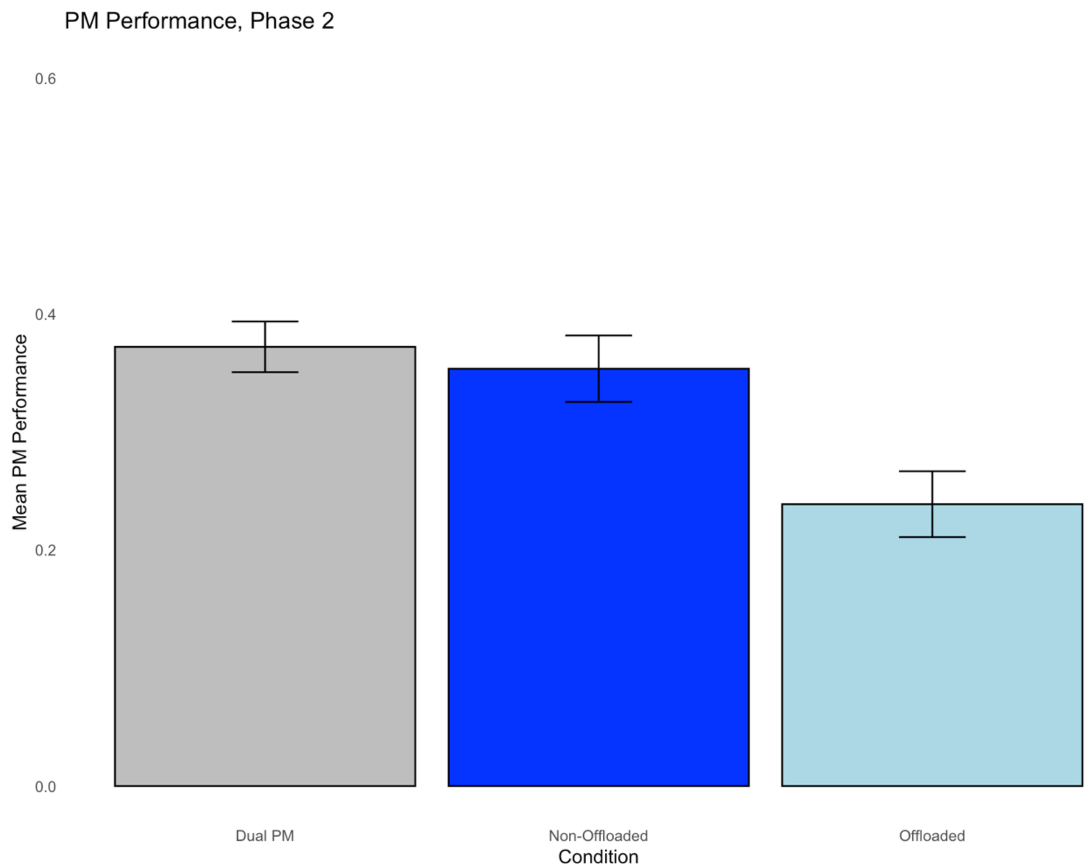


Note. In this figure, PM performance for the Dual PM condition was averaged across both counterbalance conditions. The Non-Offloaded and Offloaded PM performance was from participants in the Offload condition. Only those in the Offload condition were given a catch trial.

PM Phase 2.

Prospective Memory. Participants performed two PM tasks in PM Phase 2. In the Offload condition, one of the PMs had previously been offloaded in PM Phase 1, and the other had been remembered without a reminder. No reminders were given during PM Phase 2 (nor were they expected). A within-subjects t-test was performed on PM performance for both the PM that had *not* been offloaded in PM Phase 1 and the PM that had been offloaded. For the PM that had *not* been offloaded, the mean PM performance in the Dual PM condition (M = .35, SE = .03) was not significantly different from the Offload condition (M = .35, SE = .03), t(107) = .10, p = .920, d = .01, 95% CI [-.18, .20]. For the PM that had been offloaded, the mean PM performance in the Dual PM condition (M = .39, SE = .03) was significantly higher than in the Offload condition (M = .24, SE = .03), t(107) = 3.74, $p < .001$, d = .36, 95% CI [.17, .55]. See Figure 9. To avoid any potential effects of participants receiving the catch trial in the Offload condition, non-offloaded PM performance was compared using only the first set of materials between-subjects. The mean PM performance for the non-offloaded PM in the Dual PM condition (M = .43, SE = .05) was not significantly different from the Offload condition (M = .45, SE = .04), t(106) = .27, p = .790, d = .05, 95% CI [-.33, .43].

Figure 9. Experiment 2a Prospective Memory Performance, PM Phase 2.



Note. In this figure, PM performance for the Dual PM condition was averaged across both counterbalance conditions. The Non-Offloaded and Offloaded PM performance was from participants in the Offload condition.

Reaction Times. Average participant non-cue RT was compared, and in the Dual PM condition, participants showed similar RTs (M = 1115, SE = 41) compared to the Offload condition (M = 1098, SE = 35), $t(107) = .35$, $p = .725$, $d = .02$, 95% CI [-.16, .22].

Next, a linear regression analysis was conducted to assess the impact of RT and condition (Dual PM vs. Offload) on non-offloaded PM performance. The model significantly predicted PM performance, $F(3, 212) = 6.16$, $p < .001$, (Adjusted R^2 = .07). Controlling for condition, RT was a significant predictor of PM performance (b = .00016, SE = .00007, p = .020), indicating that longer RTs were associated with better PM performance. Controlling for RT, the main effect of condition was not significant (b = -.1317, SE = .1204, p = .275). There was no significant interaction between RT and condition (b = .00013, SE = .00010, p = .223). The positive relationship between RT and PM performance was similar across the Offload and Dual PM conditions.

Ongoing Categorization Performance. Ongoing categorization scores could only be calculated for Materials A. Participants did not perform significantly better on the ongoing categorization task in the Dual Condition (M = .88, SE = .01) than in the Offload Condition (M = .87, SE = .01), $t(106) = .53$, p = .601, d = .10, 95% CI [-.28, .48].

Discussion

Experiment 2a introduced significant changes to the study design. Unlike the previous experiment, which used a between-subjects design, this experiment employed a within-subjects approach. All participants now experienced both the Dual PM and Offload conditions, allowing for direct comparison of PM performance across conditions within the same individuals. Additionally, each condition was extended to include PM Phase 2 following PM Phase 1. During PM Phase 1, participants were given two PMs and either offloaded one of them or not depending on the condition. PM Phase 2 immediately followed, for which participants had to remember both PMs, but were told that no reminders would be provided. This design allowed for an examination the delayed effects of offloading (in PM Phase 2).

In PM Phase 1 of Experiment 2a, we found no evidence of a benefit of offloading one PM task on the performance of a concurrent non-offloaded PM task. Indeed, performance trended toward being impaired as the result of offloading the other PM. Again, this opposed the hypothesis that participants in the Offload condition would perform better due to a reduced need to monitor for the offloaded PM (and subsequent reallocation of cognitive capacity to the non-offloaded PM). However, these results did highlight the possibility for a benefit to monitoring multiple PMs internally, perhaps related to the increased cue density associated with having twice as many cues across the same number of ongoing trials (assuming that offloading mitigated the need to monitor for cues that were presented along with a reminder).

Like Experiment 1, participants had near-ceiling performance (.93) for the offloaded PM, when the previously offloaded cue was presented at the end of PM Phase 1 as a catch trial without a reminder, only 9 percent of participants responded. Indeed, performance on the catch trial was worse than that of any of the other non-offloaded PM trials. This indicates that participants had indeed offloaded and were relying on the reminder rather than the offloaded cue.

Results showed that in PM Phase 2, those in the Offload condition performed significantly worse on the PM that had been offloaded in PM Phase 1 compared to participants in the Dual condition that had remembered this PM internally. This suggests some benefits of practicing this PM were mitigated by offloading. This finding was consistent with reduction in practice effects when offloading recently observed in time-based PM (Guo et al., 2023).

In PM Phase 1, participants in the Offload PM condition had somewhat shorter reaction times for the ongoing task, indicating that fewer cognitive resources may have been allocated to the PM tasks in the Offload condition than the Dual PM condition. This is consistent with the expectation that offloading one of two PMs would reduce cognitive load needed to monitor for the PM tasks.

Experiment 2b

To create a condition in which offloading might be more likely to lead to benefits for the non-offloaded task, we modified the experimental design in Experiment 2b by increasing the difficulty of the PM tasks. Experiment 2b was similar to Experiment 2a; however, two *additional* PM tasks were given to increase the difficulty of remembering the PMs. In both the Dual PM and Offload conditions, the 4 PMs were presented to participants one at a time. In the Offload condition, following the instructions for the last PM, further instruction was provided indicating that in order to make it easier to recognize two of the PM cues, those cues will be written in bold font with the critical features underlined. A single example each offloaded PM was given to demonstrate what the reminder would be presented. This modification maintained the same ratio of offloaded to non-offloaded tasks as in Experiment 2a, while significantly increasing the overall cognitive load of the PM tasks. In the case of retrospective memory, Storm and Stone (2015) only observed a saving-enhanced memory effect when the offloaded task was more difficult. Offloading a more difficult PM task should increase the associated benefits.

Method

Participants. University of California, Santa Cruz undergraduates (N = 143) were recruited through the psychology subject pool and participated for partial class credit. The sample size was calculated a priori for .90 power to observe a Cohen's *d* of .30 with an alpha level of .05 (two-tailed) for the within-subjects difference in the PM performance. This sample size would provide sufficient power to observe practice effects similar to those observed in Experiment 2a for the offloaded PM in Phase 2.

Materials and Procedure. The materials for Experiment 2b were similar to Experiment 2a, modified to provide four PMs (and offload two in the Offload condition). The two new PMs added to the first set of materials were: items that end in 'ing' and items that can fly. For the second set of materials, the new PMs were: items spelled the same forwards and backwards and items with wheels. For the instructions, PMs were presented one at a time with participants clicking the "next" button to advance to the next PM instruction. In the Offload condition, after receiving the last PM instruction, when they clicked the "next" button, additional offloading instructions were provided. The offloading instructions read, "In order to make it easier to recognize 2 of these additional tasks, items with a double letter or that items that fly will be written in bold with the critical features underlined: ba**<u>ll</u>**ot or **<u>airplane</u>**." Participants had as much time as they needed to read instructions for each

PM. Each page contained instructions for a single PM. The additional instructions for offloading included both to-be-offloaded PMs along with an example of how both reminders would be presented.

Practice. Each practice round was expanded to 40 trials in order to include one instance of each of the four PMs with similar spacing of PM cues to what was used in Experiment 2a. The offloaded PMs were presented in bold and underlined as they had appeared in the instructions.

PM Phases 1 and 2. The total number of PM cues increased from ten in Experiment 2a to twelve in Experiment 2b to allow for three of each PM to be presented.

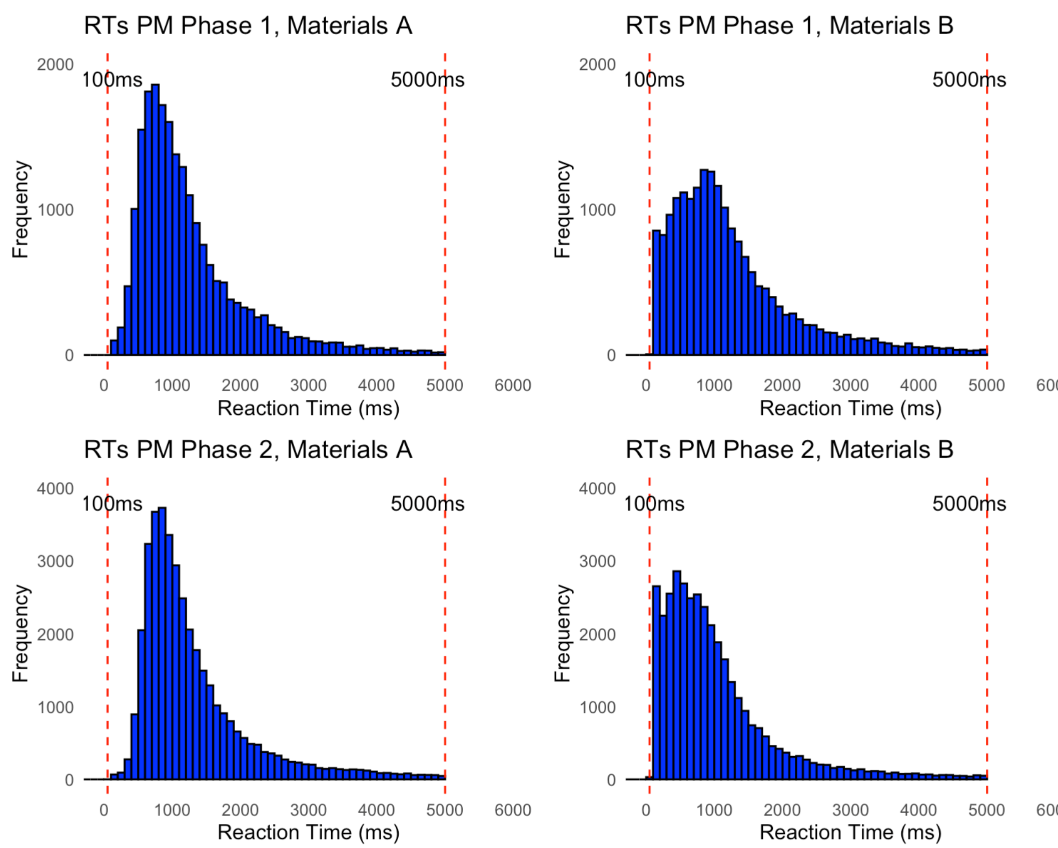
Results

Data Cleaning. Of the 143 participants who finished the practice phase, 23 were removed. Of those, 12 were from the Dual PM first condition and 11 were from the Offload first condition. Of the 23 removed, 6 did not complete the experiment and were removed. An additional 17 did not respond correctly to at least half of the attention checks and were removed.

After removing theses participants, outlier reaction times were removed. RTs were filtered based on task-specific expectations to include only RTs between 100 ms and 5000ms, as shown in Figure 10. This eliminated 7.2 percent of the total RTs. A Shapiro-Wilk test of the filtered data, which included a sampling of 5000 RT data points, indicated that those data may not be normally distributed for PM Phase 1 Materials A, $W = .84$, $p < .001$; for PM Phase 1 Materials B, $W = .87$, $p < .001$; for

PM Phase 2 Materials A, $W = .81$, $p < .001$; for PM Phase 2 Materials B, $W = .83$, $p < .001$.

Figure 10. Experiment 2b Reaction Times.



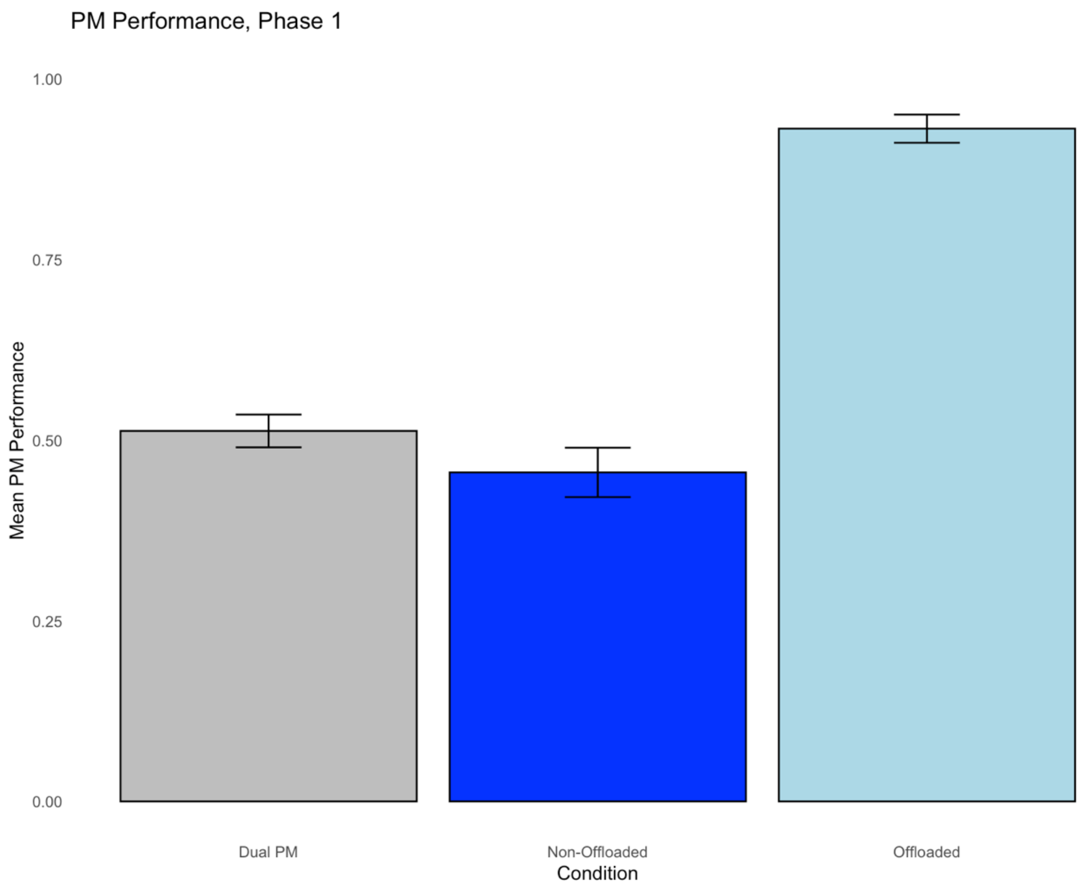
Note. Histogram represents reaction times pooled across all conditions. Outlier cutoffs indicated by red dashed lines.

Next, participants were removed who did not correctly score above chance (at least .50) on the ongoing categorization task. This removed 3 additional participants from the Dual PM condition and 1 from the Offload condition, leaving 55 from the Dual PM first condition and 61 from the Offload PM first conditions in the filtered data to be analyzed.

PM Phase 1.

Prospective Memory. A within-subjects t-test was performed on PM performance of the non-offloaded PM between the Dual PM and Offload conditions. Here and in Phase 2, in the Dual PM condition, PM performance is reported separately for each of the two PMs. The first PM is compared to the non-offloaded PM and the second is compared to the offloaded PM. With the order of the PMs counterbalanced, the Dual PM comparison to both the non-offloaded and offloaded PM contains all 4 PM tasks. The mean PM performance for the non-offloaded PM in the Dual PM condition (M = .42, SE = .03) was not significantly different than in the Offload condition (M = .40, SE = .03), t(115) = .60, p = .550, d = .06, 95% CI [-.13, .24]. The mean PM performance of the offloaded PM was .96 (SE = .01). See Figure 11. To avoid any potential effects of participants receiving the catch trial in the Offload condition, non-offloaded PM performance was compared using only the first set of materials between-subjects. The mean PM performance for the non-offloaded PM in the Dual PM condition (M = .44, SE = .04) was not significantly different from that in the Offload condition (M = .38, SE = .04), t(114) = .99, p = .323, d = .18, 95% CI [-.18, .55].

Figure 11. Experiment 2b Prospective Memory Performance, PM Phase 1.



Note. In this figure, PM performance for the Dual PM condition was averaged across both counterbalance conditions. The Non-Offloaded and Offloaded PM performance was from participants in the Offload condition.

Reaction Times. The average non-cue RT in the Dual PM condition was significantly longer (M = 1304, SE = 47) compared to the Offload condition (M = 1215, SE = 38), $t(115) = 2.13$, $p = .035$, $d = .20$, 95% CI [-.14, .38].

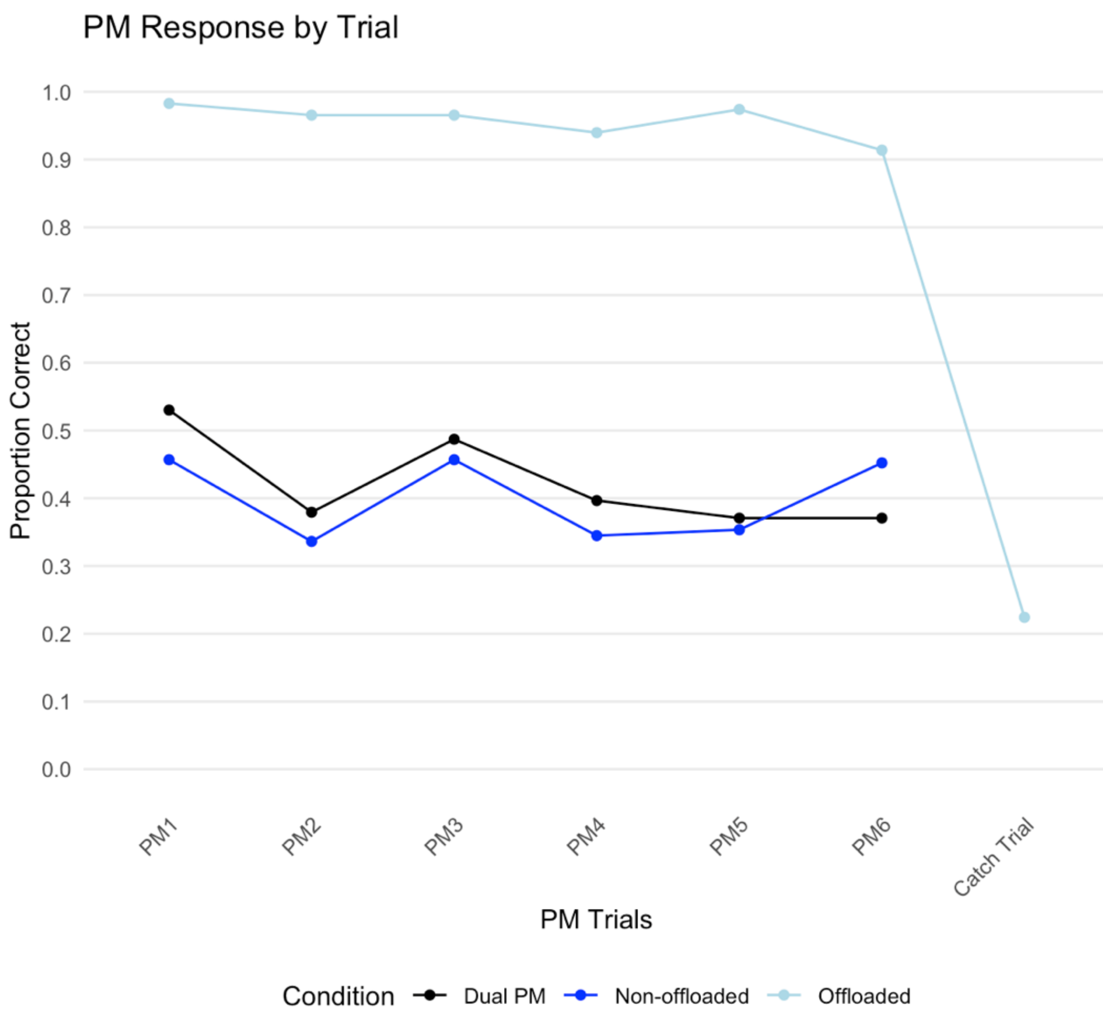
A linear regression analysis was conducted to assess the impact of RT and condition (Dual PM vs. Offload) on non-offloaded PM performance. The model significantly predicted PM performance, $F(3, 228) = 3.47$, $p = .017$, (Adjusted R^2 =

.03). Controlling for condition, RT was a significant predictor of PM performance (*b* = .00016, *SE* = .00006, *p* = .010), indicating that longer RTs were associated with better PM performance. Controlling for RT, the main effect of condition was not significant (*b* = .0132, *SE* = .1264, *p* = .917). There was no significant interaction between RT and condition (*b* = -.00002, *SE* = .00010, *p* = .846). The positive relationship between RT and PM performance was similar across the Offload and Dual PM conditions.

Ongoing Categorization Performance. Given the nature of the ongoing task, again, scores could only be calculated for Materials A, the categorization task. Participants performed better on the ongoing categorization task in the Dual condition (*M* = .88, *SE* = .01) than in the Offload Condition (*M* = .84, *SE* = .01), *t*(114) = 4.22, *p* < .001, *d* = .78, 95% CI [.40, 1.16].

Catch Trial. For the non-offloaded PM, performance varied across trials: PM1 (*M* = .46, *SE* = .05), PM2 (*M* = .34, *SE* = .04), PM3 (*M* = .46, *SE* = .05), PM4 (*M* = .34, *SE* = .04), PM5 (*M* = .35, *SE* = .04), and PM6 (*M* = .45, *SE* = .05), with all trials showing higher PM performance than the catch trial (*M* = .22, *SE* = .04). PM1: *t*(115) = 4.56, *p* < .001, *d* = .42, 95% CI [.23, .61], PM2: *t*(115) = 1.96, *p* = .052, *d* = .18, 95% CI [.00, .37], PM3: *t*(115) = 4.32, *p* < .001, *d* = .40, 95% CI [.21, .59], PM4: *t*(115) = 2.38, *p* = .019, *d* = .22, 95% CI [.04, .40], PM5: *t*(115) = 2.60, *p* = .011, *d* = .24, 95% CI [.06, .43], PM6: *t*(114) = 4.31, *p* < .001, *d* = .40, 95% CI [.21, .59]. See Figure 12.

Figure 12. Experiment 2b Trial by Trial Analysis.

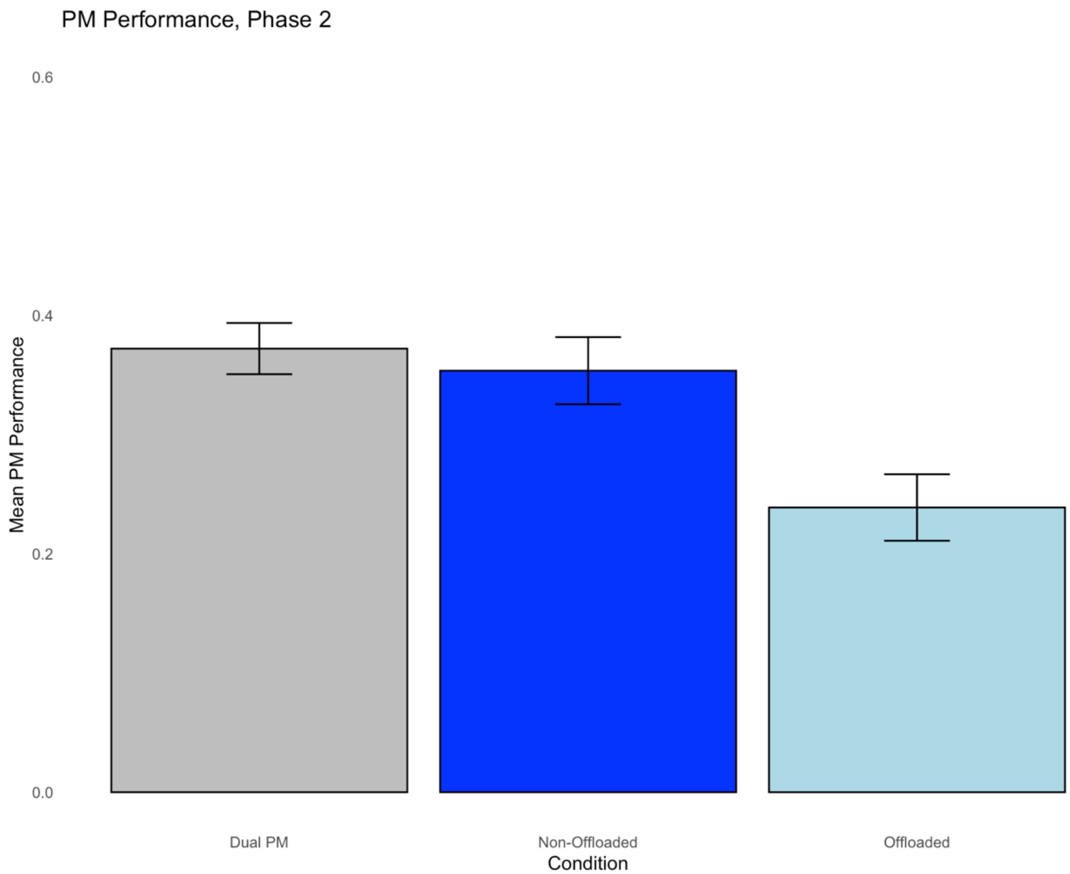


Note. In this figure, PM performance for the Dual PM condition was averaged across both counterbalance conditions. The Non-Offloaded and Offloaded PM performance was from participants in the Offload condition. Only those in the Offload condition were given a catch trial.

PM Phase 2.

Prospective Memory. A within-subjects t-test was performed on PM performance between conditions for both the PM that had *not* been offloaded in PM Phase 1 as well as for the PM that had been offloaded. For the PM that had *not* been offloaded, the mean PM performance in the Dual PM condition (M = .30, SE = .02) was not significantly different from the Offload condition (M = .30, SE = .03), t(115) = .23, p = .815, d = .02, 95% CI [-.16, .20]. For the PM that had been offloaded, the mean PM performance in the Dual PM condition (M = .29, SE = .03) was also not significantly higher than the Offload condition (M = .26, SE = .03), t(115) = .99, p = .323, d = .09, 95% CI [-.09, .27]. See Figure 13. To avoid any potential effects of participants receiving the catch trial in the Offload condition, non-offloaded PM performance was compared using only the first set of materials between-subjects. The mean PM performance for the non-offloaded PM for the Dual PM condition (M = .41, SE = .04) was not significantly different from the Offload condition (M = .42, SE = .04), t(114) = .20, p = .839, d = .04, 95% CI [-.33, .40].

Figure 13. Experiment 2b Prospective Memory Performance, PM Phase 2.



Note. In this figure, PM performance for the Dual PM condition was averaged across both counterbalance conditions. The Non-Offloaded and Offloaded PM performance was from participants in the Offload condition.

Reaction Times. The average RT was compared, and participants in the Dual PM condition (M = 1170, SE = 42) did not have significantly different reaction times compared to the Offload condition (M = 1167, SE = 37), $t(115) = .07$, $p = .943$, d = .01, 95% CI [-.18, .19].

Next, a linear regression analysis was conducted to assess the impact of RT and condition (Dual PM vs. Offload) on non-offloaded PM performance. The model significantly predicted PM performance, $F(3, 228) = 9.71$, $p < .001$, (Adjusted R^2 = .10). Controlling for condition, RT was a significant predictor of PM performance (b = .00013, SE = .00005, p = .019), indicating that longer RTs were associated with better PM performance. Controlling for RT, the main effect of condition was not significant (b = -.1918, SE = .1037, p = .066). There was a significant interaction between RT and condition (b = .00017, SE = .00008, p = .040). The positive relationship between RT and PM performance was stronger in the Offload condition compared to the Dual PM condition.

Ongoing Categorization Performance. Only ongoing categorization scores could be calculated for Materials A. There was not a significant difference in performance in the ongoing categorization task between the Dual PM condition (M = .88, SE = .01) and the Offload Condition (M = .86, SE = .01), $t(114) = 1.13$, p = .261, d = .21, 95% CI [-.16, .58].

Discussion

In Experiment 2b, the difficulty of the PM task was increased. In the Dual PM condition, participants had to remember 4 PMs concurrently, and in the Offload condition, they were given 4 PMs, and 2 were offloaded to a reminder. Similar reminders were used to those in Experiments 1 and 2a, alerting the participant when the PM cue was presented.

Experiment 2b revealed similar findings to Experiment 2a in PM Phase 1. PM performance for the non-offloaded PM did not differ significantly between the Dual and Offload conditions, and the performance on the Offloaded PMs remained high at

.96. In the Offload condition, participants performed better on all the non-offloaded PM trials compared to the catch trial.

For Experiment 2b, the more challenging PM task created a set of conditions that increased the potential benefits of offloading, as offloading two PM cues should free up more cognitive resources than offloading a single PM cue (as was used Experiments 1 and 2a). However, a reallocation of cognitive resources was not observed in terms of performance on the non-offloaded PM or on the ongoing task. Significant differences in non-offloaded PM performance were not observed between participants in the Dual PM and Offload conditions and participants in the Dual PM condition scored better on the ongoing task than those in the Offload condition. Again, however, given the finding in Experiment 1 that the Single PM condition failed to outperform the Dual PM condition (perhaps due to the increased cue density associated) it is also possible that effects of offloading may only serve to reduce potential benefits of the Dual PM condition.

A significant difference in PM performance between the Dual PM and Offload conditions was not observed in PM Phase 2 for the PM that had been offloaded in PM Phase 1. This indicated that participants in the Offload condition did not experience significantly less benefit from practicing the PM during PM Phase 1 compared to those in the Dual PM condition—who should have gotten more practice during PM Phase 1. The more complicated offloading instructions that participants received in Experiment 2b may have made the offloading process less efficient and encouraged participants to monitor for all four PM cues (both offloaded and non-offloaded).

In PM Phase 1, participants in the Offload PM condition trended towards shorter reaction times for the ongoing task (although this did not reach significance as it had in Experiment 2a). A difference would have been expected if offloading one of two PMs reduced cognitive load needed to monitor for the PM tasks. If participants were not able to effectively offload under this more complex set of conditions, any difference in reaction times in PM Phase 1 would be minimized.

Experiment 3

In the final experiment, we sought to make the offloaded PM task as easy as possible by telling subjects exactly what to do when the reminder was presented, thus increasing the potential for cognitive resource reallocation. To do this, Experiment 3 replicated Experiment 2a, adding a between-subjects variable. In the new between-subjects condition, the reminders given to participants highlighted the cue and told participants exactly what to do when it was observed. When a reminder was presented, additional instruction saying, "Press Spacebar Now" was also shown in large red font below the item. Even though the action to be performed was designed to be simple to remember, providing explicit instructions eliminated the need to remember that PM more completely. This updated procedure mitigated the possibility that presenting the cue item in bold swapped one PM task (e.g., respond to items with a double letter) for another (e.g., respond when the item is bold). Indeed, Einstein et al. (1998) found that reminders that included both the event-based cue and the intended action were most effective. As with Experiments 1, 2a, and 2b, participants experienced these reminders during a practice round to gain confidence in their efficacy and trust that they would be provided at the appropriate time.

Method

Participants. University of California, Santa Cruz undergraduates (N = 280) were recruited through the psychology subject pool and participated for partial class credit. The sample size was calculated a priori for .90 power to observe a Cohen's *d* of .30 with an alpha level of .05 (two-tailed) for the within-subjects difference in the PM performance.

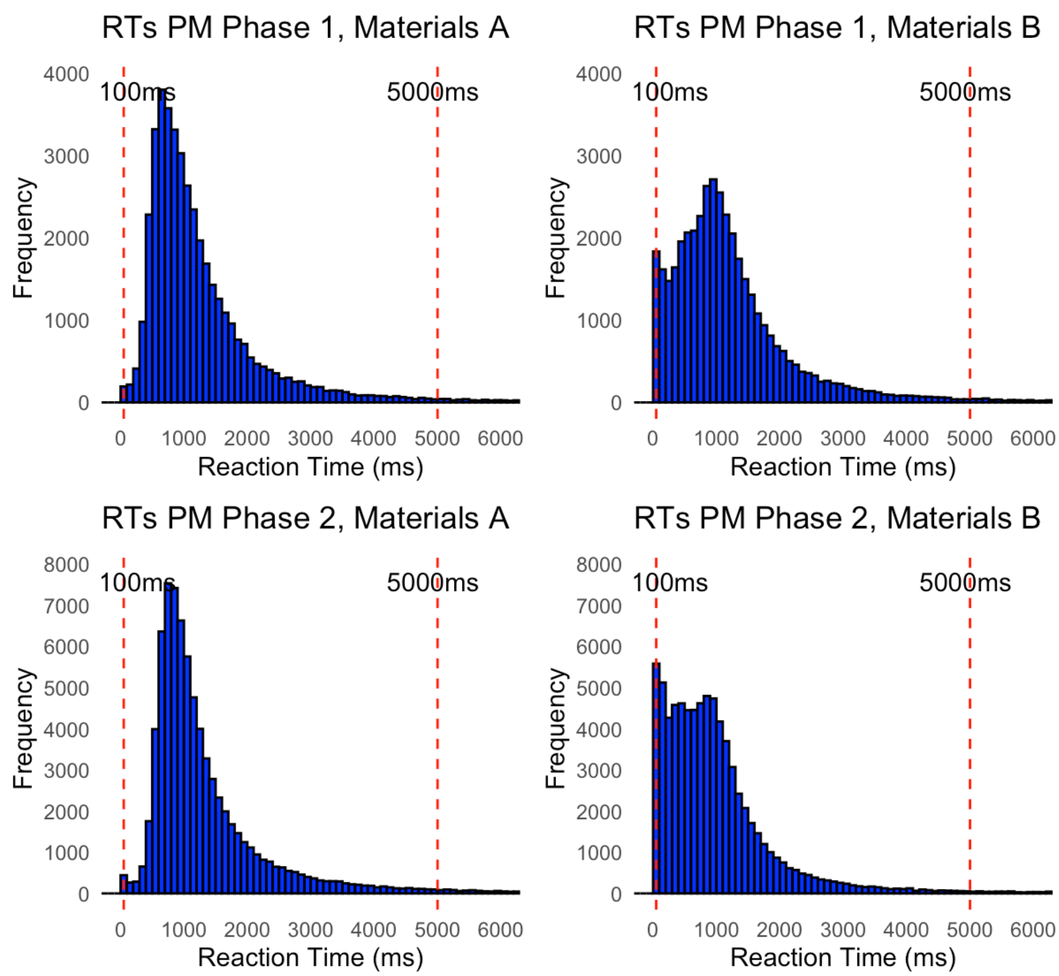
Materials and Procedure. Experiment 3 has a 2 (Reminder Type: Event + Action vs. Event Only) x 2 (Dual PM vs. Offload) design. The Event Only Reminder Type group matched Experiment 2a. (Participants in the Dual PM condition were given two PMs, and participants in the Offload condition were given two PMs along with a reminder for one of them. For the reminder, the cue items were presented in bold font.) For the Event + Action Reminder Type, the materials had a single modification to the reminder in the Offload condition (The Dual condition remained the same as Experiment 2a). For both the Practice and PM Phase 1, cues with reminders were presented with the critical orthographic features in bold and underlined (e.g., ba**ll**ot). In addition, under the cue were the instructions for the to-be-performed action, which read "Press Spacebar Now" in large red letters.

Results

Data Cleaning. Of the 280 participants who finished the practice phase, 37 were removed. Of those, 18 were from the Dual PM first condition, and 19 were from the Offload first condition. Of the participants that were removed, 13 did not complete the experiment, and 24 did not respond correctly to at least half of the attention checks.

Next, outlier reaction times were removed. RTs were filtered based on the same task-specific expectations as prior experiments to include only RTs between 100 ms and 5000ms, as shown in Figure 14.

Figure 14. Experiment 3 Reaction Times.



Note. Histogram represents reaction times pooled across all conditions. Outlier cutoffs indicated by red dashed lines.

This eliminated 6.1 percent of the total RTs. A Shapiro-Wilk test of the filtered data, which included a sampling of 5000 RT data points, indicated that those data may not be normally distributed for PM Phase 1 Materials A, $W = .83$, $p < .001$;

for PM Phase 1 Materials B, $W = .88$, $p < .001$; for PM Phase 2 Materials A, $W = .82$, $p < .001$; for PM Phase 2 Materials B, $W = .86$, $p < .001$.

Next, participants were removed who did not correctly score at least at chance (.50) on the ongoing categorization task. This removed 10 additional participants, 6 from the Dual PM first condition, 4 from the Offload first condition leaving 113 from the Dual PM first condition and 120 from the Offload PM first conditions in the filtered data to be analyzed.

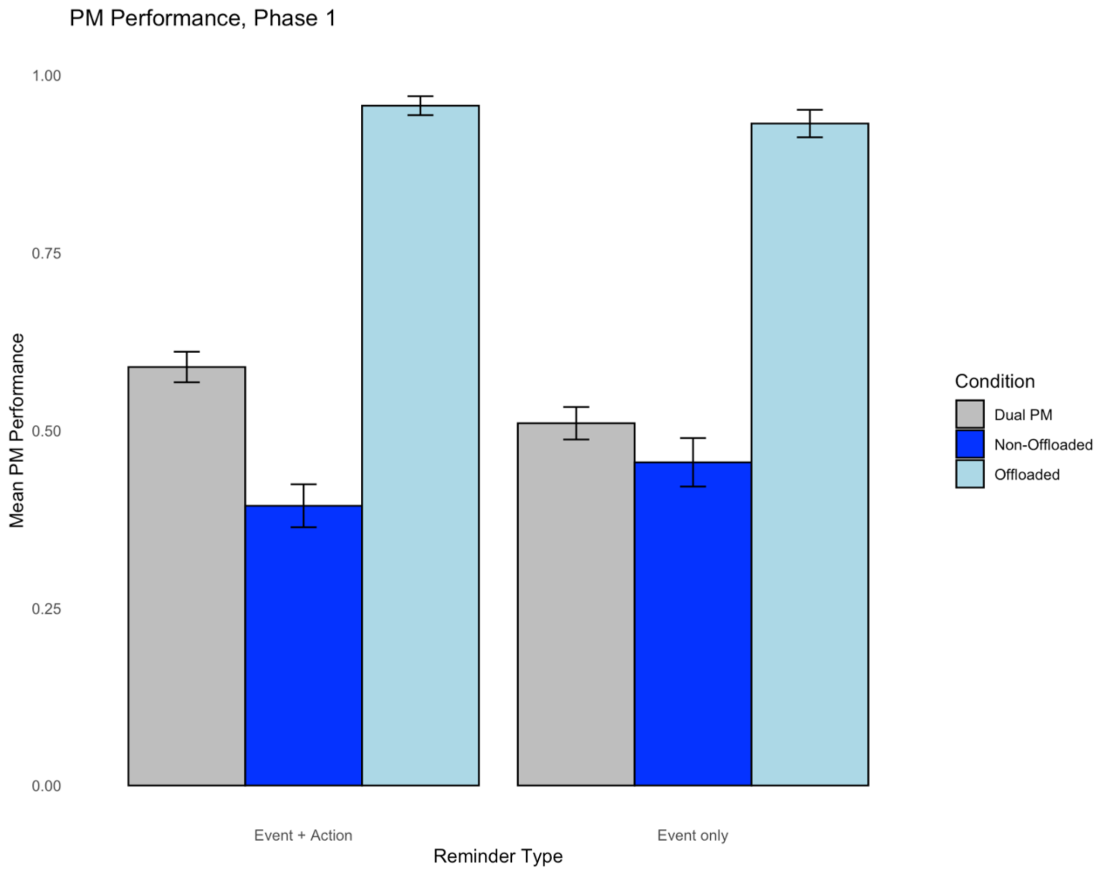
PM Phase 1.

Prospective Memory. Here and in Phase 2, in the Dual PM condition, PM performance is reported separately for each of the two PMs. The first PM is compared to the non-offloaded PM and the second is compared to the offloaded PM. With the order of the PMs counterbalanced, the Dual PM comparison to both the non-offloaded and offloaded PM both PMs contains both PM tasks (i.e., double letter and 3 syllable). First, a 2(Reminder Type: Event + Action vs. Event Only) x 2(Condition: Dual vs. Offload) mixed ANOVA was performed for performance of the non-offloaded PM. There was a significant interaction effect between reminder type and condition, $F(1, 230) = 4.64$, $p = .032$, $\eta_p^2 = .02$. The analysis also revealed a significant effect of condition, $F(1, 230) = 15.79$, $p < .001$, $\eta_p^2 = .06$. However, the main effect of reminder type was not significant, $F(1, 230) = .24$, $p = .625$, $\eta_p^2 < .001$. With a significant interaction, participants in each reminder type were evaluated separately.

For those with the Event + Action reminder, the mean performance of the non-offloaded PM in the Dual PM condition ($M = .59$, $SE = .03$) was significantly higher than in the Offload condition ($M = .41$, $SE = .03$), $t(122) = 4.51$, $p < .001$, $d = .41$, 95% CI [.22, .59]. The mean PM performance of the offloaded PM was .96 ($SE = .01$). A similar pattern was observed in a between-subjects evaluation of participants using only Materials A, which was performed to avoid possible effects of the catch trial the Offload condition. The mean performance of the non-offloaded PM in the Dual PM condition ($M = .68$, $SE = .04$) was significantly higher than the Offload condition ($M = .42$, $SE = .05$), $t(121) = 4.54$, $p < .001$, $d = .82$, 95% CI [.45, 1.19].

For those with the Event Only reminder, the mean performance of the non-offloaded PM in the Dual PM condition ($M = .51$, $SE = .03$) not significantly different from the Offload condition ($M = .46$, $SE = .03$), $t(108) = 1.24$, $p = .218$, $d = .12$, 95% CI [-.07, .31]. The mean PM performance of the offloaded PM was .93 ($SE = .01$). Again, a between-subjects evaluation of participants using only Materials A revealed similar results. The mean performance of the non-offloaded PM in the Dual PM condition ($M = .58$, $SE = .05$) was not significantly higher than in the Offload condition ($M = .52$, $SE = .04$), $t(107) = .90$, $p = .368$, $d = .17$, 95% CI [-.20, .55]. See Figure 15.

Figure 15. Experiment 3 Prospective Memory Performance, PM Phase 1.



Note. In this figure, PM performance for the Dual PM condition was averaged across both counterbalance conditions. The Non-Offloaded and Offloaded PM performance was from participants in the Offload condition.

Reaction Times. In the Event + Action group, in the Dual PM condition participants had average RTs of 1324 ms (*SE* = 41), and in the Offload condition the mean reaction time was 1170 ms (*SE* = 43). In the Event Only group, participants in the Dual PM condition had a mean RT of 1238 ms (*SE* = 44) and in the Offload condition, the mean RT was 1141 ms (*SE* = 30). A 2(Reminder Type: Event + Action vs. Event Only) x 2(condition: Dual vs. Offload) ANOVA was performed on RT.

This revealed that the interaction between reminder type and condition was not significant, $F(1, 230) = 1.15$, $p = .286$, $\eta_p^2 = .01$. There was a significant main effect of condition, $F(1, 230) = 22.18$, $p < .001$, $\eta_p^2 = .09$. There was not a significant main effect of Reminder Type, $F(1, 230) = 1.38$, $p = .241$, $\eta_p^2 = .01$.

Next, a linear regression analysis was conducted to assess the impact of RT on non-offloaded PM performance, including the factor of condition (Dual PM vs. Offload) as well as reminder type (Event + Action vs. Event Only). The model was significant for predicting PM performance, $F(4, 459) = 5.77$, $p < .001$, (Adjusted R^2 = .04). Controlling for condition and reminder type, RT was a significant predictor of PM performance ($b = .00014$, $SE = .00005$, p = .006), indicating that longer RTs were associated with better PM performance. Controlling for RT and reminder type, the main effect of condition was not significant ($b = .034$, $SE = .097$, $p = .729$). Controlling for condition and RT, the main effect of reminder type was also not significant ($b = -.005$, $SE = .032$, $p = .878$). The interaction between RT and condition was not significant ($b = -.00012$, $SE = .00008$, $p = .110$). These results suggest that while RT is a significant predictor of PM performance, neither the condition nor reminder type significantly affects PM performance in this model.

Ongoing Categorization Performance. Participants who received an Event + Action Reminder in the Dual PM condition had a mean ongoing categorization score of .88 ($SE = .01$). In the Offload condition, those with an Event + Action Reminder had a mean ongoing categorization score of .84 ($SE = .01$). For participants who received an Event Only Reminder, the mean ongoing categorization score was .88

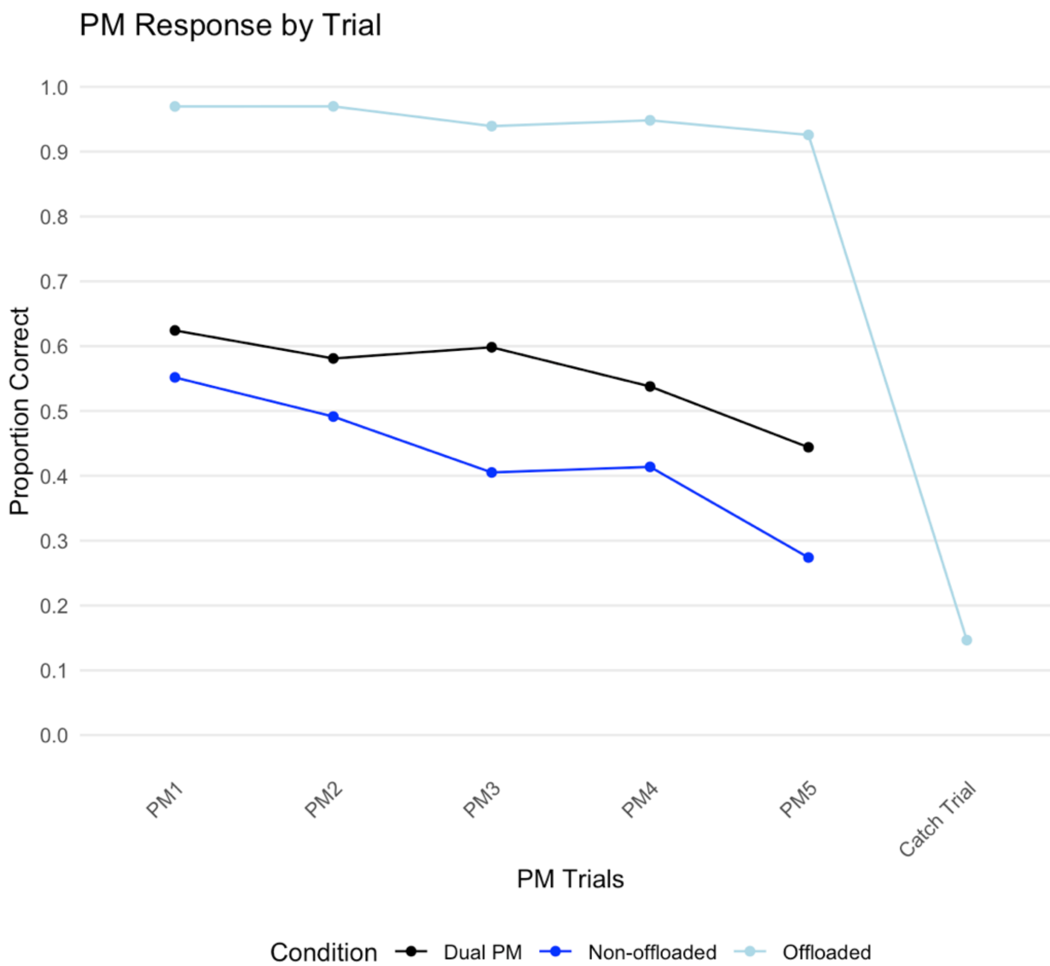
(*SE* = .01) in the Dual PM condition and .84 (*SE* = .01) in the Offload condition. A 2(Reminder Type: Event + Action Reminder vs. Event Only) x 2(Condition: Dual PM vs. Offload) ANOVA was performed for ongoing categorization score. The interaction effect between reminder type and condition was not significant, $F(1,228) = 1.31$, $p = .254$, $\eta_p^2 = .006$. There was no significant main effect of reminder type, F(1,228) = .43, $p = .514$, $\eta_p^2 < .002$ There was a significant main effect of condition, $F(1,228) = 16.47$, $p < .001$, $\eta_p^2 = .07$, with participants in the Dual PM condition showing significantly higher scores compared to the Offload condition.

Catch Trial. For participants who received the Event + Action reminder, for the non-offloaded PM, performance varied across trials: PM1 (*M* = .54, *SE* = .04), PM2 (*M* = .48, *SE* = .05), PM3 (*M* = .36, *SE* = .04), PM4 (*M* = .36, *SE* = .04), and PM5 (*M* = .26, *SE* = .04), with all trials showing higher PM performance than the catch trial (*M* = .15, *SE* = .03). PM1: $t(123) = 8.18$, $p < .001$, $d = .73$, 95% CI [.54, .93], PM2: $t(123) = 6.93$, $p < .001$, $d = .62$, 95% CI [.43, .82], PM3: $t(123) = 4.17$, $p < .001$, $d = .38$, 95% CI [.19, .56], PM4: $t(123) = 4.28$, $p < .001$, $d = .39$, 95% CI [.20, .57], PM5: $t(123) = 2.30$, $p = .012$, $d = .21$, 95% CI [.03, .39].

For participants who received the Event Only reminder, for the non-offloaded PM, performance varied across trials: PM1 (*M* = .55, *SE* = .05), PM2 (*M* = .50, *SE* = .05), PM3 (*M* = .45, *SE* = .05), PM4 (*M* = .47, *SE* = .05), and PM5 (*M* = .29, *SE* = .04), with all trials showing higher PM performance than the catch trial (*M* = .14, *SE* = .03). PM1: $t(108) = 7.05$, $p < .001$, $d = .68$, 95% CI [.47, .88], PM2: $t(108) = 6.70$, $p < .001$, $d = .64$, 95% CI [.44, .85], PM3: $t(108) = 6.72$, $p < .001$, $d = .64$, 95% CI [.44,

.85], PM4: $t(108) = 6.76$, $p < .001$, $d = .64$, 95% CI [.44, .85], PM5: $t(108) = 3.03$, $p =$ .002, $d = .29$, 95% CI [.10, .48]. See Figure 16.

Figure 16. Experiment 3 Trial by Trial Analysis.



Note. In this figure, performance is combined across reminder types. PM performance for the Dual PM condition was averaged across both counterbalance conditions. The Non-Offloaded and Offloaded PM performance was from participants in the Offload condition. Only those in the Offload condition were given a catch trial.

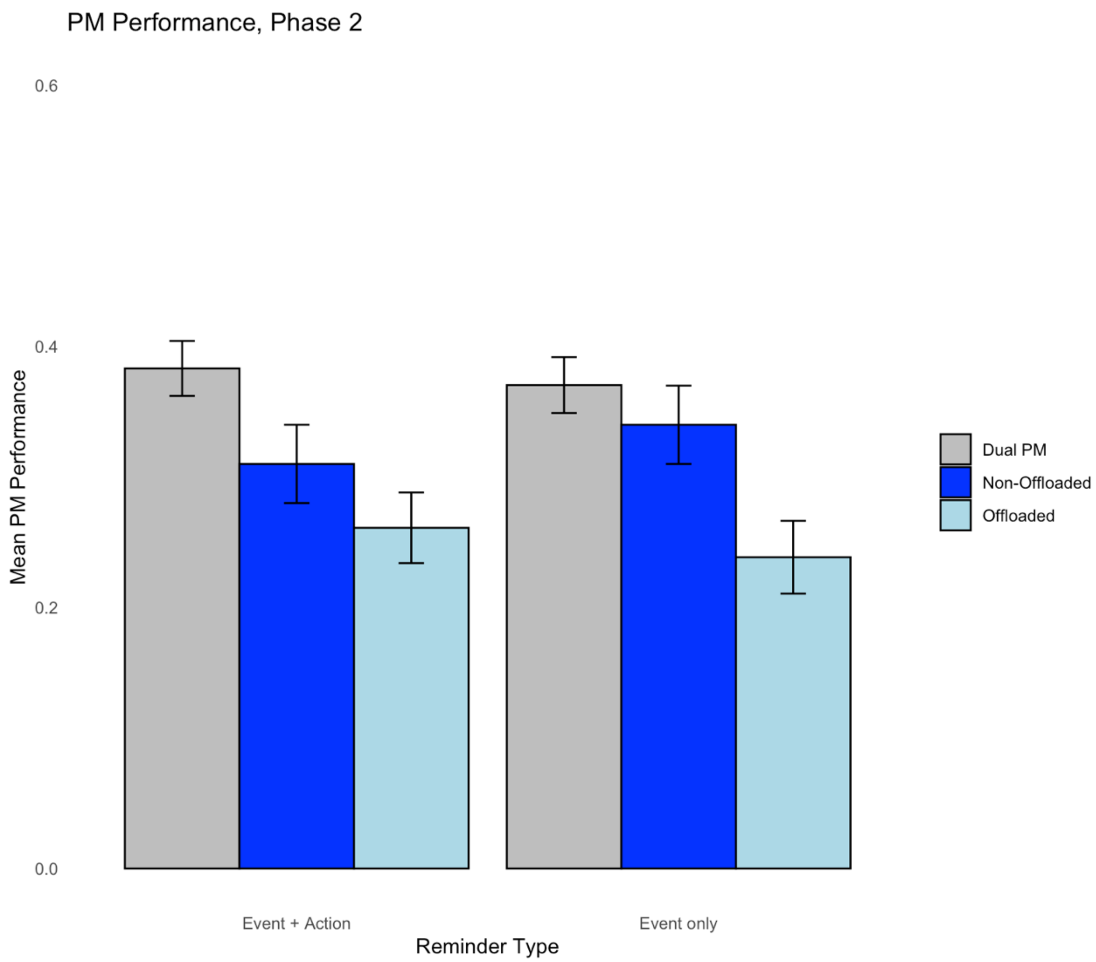
PM Phase 2.

Prospective Memory. For this analysis, a series of ANOVAs was performed. The first ANOVA examined the dependent variable of PM performance for the PM that had *not* been offloaded in PM Phase 1. The following ANOVA examined the dependent variable of PM performance for the PM that *was* offloaded in PM Phase 1. The first 2(Reminder Type: Event + Action vs. Event Only vs.) x 2(Condition: Dual vs. Offload) mixed ANOVA was performed for PM performance for the non-offloaded PM. The interaction between reminder type and condition was not significant, $F(1, 230) = 2.67$, $p = .104$, $\eta_p^2 = .01$. The main effect of condition also not significant, $F(1, 230) = 2.89$, $p = .090$, $\eta_p^2 = .01$. The main effect of reminder type was not significant, $F(1, 230) = .19$, $p = .663$, $\eta_p^2 < .001$.

While the interaction between reminder type and condition did not reach significance, when considering only those receiving the Event + Action reminder, a t-test between conditions revealed in the Dual PM condition, participants had significantly better PM performance for the non-offloaded PM ($M = .40$, $SE = 03$) than in the Offload condition ($M = .31$, $SE = .03$), $t(122) = 2.29$, $p = .024$, $d = .21$, 95% CI [.03, .38]. A similar pattern was observed in a between-subjects of participants evaluation using Materials A. The mean performance of the non-offloaded PM for participants in the Dual PM condition ($M = .56$, $SE = .04$) was significantly higher than participants in the Offload condition ($M = .43$, $SE = .05$), $t(121) = 2.42$, $p = .017$, $d = .44$, 95% CI [.08, .79]. When considering those who received an Event Only reminder, a difference in PM performance of the non-offloaded PM was not observed. Participants had similar PM performance for the non-offloaded PM in the Dual PM condition (*M* = .35, *SE* = .03) compared to the Offload condition (*M* = .34, *SE* = .03), t (108) = .05, *p* = .960, *d* = .01, 95% CI [-.18, .19]. Similar results were observed in a between-subjects evaluation of participants using Materials A. The mean performance of the non-offloaded PM for participants in the Dual PM condition (*M* = .42, *SE* = .05) was not significantly different than in the Offload condition (*M* = .44, *SE* = .04), *t*(107) = .29, *p* = .770, *d* = .06, 95% CI [-.32, .43]. See Figure 17.

Next, another 2(Reminder Type: Event + Action vs. Event Only) x 2(Condition: Dual vs. Offload) mixed ANOVA was performed for PM performance for the Offloaded PM. There was a not significant interaction effect between reminder type and condition, $F(1, 238) = .65$, $p = .422$, $\eta_p^2 = .03$. The main effect of condition was significant, $F(1, 238) = 19.51$, $p < .001$, $\eta_p^2 = .08$. The main effect of reminder type was not significant, $F(1, 238) < .01$, $p = .974$, $\eta_p^2 < .001$. For participants with the Event + Action reminder, PM performance for the previously offloaded PM in the Dual PM condition was .37 (*SE* = .03) and in the Offload condition was .27 (*SE* = .03). Similarly, for participants with the Event Only reminder, PM performance for the previously offloaded PM in the Dual PM condition was .39 (*SE* = .03) and in the Offload condition was .24 (*SE* = .03). See Figure 17.

Figure 17. Experiment 3 Prospective Memory Performance, PM Phase 2.



Note. In this figure, PM performance for the Dual PM condition was averaged across both counterbalance conditions. The Non-Offloaded and Offloaded PM performance was from participants in the Offload condition.

Reaction Times. Next, RTs were evaluated. Of those in the Event + Action reminder type, participants in the Dual PM condition had the mean PM performance of 1230 (*SE* = 38). In the Offload condition, the mean PM performance was 1166 (*SE* = 38). Of those in the Event Only reminder type, for the Dual PM condition, the mean PM performance was 1111 (*SE* = 40). For the Offload condition, the mean PM performance was 1094 (*SE* = 30). A 2(condition: Dual PM vs. Offload) x 2 (Reminder Type: Event + Action vs. Event Only) Mixed ANOVA was done on RTs. The interaction between condition and reminder type was not significant, $F(1, 230) = .48$, $p = .489$, $\eta^2 = .002$. There was not a significant effect of condition, $F(1, 230) = 1.36$, $p = .245$, $\eta^2 = .01$. The main effect of reminder type was significant, $F(1, 230) = 4.87$, $p = .028$, $\eta^2 = .02$.

A linear regression analysis was conducted to assess the impact of RT on non-offloaded PM performance, including the factor of condition (Dual PM vs. Offload) as well as reminder type (Event + Action vs. Event Only). The model was significant for predicting PM performance, $F(4, 459) = 12.07$, $p < .001$, (Adjusted $R^2 = .09$). Controlling for condition and reminder type, RT was a significant predictor of PM performance (b = .00021, SE = .00004, $p < .001$), indicating that longer RTs were associated with better PM performance. Controlling for RT and reminder type, the main effect of condition was not significant (*b* = -.072, *SE* = .082, p = .378). Controlling for RT and condition, the main effect of reminder type was also not significant (*b* = .008, *SE* = .028, *p* = .789). The interaction between RT and condition was not significant (*b* = .00003, *SE* = .00007, *p* = .675). These results suggest that while RT is a significant predictor of PM performance, neither the condition nor reminder type significantly affects PM performance in this model.

Ongoing Categorization Performance. Participants who received an Event + Action reminder had a mean ongoing assessment score of .87 (*SE* = .01) in the both the Dual PM and Offload conditions. For participants who received an Event Only reminder, the mean ongoing assessment score was also .87 (SE = .01) both the Dual PM and Offload conditions. A two-way ANOVA was conducted to examine the effect of reminder type and condition on the ongoing assessment score. The interaction effect between reminder type and condition was not significant, $F(1, 228) = .06$, $p = .803$, $\eta^2 < .001$. There was not a significant main effect of reminder type, $F(1, 228) = .04$, $p = .836$, $\eta^2 < .001$. There was not a significant main effect of condition, $F(1, 228) < .01$, $p = .998$, $\eta^2 < .001$.

Discussion

Experiment 3 added a between-subjects condition to replicate and extend the results of Experiment 2a to include an Event + Action reminder type. A more comprehensive reminder should strengthen any effects of offloading compared to an Event Only reminder such as the one that had been utilized in the previous experiments.

Indeed, in PM Phase 1 of Experiment 3, we observed an interaction effect between PM performance for the non-offloaded PM between the Dual and Offload conditions and reminder type. With the Event Only reminder, non-offloaded PM performance for participants in the Dual condition was numerically higher than for participants in the Offload condition (as was observed in Experiment 2a). However, the larger difference between the Dual PM and Offload conditions in the Event + Action group reached significance (with participants in the Dual PM condition performing better than participants in the Offload condition). In terms of the offloaded PM, participants still had near-ceiling performance (.96 for Event + Action reminders and .93 for Event Only reminders). The direction of this effect was the opposite of what was predicted by the hypothesis that by effectively offloading one PM, participants would reallocate those cognitive resources to remembering the non-offloaded PM. It appeared that there could be a benefit to remembering both PMs internally. These results were consistent with the trends from Experiments 2a and 2b as well as with the results from Experiment 1.

In PM Phase 2, when assessing the PM performance for the offloaded PM, a large reduction in the practice effect was observed for participants in the Offload condition compared to those in the Dual PM condition. This effect was found with both reminder types. This result would be expected if the participants were offloading the to-be-offloaded PM as instructed rather than monitoring for it redundantly.

When assessing the PM performance in PM Phase 2 for the non-offloaded PM, participants with the Event + Action reminder showed a larger (and statistically significant) effect of condition, with those in the Dual PM condition having better PM performance than those in the Offload condition. Participants with the Event Only reminder had similar performance between the Dual PM and Offload conditions.

In PM Phase 1, slower reaction times continued to predict PM performance showing the expected tradeoffs between the PM and ongoing tasks. Participants in the Offload condition had faster reaction times than participants in the Dual PM condition (as was found in Experiment 2a), a difference that did not vary by reminder type.

General Discussion

As we increasingly rely on digital devices to help us remember future intentions, understanding how setting reminders affects prospective memory becomes more important. This series of four experiments explored the effects of offloading on performing multiple PM tasks. The current research consistently failed to find evidence that offloading one prospective memory led to improved performance on a concurrent non-offloaded PM. In fact, offloading appeared to impair performance on non-offloaded PM in some cases. A delayed effect of offloading was also revealed. When participants were later required to perform previously offloaded tasks without reminders, their performance was worse compared to those who had practiced these tasks without offloading.

The first hypothesis was that offloading one of two PM tasks would improve performance on the non-offloaded PM compared to a condition where both PMs were remembered without offloading. This hypothesis was based on the premise that performing multiple PM tasks would require more cognitive effort than performing a single PM and that cognitive resources made available by offloading would be reallocated to monitor for the non-offloaded PM. In Experiment 1, however, we did not observe the expected difference in PM performance between the Single PM and Dual PM conditions. This finding could have been observed because including a second PM task increased the density of the cues presented to each participant, making them less of a rare event. In terms of visual search, increasing the frequency of stimuli leads to improved performance (Wolfe, et al., 2005). Future research may want to match cue density across conditions to control for this variable.

According to Kahneman and Tversky's capacity model of attention (1973), people have a limited pool of cognitive resources that can be flexibly allocated among concurrent tasks. In theory, offloading one PM task should free up cognitive resources which could be reallocated to other tasks. However, across this series of experiments, participants in the Offload condition consistently failed to show any benefit in PM performance for the non-offloaded PM task. This finding was repeated even when the difficulty of the offloaded PM was increased (Experiment 2b). Offloading a more difficult PM should increase the potential benefit of not having to remember that PM. Participants in the Offload condition also failed to show any benefit in PM performance for the non-offloaded PM when more comprehensive reminders were used (Experiment 3). The more comprehensive Event + Action reminder should have allowed participants to more completely offload by providing participants with instructions for the to-be-performed action at the appropriate time. Despite these manipulations, neither approach led to the expected reallocation of cognitive resources to the non-offloaded PM task.

A few possible mechanisms should be explored. Under the assumption that performing a single PM requires less cognitive resources than performing two PMs (although that was not observed in Experiment 1), one possibility is that the cognitive processes involved in managing multiple PMs is less flexible than previously assumed. Offloading one PM may not simply free up resources that can easily be (or are chosen to be) reallocated to other tasks. The benefits of cognitive offloading may also be partially offset by a new set of cognitive demands that weren't otherwise present. For example, when offloading one of two PM tasks, participants had to engage in several meta-level processes. Participants still needed to remember that a task has been offloaded, remember what reminders to look for, maintain awareness of the offloading strategy, and coordinate these cognitive processes. This additional cognitive overhead aligns with Smith's (2016) concept of "meta-awareness" in prospective memory which describes the higher-level cognitive processes to monitor and control how PM tasks are managed.

Another possibility is that offloading may introduce complacency, where the act of offloading one task might lead to a general reduction in attention allocated to all PM tasks. This aligns with the theoretical mechanisms proposed by Soares and Storm (2018) in their study of photo-taking and memory. Soares and Storm investigated the photo-taking-impairment effect, previously described by Henkel (2014), where participants demonstrated worse memory for objects they photograph than for those they simply observed. Henkel proposed that taking a photograph allowed participants to offload that memory to an external store allowing participants not to remember the objects being photographed. However, Soares and Storm (2018) performed a similar experiment, but had participants delete the photographs immediately after capturing them (eliminating the use of the external store for future reference). The photo-taking-impairment effect was still observed under those conditions. Soares and Storm proposed the concept of metacognitive illusion, in which the act of offloading (even when participants were aware that the offloaded items would not be available when needed) created a false sense of fluency. The concept of metacognitive illusion could also provide an explanation for the lack of improvement in non-offloaded PM tasks observed in our experiments. Participants who offloaded one PM task might have experienced a subjective sense of task completion that triggered heuristic-based thinking (Slovic & Tversky, 1982) that extended beyond the offloaded task (which was completed with above 90 percent accuracy in all experiments), leading to reduced effort or cognitive resource allocation for all PM tasks.

More speculatively, metacognitive illusion could provide a theoretical account for why offloading may have even reduced the performance for the non-offloaded PM under some conditions. In Experiment 3, enhancing the offloading manipulation in the Event + Action reminder group made the offloaded PM task easier. If differences in PM performance were driven by fluency of performing the offloaded PM task in the Offload condition, we would predict a larger difference in PM performance between the Dual PM and Offload conditions with the Event + Action reminder. This was exactly what we observed. Similarly, Fellers, Miyatsu, and Storm (2023) revealed that providing reminders for complex PM tasks involving multiple components caused an inflated sense of confidence for aspects of the PM task that did not benefit from the reminder. In the current research, metacognitive illusion may have caused participants to reduce their monitoring efforts for the non-offloaded PM, leading to a decrease in PM performance.

In line with the concept of metacognitive illusion, across Experiments 2a, 2b, and 3, participants in the Dual PM condition showed superior performance in the ongoing categorization task, even while participants in the Dual PM condition trended towards better PM performance (which reached significance in the case of Experiment 3). Prior research has revealed that PM performance generally varies inversely with ongoing task performance as participants are forced to balance finite cognitive resources (Cantarella, 2023; Meier et al., 2015; c.f., Guo, 2023). If participants in the Dual PM condition experienced a metacognitive illusion of their abilities due to the fluency with which they remembered the offloaded information, the effects of the metacognitive illusion may have also been observed in the ongoing task. Indeed, Guo (2023) recently failed to find a performance benefit for the ongoing task when participants offloaded time-based PMs.

The finding that participants did not perform better in either the non-offloaded PM task or the ongoing categorization task in the Offload condition (performance was worse in many cases) could also viewed in terms of the redistribution dimension of Risko et al.'s (in press) offloading framework. According to Risko et al., participants would have to voluntarily redistribute cognitive resources to a new task after offloading. Any saved cognitive resources would not automatically spill over into the remaining tasks but could also be strategically conserved by participants.

Another consideration is that if participants in the Offload condition were replacing the offloaded PM task with a new one (looking for the reminder itself), they may monitor for the reminder itself (although, monitoring for the salient reminder should require fewer cognitive resources than monitoring for the original PM cue). In this case, we would expect participants to perform well on the offloaded PMs since distinct cues PMs improve PM performance (Brandimonte & Passolunghi, 1994; McDaniel & Einstein, 2000; McDaniel & Einstein, 1993). Participants should also perform poorly in the catch trial (as they did) since participants would not be monitoring for the cue without the reminder. However, when the reminder was improved by adding instructions for the to-be-performed action (Experiment 3), a larger decrease in performance for the non-offloaded PM was observed for participants in the Offload condition. An improved reminder should further reduce the cognitive resources required in the Offload condition, allowing for additional cognitive resources to be reallocated to other tasks.

An alternative explanation for the reduced PM performance in the non-offloaded condition challenges the assumption that offloading always benefits cognition. There may be an advantage to remembering two PM tasks rather than one. Participants tasked with remembering two PMs encountered a higher frequency of PM cues within the ongoing task compared to those in the single PM condition. Supporting this idea, Experiment 1 showed numerically higher PM performance for participants in the Dual PM condition compared to the Single PM condition.

Offloading one of the two PM cues might, therefore, inadvertently hinder performance on the non-offloaded PM task. Rather than benefiting performance by reducing cognitive load, offloading may function to decrease the effective cue density. When participants only need to monitor for half of the cues, the frequency of internally remembered cues is halved. This reduction transforms the non-offloaded PM cue into a rarer event, potentially making it more difficult to detect and respond to effectively.

In this light, offloading may not be reducing the costs of remembering two PMs internally, but rather diminishing the benefits. This perspective suggests that the cognitive advantages of maintaining multiple PM tasks might outweigh the presumed benefits of offloading, while simultaneously gaining exposure to potential issues of meta-level processes involved in offloading or cognitive-induced complacency.

The second hypothesis was that offloading PMs is inefficient and predicted worse performance on the non-offloaded PM compared to a single PM condition, assuming that the Dual PM task would be more difficult than the Single PM task (which was not observed in Experiment 1). This result was observed when tested in Experiment 1. The hypothesis predicted a difference in PM performance would be observed between participants in the Single PM and Offload conditions because participants would redundantly monitor for the offloaded PM, and cognitive resources would be split between these two monitoring tasks. Even after offloading one PM, participants may have (either implicitly or explicitly) continued to allocate cognitive resources towards monitoring for the offloaded PM. Indeed, Murphy (2021) revealed that participants who offloaded, were still able to recall some of the offloaded items when the external stores were expectedly unavailable (even though their memory for the offloaded items was worse than for the non-offloaded items). In context of the offloading framework put forward by Risko el al. (in press), these results could be described as duplicative behavior in the substitutive-duplicative dimension, perhaps due to some perception of poor reliability of the external memory store.

The reduced performance for the non-offloaded PM compared to the PM performance in the Single PM condition would align with Smith's Preparatory Attentional and Memory process (PAM) theory (2003) which posits that successful PM relies on a resource demanding attentional process which is required to keep a state of readiness to perform the intended action. According to Smith (2003) some degree of attentional resources is always devoted to monitoring for PM cues, no matter how salient those cues may be. Presumably, even when a PM task is offloaded, the PAM theory would suggest that some attentional resources would still be allocated to monitoring for the offloaded task. In Experiments 1, 2a, and 2b, PAM theory would also suggest that cognitive resources would still need to be allocated to remembering the intended action even when a reminder is used to alert participants of the cue event.

In the case of a PM that was not offloaded with complete efficiency, however, PM performance for the non-offloaded PM would be expected to fall between performance in a Single PM condition (which should function similarly to having 2 PMs and completely offloading one) and in a Dual PM condition (having 2 PMs and not offloading at all). In the current research, participants had worse PM performance for the non-offloaded PM compared to the Single PM, although the performance for the non-offloaded PM was not within the expected range between the Single PM and Dual PM conditions. Reductions in non-offloaded PM performance compared to the

Single PM condition could still be driven by meta-level processes involved in offloading or cognitive-induced complacency.

If participants had a reduced performance in the for the non-offloaded PM for other reasons (e.g., metacognitive illusion or offloading-induced complacency), the current series of experiments would not be able to tease apart these mechanisms. Future research may want to manipulate perceived reliability of the reminders. If participants receiving reminders that were believed to be highly reliable have worse performance for the non-offloaded PM, it would provide evidence to support metacognitive illusion theory.

The third hypothesis predicted that participants would show worse performance on a PM task in PM Phase 2 when offloading was no longer available. This decreased performance was expected due to the lack of practice in monitoring for the PM cue. This hypothesis was supported by the results of both Experiment 2a and Experiment 3. While the practice effect has been well described in terms of retrospective memory (Roediger & Karpicke, 2006), it has only more recently been tested in terms of prospective memory (Guo et al., 2023; Brom & Kliegel, 2014) and mostly in the context of time-based cues.

Interestingly, a between-phase practice effect was not observed in any of the current experiments, even for participants in the Dual PM condition who performed the same prospective memory task in PM Phases 1 and 2. One possible explanation for this is the reduction in cue density in PM Phase 2. In Experiments 2a, 2b, and 3, Phase 1 comprised 160 trials with 10 total PM cue trials (5 for each PM task), resulting in a cue frequency of approximately 1 in 16 trials. Phase 2 included 320 trials but maintained the same number of cue trials, effectively halving the cue density to about 1 in 32 trials. This design aimed to reduce the duration of the experiment which was increased with an addition phase as well as by updating to a within-subjects design, but introduced a difference in cue density between phases complicates direct PM performance comparisons. With consistent cue density, we would have expected a learning effect in the Dual PM condition, with improved performance in Phase 2. This further highlights questions about the impact of cue density on PM performance discussed in the other hypotheses. Future research would benefit from maintaining a consistent cue density across phases and conditions to better isolate the effects of learning and offloading on PM performance.

The effect of setting reminders is particularly important in the context of remembering to perform important tasks in daily life. For example, people may rely on reminders to take medications before going to bed. Although using reminders for such activities enhances task completion (Henry et al., 2012), offloading may inadvertently diminish the practice effect that would otherwise strengthen PM performance over time.

In the absence of reminders, the successful execution of the PM task demands self-initiated cognitive effort, which is reduced with the use of a reminder (Huang et al., 2014). The current set of experiments provides evidence that offloading an event-based PM task to a reminder reduces the practice effect that would typically enhance future PM performance. Participants who initially relied on reminders showed reduced performance compared to those in the Dual PM condition who practiced the PM without offloading in Phase 1. This suggests that while reminders can be beneficial for immediate task completion, they may hinder the practice effects that would otherwise strengthen PM learning.

The reduced practice effect for participants in the Offload condition was not observed in Experiment 2b when four PMs were given and two were offloaded. One possibility for why this was not observed when two of four PMs were offloaded is that the additional complexity of the task made it more difficult to offload efficiently. By providing a more complicated set of tasks to remember and offload, participants may not have been able to be as precise with their monitoring and offloading strategy. Indeed, in the catch trial, participants correctly responded 23 percent of the time when the to-be-offloaded PM cues were presented without the expected reminder (numerically higher than in the other experiments using only two PMs). This revealed that participants maintained some level of monitoring for the offloaded PM. While we had initially expected that increasing the complexity of the offloaded PM would increase potential benefits from offloading, it appeared that increasing the complexity may have unexpectedly led to conditions where it was more difficult to keep track of which PMs needed to remember internally. Future research may consider having participants offload three PMs and remember one critical non-offloaded PM internally. This would reduce the complexity of the instructions while further extending any potential benefits of offloading.

In the catch trial analysis, participants in the Offload condition performed worse in the catch trial than in any of the prior PM trials for the non-offloaded PM for all four experiments. This result provided evidence that participants were in offloading as instructed (at least to some degree) and aligned with recent findings in prospective memory research. Specifically, Dupont et al. (2023), conducted an experiment where participants were given a surprise test in which offloading was unexpectedly unavailable. Participants who had been relying on offloading performed poorly when suddenly required to rely on internal memory processes. This finding supports the idea that external reminders reduce internal monitoring (Risko & Gilbert, 2016).

However, it would have been valuable to include a comparable control in the Dual PM condition (a sixth PM trial) to assess the impact of offloading more robustly. Such a control could help distinguish between the effects of offloading and other factors that might influence performance over time. The trial-by-trial analysis revealed that, in general, PM performance declined across trials. This trend aligns with resource depletion theory which posits that sustained cognitive effort leads to diminished performance over time (Muraven et al., 1998). Motivation may also wane as tasks progress, particularly in the absence of feedback or renewed goal emphasis. Future studies would benefit from incorporating matched control trials in the Dual PM condition to ensure more direct comparability and to better isolate the effects of offloading from those of task duration and repetition.

Evaluating reaction times can provide insight into how participants allocate cognitive resources between tasks. Across all four experiments, reaction time was predictive of PM performance. Consistent with prior research, this provided evidence for the expected tradeoff between cognitive resource allocation between the ongoing and PM tasks (Smith, 2003; Smith & Bayen, 2004). The differences in reaction times between conditions, however, were somewhat inconclusive. Experiments 1 and 2b showed similar reaction times between participants in the Offload and Dual conditions in PM Phase 1, while Experiments 2a and 3 showed faster reaction times in the Offload compared to the Dual PM condition. Faster reaction times for participants in the Offload condition would be expected if offloading made remembering that PM easier, even if participants maintained some level of monitoring for the reminder itself or monitored redundantly.

On the face, the trend of faster reaction times in the Offload condition is consistent with participants having lower PM performance for the non-offloaded PM. In general, when participants allocated fewer cognitive resources toward the PM task, more would be available for the ongoing task, which would be measured by faster reaction times (Smith, 2003). However, this interpretation was confounded by the fact that offloading one of the two PMs should also reduce the cognitive load required for the pair of PM tasks. Reaction time measurements gauged cognitive resources allocated to the PMs together (remembering both PMs compared to remembering both while offloading one) but couldn’t measure how resource allocation compared between the non-offloaded PM and corresponding PM in the dual condition.

Cognitive offloading has become increasingly prevalent in our daily lives with the ubiquity of digital technologies. The growing reliance on external memory raises important questions about how offloading affects our cognitive abilities, both in the short term and over extended periods. Understanding the interplay between internal cognitive processes and external aids will be important for optimizing human-technology interaction in an increasingly digital world.

The methodology of the current research, which incorporated offloading some PMs while remembering others internally, pushes the literature towards more realistic scenarios. In doing so, it helped to close a gap in the literature which predominantly focuses on single PM situations. The current research demonstrated yet another set of conditions where reminders were effective for improving immediate PM performance. However, setting reminders prevented individuals from benefitting from the practice effect that they would have experienced if PMs were not offloaded. Participants who offloaded consistently had worse performance on the offloaded PM after a delay, demonstrating that offloading also has the potential to undermine PM learning.

This research also suggests that offloading in prospective memory may function differently than in retrospective memory (Storm & Stone, 2015). The act of offloading seems to affect the non-offloaded PM task in ways not previously described in the literature. This challenges the prediction that offloading simply frees up cognitive resources for easy reallocation to other tasks. Instead, there appears to be a more complex interaction between offloading and cognitive resource allocation.

This complexity may stem from limited flexibility in PM-related cognitive processes, hidden cognitive costs associated with offloading (such as meta-level monitoring), or metacognitive illusions.

Offloading is not a simple subtraction of a PM task. The relationship between having two PM tasks and offloading one is more nuanced than the difference between managing one or two PM tasks. The current findings expanded our understanding of how offloading effects PM performance and highlighted the need for further research into its cognitive mechanisms and implications for remembering in daily life.

References

Benjamin, A. S. (2006). The effects of list-method directed forgetting on recognition memory. *Psychonomic bulletin & review*, *13*, 831-836.

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, *2*(59-68).

Brandimonte, M. A., & Passolunghi, M. C. (1994). The effect of cue-familiarity, cue-distinctiveness, and retention interval on prospective remembering. *The Quarterly Journal of Experimental Psychology*, *47*(3), 565-587.

Brom, S. S., & Kliegel, M. (2014). Improving everyday prospective memory performance in older adults: comparing cognitive process and strategy training. *Psychology and aging*, *29*(3), 744.

Cantarella, G., Mastroberardino, S., Bisiacchi, P., & Macaluso, E. (2023). Prospective memory: the combined impact of cognitive load and task focality. *Brain Structure and Function*, *228*(6), 1425-1441.

Chan, J. C., & McDermott, K. B. (2007). The testing effect in recognition memory: a dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(2), 431.

Dupont, D., Zhu, Q., & Gilbert, S. J. (2023). Value-based routing of delayed intentions into brain-based versus external memory stores. *Journal of Experimental Psychology: General*, 152(1), 175.

Einstein, G. O., Smith, R. E., McDaniel, M. A., & Shaw, P. (1997). Aging and prospective memory: the influence of increased task demands at encoding and retrieval. *Psychology and aging*, *12*(3), 479.

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, memory, and cognition*, *16*(4), 717.

Einstein, G. O., McDaniel, M. A., Thomas, R., Mayfield, S., Shank, H., Morrisette, N., & Breneiser, J. (2005). Multiple processes in prospective memory retrieval: factors determining monitoring versus spontaneous retrieval. *Journal of Experimental Psychology: General*, *134*(3), 327.

Fellers, C., Miyatsu, T., & Storm, B. C. (2023). Remembering what to do when the time comes: The effects of offloading in a complex prospective memory task. *Journal of Experimental Psychology: Applied*, *29*(3), 631.

Gilbert, S. J. (2015). 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.

Guo, Y., & Huang, X. (2021). Time-based prospective memory has plasticity in behavior under different monitoring conditions. *Current Psychology*, *40*(7), 3386-3392.

Guo, Y., Gan, J., Ping, Y., Song, T., Liu, T., Wang, E., & Li, Y. (2022). Effective external reminders impair the practice effect of time-based prospective memory. *International Journal of Psychology*, *57*(3), 372-376.

Gurol-Urganci, I., de Jongh, T., Vodopivec-Jamsek, V., Atun, R., & Car, J. (2013). Mobile phone messaging reminders for attendance at healthcare appointments. *Cochrane database of systematic reviews*, (12).

Guynn, M. J., McDaniel, M. A., & Einstein, G. O. (1998). Prospective memory: When reminders fail. *Memory & Cognition*, *26*(2), 287-298.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237.

McDaniel, M. A., & Einstein, G. O. (1993). The importance of cue familiarity and cue distinctiveness in prospective memory. *Memory*, 1(1), 23-41.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *14*(7), S127-S144.

Haynes, B. I., Kliegel, M., Zimprich, D., & Bunce, D. (2018). Intraindividual reaction time variability predicts prospective memory failures in older adults. *Aging, Neuropsychology, and Cognition*, *25*(1), 132-145.

Henkel, L. A. (2014). Point-and-shoot memories: The influence of taking photos on memory for a museum tour. *Psychological science*, *25*(2), 396-402.

Henry, J. D., Rendell, P. G., Phillips, L. H., Dunlop, L., & Kliegel, M. (2012). Prospective memory reminders: A laboratory investigation of initiation source and age effects. *Quarterly journal of experimental psychology*, *65*(7), 1274-1287.

Hu, B., & Feng, C. (2013). The different effects of PM types, motivation, and task sequence on prospective memory. *Acta Psychologica Sinica*.

Huang, T., Loft, S., & Humphreys, M. S. (2014). Internalizing versus externalizing control: different ways to perform a time-based prospective memory task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(4), 1064.

Jacoby, L. L., & Kelley, C. M. (1992). Unconscious influences of memory: Dissociations and automaticity. In E. A. D. Milner, & E. M. D. Rugg (Eds.), *The neuropsychology of consciousness* (pp. 201–233). Academic Press.

Jolicoeur, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology/Revue canadienne de psychologie*, *42*(4), 461.

Jones, W. E., Benge, J. F., & Scullin, M. K. (2021). Preserving prospective memory in daily life: A systematic review and meta-analysis of mnemonic strategy, cognitive training, external memory aid, and combination interventions. *Neuropsychology*, *35*(1), 123.

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

Lavie, N. (1995). Perceptual load as a necessary condition for selective attention. *Journal of Experimental Psychology: Human perception and performance*, *21*(3), 451.

MacLeod, C. M. (1975). Long-term recognition and recall following directed forgetting. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(3), 271.

McBride, D. M., Beckner, J. K., & Abney, D. H. (2011). Effects of delay of prospective memory cues in an ongoing task on prospective memory task performance. *Memory & cognition*, *39*, 1222-1231.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *14*(7), S127-S144.

Meier, B., & Zimmermann, T. D. (2015). Loads and loads and loads: The influence of prospective load, retrospective load, and ongoing task load in prospective memory. *Frontiers in human neuroscience*, 9, 322.

Muraven, M., Tice, D. M., & Baumeister, R. F. (1998). Self-control as a limited resource: regulatory depletion patterns. *Journal of personality and social psychology*, *74*(3), 774.

Murphy, D. H. (2023). Strategic offloading: How the value of to-be-remembered information influences offloading decision-making. *Applied Cognitive Psychology*, acp.4051.

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

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

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in cognitive sciences*, *20*(9), 676-688.

Risko, E. F., Kelly, M. O., Lu, X., & Pereira, A. (in press). Varieties of Offloading Memory: A Framework. In Q. Wang & A. Hoskins, *The Remaking of Memory in the Age of Social Media and the Internet.* Oxford University Press.

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in cognitive sciences*, 15(1), 20-27.

Roediger III, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological science*, 17(3), 249-255.

Runge, Y., Frings, C., & Tempel, T. (2019). Saving-enhanced performance: Saving items after study boosts performance in subsequent cognitively demanding tasks. *Memory*, *27*(10), 1462-1467.

Scullin, M. K., McDaniel, M. A., & Shelton, J. T. (2013). The Dynamic Multiprocess Framework: Evidence from prospective memory with contextual variability. *Cognitive psychology*, *67*(1-2), 55-71.

Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge University Press.

Smith, R. E. (2016). 12 Prospective Memory: A Framework for Research on Metaintentions. *The Oxford handbook of metamemory*, 217.

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.

Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 756.

Soares, J. S., & Storm, B. C. (2018). Forget in a flash: A further investigation of the photo-taking-impairment effect. *Journal of Applied Research in Memory and Cognition*, *7*(1), 154-160.

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *science*, *333*(6043), 776-778.

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.

Tam, H. L., Wong, E. M. L., Cheung, K., & Chung, S. F. (2021). Effectiveness of text messaging interventions on blood pressure control among patients with hypertension: systematic Review of Randomized Controlled Trials. *JMIR mHealth and uHealth, 9*(9), e24527.

Tran, N., Coffman, J., Sumino, K., Cabana, M.D. (2019). Patient reminder systems and asthma medication adherence: a systematic review. *Journal of Asthma, 51*(5), 536-543.

Graf, P., & Uttl, B. (2001). Prospective memory: A new focus for research. *Consciousness and cognition*, 10(4), 437-450.

Wegner, D. M., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of personality and social psychology*, *61*(6), 923.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, 435(7041), 439-440.

Woodward Jr, A. E., Bjork, R. A., & Jongeward Jr, R. H. (1973). Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior*, 12(6), 608-617.