2023

Great Expectations: Anticipating a Reminder Influences Prospective Memory Encoding and Unaided Retrieval

Philip Peper

GREAT EXPECTATIONS: ANTICIPATING A REMINDER INFLUENCES PROSPECTIVE MEMORY ENCODING AND UNAIDED RETRIEVAL

by

Philip Peper

DISSERTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Experimental Psychology at the University of Texas at Arlington
May 2023

Arlington, Texas

Supervising Committee:
B. Hunter Ball, Supervising Professor
Tracy Greer
Daniel Levine
Matthew K. Robison
George Siemens

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I thank my Supervising Committee for the feedback on this project. Special thanks to my Supervising Professor, Hunter Ball, for his critical feedback and support that shaped me into the scientist I am today. My friends in the program – Cassie Argenbright, Rylee Linhardt, and Durna Alakbarova – provided emotional, technical, and theoretical advice that was a great help. Connor Dupre – research assistant, lab manager, and my soon-to-be replacement – allowed me to focus on this dissertation by keeping our lab running smoothly. I lastly thank all the wonderful research assistants over the years that assisted in data collection, with special thanks to Anisha Sharma and Kelly Pham for collecting most of the data included in this project.

DEDICATION

I dedicate this dissertation to those I love most. All of this would not have been possible if it were not for my parents, Rick and Lori Peper, and my sisters, Nici and Annie Peper, whose constant support and belief in me helped me through my toughest years. And McCall Lemmons, who was a constant source of joy and positivity and who always listened patiently. Lastly, I dedicate this to my best friends Richard Busker, Dillon Holm, Joe Poulin, Chris Halvorson, Aaron Johnson, Andy Mitchell, Dan Jeffries, Shane Solinger, and Bob Anfinson who kept me laughing all this time.

LIST OF FIGURES

Figure Page

Figure Page

LIST OF TABLES

ABSTRACT

Great Expectations: Anticipating A Reminder Influences Prospective Memory Encoding and Unaided Retrieval

Philip Peper, Ph.D.

The University of Texas at Arlington, 2023

Supervising Professor: B. Hunter Ball

Reminders are effective ways to improve prospective memory (PM) – our ability to remember to complete a future action – but reminder use may have unintended consequences. Recent work in retrospective memory has shown that expecting a reminder reduces unaided memory retrieval by reducing encoding effort (i.e., encoding effort hypothesis). However, previous research in PM varies as to whether encoding effort influences PM retrieval. We measured study duration (Experiment 1) and pupil size (Experiment 2), and manipulated depth of processing (Experiment 3) at encoding to examine whether encoding effort influences PM retrieval (i.e., encoding effort hypothesis). Across all experiments, we had participants complete four PM task blocks followed by a recognition memory task. Two reminder conditions had reminders for the first three blocks, but not on the fourth. A no reminder control never had a reminder. Critically, a non-expecting reminder condition was told they would *not* have a reminder prior to encoding targets in the fourth block, while an expecting reminder condition was told they would have a reminder. The encoding effort hypothesis was supported by showing expecting a reminder in the fourth block reduced unaided PM retrieval and target recognition (Experiments 1 and 2) and deep processing at encoding improved unaided PM retrieval and target recognition while negating the effect of expecting a reminder (Experiment 3). Our results suggest reminder expectations reduce encoding effort, and greater encoding effort improves unaided retrieval, but having reminders at retrieval offsets the negative effects of reduced encoding effort. We propose the PM Effort Monitoring and Control Framework that describes how when participants experience the low effort and effectiveness of retrieval with reminders, this awareness leads to a less effortful encoding strategy when they expect another reminder during a subsequent PM task

Great Expectations: Anticipating A Reminder Influences Prospective Memory Encoding and Unaided Retrieval

Prospective memory (PM) is the process of remembering to complete a planned action at the appropriate time in the future, such as taking a medication with breakfast. Successful functioning in the modern world requires managing countless intentions over a given week. For example, one must be able to juggle work deadlines, social plans, doctor's appointments, and take medications. While previous research has shown that effortful PM encoding strategies lead to better PM retrieval (strategic view; e.g., Gollwitzer, 1999), recent work has suggested that low-effort encoding can still be effective (perfunctory view; Scullin et al., 2018). PM reminders available at retrieval can effectively improve PM retrieval (e.g., Einstein & McDaniel, 1990; Peper et al., 2022) despite recent work in the retrospective memory domain that found expecting a reminder reduces effort at encoding (e.g., Kelly & Risko, 2019). The primary goal of the current study is to use reminder expectations to manipulate encoding effort and compare the strategic view of encoding with the perfunctory view (i.e., encoding effort hypothesis).

Successful PM requires effectively *encoding* new intentions into memory, *maintaining* the intention in long-term memory over a delay, and *retrieving* the intention from memory at the appropriate time (Kliegel et al., 2000). In a typical laboratory event-based PM task, participants complete an ongoing task (e.g., one versus two syllable judgments) and form an intention to make a special keypress (e.g., *spacebar*) whenever they see any words from a specific set within the next ongoing task. Participants then encode the target words they should remember for later (e.g., *pepper, shoe, horse*, etc.). The intention must then be maintained over a retention interval (e.g., 60 second math distractor), after which a participant completes a PM task with target words embedded in the ongoing task. The intention must then be retrieved upon seeing one of the target words. PM retrieval is measured by the proportion of targets that receive a successful PM response. Event-based PM tasks are essentially a dual-task that requires allocating attention to both the ongoing task and the PM intention. Ongoing task performance can indicate how much attention is allocated toward retrieving PM intention, such that a cost to ongoing task performance (worse accuracy or longer response times) indicates more attention devoted to the PM intention (Smith, 2003). While competing theories of prospective memory differ in the proposed mechanisms underlying retrieval (Einstein & McDaniel, 2005; Smith, 2003; Strickland et al., 2018), all agree encoding is important for successful retrieval.

PM Encoding

Several studies have found strategic and effortful encoding strategies improve PM retrieval. For example, an implementation intention is an effortful encoding strategy that improves PM retrieval (e.g., Gollwitzer, 1999) by having participants make a verbal statement about their intention, form a mental image of their intention, or both (Scullin et al., 2017). A verbal statement includes both the event (i.e., target) and the action (i.e., PM response) relevant for the intention (e.g., "When I eat my breakfast, then I will take my medication"). In the context of a laboratory event-based PM task, that verbal statement would be "When I see the word 'pepper' during the next syllable judgment task, then I will remember to press the spacebar." Implementation intentions are believed to improve PM by strengthening the association between the PM targets and intended action. Another effortful encoding strategy that can strengthen the memory trace of targets is semantic encoding. McDaniel et al. (1998) had participants in one condition generate synonyms of the PM targets at encoding and found this improved PM retrieval compared to a condition that did not semantically encode the targets. Critically both implementation intentions and semantic processing recruit greater attention and effort at encoding. Conversely, multiple studies have showed that dividing attention at encoding impairs PM retrieval (Einstein et al., 1997; McDaniel et al., 1998). These studies have clearly shown strategic encoding can improve PM retrieval.

More recent evidence, however, suggests that PM encoding need not always be strategic (i.e., effortful) to improve performance (Scullin et al., 2018). Scullin et al. tested whether PM encoding could also occur in a perfunctory manner (i.e., with minimal thought or effort) by inserting thought probes during encoding of PM intentions across eight studies. They gave participants a categorical PM target (e.g., *fruits*), participants formed the intention to respond to exemplars from that category (e.g., *apple*). After studying the target, participants reported what they were thinking about during encoding. An example of strategic encoding in the lab would be spending a longer time studying the PM targets. Overall, participants reported mind-wandering 42.9% of the time during encoding, while 22.5% of participants barely thought about the PM task. This suggests perfunctory encoding occurs at a surprisingly high frequency. In their final experiment, Scullin et al. found that longer study duration (i.e., effortful encoding) did not influence PM retrieval, suggesting perfunctory encoding is sufficient for successful retrieval.

One way to conceptualize strategic and perfunctory encoding is along a spectrum of encoding effort. That is, strategic encoding falls on the higher end of encoding effort, and perfunctory encoding involves lower effort. While previous research has manipulated encoding strategies and observed the outcome on PM retrieval, relatively little work has directly measured encoding effort to observe the consequences on PM retrieval. In the present study, we directly measure encoding effort to compare the strategic and perfunctory views (Gollwitzer 1999; McDaniel et al., 1998; Scullin et al., 2018).

Offloading Memory

One way to compensate for possible consequences of low effort encoding is through cognitive offloading, which refers to using the external environment to reduce internal cognitive demands (Risko & Gilbert, 2016). Past studies have shown reminders improve PM retrieval in both the laboratory (e.g., Chen et al., 2017; Gilbert, 2015a; Peper et al., 2022; Vortac et al., 1995) and naturalistic tasks (Ihle et al., 2012; Schnitzspahn et al., 2020). Recently, we showed that reminders improve PM retrieval by facilitating the maintenance and retrieval processes (Peper et al., 2022). In a series of experiments, we gave participants reminders by presenting the PM targets at the top of the screen throughout the duration of the PM task. Replicating across four studies, we found that reminders improve PM retrieval without changing ongoing task cost, particularly under high memory (i.e., target) load. However, because the focus of that study was on maintenance and retrieval, it is unclear how offloading influences encoding. The present study will use a similar paradigm to test the effect of reminder expectations on PM encoding.

Despite the abundance of evidence that reminders can be used to improve memory, reminders can be lost, destroyed, or otherwise fail. People increasingly rely on technology to offload their intentions by, for example, setting reminders on their phones. But what would happen to PM retrieval performance if that phone died when one needed the reminder? Within the retrospective memory domain, Kelly and Risko (2019) tested how a failed reminder affects memory performance. In their experiment, participants completed four blocks of a free recall task. During the first three blocks, participants were able to offload (i.e., write down memory items on a list) during encoding and access the external store (i.e., reminder list) at retrieval. Critically, participants were not able to access the external store at retrieval during the final block. Half of participants learned they would not have access and did not expect to have the store (i.e., non-expecting reminder condition), whereas the other half expected the store (i.e., expecting reminder condition). The expecting reminder condition captured a situation when a reminder failed (i.e., when one set and anticipated a reminder but later did not have access to it). Their results indicated that participants in the non-expecting reminder condition had better memory performance compared to the expecting reminder condition.

Kelly and Risko (2019) also examined the serial position effects between the non-expecting and expecting reminder conditions. The primacy effect refers to the general finding that words at the beginning of a list are remembered better than words in the middle due to strategic encoding strategies (e.g., rehearsal). The recency effect refers to the finding that words at the end of a list are remembered better than words in the middle due to the words at the end remaining in working memory. Expecting a reminder reduced the primacy effect, but not the recency effect, compared to the non-expecting reminder condition. They argued that expecting a reminder leads to less effortful encoding strategies, which in turn reduces unaided retrieval. This is referred to as the encoding effort hypothesis. In a more recent study, Kelly and Risko (2022) compared study durations between the expecting and non-expecting reminder conditions. They found that the expecting reminder condition spent less time encoding compared to the non-expecting reminder condition during the final block. These studies provide strong evidence that expecting a reminder reduces encoding effort and unaided retrieval in retrospective memory. It is important to note that whenever the reminder list was available to participants, it improved memory performance, so the benefit of reminders appears to offset the reduced encoding effort when the reminder functions properly.

Use It or Lose It Hypothesis

A popular belief about the use of reminders is that overreliance on offloading will lessen internal cognitive performance due to fewer opportunities to practice or maintain necessary cognitive operations (e.g., *use it or lose it* hypothesis; Baldwin et al., 2011). However, empirical observations of this phenomenon are equivocal. Support for the use it or lose it hypothesis has come from retrospective memory studies showing that memory for to-be-remembered images is reduced when participants take photographs of the images (Henkel, 2014; Soares & Storm, 2018) and that offloading experiences onto social media reduces internal memory recall of those experiences (Tamir et al., 2018). In the context of PM, Scarampi and Gilbert (2020) tested the use it or lose it hypothesis using the intention offloading task. The goal of this task was to drag 10 numbered circles to the bottom of the screen in sequential order. Before participants began this task, they formed an intention to drag three target circles to different parts of the screen (e.g., left, top, or right). All participants completed 20 blocks of this task in two phases (i.e., 10 in each). Participants in the reminder condition offloaded their intentions in the first 10 blocks by dragging the target circles next to the target location, which served as an external cue (i.e., reminder) to drag the target circle to the target location. Participants in the no reminder condition participants had to rely on their own memory ability (i.e., no offloading) during the first 10 blocks. During the second phase, participants in both conditions were unable to offload. By comparing performance in the second phase of the task, Scarampi and Gilbert found no differences between the reminder and no reminder conditions, which is inconsistent with the use it or lose it hypothesis.

While the intention offloading task resembles a traditional PM task (i.e., remembering to perform an action in the future), there are key differences related to the goal of the current study.

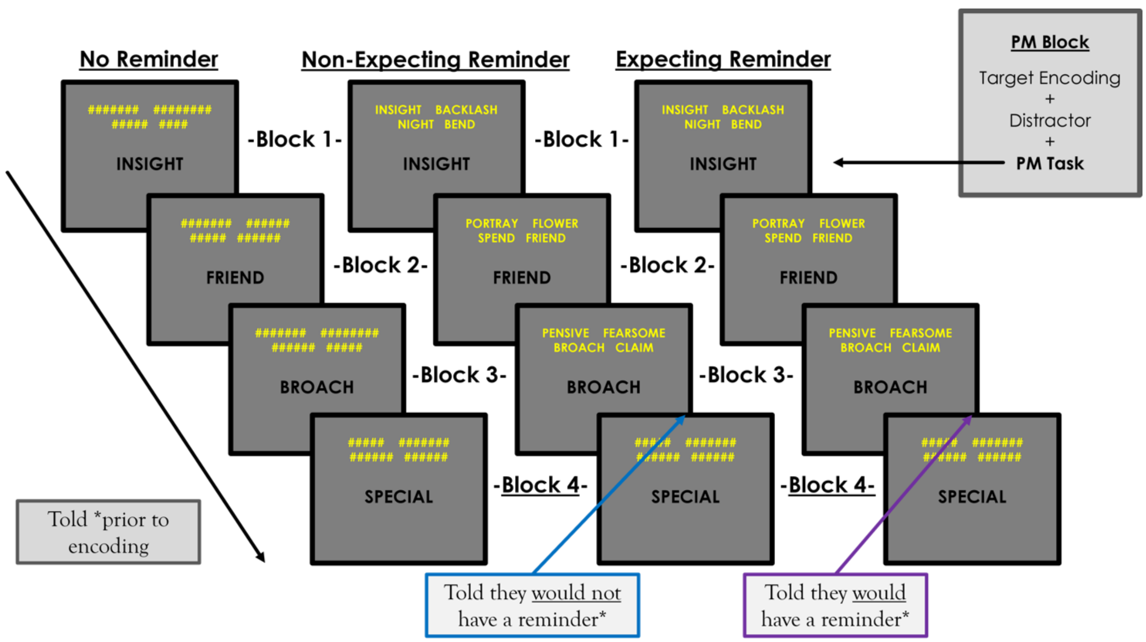
First, the intention offloading task occurs over short durations (10-20 seconds), and an argument can and has been made that this task captures a vigilance or working memory process (Gilbert, 2015a; Graf & Uttl, 2001). Second, the nature of the of the intention offloading task makes it difficult to test how offloading affects the encoding process. That is, all the target circles are numbers and may repeat across blocks, confounding any attempt at assessing recognition memory of the target circles. Lastly, theoretical processes underlying PM have been well-established in the 33 years of using the event-based PM task paradigm described above (Einstein & McDaniel, 1990). For these reasons, the use it or lose hypothesis in PM would be best tested using a traditional event-based PM paradigm.

Current Study

The present study tested the encoding effort hypothesis by examining study duration (Experiment 1) and pupil size (Experiment 2), and manipulating depth of processing (Experiment 3). Expanding on the procedure from Kelly and Risko (e.g., 2022), each experiment contained four PM blocks followed by a recognition memory task. Each PM block included a target encoding period, a math distractor, and a PM task. Three between-subjects conditions completed this overall procedure with minor distinguishing differences. Two reminder conditions had reminders during the first three PM tasks, while a no reminder condition did not. Critically, all three conditions did not have a reminder for the final block. The non-expecting reminder condition was explicitly told *prior to fourth block encoding* that they would not have a reminder during the final PM task. The expecting reminder condition was explicitly told *prior to encoding* that they would have a reminder during the final PM task (see Figure 1). We predicted based on Kelly and Risko's (2019; 2022) findings that expecting a reminder reduces encoding effort and that reduces unaided retrieval (i.e., encoding effort hypothesis). If unaided PM retrieval in the fourth block did not differ between the expecting and non-expecting conditions, then that would suggest effortful encoding is not necessary for PM retrieval (i.e., perfunctory encoding is sufficient). This procedure also allows us to test the use it or lose it hypothesis, which would predict better unaided PM retrieval in the fourth block for the no reminder condition compared to the non-expecting condition due to the former practicing PM retrieval without reminders in the first three blocks.

Figure 1

General PM Task Appearance in Each Condition for all Four Blocks in Experiments 1, 2, and 3



Note. Squares on the left, middle, and right represent the appearance of the PM task for each block in the no reminder condition, non-expecting reminder condition, and expecting condition, respectively. Critically, in every experiment all 8 targets (or 8 target-length-matched masks) encoded for that block were listed at the top of the screen during the entire PM task. While each PM block consisted of target encoding, a distractor, and the PM task, the dark gray squares for each condition and block only show how the PM task appeared. In Experiment 3, there were only non-expecting and expecting reminder conditions

Experiments 1 and 2 assessed how the degree of encoding effort is related to PM retrieval and target recognition with and without reminders. Experiment 1 assessed encoding effort directly by measuring study duration while participants studied the target words, similar to Kelly and Risko (2022). Experiment 2 used an eye-tracker to compare task-evoked pupillary responses during encoding for a conceptual replication of Experiment 1 with a physiological index of mental effort at encoding (e.g., Miller & Unsworth, 2021). Experiment 3 aimed to remove the effect of reminder expectations on encoding effort, unaided PM, and target recognition by manipulating depth of processing during block 4 target encoding. Participants either engaged in shallow processing (i.e., counting letters) or deep processing (i.e., rating pleasantness). The encoding effort hypothesis predicts better PM retrieval under deep processing regardless of reminder expectations. If expecting a reminder (expecting reminder condition) or overreliance on reminders (non-expecting reminder condition) affected encoding effort, recognition memory for PM targets should follow the same pattern as encoding effort. The encoding effort hypothesis therefore predicts PM retrieval would mirror target recognition memory. Specific predictions are listed before each individual experiment.

Preregistration Statement

The research in the present study was conducted following ethical guidelines approved by the Institutional Review Board at the University of Texas at Arlington. We preregistered how our sample sizes were determined, all exclusionary criteria, all data transformations, and all manipulations. Open Science Framework preregistrations are linked in the introductions for each experiment.

Experiment 1

Experiment 1 directly assessed the effect of reminder expectations on PM encoding effort by measuring study duration. Participants completed four PM blocks where they encoded a set of eight targets (self-paced) unique to each block before completing an event-based PM task where four of the targets were embedded. Some participants had reminders (expecting and non-expecting conditions) for the first three PM tasks, while others did not (no reminder condition). The reminder consisted of all eight targets listed at the top of the screen in yellow font for the duration of the PM task. In the fourth block, the expecting condition believed they would have a reminder, while the non-expecting reminder and no reminder conditions believed they would not.

Based on previous research on the efficacy of reminders in event-based PM (Peper et al., 2022), we predicted reminders would improve PM retrieval in the first three blocks. According to the encoding effort hypothesis, participants should change their encoding strategies when they expect to have a reminder. This should be particularly noticeable after they successfully retrieve the intention in block 1. That is, participants with and without reminder should encode for similar durations during the first block, but those in the reminder conditions should encode targets for less time than participants in the no reminder condition in blocks 2 and 3. The encoding effort hypothesis is tested by the critical comparison between conditions in block 4. Those in the non-expecting condition should update their encoding strategy in block 4, resulting in longer study durations (comparable to the no reminder condition). Those in the expecting reminder condition should continue to encode for less time. Critically, the encoding effort hypothesis would predict the expecting condition to have worse unaided PM retrieval in block 4 than the non-expecting condition, which would extend Kelly and Risko’s (e.g., 2022) work into the PM domain. If we observed no differences in block 4 PM retrieval between the expecting and non-expecting conditions, then this would go against the encoding effort hypothesis and suggest perfunctory encoding is sufficient for successful PM retrieval. It is important to note that whenever encoding effort differs between conditions, the same pattern of results should emerge in retrospective recognition memory for PM targets.

In the first three PM blocks, participants with reminders may not practice the cognitive operations necessary to complete the tasks with their internal memory alone. The use it or lose it hypothesis argues that cognitive processes can wane in effectiveness if they are not practiced or maintained (e.g., Baldwin et al., 2011). According to this view, even participants aware of not having a reminder in block 4 would have worse unaided PM after using reminders in the first three blocks compared to participants who never had reminders. In Experiment 1, that means the non-expecting condition would have worse block 4 PM retrieval compared to the no reminder condition. If the use it or lose it phenomenon exists and acts on encoding processes, block 4 study duration should be less in the non-expecting condition compared to the no reminder condition. If the use it or lose it phenomenon exists and acts uniquely on PM retrieval processes, fourth block ongoing task performance should also be worse in the non-expecting compared to the no reminder condition. If the use it or lose it phenomenon exists and acts on general retrieval processes, recognition memory should also be worse in the non-expecting condition compared to the no reminder condition. These hypotheses were preregistered. However, predictions for encoding effort during blocks 1-3 were described as "the self-regulated learning account" in the preregistration. The preregistration for Experiment 1 can be found here (https://osf.io/wnpcf).

Methods

Participants

The preregistered power analysis was conducted with G*Power based on the medium effect size, which recommended 159 participants. The goal was to obtain .80 power at .05 alpha probability for an effect size of $f = .25$ in a one-way ANOVA with three groups. We collected data until we reached a final sample of 169 participants after exclusions detailed below. The initial sample was 175 before exclusions. Participants (aged 17-32) were undergraduates at the

University of Texas at Arlington awarded with class credit for their participation. All participants were randomly assigned to either the expecting reminder condition (*n* = 57), non-expecting reminder condition (*n* = 57), or no reminder condition (*n* = 55).

Design

A 3 (condition: expecting, non-expecting, and no reminder; *between*) x 4 (block: 1, 2, 3, and 4; *within*) mixed-method design was used (Figure 1). Reminder conditions were manipulated between subjects and differed by the presence of a reminder in the first three blocks and the instructions prior to encoding of the fourth block. Block was manipulated within subjects such that all participants completed four PM blocks. During the first three blocks, the expecting and non-expecting reminder conditions had reminders for the targets during the PM task and the no reminder condition had no reminder. Before fourth block encoding the expecting condition was told they would have a reminder, the non-expecting condition was told they would not have a reminder, and the no reminder condition received the same instructions they had for previous blocks.

Materials

There were 380 words selected from the English Lexicon Project that served as ongoing task stimuli (Balota et al., 2007). Half of the words had one-syllable and half had two-syllables, each ranging from 5-7 letters in length. Twenty words were used for the first practice and reused for the second practice. For the PM blocks, 360 words made up the ongoing task stimuli. Thirty-two additional words were selected as PM targets and matched the ongoing task stimuli in terms of syllable count and letter length. That is, five targets had five letters and one syllable, five had five letters and two syllables, six had six letters and one syllable, six had six letters and two syllables, five had seven letters and one syllable, and five had seven letters and two syllables.

Thirty-two unique words with characteristics matching the PM targets were selected as *new* words in the recognition memory task. All stimuli were presented in uppercase black font, centered on the screen with a gray background. Reminders for all eight targets were centered around the top of the screen in yellow font. Whenever participants had no reminder, masks (e.g., ######) were centered around the top of the screen in yellow font of the same letter length as the targets for that block. This was done because to equate screen luminosity across conditions and blocks, which was necessary in Experiment 2 when pupillometry was measured. The math distractor consisted of multiplication problems with answers ranging from 1-500.

Data collection was completed in person. Participants consented before answering demographic questions. The experiment was programmed using Python.

Procedure

Broadly, the experiment consisted of two ongoing task practices, four PM blocks, a post-experimental questionnaire, and a recognition memory task. Participants were instructed to make their syllable judgments on English words as quickly and accurately as they could by pressing the “F” key for 1-syllable words (e.g., *storm*) and the “J” key for 2-syllable words (e.g., *pepper*). For the PM task, participants completed the syllable judgment ongoing task with embedded PM target words. The ongoing task and PM tasks were self-paced. Between trials there was a 500ms fixation cross before the next word appeared.

After consenting and completing the demographics form, participants sat down at a computer and the lights turned off in the experiment room. Lights were off to equate the procedures across all three experiments in anticipation of using pupillometry in Experiment 2. The experiment began with ongoing task instructions before participants completed a 20-trial practice with accuracy and response time feedback. We intended to have a minimum required accuracy of 75% before participants could move onto the second practice, but an error in the program allowed participants to continue if their accuracy was below that threshold. Then participants completed a second practice of 20 ongoing task trials without feedback.

The instructions for the PM task were given to the participants after they finished both practices. Participants were instructed that we were also interested in their ability to remember to perform an action in the future. Instructions stated they would learn eight target words and they were to press the *spacebar* instead of making a syllable judgement if they ever saw one of the eight words in the next ongoing task. They were also told they would complete four of these PM blocks. In the two reminder conditions, participants learned they would have a reminder to help them remember the targets. After the PM instructions, all participants summarized the instructions to an experimenter and then took a quiz on the instructions to verify they understood the PM task. We have found in past online studies that instruction quizzes helped to facilitate understanding (Peper et al., 2022). Accuracy on the quiz needed to be 100% for the participants to move on. Otherwise, participants had to go back and read the PM instructions and complete the quiz again.

PM blocks consisted of an encoding period, a math distractor, and a PM task. The 32 PM targets were randomly assigned eight to a block for each participant. During encoding, participants studied eight target words one at a time for as long as they wished (i.e., self-paced study). After studying the targets, participants were reminded of the PM task instructions before beginning a math distractor. The math distractor had participants complete multiplication problems for 60-seconds. The purpose of this was to ensure memory for the intention was moved to long-term memory instead of being maintained in working memory (Graf & Uttl, 2001). After finishing the math distractor, participants began the 84 trial PM task with four of the eight targets appearing on trials 20, 40, 60, and 80. The reminder conditions had the eight targets listed at the top of the screen throughout the entire PM task while the no reminder control had to rely on internal memory alone. Upon completing the PM task, participants were told they finished that block and did not need to remember the eight targets from that block any longer.

The first three PM blocks occurred exactly as they were described above, but the critical manipulation happened during the instruction period immediately before fourth block encoding. In the expecting reminder condition, participants were told they would have a reminder for the target words (i.e., same instructions as the previous blocks). Participants in the non-expecting reminder condition were told they would not have a reminder to help them remember the target words. That information was highlighted in a different color to draw attention to this fact. The no reminder condition received the same instructions they had previously with no mention of a reminder. Importantly, after the math distractor, all participants completed the fourth PM task without a target reminder.

After the four PM blocks, participants completed a post-experimental questionnaire. They answered questions assessing their retrospective memory for the PM task, the reminder conditions answered questions about the reminders, including a manipulation check that assessed whether they believed they would have a reminder in the fourth block. Two multiple-choice questions asked participants whether they remembered they were supposed to do a secondary task in addition to the ongoing task and what they were supposed to do for the secondary task. An incorrect response for that question and having a zero for PM retrieval across all four blocks signified a retrospective memory error (see exclusionary criteria). The reminder conditions were also asked whether they expected a reminder on the final block (i.e., manipulation check).

Lastly, participants completed a 64-trial recognition task after receiving instructions that told them to press the 1 key for “old” items (i.e., PM targets) or the 2 key for “new” items. All 32 PM targets and 32 words matched in letter length and syllable count appeared in the recognition task. Each word appeared one at a time until the participants made a response. Participants were then thanked for their participation and debriefed.

Exclusionary Criteria

The following exclusions were preregistered. Participants were excluded for the following: a) both failing to detect any PM targets *and* forgetting the prospective memory task (*n* = 0); b) getting below 60% accuracy on the PM block ongoing task (*n* = 4); c) having PM block ongoing task response times greater than 3 or less than -3 standard deviations from the group mean (*n* = 2); d) making false alarms (i.e., PM response on nontarget trials) on over 15% of trials (*n* = 0).

Results

For the first set of analyses, participants in the expecting and non-expecting conditions were combined into a single *reminder* condition because the first three blocks were identical for these two conditions. A series of 2 (reminder: reminder vs no reminder) x 3 (Block: 1, 2, and 3) mixed-method ANOVAs were conducted for PM retrieval, study duration, uncontaminated recognition, and ongoing task performance (accuracy and response times). To test the effect of reminders on PM retrieval and ongoing task performance, we reported the main effect of reminders collapsed across the first three blocks. The effect of reminders on encoding effort across time was tested by the reminder by block interaction for study duration. All main effects and interactions were reported. We probed the main effect of block by comparing the first and second block, the first and third block, and the second and third block separately and used

Bonferroni-corrected p-values of .017. However, we did not probe a main effect of block if it was qualified by a significant reminder by block interaction. We probed the interaction by comparing the reminder and no reminder conditions separately for each of the first three blocks and used Bonferroni-corrected p-values of .017. We interpreted the Greenhouse-Geisser corrected values in instances in which Mauchly's Test indicated that the assumption of sphericity was violated (in which case the F-test is subscripted using "F_{GG}"). These analyses deviated from the preregistered 3 (condition: no reminder, expecting reminder, and non-expecting reminder) x 4 (Block: 1, 2, 3, and 4) mixed ANOVAs that we initially planned for each dependent variable. We opted for the analyses reported in the manuscript for more power and simplicity when interpreting the effects of reminders when the two reminder conditions were identical in the first three blocks. Figures for each dependent variable are plotted separately for each condition across all four blocks to provide a comprehensive visualization. The same change was made for Experiment 2.

For the second set of analyses, we compared performance separately for the three conditions only in block 4. More specifically, we compared the expecting and non-expecting conditions (i.e., encoding effort hypothesis) and the non-expecting and no reminder conditions (i.e., use it or lose it hypothesis) for PM retrieval, study durations, uncontaminated recognition, and ongoing task performance.

To verify that our manipulation was working correctly, we looked at the frequency of participants in the two reminder conditions who failed the manipulation check and reanalyzed the PM retrieval data without those participants. Participants in the expecting reminder condition failed the manipulation check if they answered "no" to the question of whether they expected to have a reminder during the fourth PM task. Participants in the non-expecting reminder condition failed the manipulation check if they answered "yes" to the question of whether they expected to have a reminder during the fourth PM task. See Table 1 for frequencies of manipulation check failures broken down by experiment.

Table 1

Frequency of manipulation check failures and means and standard errors for contaminated recognition hits and new item recognition accuracy

Experiment	Condition	Manipulation Check Failures	Contaminated Recognition Hits				New Item Accuracy
			Block 1	Block 2	Block 3	Block 4	
Exp. 1	No Reminder	--	0.78 (0.03)	0.76 (0.03)	0.7 (0.04)	0.78 (0.03)	0.90 (0.01)
	Non-Expecting	9	0.79 (0.04)	0.70 (0.04)	0.83 (0.03)	0.79 (0.03)	0.85 (0.02)
	Expecting	17	0.85 (0.02)	0.80 (0.03)	0.74 (0.04)	0.59 (0.04)	0.87 (0.02)
Exp. 2	No Reminder	--	0.76 (0.04)	0.82 (0.03)	0.81 (0.03)	0.88 (0.03)	0.85 (0.02)
	Non-Expecting	4	0.88 (0.03)	0.87 (0.03)	0.84 (0.03)	0.78 (0.03)	0.81 (0.03)
	Expecting	13	0.87 (0.03)	0.85 (0.03)	0.85 (0.03)	0.67 (0.05)	0.78 (0.03)
Exp. 3	Non-Expecting Shallow	12	0.79 (0.04)	0.78 (0.03)	0.72 (0.04)	0.77 (0.04)	0.80 (0.03)
	Non-Expecting Deep	9	0.71 (0.04)	0.72 (0.04)	0.68 (0.04)	0.91 (0.03)	0.84 (0.02)
	Expecting Shallow	7	0.77 (0.03)	0.67 (0.04)	0.69 (0.04)	0.63 (0.04)	0.86 (0.02)
	Expecting Deep	5	0.76 (0.03)	0.71 (0.04)	0.73 (0.04)	0.85 (0.04)	0.91 (0.02)

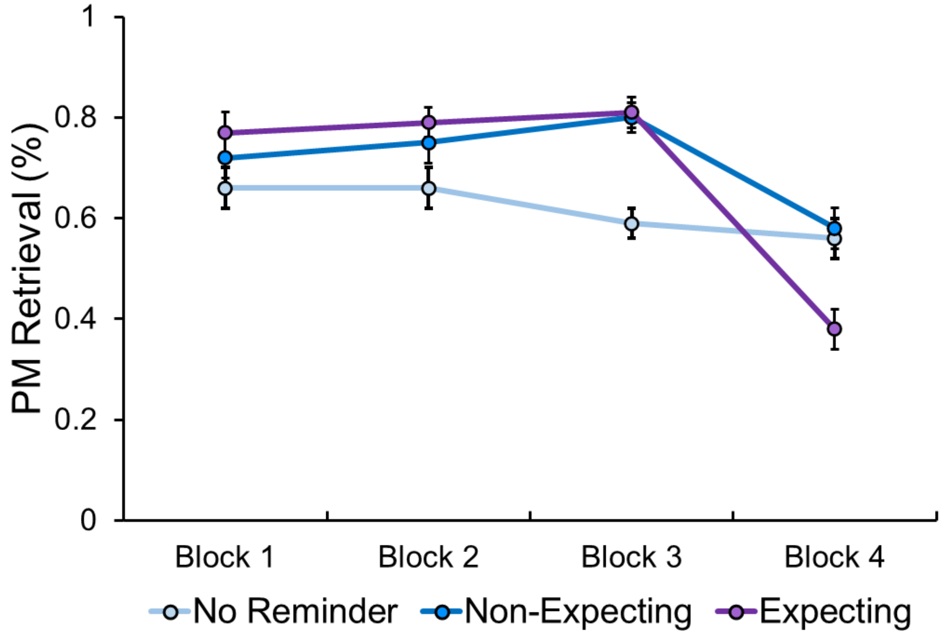
Note. Standard errors in parentheses; manipulation check failures refer to the number of participants in each reminder condition that failed the manipulation check in the post-experimental questionnaire; contaminated recognition hits refer to the proportion of contaminated targets (out of four) correctly identified as "old" during the recognition memory task; new item accuracy refers to the proportion of new items (out of 32) correctly identified as "new" during the recognition memory task.

PM Retrieval

PM retrieval was calculated as the proportion of target trials (out of four total) on which participants pressed the spacebar rather than making an ongoing task response in each block. Figure 2 presents the PM retrieval results separately for each condition and block.

Figure 2

PM Retrieval Across All Four Blocks Separated by Condition in Experiment 1



Note. PM retrieval refers to the proportion of PM targets detected (out of four). The circles represent mean performance, and the error bars reflect standard error.

Blocks 1-3. In the first three blocks, reminders improved PM retrieval overall [$F(1, 167) = 18.21$, $p < .001$, $\eta_p^2 = .098$]. There was no effect of block [$F < 1$]. There was a significant reminder by block interaction [$F(2, 334) = 3.73$, $p = .025$, $\eta_p^2 = .022$]. Post-hoc comparisons to explore the interaction used a Bonferroni-adjusted significance level of $p < .017$. While there was no PM retrieval difference between reminder and no reminder conditions in block 1 [$F(1, 167) = 4.25$, $p = .041$, $\eta_p^2 = .025$] reminders improved PM retrieval in block 2 [$F(1, 167) = 6.11$, $p = .014$, $\eta_p^2 = .035$] and block 3 [$F(1, 167) = 26.60$, $p < .001$, $\eta_p^2 = .137$].

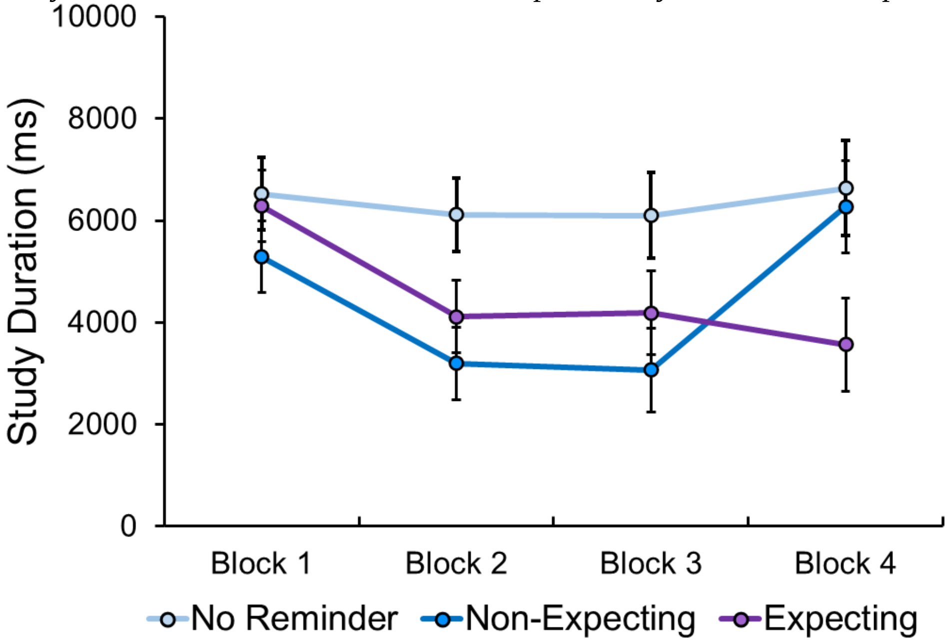
Block 4. Block 4 analyses supported the encoding effort hypothesis by showing PM retrieval was better in non-expecting condition compared to the expecting condition [$F(1, 113) = 13.38$, $p < .001$, $\eta_p^2 = .106$]. However, block 4 analyses did support the use it or lose it hypothesis, showing PM retrieval was no different between the non-expecting condition and the no reminder condition [$F < 1$]. Results were the same after excluding participants who failed the manipulation check, so they were not reported.

Study Duration

Study durations were calculated by averaging the length of time each target was studied. Figure 3 presents the results separately for each condition and block.

Figure 3

Study Duration Across All Four Blocks Separated by Condition in Experiment 1



Note. Study duration refers to the average time in milliseconds spent studying each target. The circles represent mean study duration, and the error bars reflect standard error.

Blocks 1-3. In the first three blocks, reminders reduced study durations [$F(1, 167) = 26.60$, $p < .001$, $\eta_p^2 = .137$]. Study durations also decreased across block [$F_{GG}(1.39, 232.86) = 14.29$, $p < .001$, $\eta_p^2 = .079$]. Critically, there was an interaction between reminder and block [$F_{GG}(1.39, 232.86) = 6.50$, $p = .005$, $\eta_p^2 = .037$]. Post-hoc comparisons to explore the interaction (Bonferroni $p < .017$) showed no differences between reminder and no reminder conditions in block 1 [$F < 1$]. However, reminders reduced study durations in block 2 [$F(1, 168) = 7.99$, $p = .005$, $\eta_p^2 = .045$] and block 3 [$F(1, 168) = 5.96$, $p = .016$, $\eta_p^2 = .034$]. This is consistent with the encoding effort hypothesis that expecting a reminder reduces encoding effort after the first experience of retrieval with a reminder.

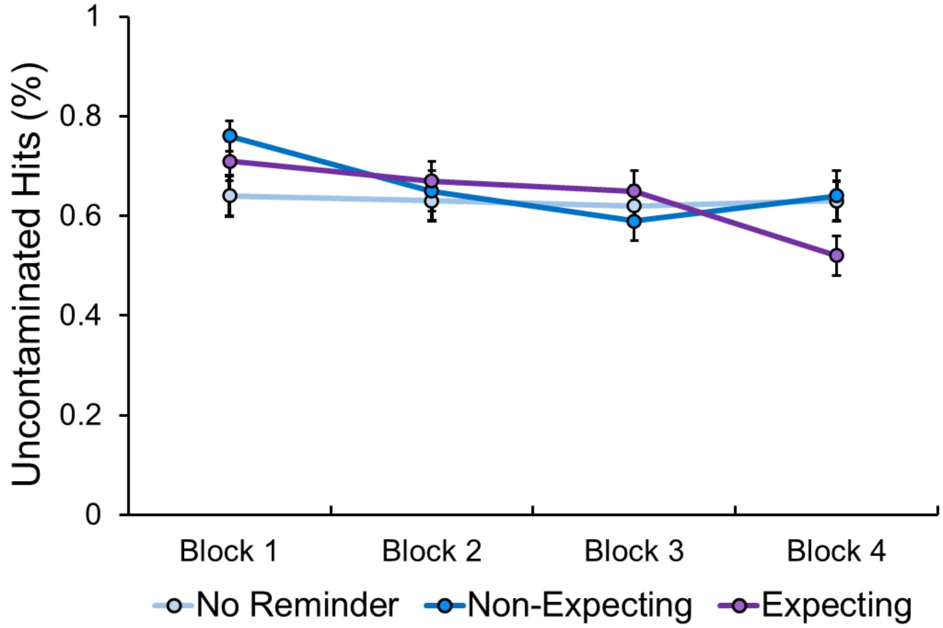
Block 4. Block 4 analyses supported the encoding effort hypothesis by showing the non-expecting condition had longer study durations than the expecting condition [$F(1, 112) = 11.32$, $p = .001$, $\eta_p^2 = .092$]. However, block 4 analyses showed study durations were no different between the non-expecting condition and the no reminder condition [$F < 1$].

Uncontaminated Recognition

As described previously, participants studied eight targets, but only four of those appeared during the task. Because retrieving the intention during the task serves as an additional encoding opportunity that should occur more frequently in the reminder conditions due to better PM retrieval, recognition memory was assessed for the four “uncontaminated” items that were encoded but never presented during the PM task. Uncontaminated recognition was calculated by the taking the proportion of correct (i.e., “old”) responses made for uncontaminated PM targets. Figure 4 presents the results separately for each condition and block.

Figure 4

Uncontaminated Target Recognition Across All Four Blocks Separated by Condition in Experiment 1



Note. Uncontaminated hits refer to the proportion of uncontaminated targets (out of four) correctly identified as "old" during the recognition memory portion of the task. The circles represent mean performance, and the error bars reflect standard error.

Blocks 1-3. In the first three blocks, there was no effect of reminders [$F(1, 167) = 1.42$, $p = .235$, $\eta_p^2 = .008$]. There was a main effect of block [$F(2, 334) = 3.12$, $p = .046$, $\eta_p^2 = .018$]. Post-hoc comparisons of the block effect (Bonferroni $p < .017$) showed that while memory did not differ between blocks 1 and 2 [$F(1, 168) = 4.73$, $p = .031$, $\eta_p^2 = .027$] or between blocks 2 and 3 [$F(1, 168) = 4.73$, $p = .031$, $\eta_p^2 = .008$], performance was better in block 1 than block 3 [$F(1, 168) = 9.98$, $p = .002$, $\eta_p^2 = .056$]. This suggests a possibly primacy effect for uncontaminated items. There was no reminder by block interaction [$F(2, 334) = 1.45$, $p = .235$, $\eta_p^2 = .009$].

Block 4. Block 4 analyses supported the encoding effort hypothesis by showing uncontaminated recognition was better in the non-expecting condition compared to the expecting condition [$F(1, 113) = 4.31$, $p = .040$, $\eta_p^2 = .037$]. However, block 4 analyses did not support the use it or lose it hypothesis, showing uncontaminated recognition was no different between the non-expecting condition and the no reminder condition [$F < 1$].

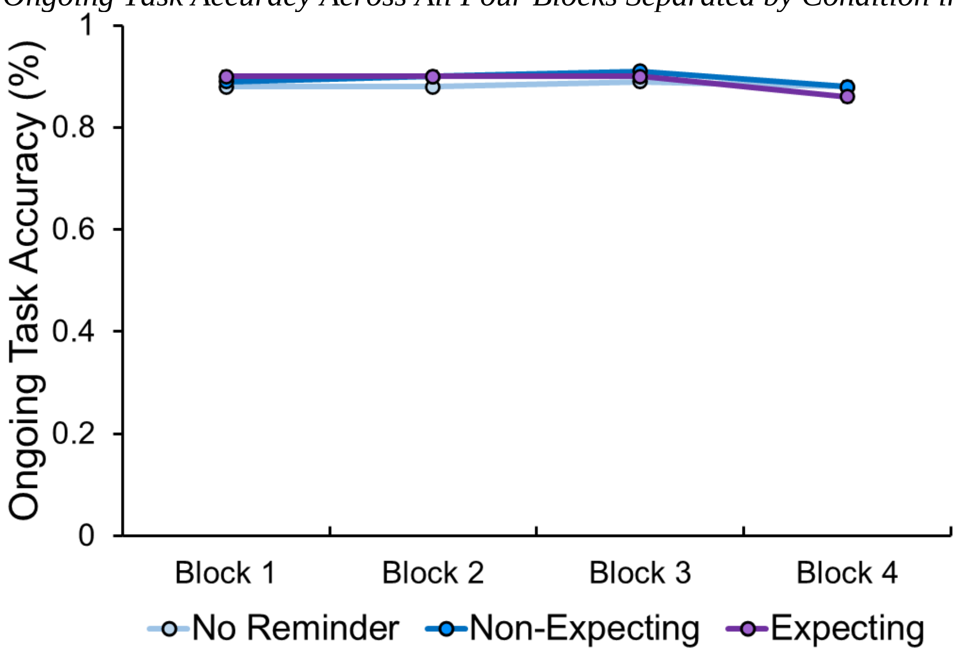
Ongoing Task Performance

The first three trials of each block and the three trials following a target trial were excluded from ongoing task analyses (Ball & Bugg, 2018). Trials following target trials were removed due to after-effects of longer response times following target trials (Meier & Rey-Mermet, 2012). Response times were calculated for accurate ongoing task trials only. Ongoing task trials were excluded if response times fell outside +/-3.0 standard deviations of each participant's overall mean.

Accuracy. Figure 5 presents the results separately for each condition and block. *Blocks 1-3.* In the first three blocks, there was no effect of reminders [$F(1, 167) = 1.91$, $p = .169$, $\eta_p^2 = .011$], no effect of block [$F < 1$], and no interaction between reminder and block [$F < 1$].

Figure 5

Ongoing Task Accuracy Across All Four Blocks Separated by Condition in Experiment 1



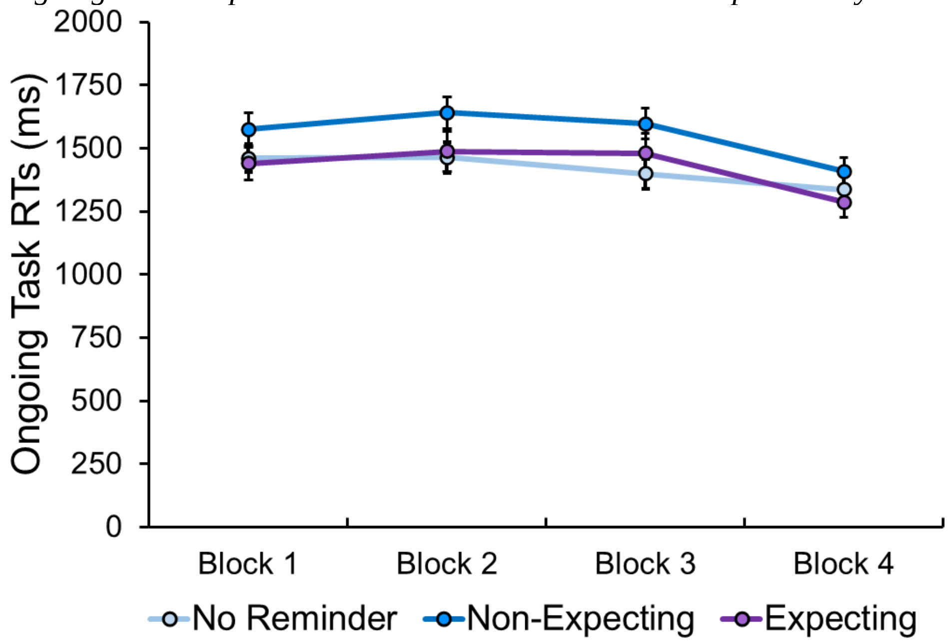
Note. Ongoing task accuracy refers to the proportion of accurate responses on non-target trials. The circles represent mean performance, and the error bars reflect standard error.

Block 4. Block 4 analyses showed no difference in ongoing task accuracy between the non-expecting and expecting conditions [$F(1, 112) = 1.64$, $p = .203$, $\eta_p^2 = .014$]. Additionally, block 4 analyses showed no difference in ongoing task accuracy between the non-expecting and no reminder conditions [$F < 1$].

Response Times. Figure 6 presents the results separately for each condition and block. *Blocks 1-3.* In the first three blocks, there was no effect of reminders [$F(1, 167) = 1.56$, $p = .214$, $\eta_p^2 = .009$], no effect of block [$F_{GG}(1.88, 315.26) = 2.49$, $p = .088$, $\eta_p^2 = .015$], and no interaction between reminder and block [$F_{GG}(1.88, 315.26) = 2.63$, $p = .077$, $\eta_p^2 = .016$].

Figure 6

Ongoing Task Response Times Across All Four Blocks Separated by Condition in Experiment 1



Note. Ongoing task RTs refers to the average response times on non-target trials. The circles represent mean performance, and the error bars reflect standard error.

Block 4. Block 4 analyses showed no difference in ongoing task response times between the non-expecting and expecting conditions [$F(1, 112) = 2.24$, $p = .137$, $\eta_p^2 = .020$]. Additionally, block 4 analyses showed no difference in ongoing task response times between the non-expecting and no reminder conditions [$F < 1$].

Discussion

Experiment 1 tested the effect of reminders and reminder expectations on prospective memory and encoding effort. Replicating previous work using a similar task (Peper et al., 2022), reminders improved PM retrieval without impacting retrospective recognition memory for the targets. Of secondary interest, ongoing task performance results suggested neither reminders nor the expectations of reminders influenced how much attention was allocated to the PM task, consistent with our previous work on PM reminders (Peper et al., 2022).

Critically, after a single experience with reminders, participants expecting a reminder spent less effort encoding the next set of targets, as evidenced by shorter study durations in the reminder conditions selectively during the second and third blocks. This is consistent with the encoding effort hypothesis that participants noticed the success of PM retrieval in the first block and then encoded with less effort in the second and third block. While the reduction in effort did not reduce PM retrieval when the reminders were present, lower encoding effort worsened PM retrieval and uncontaminated recognition memory when the reminders were unexpectedly unavailable for the expecting condition in block 4. This pattern of results is consistent with the encoding effort hypothesis, replicating in PM the pattern observed in retrospective memory (e.g., Kelly & Risko, 2022) and suggesting shared mechanisms underly the effect of expecting reminders in both retrospective memory and PM. That is, expecting a reminder reduces encoding effort but their presence at retrieval offsets performance deficits that may otherwise occur. The results of Experiment 1 also found no support of the use it or lose it hypothesis, as practice using reminders did not impact block 4 unaided PM or uncontaminated recognition when participants did not expect a reminder (see also Scarampi and Gilbert, 2021).

Experiment 2

Experiment 2 was designed as a conceptual replication of Experiment 1 using a physiological measurement of encoding effort that provided an assessment of how attentional effort at encoding influences both PM retrieval and the physiological parallels during PM retrieval and recognition. Pupillometry is a methodological technique that can be used to assess attentional effort (Unsworth & Miller, 2021). A task-evoked pupillary response (TEPR) is the degree of pupil dilation in response to a stimulus and is a reliable measure of attentional effort and task engagement (Beatty & Lucero-Wagoner, 2000; Just & Carpenter, 1993). Past research has shown those with higher working memory capacity exhibit higher attentional effort (i.e., larger TEPRs) during a working memory task (e.g., Unsworth & Robison, 2015). Larger TEPRs at encoding would suggest participants are devoting greater attentional effort to learning the items. Extant research has observed that attentional effort at encoding is positively related to recall in a paired associates task (Miller & Unsworth, 2021), a free recall task (Ariel & Castel, 2014), and a recognition task (Papesh et al., 2012). Miller and Unsworth had participants encode word pairs (e.g., *pepper-table*) and at test presented the first word ("*pepper-?*") to prompt retrieval of the associated second word ("*table"*). They found participants with higher rates of successful retrieval showed larger TEPRs at encoding. A similar result was found by Ariel and Castel (2014) who looked at TEPRs during encoding of items with different point values. Their results indicated that items with higher point values had larger TEPRs at encoding and had better recall at test. Consistent with this idea, Papesh et al. (2012) used a recognition memory task and found larger TEPRs during encoding for correctly identified old items compared to old items judged as new, especially for high confidence old items, suggesting larger TEPRs at encoding led to stronger memory traces for those items. While these findings clearly demonstrate a relationship between attentional effort at encoding and memory retrieval, these findings have not been extended into the PM domain. Experiment 2 will employ pupillometry to obtain a physiological measure encoding effort.

During encoding in each block of Experiment 2, every target was encoded for 5-seconds each. Reminder condition was manipulated between subjects with all participants randomly assigned to either the no reminder condition, non-expecting condition, or expecting condition. According to the encoding effort hypothesis, encoding TEPRs should not differ in the first block, but the reminder conditions should have smaller encoding TEPRs in blocks 2 and 3. Then in block 4, those in the expecting condition would encode targets less effortfully (i.e., smaller TEPRs) than the non-expecting condition in block 4 and that this would lead to worse PM retrieval and target recognition. Once again, whenever encoding effort differs between conditions in block 4, the same pattern of results should emerge in recognition memory.

The use it or lose it hypothesis would predict worse fourth block PM retrieval in the non-expecting condition compared to the no reminder condition. If the use it or lose it phenomenon exists and acts on encoding processes, encoding TEPRs during the fourth block should be smaller in the non-expecting condition compared to the no reminder condition. If the use it or lose it phenomenon exists and acts on PM retrieval processes, fourth block ongoing task performance should also be worse (and/or ongoing task TEPRs should be smaller) in the non-expecting condition compared to the no reminder condition. If the use it or lose it phenomenon exists and acts on general retrieval processes, recognition memory should also be worse (and/or recognition TEPRs should be smaller) in the non-expecting condition compared to the no reminder condition. These hypotheses were preregistered. However, the predictions made for encoding effort across the first three blocks are described as “the self-regulated learning account” in the preregistration. The preregistration for Experiment 2 can be found here (https://osf.io/t8xwz).

Methods

Participants

The preregistered power analysis was conducted with G*Power based on the large effect size (*d* = .80) found in Ariel and Castel (2014) between high-value and medium/low-value words at encoding, which recommended 123 participants. The goal was to obtain .90 power at .025 alpha probability for an effect size of *d* = .80 between the expecting and non-expecting conditions and the non-expecting and no reminder conditions in block 4. We collected data until we reached a final sample of 133 participants after exclusions detailed below. The initial sample was 140 before exclusions. Participants (aged 17-37) were undergraduates at the University of Texas at Arlington awarded with class credit for their participation. All participants were randomly assigned to either the expecting reminder condition (*n* = 45), non-expecting reminder condition (*n* = 43), or the no reminder condition (*n* = 45).

Design and Materials

The overall design and all materials were identical to those in Experiment 1.

Procedure

While most the procedure was identical to Experiment 1, there were a few differences in Experiment 2 that primarily addressed issues of timing and luminosity that are necessary to consider when using pupillometry. During encoding, target study duration was 5-seconds for each target. A mask fixation of seven characters, the maximum letter length for each target (Miller & Unsworth, 2021), was placed between each target during encoding. Each ongoing and PM task trial lasted 3000 milliseconds, and a mask (i.e., #####) appeared after a response was made with the same number of characters as the word stimulus on that trial. For example, if a participant pressed the F key for the word *broad* after 1000 milliseconds, then a five-character mask appeared for 2000 milliseconds before the 500 milliseconds fixation appeared that demarcated the next trial. For the recognition memory task, each trial consisted of a mask fixation (1000 milliseconds), the word stimuli (2000 milliseconds), another mask fixation (1000 milliseconds), and then an “old” vs. “new” prompt in which participants had unlimited to respond (Papesh et al., 2012).

Eye-tracking and Pupillometry

Pupillometry data was collected by a GazePoint GP3 eye-tracker. Participants sat approximately 70 centimeters away from the eye-tracker with their chin resting in a chin rest to prevent movement. Before starting the task, participants completed a 5-point calibration. The eye-tracker sampled pupil dilation and gaze binocularly at 60 Hz. Experiment rooms were windowless, and the lights were turned off so that pupil dilation was more sensitive to variations in attentional effort (Miller & Unsworth, 2021). Pupillometry data was calculated separately for encoding trials (5000 milliseconds), PM block trials (3000 milliseconds), and recognition memory trials (2000 milliseconds).

Exclusionary Criteria

The following exclusions were preregistered. Participants were excluded from all analyses for the following: a) both failing to detect any PM targets *and* forgetting the prospective memory task (n = 0); b) getting below 60% accuracy on the PM block ongoing task (n = 6); c) having PM block ongoing task response times greater than 3 or less than -3 standard deviations from the group mean (n = 1); d) making false alarms (i.e., PM response on nontarget trials) on over 15% of trials (n = 0).

Results

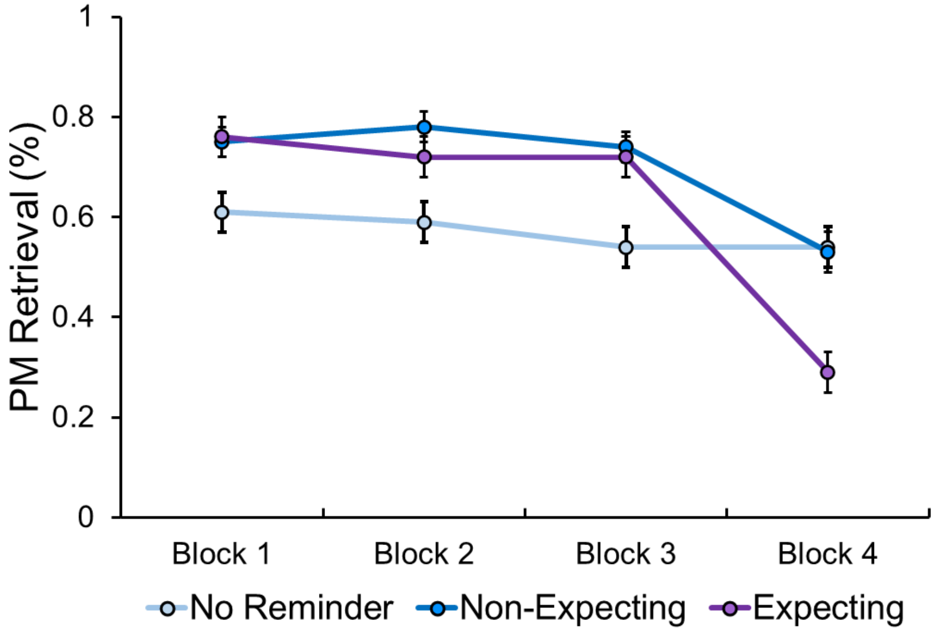
Behavioral Analyses

For the behavioral data, the analytic approach was identical to Experiment 1, except there were no study durations. PM retrieval, uncontaminated recognition hits, and ongoing task performance (accuracy and response times) were submitted first to a 2 (reminder: reminder vs no reminder) x 3 (Block: 1, 2, and 3) mixed-method ANOVA. Subsequently, block 4 comparisons included the expecting and non-expecting conditions (i.e., encoding effort hypothesis) and the non-expecting and no reminder conditions (i.e., use it or lose it hypothesis) for all variables.

PM Retrieval

Figure 7 presents the results separately for each condition and block.

Figure 7
PM Retrieval Across All Four Blocks Separated by Condition in Experiment 2



Note. PM retrieval refers to the proportion of PM targets detected (out of four). The circles represent mean performance, and the error bars reflect standard error.

Blocks 1-3. In the first three blocks, reminders improved PM retrieval overall [$F(1, 131) = 19.52$, $p < .001$, $\eta_p^2 = .130$]. There was no effect of block [$F(2, 262) = 1.63$, $p = .198$, $\eta_p^2 = .012$] and no reminder by block interaction [$F < 1$].

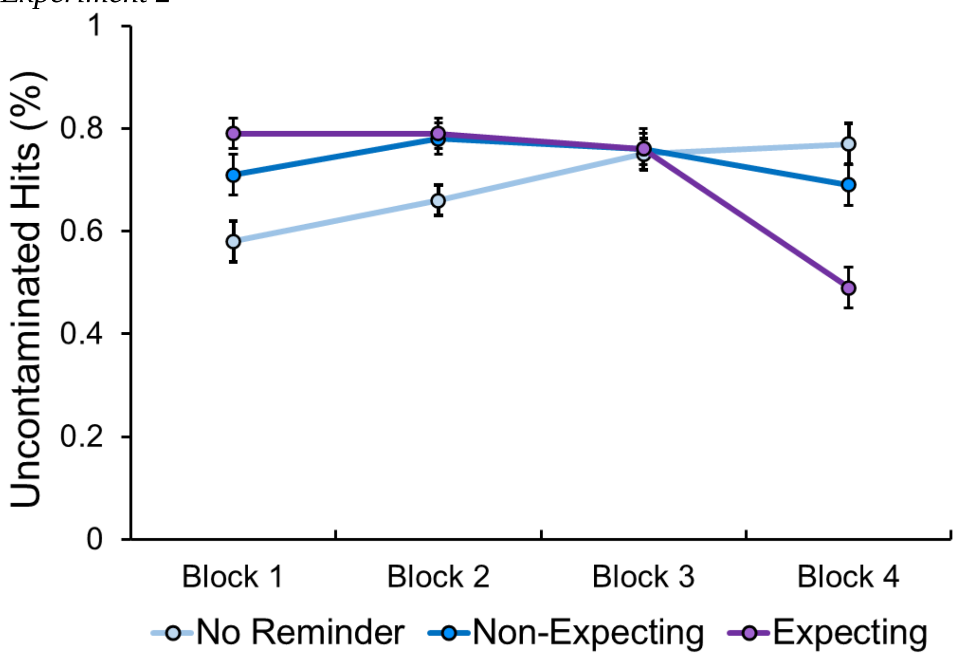
Block 4. Block 4 analyses supported the encoding effort hypothesis by showing PM retrieval was better in non-expecting condition compared to the expecting condition [$F(1, 86) = 19.49$, $p < .001$, $\eta_p^2 = .185$]. However, Block 4 analyses did support the use it or lose it hypothesis, showing PM retrieval was no different between the non-expecting condition and the no reminder condition [$F < 1$]. Results were the same after excluding participants who failed the manipulation check, so they were not reported.

Uncontaminated Recognition

Figure 8 presents the results separately for each condition and block.

Figure 8

Uncontaminated Target Recognition Across All Four Blocks Separated by Condition in Experiment 2



Note. Uncontaminated hits refer to the proportion of uncontaminated targets (out of four) correctly identified as “old” during the recognition memory portion of the task. The circles represent mean performance, and the error bars reflect standard error.

Blocks 1-3. In the first three blocks, reminders improved uncontaminated recognition [$F(1, 167) = 1.42$, $p = .235$, $\eta_p^2 = .008$]. Uncontaminated recognition also improved across block [$F(2, 262) = 5.90$, $p = .003$, $\eta_p^2 = .043$]. Critically, there was an interaction between reminder by block [$F(2, 262) = 5.29$, $p = .006$, $\eta_p^2 = .039$]. Post-hoc comparisons to explore the interaction used a Bonferroni-adjusted significance level of $p < .017$. While reminders improved uncontaminated recognition in Block 1 [$F(1, 131) = 15.60$, $p < .001$, $\eta_p^2 = .106$] and Block 2 [$F(1, 131) = 9.38$, $p = .003$, $\eta_p^2 = .067$], reminders did not influence uncontaminated recognition in Block 3 [$F < 1$].

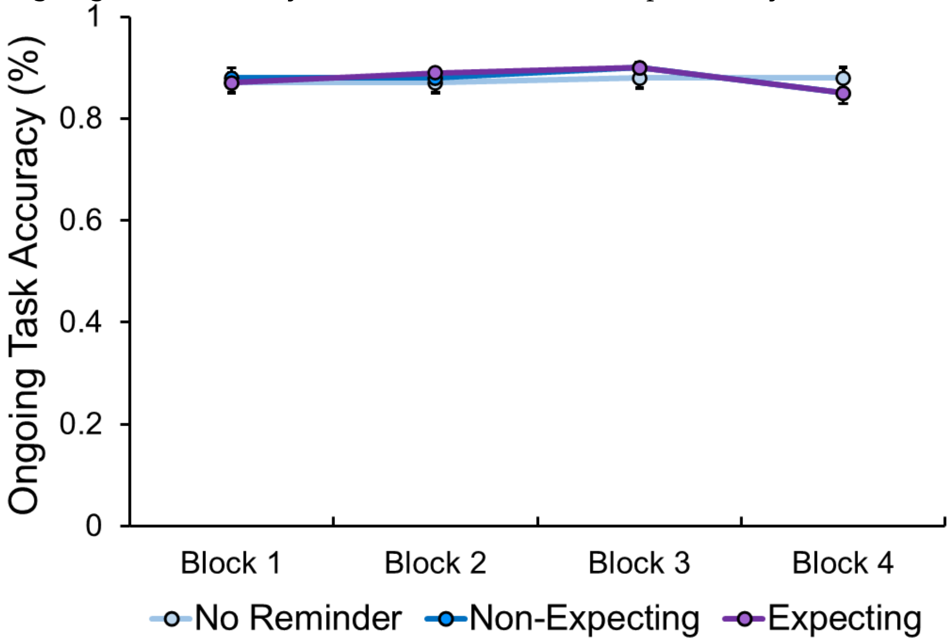
Block 4. Block 4 analyses supported the encoding effort hypothesis by showing uncontaminated recognition was better in non-expecting condition compared to the expecting condition [$F(1, 86) = 11.38$, $p = .001$, $\eta_p^2 = .117$]. However, block 4 analyses did not support the use it or lose it hypothesis, showing uncontaminated recognition was no different between the non-expecting condition and the no reminder condition [$F(1, 86) = 2.05$, $p = .156$, $\eta_p^2 = .023$].

Ongoing Task Performance

Accuracy. Figure 9 presents the results separately for each condition and block. *Blocks 1-3.* In the first three blocks, there was no effect of reminders [$F < 1$]. There was also no effect of block [$F_{GG}(1.24, 161.82) = 2.15$, $p = .140$, $\eta_p^2 = .016$]. There was no interaction between reminder and block [$F < 1$].

Figure 9

Ongoing Task Accuracy Across All Four Blocks Separated by Condition in Experiment 2



Note. Ongoing task accuracy refers to the proportion of accurate responses on non-target trials. The circles represent mean performance, and the error bars reflect standard error.

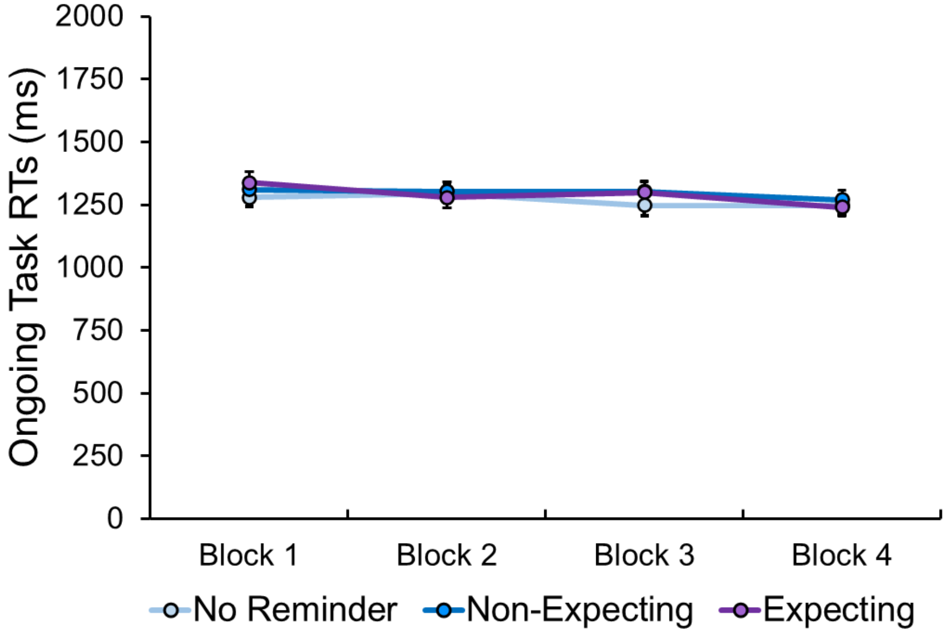
Block 4. Block 4 analyses showed no difference in ongoing task accuracy between the non-expecting and expecting conditions [$F < 1$]. Additionally, Block 4 analyses showed no difference in ongoing task accuracy between the non-expecting and no reminder conditions [$F(1, 86) = 1.17$, $p = .282$, $\eta_p^2 = .013$].

Response Times. Figure 10 presents the results separately for each condition and block.

Blocks 1-3. In the first three blocks, there was no effect of reminders [$F < 1$]. There was also no main effect of block [$F_{GG}(1.70, 222.53) = 1.37$, $p = .257$, $\eta_p^2 = .010$]. There was no interaction between reminder and block [$F_{GG}(1.70, 222.53) = 1.83$, $p = .168$, $\eta_p^2 = .014$].

Figure 10

Ongoing Task Response Times Across All Four Blocks Separated by Condition in Experiment 2



Note. Ongoing task response times refers to the average response times on non-target trials. The circles represent mean performance, and the error bars reflect standard error.

Block 4. Block 4 analyses showed no difference in ongoing task response times between the non-expecting and expecting conditions [$F < 1$]. Additionally, Block 4 analyses showed no difference in ongoing task response times between the non-expecting and no reminder conditions [$F < 1$].

Eye Tracking Analyses

Observations were removed for falling outside of plausible range for a human pupil diameter (< 2 millimeters or > 8 millimeters; Mathôt, 2018). Participants were excluded from pupil analyses (but not behavioral analyses) if they had missing data (e.g., missing values during blinks) on 50% or more of the trials. This resulted in 11 participant exclusions for encoding, seven exclusions for the PM block, and nine exclusions for recognition memory. We handled missing data for included participants with linear interpolation that filled in observations with values forming the best fitting line between valid observations within a trial (Gross & Dobbins, 2021). A low-pass Butterworth filter was applied to the data to preserve the overall structure while rounding any erratic observations across 10 Hz to minimize potentially erroneous variability between observations (Bowling et al., 2019). After interpolation and smoothing, pupil dilations for each trial were averaged into 200 millisecond bins and baseline-corrected to calculate the TEPRs. The baseline correction involved subtracting pupil diameter from the final 200 milliseconds of the intertrial interval (fixation) from the bins in each trial. For trials that required a response, TEPRs were only calculated for accurate responses (accurate non-target trials, PM target hit trials, and uncontaminated recognition hits), because inaccurate responses typically have smaller TEPRs than accurate responses (e.g., Papesh et al., 2012). In the same way we calculated ongoing task accuracy and response times, we excluded the first three trials of each block and the three trials following a target from the non-target TEPR analyses. Reminder checks were calculated by noting the trials when a participant both had a valid pupil measurement and a gaze that was in the top quartile of the screen where the reminder appeared (or the reminder mask in the no reminder blocks).

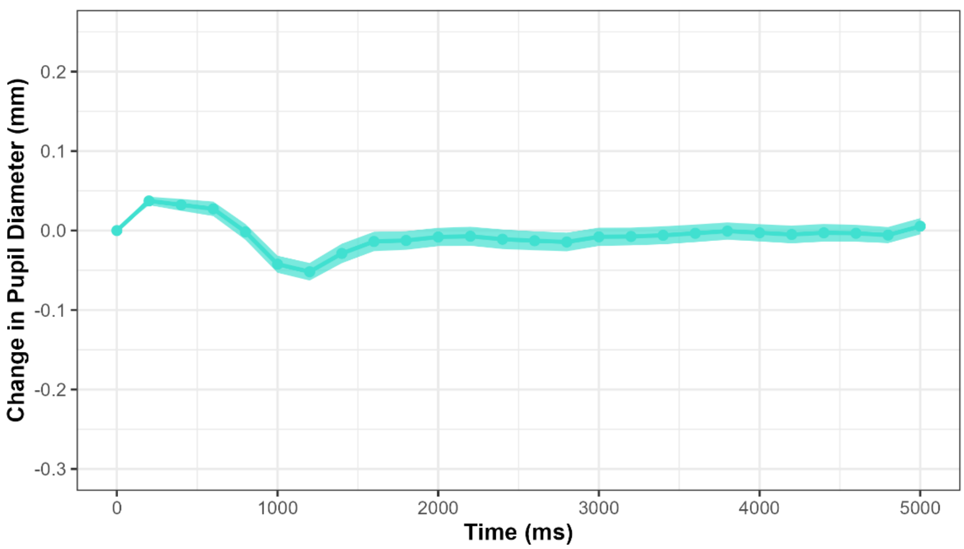
As anticipated, those in the reminder conditions checked the reminders more frequently than in the no reminder condition. Unfortunately, saccades to the reminders increases pupil dilation (Jainta et al., 2011), resulting in artifactual differences between conditions. We therefore calculated TEPRs for ongoing task non-target trials where participants did not check the reminder (i.e., no-check trials). However, we were unable to analyze PM hit trials because most participants in the reminder conditions checked the reminder on target trials so we were left with too few participants (n = 5 in the reminder conditions) to make valid inferences.

The average TEPR for all participants collapsed across condition and block were plotted separately for encoding trials (Figures 11 and 12), uncontaminated recognition hit trials (Figure

13), and non-target trials (Figure 14). As can be seen in Figure 11, TEPRs for encoding trials generally declined until about 1200ms, which appears to reflect a pupillary light reflex (Binda et al., 2013). We therefore calculated encoding TEPRs by using the 1200ms bin as the baseline to correct for the pupillary light reflex (Figure 12), similar to other researchers who examined TEPRs at encoding (Miller & Unsworth, 2021). In the end, we conducted analyses on encoding trials corrected for the pupillary light reflex, non-target trials without reminder checks, and uncontaminated recognition memory hits. The results were generally similar for encoding trial TEPRs whether we corrected for the pupillary light reflex or not, so we only reported results for TEPRs corrected for the pupillary light reflex.

Figure 11

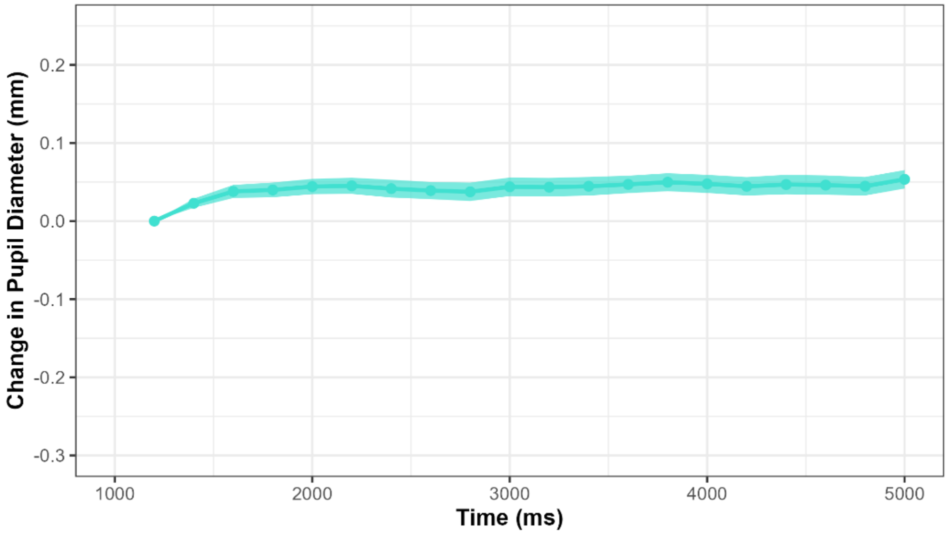
TEPR on Encoding Trials Not Corrected for Pupillary Light Reflex Collapsed Across Condition and Block in Experiment 2



Note. Mean encoding TEPR refers to the change in average baseline-corrected pupil size in millimeters across the length of each encoding trial before correcting for the pupillary light reflex (i.e., average pupil size from 0 milliseconds to 5000 milliseconds). The circles represent mean pupil size, and the error bars reflect standard error.

Figure 12

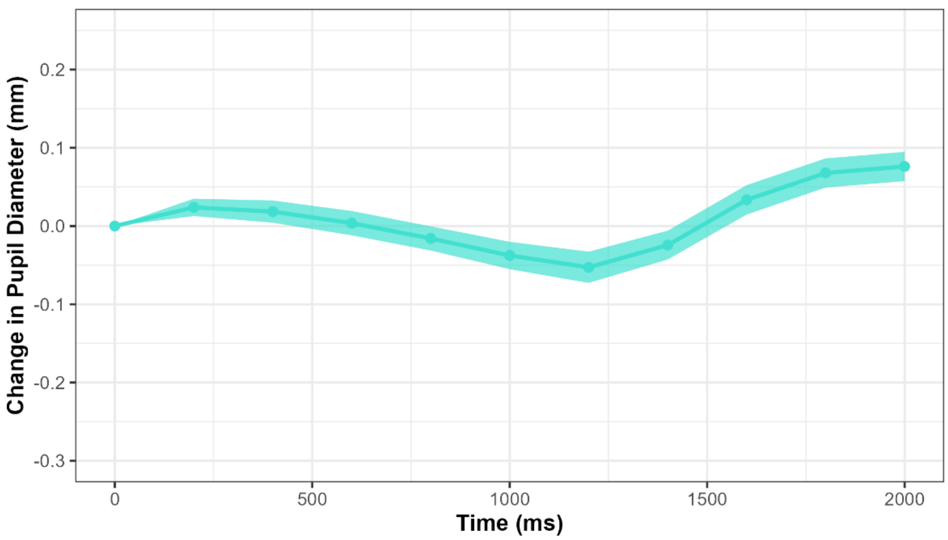
TEPR on Encoding Trials After Correcting for Pupillary Light Reflex Collapsed Across Condition and Block in Experiment 2



Note. Mean encoding TEPR refers to the change in average pupil size in millimeters across the length of each encoding trial after correcting for the pupillary light reflex (i.e., average pupil size from 1200 milliseconds to 5000 milliseconds). The circles represent mean pupil size, and the error bars reflect standard error.

Figure 13

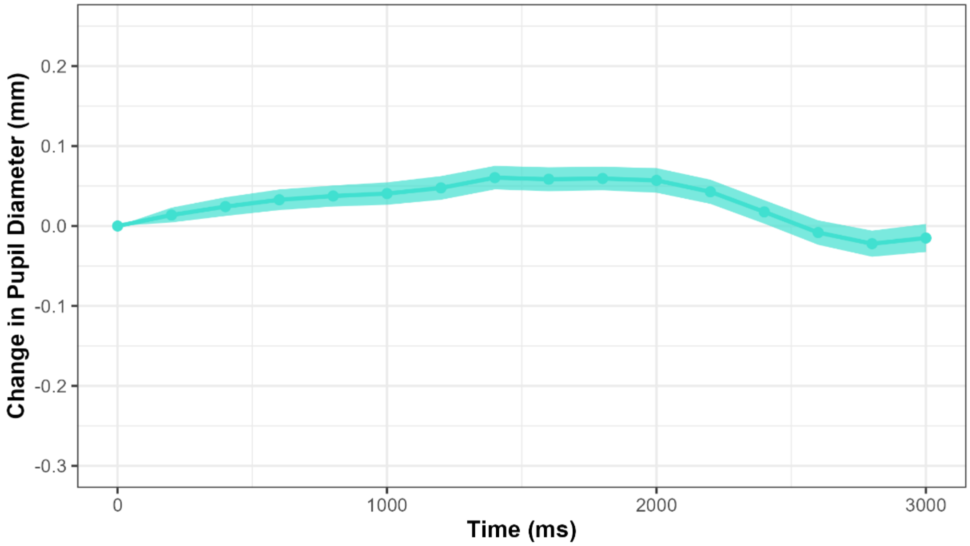
TEPR on Uncontaminated Recognition Hit Trials Collapsed Across Condition and Block in Experiment 2



Note. Mean uncontaminated recognition TEPR refers to the change in average baseline-corrected pupil size in millimeters across the length of each uncontaminated recognition trial that received a correct "old" response (i.e., average pupil size from 0 milliseconds to 2000 milliseconds). The circles represent mean pupil size, and the error bars reflect standard error.

Figure 14

TEPR on Ongoing Task Trials Collapsed Across Condition and Block in Experiment 2



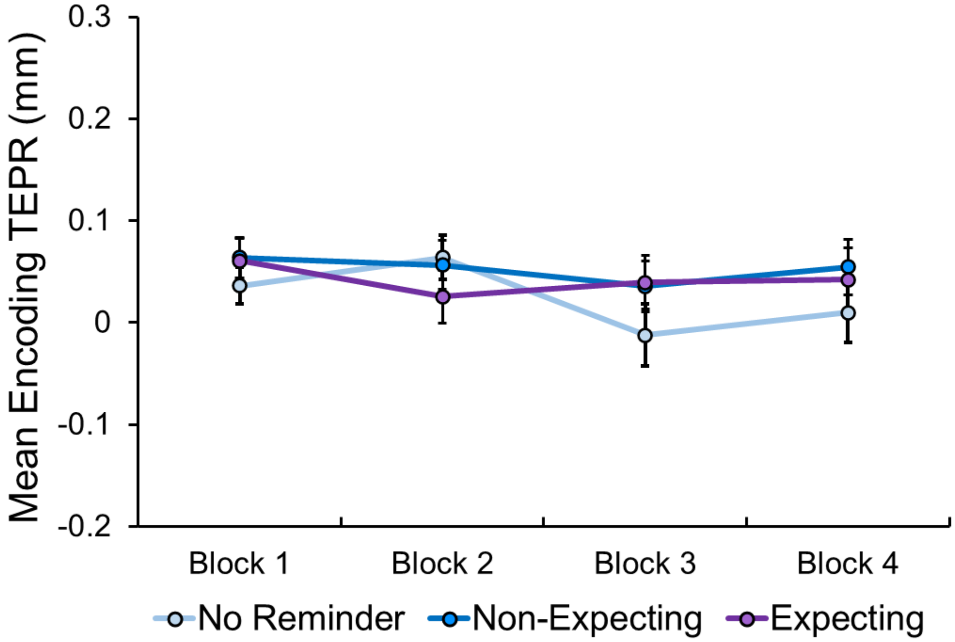
Note. Mean ongoing task TEPR refers to the change in average baseline-corrected pupil size in millimeters across the length of each accurate ongoing task trial when participants did not check the reminder (i.e., average pupil size from 0 milliseconds to 3000 milliseconds). The circles represent mean pupil size, and the error bars reflect standard error.

Eye-tracking data was also submitted to the same analytical approach as the behavioral data. That is, encoding TEPRs, uncontaminated recognition hit TEPRs, ongoing task trial TEPRs, target trial reminder checks, and non-target reminder checks were first submitted first to a 2 (reminder: reminder vs no reminder) x 3 (Block: 1, 2, and 3) mixed-method ANOVA. Block 4 comparisons included the expecting and non-expecting conditions (i.e., encoding effort hypothesis) and the non-expecting and no reminder conditions (i.e., use it or lose it hypothesis) for encoding and recognition data.

Encoding TEPRs. Figures 15 and 16 present the results separately for each condition and block. *Blocks 1-3.* In the first three blocks, there was no main effect of reminder [$F < 1$], no main effect of block [$F(2, 240) = 2.73$, $p = .067$, $\eta_p^2 = .022$], and no interaction between reminder and block [$F(2, 240) = 1.95$, $p = .145$, $\eta_p^2 = .016$]. This is not consistent with the encoding effort hypothesis that expecting a reminder reduces encoding effort after the first experience of retrieval with a reminder.

Figure 15

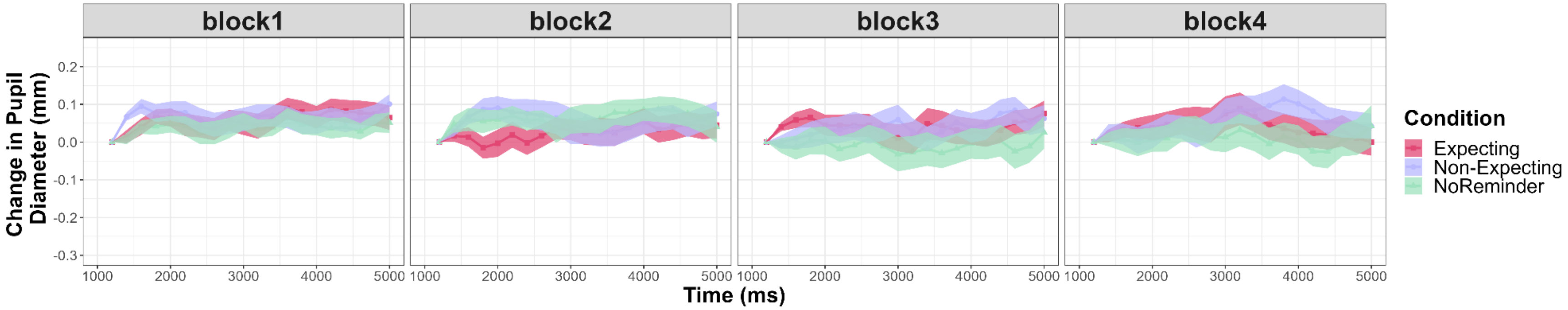
Mean TEPRs on Encoding Trials Across All Four Blocks Separated by Condition in Experiment 2



Note. Mean encoding TEPR refers to the average baseline-corrected pupil size in millimeters across the length of each encoding trial after correcting for the pupillary light reflex (i.e., average pupil size from 1200 milliseconds to 5000 milliseconds). The circles represent mean pupil size, and the error bars reflect standard error.

Figure 16

TEPRs on Encoding Trials Across All Four Blocks Separated by Condition in Experiment 2



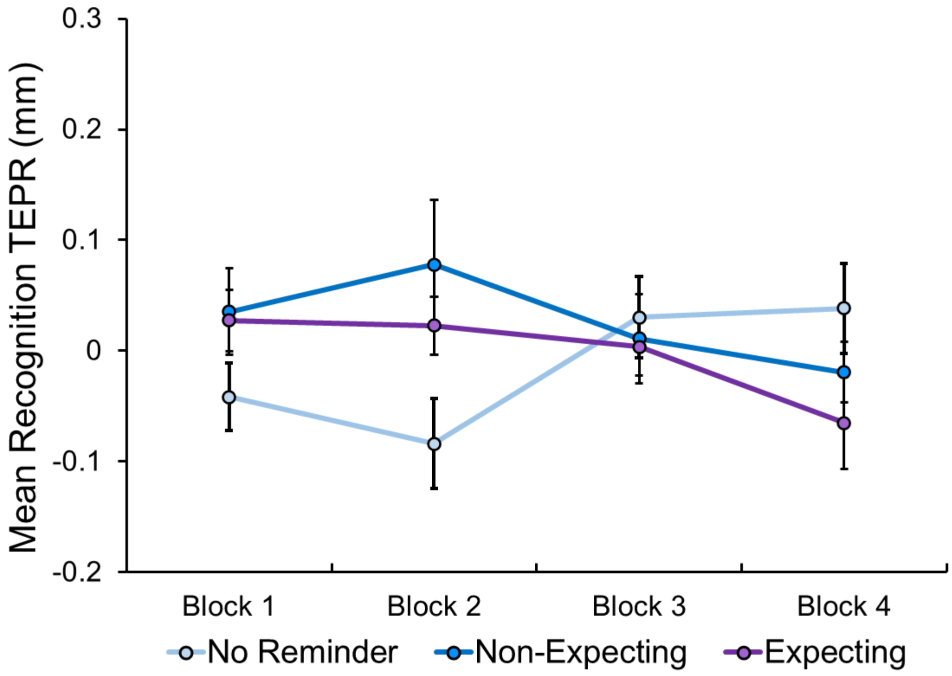
Note. Change in pupil diameter refers to the change in average pupil size in millimeters across the length of each encoding trial after correcting for the pupillary light reflex (i.e., average pupil size from 1200 milliseconds to 5000 milliseconds). The circles represent mean pupil size, and the error bars reflect standard error.

Block 4. Block 4 analyses showed no difference in encoding TEPRs between the non-expecting and expecting conditions [$F < 1$]. Additionally, block 4 analyses showed no difference in encoding TEPRs between the non-expecting and no reminder conditions [$F(1, 79) = 1.25$, $p = .267$, $\eta_p^2 = .016$].

Uncontaminated Recognition Hit TEPRs. Figures 17 and 18 present the results separately for each condition and block. *Blocks 1-3.* In the first three blocks, reminders increased uncontaminated hit TEPRs [$F(1, 114) = 5.23$, $p = .024$, $\eta_p^2 = .044$]. There was no effect of block [$F < 1$] and no interaction between reminder and block [$F(2, 228) = 2.72$, $p = .068$, $\eta_p^2 = .023$].

Figure 17

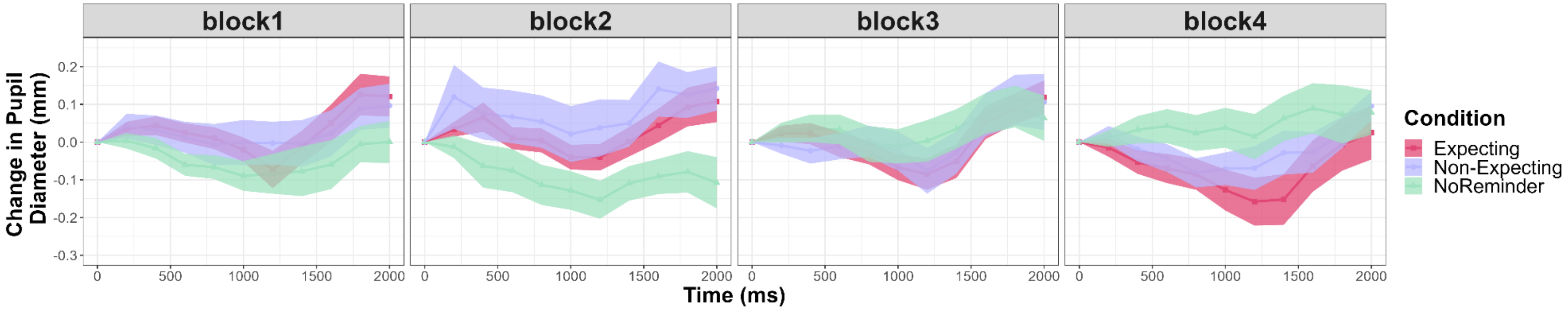
Mean TEPRs on Uncontaminated Recognition Hit Trials Across All Four Blocks Separated by Condition in Experiment 2



Note. Mean recognition TEPR refers to the average baseline-corrected pupil size in millimeters across the length of each uncontaminated recognition trial that received a correct "old" response (i.e., average pupil size from 0 milliseconds to 2000 milliseconds). The circles represent mean pupil size, and the error bars reflect standard error.

Figure 18

TEPRs on Uncontaminated Recognition Hit Trials Across All Four Blocks Separated by Condition in Experiment 2



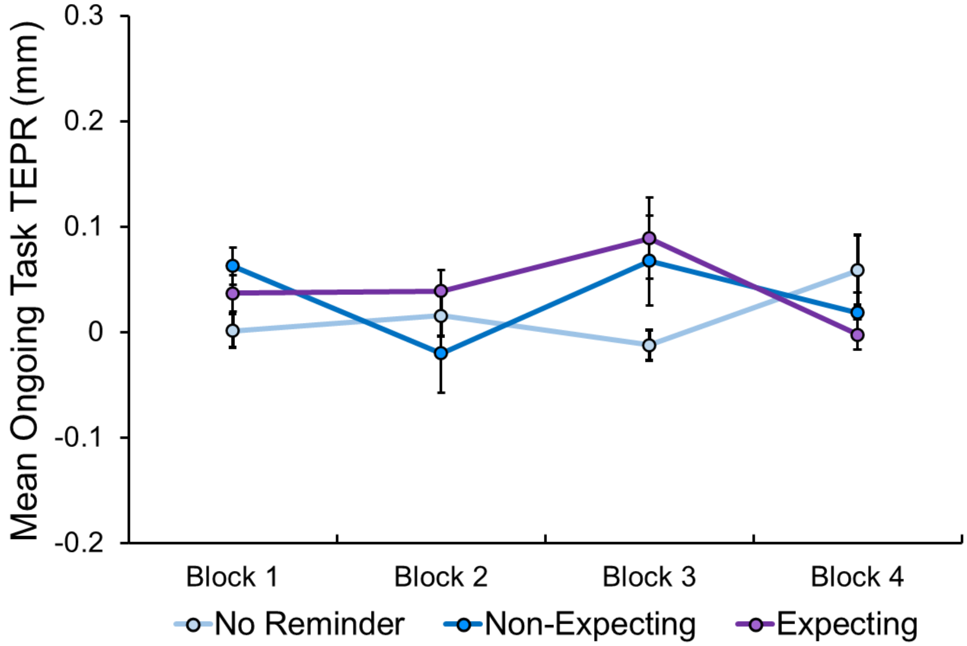
Note. Change in pupil diameter refers to the change in average baseline-corrected pupil size in millimeters across the length of each uncontaminated recognition trial that received a correct "old" response (i.e., average pupil size from 0 milliseconds to 2000 milliseconds). The circles represent mean pupil size, and the error bars reflect standard error.

Block 4. Block 4 analyses showed no difference in uncontaminated recognition hit TEPRs between the non-expecting and expecting conditions [$F < 1$]. Additionally, block 4 analyses showed no difference in uncontaminated recognition hit TEPRs between the non-expecting and no reminder conditions [$F(1, 75) = 1.38$, $p = .243$, $\eta_p^2 = .018$].

Ongoing Task Trial TEPRs. Figures 19 and 20 present the results separately for each condition and block *Blocks 1-3.* In the first three blocks, reminders increased ongoing task trial TEPRs [$F(1, 94) = 4.45$, $p = .038$, $\eta_p^2 = .045$]. However, there was no main effect of block [$F_{GG} < 1$] and no interaction between reminder and block [$F_{GG} < 1$].

Figure 19

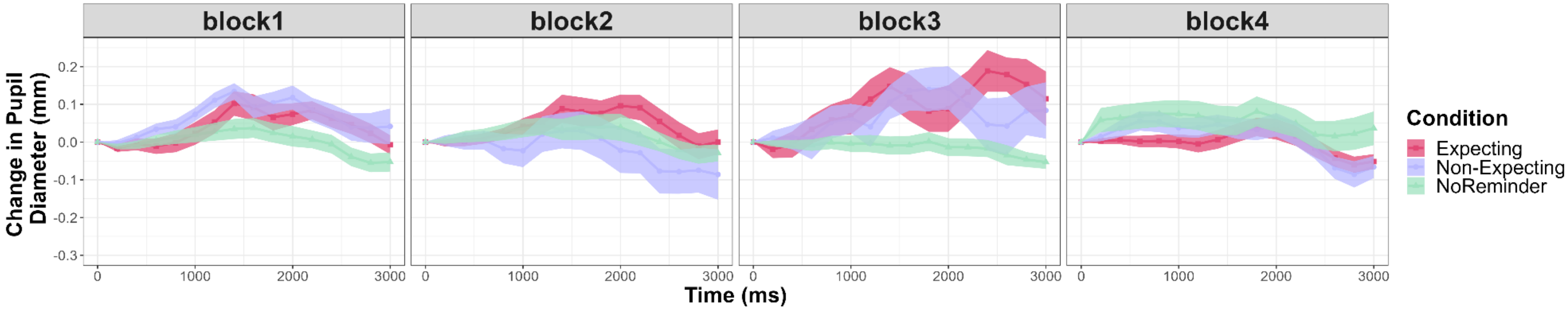
Mean TEPRs on Ongoing Task Trials Across All Four Blocks Separated by Condition in Experiment 2



Note. Mean ongoing task TEPR refers to the average baseline-corrected pupil size in millimeters across the length of each accurate ongoing task trial (i.e., average pupil size from 0 milliseconds to 3000 milliseconds). The circles represent mean pupil size, and the error bars reflect standard error.

Figure 20

TEPRs on Ongoing Task Trials Across All Four Blocks Separated by Condition in Experiment 2



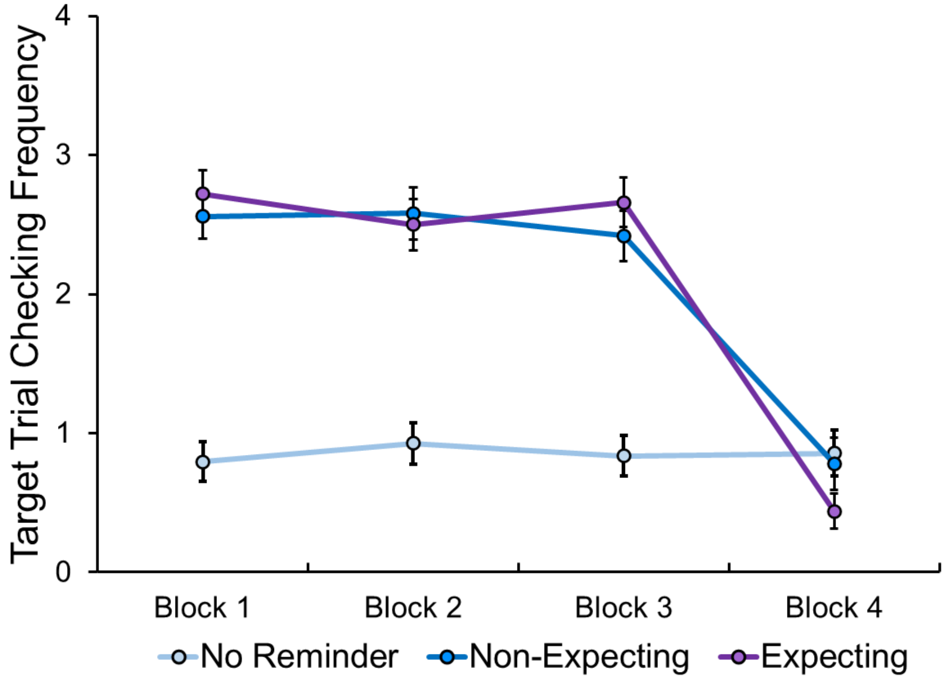
Note. Change in pupil diameter refers to the change in average baseline-corrected pupil size in millimeters across the length of each accurate ongoing task trial (i.e., average pupil size from 0 milliseconds to 3000 milliseconds). The circles represent mean pupil size, and the error bars reflect standard error.

Block 4. Block 4 analyses showed no difference in ongoing task trial TEPRs between the non-expecting and expecting conditions [$F < 1$]. Additionally, block 4 analyses showed no difference in ongoing task trial TEPRs between the non-expecting and no reminder conditions [$F(1, 78) = 1.07$, $p = .305$, $\eta_p^2 = .013$].

PM Target Trial Reminder Checks. Figure 21 presents the results separately for each condition and block *Blocks 1-3.* In the first three blocks, the reminder conditions checked the reminders on target trials more often than the no reminder condition [$F(1, 124) = 107.95$, $p < .001$, $\eta_p^2 = .465$]. However, there was no effect of block [$F < 1$], and no interaction between reminder and block [$F < 1$].

Figure 21

Frequency of Reminder Checks on Target Trials Across All Four Blocks Separated by Condition in Experiment 2



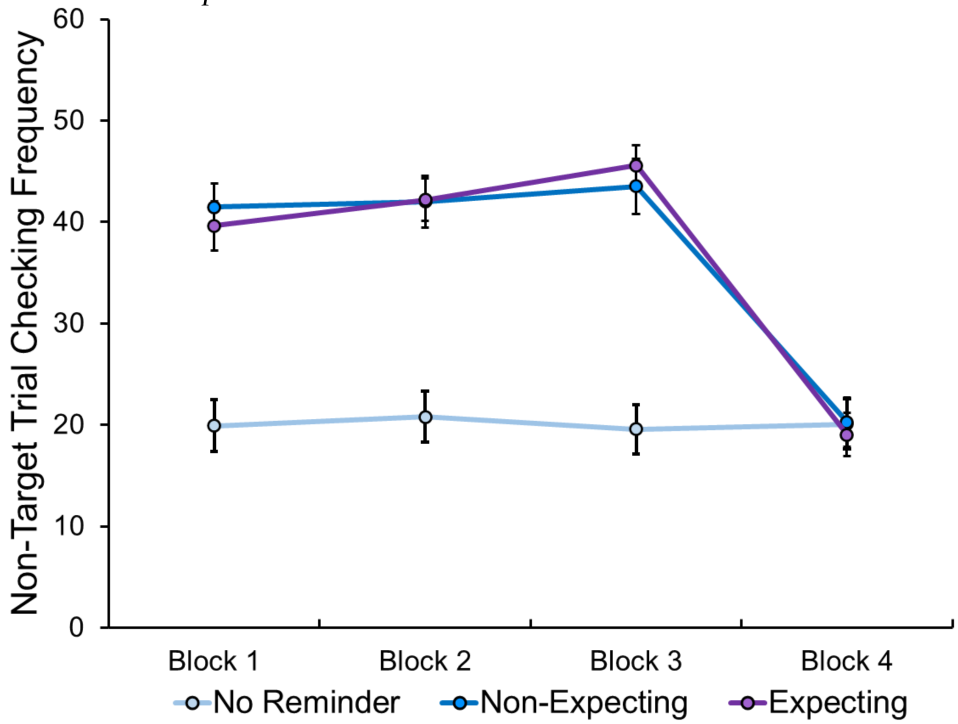
Note. Mean frequency of reminder checks on PM task target trials (out of four target trials in each block). The circles represent mean checking frequency, and the error bars reflect standard error.

Block 4. Block 4 analyses showed no difference in target trial reminder checks between the non-expecting and expecting conditions [$F(1, 71) = 2.06$, $p = .156$, $\eta_p^2 = .028$]. Additionally, block 4 analyses showed no difference in target trial reminder checks between the non-expecting and no reminder conditions [$F < 1$]. This is not surprising, as the "reminders" consisted of a series of "######'s" at the top of the screen in all conditions during block 4.

Non-Target Trial Reminder Checks. Figure 22 presents the results separately for each condition and block *Blocks 1-3.* In the first three blocks, the reminder conditions checked the reminders on non-target trials more often than the no reminder condition [$F(1, 122) = 89.78$, $p < .001$, $\eta_p^2 = .424$]. However, there was no effect of block [$F_{GG}(1.75, 213.27) = 2.00$, $p = .141$, $\eta_p^2 = .015$], and no interaction between reminder and block [$F_{GG}(1.75, 213.27) = 1.62$, $p = .203$, $\eta_p^2 = .013$].

Figure 22

Frequency of Reminder Checks on Non-Target Trials Across All Four Blocks Separated by Condition in Experiment 2



Note. Mean frequency of reminder checks on non-target ongoing task trials (out of 80 non-target trials in each block). The circles represent mean checking frequency, and the error bars reflect standard error.

Block 4. Block 4 analyses showed no difference in non-target trial reminder checks between the non-expecting and expecting conditions [$F < 1$]. Additionally, block 4 analyses showed no difference in non-target trial reminder checks between the non-expecting and no reminder conditions [$F < 1$]. This is not surprising, as the “reminders” consisted of a series of ““######’s” at the top of the screen in all conditions during block 4.

Discussion

Experiment 2 was a conceptual replication of Experiment 1 to test the effect of reminders and reminder expectations on PM retrieval and encoding effort using pupillometry. The behavioral results in Experiment 2 replicated the pattern of results observed in Experiment1, such that reminders improved PM retrieval and did not influencing ongoing task performance. However, in Experiment 2 reminders also improved uncontaminated recognition memory in blocks 1 and 2. It is possible that participants in Experiment 2 were able to check the reminder more often than Experiment 1 due to each PM task trial lasting three seconds, and these additional reminder checks led to the pattern of improved recognition memory that was not observed in Experiment 1.

Importantly, our behavioral results in Experiment 2 were consistent with the encoding effort hypothesis. Participants expecting a reminder had reduced unaided PM and uncontaminated recognition in block 4, necessitating further examination of the encoding effort hypothesis. We also replicated the pattern of results from Experiment 1 inconsistent with the use it or lose it hypothesis, that practice with reminders had no impact on unaided PM, ongoing task performance, or uncontaminated recognition memory.

Critically, using encoding TEPRs as an index of attentional effort at encoding, reminder experience and expectations did not influence encoding effort. This is not consistent with

Experiment 1 or the encoding effort hypothesis. It is possible that reminders expectations did not influence encoding effort, but an alternative possibility is that eight encoding trials in each block did not provide enough observations to reliably assess encoding effort. For example, Miller et al. (2019) had 50 encoding trials to observe the relationship between encoding TEPRs and memory retrieval. Figure 16 shows encoding TEPRs separated by block and condition. The large standard errors suggest eight encoding trials may have been too few to observe any stable differences between reminder conditions or across blocks. A third explanation for the encoding TEPR results is discussed in greater detail in the General Discussion.

The uncontaminated hit TEPR findings in Experiment 2 were consistent with previous research on recognition memory TEPRs (e.g., Kafkas & Montaldi, 2015) and our own research on reminders (Peper et al., 2022). Reminders in Experiment 2 of the present study increased uncontaminated hit TEPRs with a pattern similar to uncontaminated recognition hits (compare Figure 17 with Figure 8). Past research has found pupil responses are sensitive to the familiarity of (i.e., the *feeling* of remembering) old items during recognition, with larger pupil responses observed for familiar items than novel items (Kafkas & Montaldi, 2015). We have argued that PM reminders improve target detection by strengthening the familiarity of targets, which can help participants discriminate PM targets from non-target stimuli (Peper et al., 2022). The enhanced familiarity of targets engendered by the reminders could, in some cases, carry over to target recognition. However, these results should be interpreted with caution considering there were only four uncontaminated trials for each block.

The other additional pupillometry and eye-tracking measures yielded interesting results. Reminders increased non-target trial TEPRs, suggesting reminders may have increased attention allocated to the ongoing task without affecting ongoing task performance. This is inconsistent with the behavioral results for ongoing task performance. However, it is possible that participants with reminders allocated more attention to the ongoing task, but the high frequency of reminder checking washed out any differences in ongoing task response times. The reminder checking analyses suggest that reminders were checked to maintain the intention (on non-target trials) and to verify the intention (on target trials). This is consistent with previous research in our lab suggesting intention maintenance and target verification are two mechanisms by which reminders can improve PM retrieval (Peper et al., 2022).

Experiment 3

The first two experiments clearly showed PM retrieval was worse when one expected to have a reminder that was not available when anticipated (i.e., encoding effort hypothesis). While Experiment 1 provided additional behavioral support in terms of study duration for the idea that participants encoded information less effortfully across blocks when expecting reminders, the pupillary analyses of Experiment 2 failed to find any differences across blocks or conditions (i.e., metacognitive view). Experiment 3 was designed to further test the encoding effort hypothesis by manipulating depth of processing at block 4 encoding to remove the effect of reminder expectations. All participants in Experiment 3 were in either the expecting or non-expecting condition and completed four PM blocks. Critically, depth of processing was manipulated during block 4 encoding by having participants either count the number of letters in each target (i.e., shallow condition) or rate the pleasantness of each target (i.e., deep condition).

McDaniel et al. (1998) employed an event-based PM task and manipulated whether participants generated a word that rhymed with the target at encoding (i.e., shallow processing) or generated a synonym for the target (i.e., deep processing). Their results showed that deep processing at encoding improved PM retrieval compared to shallow processing. We anticipated support of the encoding effort hypothesis by showing better fourth block PM retrieval with deep processing than shallow processing. If expecting a reminder lowers unaided memory retrieval by reducing encoding effort (Kelly and Risko, 2019, 2022), we should find no difference in fourth block PM retrieval and target recognition when encoding is deep. These hypotheses were preregistered. The preregistration for Experiment 3 can be found here (https://osf.io/yhg3n).

Methods

Participants

The preregistered power analysis was conducted with G*Power and recommended 160 participants for a large effect size (d = .80) based on retrieval differences between deep and shallow encoding. The goal was to obtain .80 power at .025 alpha probability for the follow-up comparison to probe the interaction between processing and expectation in block 4. We collected data until we reached a final sample of 172 participants after exclusions detailed below. The initial sample was 182 before exclusions. Participants (aged 17-49) were undergraduates at the University of Texas at Arlington awarded with class credit for their participation. All participants were randomly assigned to either the shallow expecting condition (n = 43), shallow non-expecting condition (n = 44), deep expecting condition (n = 41), or deep non-expecting condition (n = 42).

Design and Materials

A 2 (processing: shallow vs. deep; *between*) x 2 (reminder expectations: expecting vs. non-expecting; *between*) x 4 (block: 1, 2, 3, and 4; *within*) mixed-method design was used. All four between subjects conditions were identical for the first three blocks, such that all participants had reminders during the PM tasks. Like the previous experiments, no participants had a reminder in the fourth block. Conditions only differed during encoding of the fourth block.

Prior to fourth block encoding, the expecting conditions were told they would have a reminder in during the next PM task, and the non-expecting conditions were told there would be no reminder. During fourth block encoding, participants in the shallow processing condition counted the number of letters in each target, and participants in the deep processing condition rated the pleasantness of each target. All materials were identical to those in Experiments 1 and 2.

Procedure

The procedure for everyone was identical to the expecting and non-expecting conditions in Experiment 1 through block 3, except that participants had 5-seconds to encode each target (like Experiment 2). After completing block 3, a new set of instructions appeared to describe the differences in block 4 encoding. These instructions had a salient *IMPORTANT INSTRUCTIONS* notice at the top of the screen explaining the processing manipulation. Instructions for shallow processing informed participants they would study each target for 5-seconds and then a second screen would appear prompting them to report the number of letters in that target word by pressing the 5-, 6-, or 7-key. The deep processing conditions were informed they would study each target for 5-seconds and then a second screen would appear prompting them to rate the pleasantness of that target word by pressing the 1-key (unpleasant), 2-key (neutral), or 3-key (pleasant). All participants then had to explain these new instructions to the experimenter before continuing. During encoding, the target appeared for 5-seconds, followed by a fixation, before the screen appeared prompting participants to report the number of letters or the pleasantness of the target. For the shallow condition, the prompting screen asked participants to report how many letters the previous target had from 5-7, whereas the prompting screen for the deep conditions asked participants to rate how pleasant the target was with a note to press the 1-key for *unpleasant*, the 2-key for *neutral*, and the 3-key for *pleasant*. After block 4 encoding, the rest of the procedure was the exact same as the expecting and non-expecting conditions in Experiment 1.

Exclusionary Criteria

The following exclusions were preregistered. Participants were excluded for the following: a) both failing to detect any PM targets *and* forgetting the prospective memory task (n = 1); b) getting below 60% accuracy on the PM block ongoing task (n = 7); c) having PM block ongoing task response times greater than 3 or less than -3 standard deviations from the group mean (n = 3); d) making false alarms (i.e., PM response on nontarget trials) on over 15% of trials (n = 1).

Results

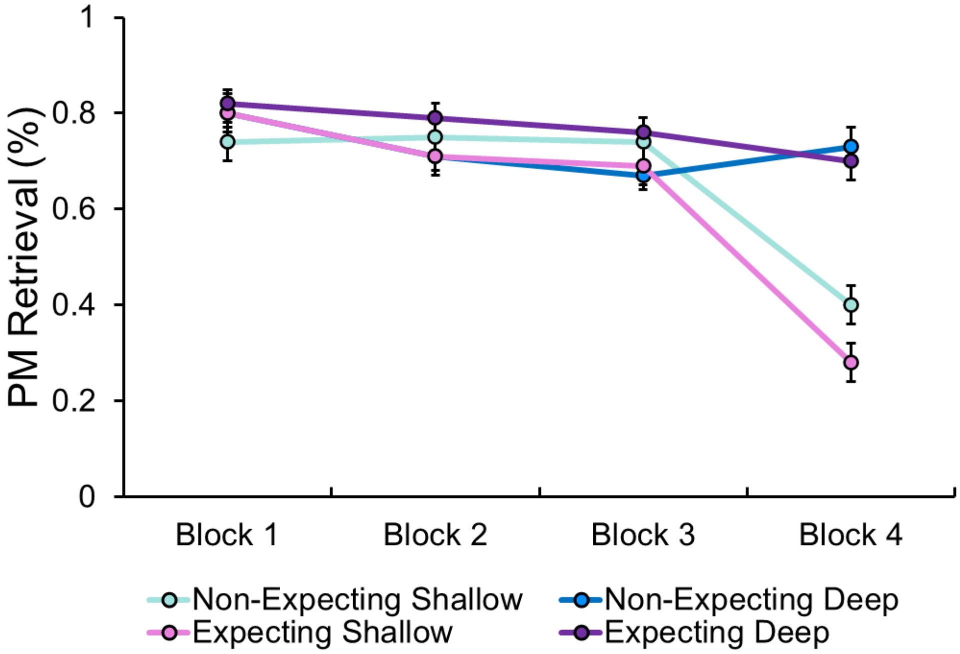
All four conditions were identical in the first three blocks. To verify no performance differences between conditions, we first submitted PM retrieval, uncontaminated recognition, and ongoing task performance (accuracy and response times) to 2 (Processing: shallow vs deep) x 2 (Expectation: expecting vs non-expecting) x 3 (Block: 1, 2, and 3) mixed-methods ANOVAs. Bonferroni-corrections ($p < .017$) were applied to post-hoc comparisons after a main effect of block. Block 4 analyses were preregistered conducted using a series of 2 (Processing: shallow vs deep) x 2 (Expectation: expecting vs non-expecting) to test the encoding effort hypothesis.

PM Retrieval

Figure 23 presents the results separately for each condition and block.

Figure 23

PM Retrieval Across All Four Blocks Separated by Condition in Experiment 3



Note. PM retrieval refers to the proportion of PM targets detected (out of four). The circles represent mean performance, and the error bars reflect standard error.

Blocks 1-3. In the first three blocks, there was only a significant main effect of block [$F(2, 332) = 4.67$, $p = .010$, $\eta_p^2 = .027$], confirming that there were no differences between identical between-subjects conditions. Post-hoc comparisons of the block effect (Bonferroni $p <$ .017) showed that while PM retrieval did not differ between Blocks 1 and 2 [$F(1, 169) = 4.09$, $p = .045$, $\eta_p^2 = .024$] or between Blocks 2 and 3 [$F(1, 169) = 1.05$, $p = .307$, $\eta_p^2 = .006$], PM retrieval was better in Block 1 than Block 3 [$F(1, 169) = 7.93$, $p = .005$, $\eta_p^2 = .045$]. This suggests a possibility of fatigue effects that worsened PM retrieval over time. There was no effect of expectation [$F < 1$], no effect of processing [$F < 1$], no interaction between expectation and processing [$F(1, 166) = 1.46$, $p = .229$, $\eta_p^2 = .009$], no interaction between expectation and block [$F < 1$], no interaction between processing and block [$F < 1$], and no three-way interaction between expectation, processing, and block [$F(2, 332) = 2.05$, $p = .130$, $\eta_p^2 = .012$].

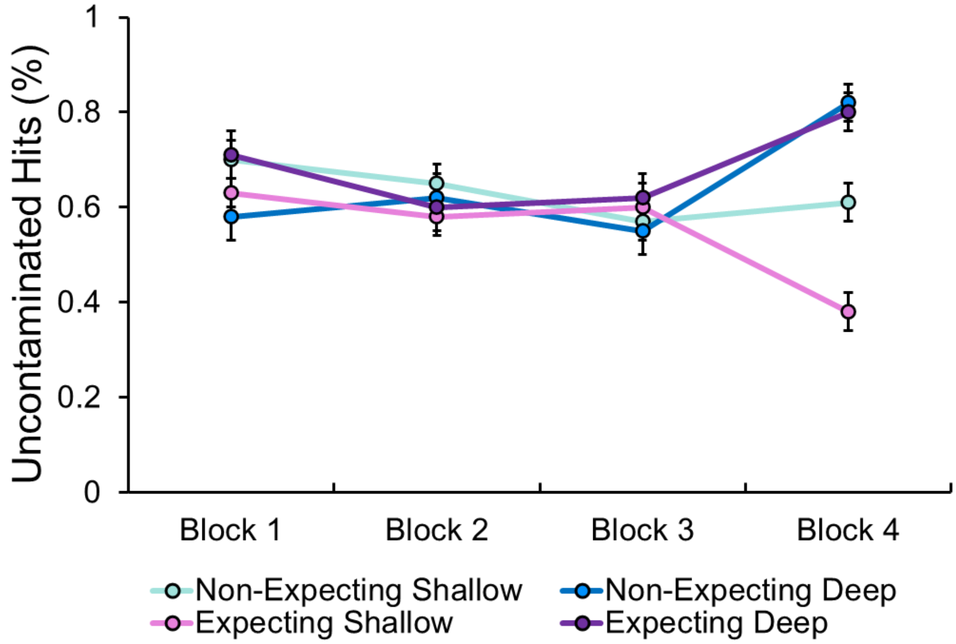
Block 4. While deep processing at encoding improved PM retrieval in Block 4 [$F(1, 166) = 69.18$, $p < .001$, $\eta_p^2 = .294$], there was no effect of expectation [$F(1, 166) = 2.63$, $p = .107$, $\eta_p^2 = .016$] and no interaction between processing and expectation [$F < 1$]. The encoding effort hypothesis was supported by finding deep processing at encoding removed the effect of reminder expectation on PM retrieval. Results were the same after excluding participants who failed the manipulation check, so they were not reported.

Uncontaminated Recognition

Figure 24 presents the results separately for each condition and block

Figure 24

Uncontaminated Target Recognition Across All Four Blocks Separated by Condition in Experiment 3



Note. Uncontaminated hits refer to the proportion of uncontaminated targets (out of four) correctly identified as "old" during the recognition memory portion of the task. The circles represent mean performance, and the error bars reflect standard error.

Blocks 1-3. In the first three blocks, there was only a significant main effect of block [$F(2, 332) = 4.20$, $p = .016$, $\eta_p^2 = .025$], confirming that there were no differences between identical between-subjects conditions. Post-hoc comparisons of the block effect (Bonferroni $p <$

.017) showed that while uncontaminated recognition did not differ between Blocks 1 and 2 [F(1, 169) = 2.99, p = .086, η_p^2 = .017] or between Blocks 2 and 3 [F(1, 169) = 1.32, p = .253, η_p^2 = .008], uncontaminated recognition was better in Block 1 than Block 3 [F(1, 169) = 8.47, p = .004, η_p^2 = .048]. There was no effect of expectation [F < 1], no effect of processing [F < 1], no interaction between expectation and processing [F(1, 166) = 2.03, p = . 156, η_p^2 = .012], no interaction between expectation and block [F(2, 332) = 1.89, p = .152, η_p^2 = .011], no interaction between processing and block [F < 1], and no three-way interaction between expectation, processing, and block [F(2, 332) = 1.35, p = .261, η_p^2 = .008].

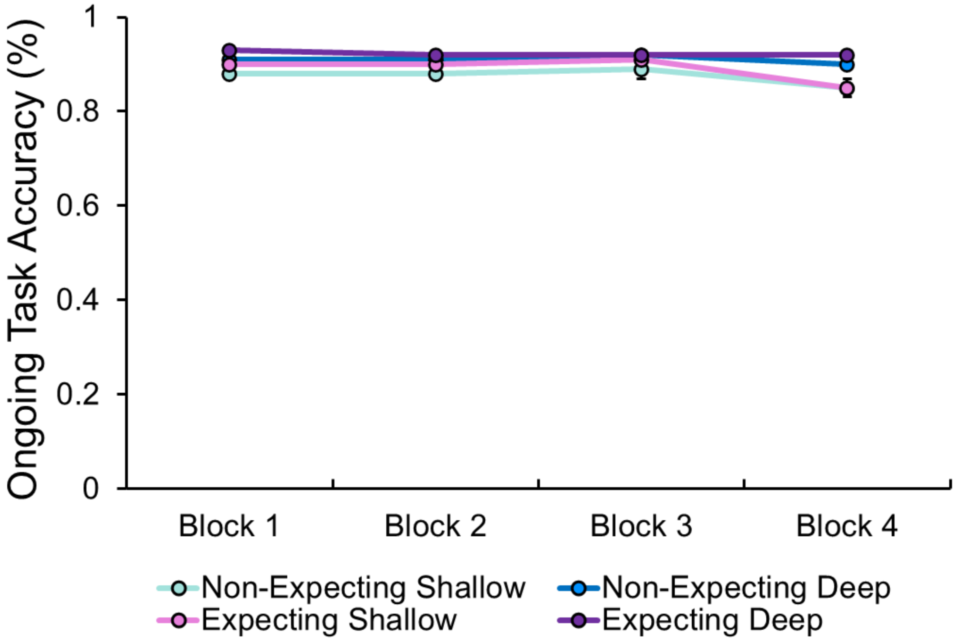
Block 4. Block 4 analyses supported the encoding effort hypothesis by showing uncontaminated recognition was better for deeply encoded targets [F(1, 166) = 59.28, p < .001, η_p^2 = .263] and better in the non-expecting conditions compared to the expecting conditions [F(1, 166) = 8.79, p = .003, η_p^2 = .050]. There was also a processing by expectation interaction [F(1, 166) = 6.59, p = .011, η_p^2 = .038]. Post-hoc comparisons of the block effect (Bonferroni p < .017) showed that while uncontaminated recognition did not differ when encoding was deep [F < 1], uncontaminated recognition was better in the non-expecting condition than the expecting condition when encoding was shallow [F(1, 85) = 14.03, p < .001, η_p^2 = .142]

Ongoing Task Performance

Accuracy. Figure 25 presents the results separately for each condition and block. *Blocks 1-3.* In the first three blocks, there was no effect of block [F_{GG} (1.88, 301.79) = 1.25, p = .287, η_p^2 = .007], no effect of expectation [F(1, 166) = 1.35, p = .248, η_p^2 = .008], no effect of processing [F(1, 166) = 1.46, p = .229, η_p^2 = .009], no interaction between expectation and processing [F < 1], no interaction between expectation and block [F < 1], no interaction between processing and block [$F < 1$], and no three-way interaction between expectation, processing, and block [$F < 1$].

Figure 25

Ongoing Task Accuracy Across All Four Blocks Separated by Condition in Experiment 3



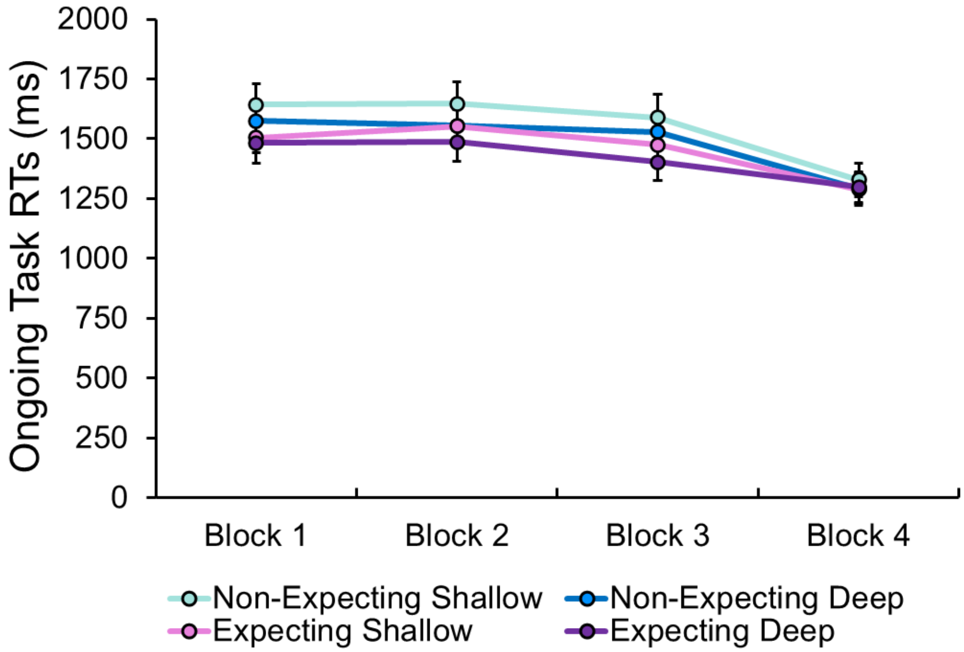
Note. Ongoing task accuracy refers to the proportion of accurate responses on non-target trials. The circles represent mean performance, and the error bars reflect standard error.

Block 4. While deep processing at encoding improved ongoing task accuracy in block 4 [$F(1, 166) = 11.87$, $p < .001$, $\eta_p^2 = .067$], there was no effect of expectation [$F < 1$] and no interaction between processing and expectation [$F < 1$]. This suggests that participants were able to allocate more attention to the ongoing task when they deeply encoded PM targets.

Response Times. Figure 26 presents the results separately for each condition and block. *Blocks 1-3.* In the first three blocks, there was only a significant main effect of block [F_{GG} (1.83, 302.88) = 5.30, $p = .005$, $\eta_p^2 = .031$], confirming that there were no differences between identical between-subjects conditions. Post-hoc comparisons of the block effect (Bonferroni $p < .017$) showed that while ongoing task response times did not differ between Blocks 1 and 2 [$F < 1$] or between Blocks 1 and 3 [$F(1, 169) = 5.16$, $p = .024$, $\eta_p^2 = .030$], ongoing task response times were faster in Block 3 than Block 2 [$F(1, 169) = 11.86$, $p < .001$, $\eta_p^2 = .066$]. This suggests either participants experienced practice effects or they allocated more attention to the ongoing task over time. There was no effect of expectation [$F(1, 166) = 2.01$, $p = .158$, $\eta_p^2 = .012$], no effect of processing [$F < 1$], no interaction between expectation and processing [$F < 1$], no interaction between expectation and block [$F < 1$], no interaction between processing and block [$F < 1$], and no three-way interaction between expectation, processing, and block [$F < 1$].

Figure 26

Ongoing Task Response Times Across All Four Blocks Separated by Condition in Experiment 3



Note. Ongoing task RTs refers to the average response times on non-target trials. The circles represent mean performance, and the error bars reflect standard error.

Block 4. Block 4 analyses showed no main effect of processing [$F < 1$], no main effect of expectation [$F < 1$], and no interaction between processing and expectation [$F < 1$].

Discussion

Experiment 3 tested the encoding effort hypothesis by manipulating depth of processing and reminder expectations at encoding in block 4. Importantly, PM retrieval, uncontaminated recognition, and ongoing task performance did not differ by condition in the first three blocks when the conditions were identical. Ongoing task performance was also unaffected by processing and expectation in block 4. Consistent with Experiments 1 and 2, uncontaminated recognition was better in the non-expecting condition when processing was shallow, suggesting participants in the non-expecting condition realized they would need to encode more effortfully without reminders even in the shallow processing condition. While there was no processing by expectation interaction for PM retrieval, Figure 23 shows the non-expecting condition had numerically better PM than the expecting condition. Critically, deep processing at block 4 encoding improved PM retrieval and uncontaminated recognition and removed differences between expecting and non-expecting conditions, supporting the encoding effort hypothesis. One potential limitation is that the shallow expecting condition had below chance uncontaminated recognition hits in block 4 ($p = .009$), despite valid contaminated recognition hits and correct rejections on new items (see Table 1). A possible explanation for this is discussed in greater detail in the General Discussion. Overall, the results of Experiment 3 suggest the detrimental effect of expecting a reminder on unaided memory can be attributed to processing at encoding.

General Discussion

The present study aimed to understand why reminder expectations reduce PM encoding effort and the possible negative ramifications this can have on intention fulfilment. The current study addressed the following gaps in the literature: 1) the relationship between encoding effort and PM retrieval with and without offloading (i.e., encoding effort hypothesis), 2) how reminder expectations influence encoding effort across time, and 3) whether there is evidence of the use it or lose it hypothesis (and the underlying mechanisms) in event-based PM. Behavioral results supported the encoding effort hypothesis by showing that expecting a reminder reduced study durations (Experiment 1), expecting a reminder lowered unaided PM retrieval and target recognition in block 4 (Experiments 1 and 2), and deep processing at encoding improved unaided PM (Experiment 3). Importantly, we found no evidence that using reminders previously decreased unaided memory (i.e., use it or lose it hypothesis). Below we discuss the theoretical implications of the present findings, limitations of the project, and future directions.

Previous research has shown consistent evidence for reminders improving memory when available at retrieval (e.g., Gilbert, 2015a, 2015b; Gilbert et al., 2020; Kelly & Risko, 2019, 2022; Peper et al., 2022; Vortac et al., 1995). We replicated these findings in Experiments 1 and 2 of the present study and found reminders helped PM retrieval without cost to ongoing task performance or recognition memory (Peper et al., 2022). In Experiment 2, reminders during the PM task also improved recognition memory for targets uncontaminated by appearing in the PM task. While this pattern was not observed in Experiment 1, it is possible that reminders can provide additional encoding opportunities that improve memory even when the reminder is no longer available. Eye-tracking data from Experiment 2 indicated that participants frequently checked the reminder on both non-target trials and target trials. The discrepancy in recognition performance between Experiments 1 and 2 could be explained by methodological differences. In Experiment 1, PM task trials were self-paced, such that the next trial began immediately after a response was made, whereas in Experiment 2, each PM task trial lasted 3000 milliseconds despite average response times ranging from 1200 to 1350 milliseconds. The additional time participants had between trials in Experiment 2 could have given those participants more opportunities to check the reminder and thus enhance the internal memory representation and the accessibility of the target items.

The present study adds to our previous work with this paradigm that uncovered theoretical mechanisms by which reminders can improve PM (Peper et al., 2022). Peper et al.

showed that specific reminders (i.e., for the exact PM targets) can improve target detection by making targets easier to *discriminate* from non-targets and categorical reminders (i.e., category PM targets) can help with the target *verification* process. That is, participants can check the reminder on a target trial to verify whether the stimulus is relevant for the intention. The results of Experiment 2 in the present study showed that participants with reminders checked them on over 50% of target trials, suggesting that even specific reminders help the target verification process. Interestingly, the present Experiment 2 also showed that participants with reminders had larger non-target TEPRs without impacting behavioral indices of ongoing task performance. There are two possible interpretations of this result that warrant further study. The first possibility is that reminders enhance the maintenance of the intention in working memory without compromising ongoing task performance. A second and more likely explanation is that reminders allowed participants to allocate more attention to the ongoing task, but this was not observed in ongoing task response times due to the frequency of reminder checks on non-target trials. Future research should address this possibility by removing reminder checking trials from ongoing task response times to determine whether reminders improve PM retrieval while simultaneously increasing the attention allocated to the ongoing task.

Our results generally supported the encoding effort hypothesis that pitted the strategic view of PM encoding against the perfunctory view. While the strategic view argues that more effortful encoding strategies improve PM retrieval (e.g., Gollwitzer 1999; McDaniel et al., 1998), the perfunctory view argues low effort encoding can sufficiently illicit successful PM retrieval (Scullin et al., 2018). All three experiments in the present study provide indirect evidence in favor of the encoding effort hypothesis and strategic view, with direct encoding evidence coming from Experiments 1 and 3. In block 4 of Experiment 1, the expecting condition encoded targets less effortfully (i.e., shorter study durations) and had worse unaided PM retrieval and recognition memory than the non-expecting condition. By forcing participants to strategically encode (i.e., engage in deep processing) in Experiment 3, PM retrieval and recognition memory were improved, and the difference between the expecting and non-expecting conditions was removed. Overall, deep processing improved PM retrieval and recognition memory over shallow processing with large effect sizes (both $\eta_p^2 > .260$). Experiment 2 replicated the behavioral results in Experiment 1 by showing the expecting condition had worse unaided PM retrieval and recognition memory than the non-expecting condition in block 4, but the pupillometry index of encoding effort made support of the encoding effort hypothesis inconclusive. Taking the encoding TEPRs in block 4 at face value, participants in the expecting condition encoded the targets with the same amount of effort as the non-expecting condition, despite memory differences. However, there is a limitation that constrained our encoding pupillometry analyses.

Few studies to date have applied pupillometry during a PM task, and the only ones to do so have looked at pupil dilation during PM maintenance and/or retrieval processes (Ball et al., 2022; Christopher, 2019; Moyes et al., 2019). Previous research has shown that TEPRs are a robust measure of attentional effort (Beatty & Lucero-Wagoner, 2000; Just & Carpenter, 1993; Kahneman & Beatty, 1966) and larger TEPRs at encoding are related to better memory retrieval (Ariel & Castel, 2014; Miller & Unsworth, 2021; Papesh et al., 2012). A primary methodological difference between those experiments and the present study is in the number of encoding trials used to calculate the TEPRs. We presented eight target items in each encoding block. Papesh et al. (2012) had participants encode 80 items, Miller et al. (2019) used 50 items, and Miller and Unsworth (2021) had 50 items and 90 items in their Experiment 1 and 90 items in their Experiment 2. Figure 16 in the current study shows the eight encoding TEPRs in each block separated by condition. Comparing that figure to Figure 8 that shows encoding TEPRs collapsed across condition and block (32 encoding trials), it is evident that the greater number of trials substantially reduces the standard errors in encoding TEPRs. While the few trials and large standard errors make it difficult to form solid conclusions about pupillary responses at encoding data, we found reminders increased uncontaminated recognition hit TEPRs with half the trials, and the pattern looked similar to behavioral uncontaminated recognition hits. However, considering the limitation of measuring pupillary responses over so few trials, both encoding TEPRs and uncontaminated hit TEPRs should be interpreted with caution. Alternatively, Gross and Dobbins (2021) tested whether encoding TEPRs represent attentional effort at encoding by manipulating depth of processing and time pressure at encoding. They found that deep compared to shallow processing at encoding did not influence encoding TEPRs but increasing time pressure increased TEPRs, suggesting encoding TEPRs better reflect time pressure at encoding rather depth of processing. In Experiment 2 of the present study, all participants encoded targets for five-seconds each so the time pressure was equal across all conditions and blocks. It is possible then that encoding TEPRs were not a valid assessment of encoding effort in the current study.

One apparent limitation that may serve as another interesting avenue of future study was observing below chance uncontaminated recognition hits for the shallow expecting condition in the fourth block of Experiment 3. An interesting explanation for this result comes from the phenomenon of retrieval-induced forgetting (Anderson et al., 2000; Anderson et al., 1994; Hicks & Starns, 2004), which refers to the finding that retrieving some information from a list impairs memory for unretrieved items from the list. In their original study, Anderson et al. (1994) had participants learn eight categories, then retrieved half of the categories during a practice test before a delayed test with all eight categories. Their results showed that impaired memory on the delayed test for the categories that were not retrieved during the practice test. Hicks and Starns (2004) also observed the retrieval-induced forgetting effect in recognition memory. Anderson et al.'s (1994) design resembled our own and has interesting implications for PM intention retrieval. In all three experiments, participants learned eight target words in each block and only half the targets appeared in each PM task. To control for PM retrieval effects, we only analyzed uncontaminated recognition hits (i.e., hits for targets that never appeared in the PM block). We report analyses in the appendix that found recognition memory was better for contaminated targets than uncontaminated targets in every block in every experiment regardless of reminder condition. It is therefore possible that when participants 1) expected to have a reminder and 2) encoded targets with shallow processing, retrieval-induced forgetting led participants to genuinely perceive uncontaminated targets as new items. To our knowledge no one has ever published a test of retrieval-induced forgetting in PM. In the real world, it is possible that retrieval-induced forgetting could exacerbate the effects of expecting to have a reminder when there is none at retrieval. For example, one who writes down their intentions on a to-do list (i.e., reminder) may complete some of them before their list is lost or destroyed. In this case, memory for the intentions left uncompleted may be worse than if none of the intentions were completed at all.

While our results revealed failed reminders are costly to unaided PM retrieval, the present study found no evidence for the use it or lose it hypothesis. In both Experiment 1 and 2, the non-expecting reminder condition showed levels of PM retrieval and recognition memory no different than the no reminder condition, despite relying on reminders previously. While a few studies found reliance on offloading impeded unaided performance (Henkel, 2014; Soares &

Storm, 2018; Tamir et al., 2018), others found past offloading did not impact performance (e.g., Scarampi & Gilbert, 2020) or even improved memory for items in a list that were not offloaded ("saving-enhanced memory"; Gilbert, 2023). Scarampi and Gilbert (2020) found no evidence of the use it or lose it hypothesis in the conceptually related intention offloading task. There are two possibilities for our results. The first is that the use it or lose it phenomenon does not exist. The second is that the use it or lose it hypothesis exists only across much longer time scales than we tested in the lab. Past research has found evidence for the use it or lose it hypothesis across shorter time scales (e.g., Soares & Storm, 2018), but it seems possible that the brain's plasticity could adapt to rely on offloading given enough time and use. This would be particularly interesting to study in children as their neural pathways more readily reorganize based on environmental stimulation (Johnston, 2009).

The Metacognitive View of PM Processing

Two primary processes underlying metacognition play a key role in theories of cognitive offloading (Risko & Gilbert, 2015) and can be applied to understand why reminder expectations may reduce encoding effort. Monitoring refers to one's awareness of their own cognitive performance, whereas control refers to the decision to update or maintain a cognitive strategy based on the assessment made by the monitoring process (Nelson & Narens, 1990). The knowledge updating framework describes four core processes involved in encoding strategy selection (Dunlosky & Hertzog, 2000). Strategies vary in *effectiveness*, and people can *monitor* differences in effectiveness and *update* their knowledge about how effective that strategy is before they *utilize* the newly acquired knowledge. One way people do this is by making confidence judgments about their performance during retrieval that are used to update their knowledge and influence which encoding strategy they select next. Previous research on cognitive offloading has shown having access to an external store (e.g., the internet) increases confidence in retrieval (Dunn et al., 2021; Pieschl, 2021). Experiencing the ease or effectiveness of retrieval with reminders (i.e., *monitoring*) could cause participants to *update* their knowledge about their belief in the *effectiveness* of reminders, thereby changing their subsequent encoding strategy (i.e., *utilization*) to engage in less effortful processing and earlier termination of study when expecting to have another reminder at retrieval (Kelly & Risko, 2022). We put forth this metacognitive view to explain *why* expecting a reminder leads to less effortful encoding.

Behavioral study durations in Experiment 1 were a sound way to assess the metacognitive view of reminder usage. Due to the limitations of our encoding TEPR analysis and the alternative explanation for what TEPRs at encoding measures, we discuss the metacognitive view in terms of the results from Experiment 1 and 3. Our results support the metacognitive view of reminder usage by showing how study durations changed differentially across blocks depending on whether participants expected a reminder. That is, participants with and without reminders encoded targets for the same duration in the first block, but participants with reminders encoded targets for less time in the second and third blocks, suggesting participants with reminders utilized their knowledge of the ease or effectiveness of retrieval in the first block to reduce encoding effort in subsequent blocks. Furthermore, the expecting reminder condition believed they would have a reminder in the fourth block and encoded targets for less time than the non-expecting condition, which led to the expecting condition having worse PM retrieval and recognition memory than the non-expecting condition. Study durations measured in Experiment 1 specifically map onto the “termination of study” aspect of metacognitive control during encoding (Nelson & Narens, 1994). That is, participants expecting a reminder terminated study sooner after their first experience using reminders at retrieval.

Another aspect of metacognitive control during encoding strategy selection is referred to as the “selection of kind of processing” (Bjork et al., 2013; Nelson et al., 1994; Nelson & Narens, 1990). Experiment 3 manipulated the type of processing participants used during block 4 encoding and found that reminder expectations did not influence PM retrieval or recognition memory when targets were encoded deeply. The design of the present study did not allow us to determine whether reminder expectations influenced what type of processing participants selected during encoding, but it is possible that the non-expecting condition engaged in deeper processing than the expecting condition when given the freedom to do so in block 4 of Experiments 1 and 2. A future study that employed the same design as Experiment 1 could test whether participants expecting a reminder utilized knowledge about retrieval effectiveness to select less effortful but inferior kinds of processing by asking participants about their strategies immediately after encoding.

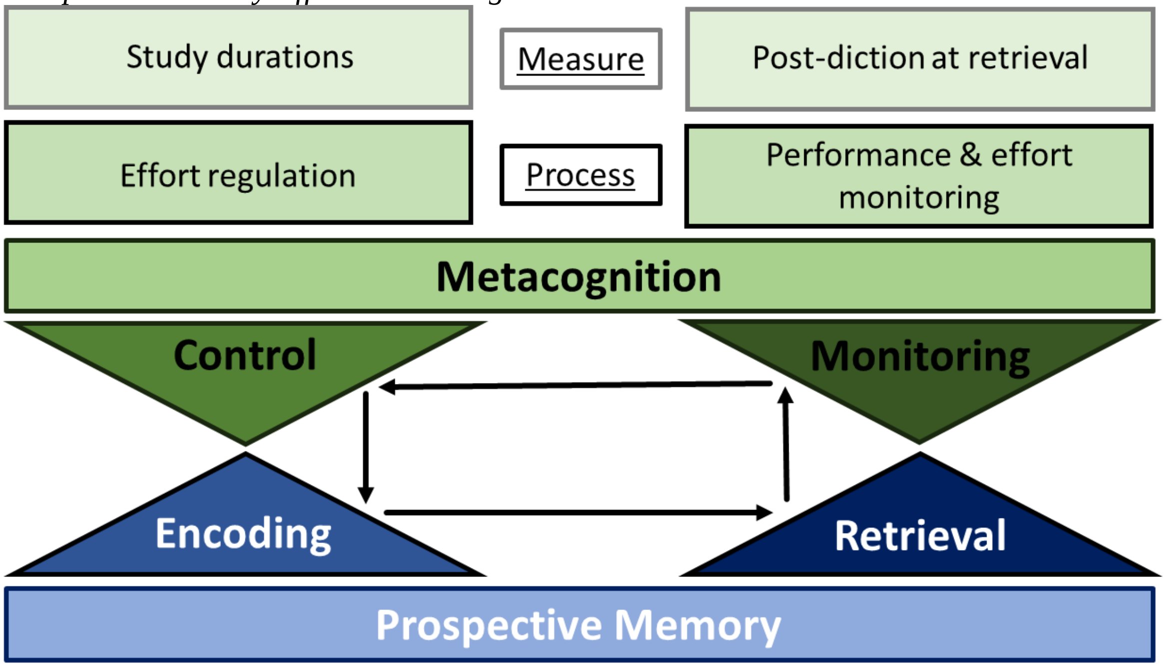
We can build upon the metacognitive view by including the effort monitoring and regulation framework (de Bruin et al., 2020). The effort monitoring and regulation framework is a marriage of cognitive load theory and self-regulated learning. This framework is particularly relevant for studies involving reminders due to the role of reminders in reducing cognitive load (e.g., Risko & Gilbert, 2016). The main idea of the effort monitoring and regulation framework is that cognitive load during encoding and retrieval inform monitoring and control processes. Cognitive load influences monitoring based on effort cues and one’s beliefs about effort, and load influences control by necessitating decisions about how to regulate one’s effort. In Experiment 1 of the current study, participants with reminders monitored the low effort cues during retrieval in block 1. Those same participants then used control to downregulate the effort they applied during encoding in block 2.

PM Effort Monitoring and Control Framework

Researchers have recently recognized the importance of including metacognitive processes in the study of PM (Kuhlmann, 2019). We developed the PM effort monitoring and control framework that builds on theories of metacognition and the findings in the current study (Figure 27). The main idea here is that people form expectations about the difficulty of retrieval via monitoring, and this changes the amount of effort spent encoding prospective memory targets via control. While most previous research in PM focuses on single blocks of encoding and retrieval, it is important to study multiple PM task blocks to understand how processes change over time and consider the metacognitive processes responsible for change (Dunlosky & Hertzog, 2000). In doing so, we can begin to understand how and when people learn (or fail to learn) from their PM errors. For this framework, monitoring at retrieval informs effort regulation (i.e., control) during the next encoding episode. Effort regulation can be measured with study durations. Monitoring performance and effort at retrieval could be measured with post-dictions after each prospective task by asking participants how many targets they detected and how much effort they put into retrieval.

Figure 27

Prospective Memory Effort Monitoring and Control Framework



Note. The PM effort monitoring and control framework describes the metacognitive processes operating at encoding and retrieval during multiple PM task blocks. We propose that 1) effort at encoding influences unaided PM retrieval, and 2) effort and performance at retrieval guide effort regulation during a subsequent encoding period. The framework details the specific metacognitive process engaged during monitoring and control and examples of how they can be measured. Critically factors that influence one component of the model (e.g., retrieval effort and performance) should have downstream effects on the other components (e.g., post-diction at retrieval and then study durations).

The PM effort monitoring and control framework is generative and produces testable predictions. When people are metacognitively calibrated, (i.e., when people have an accurate sense of their performance and effort at retrieval) their retrieval post-dictions should correlate with study durations. At retrieval, as effort declines and performance increases, study durations should also decline. Not only are reminders handy tools to improve our memory, reminders can also be used to test predictions theoretically-derived from this framework. Reminders reduce cognitive load and mental effort at retrieval, which we saw in the present study resulted in lower study durations. A next study could validate this framework using a simplified version of

Experiment 1, but also collecting post-dictions after the PM task in each block. We could also add a fifth block to see how the expecting condition reacted to having their reminder expectations violated. This framework would predict that the expecting condition would increase their study durations after experiencing the greater effort and worse retrieval without reminders in block 4.

Importantly, the PM effort monitoring and control framework can be applied to prospective memory generally. We can further test the framework by manipulating individual components within the framework and observing the downstream effects on subsequent components. For example, by making retrieval worse and more effortful, the next round of study durations should be longer. Giving participants false feedback to impact their retrieval monitoring post-dictions should influence the next round of study durations. By experimentally reducing effort at encoding, we should see the lower prospective memory retrieval reflected in the post-dictions. Once this model has been fully validated, it can be used to identify the mechanisms underlying prospective memory deficits in clinical populations. For example, people with ADHD show deficits in prospective memory(e.g., Altgassen et al., 2014), effort regulation (Wiersema et al., 2006), and metacognition (e.g., Butzbach et al., 2021) due to the role of attention in all three. It is possible that, compared to neurotypical individuals, people with ADHD show disruption in one or more components in the prospective memory effort monitoring and control framework. By pinpointing the mechanisms, clinicians would be able to create interventions designed to address specific deficits, such as reminders (retrieval), performance feedback (retrieval monitoring), and deep processing strategies (encoding control).

Conclusions

The current study replicated previous work showing reminders improved PM and without reminders, more effortful encoding produces better memory retrieval. Our results suggest people encode intentions less effortfully when they expect to have reminders at retrieval, which points to the importance of metacognition in the study of cognitive offloading. The present study also contributes to the theoretical understanding of how reminders improve PM and provides avenues for future research. For populations with PM deficits, setting reminders with technology has recently been recommended as a strategy for improving PM (e.g., Scullin et al., 2022). The results of the present study show that reminder recommendations should contain nuance to avoid unintended consequences of depending on reminders. As technology becomes more popular as a tool for offloading, it is clear from the current project that the technology must be reliable or the user could make critical errors if the technology fails, such as forgetting to take a medication when one’s phone dies.

APPENDIX

Supplementary Data Analysis

Contaminated Recognition Versus Uncontaminated Recognition

A series of paired-samples t-tests were conducted to compare contaminated recognition hits to uncontaminated recognition hits in each block for all three experiments. As described previously, participants studied eight targets, but only four of those appeared during the task. Because retrieving the intention during the task serves as an additional encoding opportunity that should occur more frequently in the reminder conditions due to better PM retrieval, targets were separated by *contaminated* (targets that appeared in the PM task) and *uncontaminated* (targets that did not appear in the PM task). Recognition hits were calculated by the taking the proportion of correct (i.e., “old”) responses made for PM targets during the recognition memory task at the end of the experiment.

In Experiment 1, recognition memory was better for contaminated targets than uncontaminated targets in block 1 [$t(168) = 4.62$, two-sided $p < .001$, $d = .355$], block 2 [$t(168) = 4.42$, two-sided $p < .001$, $d = .340$], block 3 [$t(168) = 5.58$, two-sided $p < .001$, $d = .429$], and block 4 [$t(168) = 5.39$, two-sided $p < .001$, $d = .414$]. In Experiment 2, recognition memory was better for contaminated targets than uncontaminated targets in block 1 [$t(132) = 5.79$, two-sided $p < .001$, $d = .502$], block 2 [$t(132) = 4.26$, two-sided $p < .001$, $d = .369$], block 3 [$t(132) = 3.52$, two-sided $p < .001$, $d = .305$], and block 4 [$t(132) = 4.82$, two-sided $p < .001$, $d = .418$]. In Experiment 3, recognition memory was better for contaminated targets than uncontaminated targets in block 1 [$t(169) = 4.10$, two-sided $p < .001$, $d = .315$], block 2 [$t(169) = 5.04$, two-sided $p < .001$, $d = .387$], block 3 [$t(169) = 5.57$, two-sided $p < .001$, $d = .427$], and block 4 [$t(169) = 6.77$, two-sided $p < .001$, $d = .519$]. Overall, these results suggest that the additional retrieval opportunity provided by targets appearing in the PM task (i.e., *contaminated* targets) strengthens the memory representation and accessibility of those targets.

Block 4 Uncontaminated Recognition Hits in Experiment 3

A one-samples t-test was conducted to compare uncontaminated recognition hits with chance (50%) for the expecting shallow condition in Experiment 3. Uncontaminated recognition hits were calculated by the taking the proportion of correct (i.e., “old”) responses made for block 4 PM targets that did not appear in the PM task during the recognition memory task at the end of the experiment. The results determined that uncontaminated recognition (M = .38, SE = .04) was significantly lower than chance [$t(42)$ = 2.76, two-sided p = .009, d = .420].

References

Altgassen, M., Kretschmer, A., & Kliegel, M. (2014). Task dissociation in prospective memory performance in individuals with ADHD. *Journal of Attention Disorders, 18*(7), 617-624.

Anderson, M. C., Bjork, E. L., & Bjork, R. A. (2000). Retrieval-induced forgetting: Evidence for a recall-specific mechanism. *Psychonomic bulletin & review, 7*, 522-530.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063.

Ariel, R., & Castel, A. D. (2014). Eyes wide open: enhanced pupil dilation when selectively studying important information. *Exp Brain Res*, *232*(1), 337-344. https://doi.org/10.1007/s00221-013-3744-5

Baldwin, V. N., Powell, T., & Lorenc, L. (2011). Factors influencing the uptake of memory compensations: A qualitative analysis. *Neuropsychological rehabilitation*, *21*(4), 484-501.

Ball, B. H., & Bugg, J. M. (2018). Context cue focality influences strategic prospective memory monitoring. *Psychonomic bulletin & review*.

Ball, B. H., Peper, P., & Robison, M. K. (2022). Aging and Prospective Memory Offloading. *Talk given at the Cognitive Aging Conference, Atlanta, Georgia.*

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007, Aug). The English Lexicon Project. *Behav Res Methods*, *39*(3), 445-459. https://doi.org/10.3758/bf03193014

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of psychophysiology*, *2*(142-162).

Binda, P., Pereverzeva, M., & Murray, S. O. (2013). Attention to bright surfaces enhances the pupillary light reflex. *Journal of Neuroscience, 33*(5), 2199-2204.

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual review of psychology*, *64*, 417-444.

Bowling, D. L., Graf Ancochea, P., Hove, M. J., & Fitch, W. T. (2019). Pupillometry of groove: evidence for noradrenergic arousal in the link between music and movement. *Frontiers in Neuroscience, 12*, 1039.

Butzbach, M., Fuermaier, A. B., Aschenbrenner, S., Weisbrod, M., Tucha, L., & Tucha, O. (2021). Metacognition, psychopathology and daily functioning in adult ADHD. *Journal of Clinical and Experimental Neuropsychology*, *43*(4), 384-398.

Chen, Y., Lian, R., Yang, L., Liu, J., & Meng, Y. (2017). Working memory load and reminder effect on event-based prospective memory of high-and low-achieving students in math. *Journal of Learning Disabilities*, *50*(5), 602-608.

Christopher, E. A. (2019). Using pupillometry to observe covert mental activity during prospective memory tasks. *Unpublished Doctoral Dissertation*.

de Bruin, A. B., Roelle, J., Carpenter, S. K., Baars, M., & EFG-MRE. (2020). Synthesizing cognitive load and self-regulation theory: a theoretical framework and research agenda. *Educational Psychology Review*, *32*, 903-915.

Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and aging, 15*(3), 462.

Dunn, T. L., Gaspar, C., McLean, D., Koehler, D. J., & Risko, E. F. (2021). Distributed metacognition: Increased bias and deficits in metacognitive sensitivity when retrieving information from the internet. *Technology, Mind, and Behavior*.

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

Einstein, G. O., & McDaniel, M. A. (2005). Prospective memory: Multiple retrieval processes. *Current Directions in Psychological Science, 14*(6), 286-290.

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), 479-488. http://www.ncbi.nlm.nih.gov/pubmed/9308095

Gilbert, S. (2023). Cognitive offloading is value-based decision making: Modelling cognitive effort and the expected value of memory. *[Preprint] PsyArxiv*.

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

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

Gilbert, S. J., Bird, A., Carpenter, J. M., Fleming, S. M., Sachdeva, C., & Tsai, P.-C. (2020). Optimal use of reminders: Metacognition, effort, and cognitive offloading. *Journal of Experimental Psychology: General, 149*(3), 501.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American psychologist*, *54*(7), 493.

Graf, P., & Uttl, B. (2001). Prospective memory: a new focus for research. *Consciousness and Cognition*, *10*(4), 437-450. https://doi.org/10.1006/ccog.2001.0504

Gross, M. P., & Dobbins, I. G. (2021). Pupil dilation during memory encoding reflects time pressure rather than depth of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *47*(2), 264.

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.

Hicks, J. L., & Starns, J. J. (2004). Retrieval-induced forgetting occurs in tests of item recognition. *Psychonomic bulletin & review*, *11*, 125-130.

Ihle, A., Schnitzspahn, K., Rendell, P. G., Luong, C., & Kliegel, M. (2012). Age benefits in everyday prospective memory: the influence of personal task importance, use of reminders and everyday stress. *Aging, Neuropsychology, and Cognition*, *19*(1-2), 84-101. https://doi.org/10.1080/13825585.2011.629288

Jainta, S., Vernet, M., Yang, Q., & Kapoula, Z. (2011). The pupil reflects motor preparation for saccades-even before the eye starts to move. *Frontiers in Human Neuroscience*, 97.

Johnston, M. V. (2009). Plasticity in the developing brain: implications for rehabilitation. *Developmental disabilities research reviews*, *15*(2), 94-101.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology*, *47*(2), 310.

Kafkas, A., & Montaldi, D. (2015). The pupillary response discriminates between subjective and objective familiarity and novelty. *Psychophysiology*, *52*(10), 1305-1316.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583-1585.

Kelly, M. O., & Risko, E. F. (2019). Offloading memory: Serial position effects. *Psychonomic bulletin & review*, *26*(4), 1347-1353.

Kelly, M. O., & Risko, E. F. (2022). Study effort and the memory cost of external store availability. *Cognition, 228*, 105228.

Kliegel, M., McDaniel, M. A., & Einstein, G. O. (2000). Plan formation, retention, and execution in prospective memory: A new approach and age-related effects. *Memory & cognition*, *28*(6), 1041-1049.

Kuhlmann, B. G. (2019). Metacognition of prospective memory: Will I remember to remember? In *Prospective Memory* (pp. 60-77). Routledge.

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition*, *1*(1).

McDaniel, M. A., Robinson-Riegler, B., & Einstein, G. O. (1998). Prospective remembering: Perceptually driven or conceptually driven processes? *Memory & cognition*, *26*(1), 121-134.

Meier, B., & Rey-Mermet, A. (2012). Beyond monitoring: After-effects of responding to prospective memory targets. *Consciousness and Cognition*, *21*(4), 1644-1653.

Miller, A. L., & Unsworth, N. (2021). Attending to encode: The role of consistency and intensity of attention in learning ability. *Journal of Memory and Language*, *121*, 104276.

Moyes, J., Sari-Sarraf, N., & Gilbert, S. J. (2019). Characterising monitoring processes in event-based prospective memory: Evidence from pupillometry. *Cognition*, *184*, 83-95.

Nelson, T. O., Dunlosky, J., Graf, A., & Narens, L. (1994). Utilization of metacognitive judgments in the allocation of study during multitrial learning. *Psychological Science, 5*(4), 207-213.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In *Psychology of learning and motivation* (Vol. 26, pp. 125-173). Elsevier.

Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. *Metacognition: Knowing about knowing, 13*, 1-25.

Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012, Jan). Memory strength and specificity revealed by pupillometry. *Int J Psychophysiol, 83*(1), 56-64. https://doi.org/10.1016/j.ijpsycho.2011.10.002

Peper, P., Alakbarova, D., & Ball, B. H. (2022). Differential benefits of prospective memory reminders depending on cognitive load [Preprint]. *Journal of Experimental Psychology: Learning, Memory, and Cognition*(Accepted). https://doi.org/10.31234/osf.io/6rms2

Pieschl, S. (2021). Will using the Internet to answer knowledge questions increase users' overestimation of their own ability or performance? *Media Psychology, 24*(1), 109-135.

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

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

Schnitzspahn, K. M., Kvavilashvili, L., & Altgassen, M. (2020, Jul). Redefining the pattern of age-prospective memory-paradox: new insights on age effects in lab-based, naturalistic, and self-assigned tasks. *Psychol Res, 84*(5), 1370-1386. https://doi.org/10.1007/s00426-018-1140-2

Scullin, M. K., Jones, W. E., Phenis, R., Beevers, S., Rosen, S., Dinh, K., Kiselica, A., Keefe, F. J., & Benge, J. F. (2022). Using smartphone technology to improve prospective memory functioning: A randomized controlled trial. *Journal of the American Geriatrics Society, 70*(2), 459-469.

Scullin, M. K., Kurinec, C. A., & Nguyen, K. (2017). The effects of implementation intention strategies on prospective memory cue encoding. *Journal of Cognitive Psychology, 29*(8), 929-938.

Scullin, M. K., McDaniel, M. A., Dasse, M. N., Lee, J. H., Kurinec, C. A., Tami, C., & Krueger, M. L. (2018). Thought probes during prospective memory encoding: Evidence for perfunctory processes. *PLoS One, 13*(6), e0198646. https://doi.org/10.1371/journal.pone.0198646

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

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.

Strickland, L., Loft, S., Remington, R. W., & Heathcote, A. (2018). Racing to remember: A theory of decision control in event-based prospective memory. *Psychological review, 125*(6), 851-887. https://doi.org/10.1037/rev0000113

Tamir, D. I., Templeton, E. M., Ward, A. F., & Zaki, J. (2018). Media usage diminishes memory for experiences. *Journal of Experimental Social Psychology, 76*, 161-168.

Unsworth, N., & Miller, A. L. (2021). Individual Differences in the Intensity and Consistency of Attention. *Current Directions in Psychological Science*, 09637214211030266.

Unsworth, N., & Robison, M. K. (2015). Individual differences in the allocation of attention to items in working memory: Evidence from pupillometry. *Psychonomic bulletin & review, 22*(3), 757-765.

Vortac, O. U., Edwards, M. B., & Manning, C. A. (1995). Functions of external cues in prospective memory. *Memory*, *3*(2), 201-219.

Wiersema, R., Van Der Meere, J., Roeyers, H., Van, C., Rudy, & Baeyens, D. (2006). Event rate and event-related potentials in ADHD. *Journal of Child Psychology and Psychiatry, 47*(6), 560-567.