USING PUPILLOMETRY TO OBSERVE COVERT MENTAL ACTIVITY DURING PROSPECTIVE MEMORY TASKS

by

Edward A. Christopher

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Psychological Sciences

West Lafayette, Indiana

May 2019

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Thomas S. Redick, Chair

Department of Psychological Sciences

Dr. Darryl W. Schneider

Department of Psychological Sciences

Dr. Jeffrey D. Karpicke

Department of Psychological Sciences

Dr. Robert W. Proctor

Department of Psychological Sciences

Approved by:

Dr. David Rollock

Head of the Graduate Program

"… if you have faith as small as a mustard seed, you can say to this mountain, 'Move from here to there,' and it will move. Nothing will be impossible for you."

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Author: Christopher, Edward A. PhD
Institution: Purdue University
Degree Received: May 2019
Title: Using Pupillometry to Observe Covert Mental Activity During Prospective Memory Tasks
Committee Chair: Thomas S. Redick

Remembering to complete some future intention (i.e., prospective remembering) is a frequent requirement of everyday activities. Prospective memory failures (e.g., forgetting to take one's medication) can have devastating consequences. Cognitive psychologists have sought to understand how individuals can successfully fulfill their prospective memory intentions. Unfortunately, it has been difficult to find evidence for specific cognitive mechanisms that could feasibly account for prospective memory behaviors. In part, this is because many theories of prospective memory stipulate that prospective remembering is accomplished through discrete/covert mental processes. In the current set of experiments, eye-tracking technology was used to test these various mechanistic explanations. Using an eye-tracking computer to measure pupillary responses to prospective memory task characteristics allowed for the observation of changes in discrete mental activity during the course of a prospective memory task scenario. Across two experiments, I observed elevated pupil dilation when participants were given additional prospective memory demands. Furthermore, when participants correctly recognized the presentation of a prospective memory target, it appeared that their pupil dilation increased dramatically, and elevated dilation persisted for several trials. This pattern of pupil dilation is consistent with an account of prospective remembering that suggests individuals sometimes engage in actively monitoring for an opportunity to

complete their prospective memory intention, and that at other times, individuals will reduce or discontinue monitoring activity until some cue brings the prospective memory intention back into mind. Consistent with such an account, individual differences in working memory were positively associated with pupil size only when the prospective memory task afforded monitoring. This was in line with recent research implicating the working memory system in facilitating active monitoring during certain prospective memory contexts. Finally, the current set of experiments demonstrated the utility of pupillometric methods for measuring active monitoring in a prospective memory scenario.

INTRODUCTION

The goal of this project was to elucidate exactly how individuals successfully remember to do things in the future. The ability to remember to do something in the future, known as *prospective remembering*, is crucial to everyday functioning. Imagine being at work, and learning that at the end of the workday, several hours later, you will need to retrieve your child from their school, which is not part of your normal routine. It would be inefficient to spend the rest of the day thinking about, and preparing for this future intention. Conversely, failing to recall the intention at all until the next day will have dire consequences. To successfully complete this prospective memory task, one must remember the intention at the appropriate time. Prospective memory tasks can be intuitively separated from retrospective memory tasks (i.e., retrieving a past event; Einstein & McDaniel, 2005; Maylor, 1990). For example, an individual with a heart condition can likely recall their having that condition when asked about it. Their ability to remember to take their heart medication at the prescribed time however, represents a separate, though clearly important task for their cognitive system. Indeed, one can easily bring to mind many such prospective memory tasks that are essential to functioning.

Prospective Memory

The study of prospective memory has most often been conducted in a laboratory environment, though experiments outside of the lab have also produced interesting results (Harris, 1984). For example, Kvavilashvili and Fisher (2007) tasked participants with remembering to call the experimenter at later dates, and further asked participants to keep a diary related to their prospective memory successes and failures. Eventually, they demonstrated that contrary to the popular belief among researchers at the time (Henry, MacLeod, Phillips, & Crawford, 2004), older adults maintained prospective memory abilities indistinguishable from the abilities of younger adults. The more common approach, however, when studying prospective memory, has been to recruit participants for a computerized task to be completed in the lab. This task almost always involved participants making keypress responses to stimuli presented rapidly on screen as part of some ongoing task, and additionally, some or all participants would have been instructed to make a different keypress response when they detected a prospective memory target stimulus, which occurred only infrequently. Typically, participants would practice the ongoing task prior to receiving the prospective memory instructions, and after prospective memory instructions would complete some sort of delay task prior to beginning the critical ongoing task trials where a prospective memory target could occur (Einstein & McDaniel, 2005).

For example, in a seminal paper describing their novel paradigm for studying prospective memory in the lab, Einstein and McDaniel (1990) had participants complete an immediate recall task, during which a prospective memory target was occasionally presented. On each trial of the immediate recall task, participants were presented with a list of words, and after the conclusion of each list of words, participants were prompted to recall the words they remembered. Participants practiced this immediate recall task for two trials, and then were given the prospective memory instructions. The experimenter instructed participants that they should press a particular key on the keyboard if they encountered the word “rake” during the task. The prospective memory target (i.e., the word “rake”) only appeared 3 times during the subsequent 42 trials, though participants were not made aware beforehand of the frequency with which the target word would be presented. Participants in this experiment correctly completed their prospective memory intention (i.e., pressed the key when presented with the word "rake") approximately half of the time (47%). Both the general design of Einstein and McDaniel's (1990) experiment, and the relative success of their participants in completing their prospective memory intention, was characteristic of modern prospective memory studies conducted in the lab.

Unfortunately, the mental processes supporting prospective memory were still not well understood. Many theories of prospective memory had identified the process of monitoring for the opportunity to complete one's prospective memory goal as a critical mechanism that could facilitate successful prospective remembering. Smith (2010) argued for a *preparatory attention and memory (PAM)* model, positing that for future intentions to be carried out successfully, an individual would always need to be actively monitoring for the appropriate opportunity to complete their prospective memory goal. *Multiprocess* theories of prospective memory have contended that monitoring processes were only situationally necessary for successful prospective remembering, with certain prospective memory contexts facilitating prospective memory success irrespective of monitoring processes (McDaniel & Einstein, 2000; Scullin, McDaniel, & Shelton, 2013).

In the past, the type of prospective memory target that participants were tasked with identifying, either focal or non-focal, was argued by these researchers (i.e., McDaniel & Einstein, 2000; Scullin et al., 2013) to determine the extent to which participants would engage in active monitoring. Prospective memory target focality has been defined as the extent to which the processing of ongoing task or environmental stimuli will facilitate the processing and detection of the prospective memory target. For example, if my prospective memory intention is to tell Frank happy birthday when I see him next, and while scanning the cafeteria for a friend to sit with at lunch I spot Frank, Frank would serve as a focal prospective memory cue. Conversely, if my prospective memory intention is to locate and ask someone from class for today's notes, and while looking for someone to sit with in the cafeteria I see Frank, who is in my class, Frank the stimulus would serve as a non-focal cue. In the later scenario, ongoing task processing is unlikely to facilitate recall of the prospective memory intention relative to the former task scenario. A focal prospective memory target, McDaniel and Einstein (2000) argued, would not require active monitoring processes, as participants could rely on a spontaneous retrieval mechanism. Under multiprocess theory, a spontaneous retrieval mechanism was believed to support prospective remembering in certain situations by bringing the intention to mind with little or no effort on the part of the individual remembering. Exactly how this occurs has not been well defined, but McDaniel and Einstein (2000) argued for the existence of this critical second mechanism that can support prospective memory without introducing costs to ongoing task performance.

Smith (2010) debated this point, and has maintained in PAM theory that monitoring is equally necessary for both focal and non-focal targets. In their *delay theory*, Heathcote, Loft, and Remington (2015) went even further than multiprocess theorists, arguing that prospective remembering is not a matter of monitoring or other target-processing mechanisms, but rather that prospective memory targets would be successfully detected when participants had sufficient time for processing the stimulus. In this theory, the necessary time for stimulus processing was determined by information processing thresholds that varied between participants depending on which components of the task they had prioritized.

Convincingly testing one or more of these differing accounts of prospective memory has proven difficult. Empirically testing these theories is hampered by the nature of monitoring. Monitoring has most typically been measured by comparing response times (RTs) on a task in either the presence or absence of a prospective memory intention. For example, Smith, Hunt, McVay, and McConnell (2007) presented participants with strings of letters, and instructed them to determine which letter strings represented words as opposed to non-words (i.e., a lexical decision task). As part of the experiment, participants were divided into one of three conditions. One group, those participants in the control condition, did not receive a prospective memory intention. Some participants were further told that if they saw a particular word (e.g., table), they should make a special keypress response. Finally, the remaining participants were told to make a special keypress response if they noticed a string of letters presented in red ink, as opposed to the white ink all other stimuli were to be presented in. Using this design, Smith et al. (2007) endeavored to test their theory that successfully performing a prospective memory task would incur RT costs (slowing), because participants would need to rely on active monitoring processes. They predicted this would be the case even when the prospective memory targets were focal and salient cues, respectively. Participants given one of these additional prospective memory tasks were indeed significantly slower throughout the lexical decision task compared to participants in the control condition. Their slower RTs were taken to be evidence that covert monitoring processes were interfering with completion of the lexical decision task.

Prospective Memory Theories

Using a task similar to the one employed by Smith et al. (2007), Einstein et al. (2005) demonstrated that when participants were presented with a more focal target for the prospective memory task (in this case, participants were responding to a specific word during a lexical decision task), they did not slow down during the ongoing task to an extent that was statistically significant. Based on this observation of a null result (Einstein et al., 2005), they proposed a multiprocess theory (Einstein & McDaniel, 2005; McDaniel & Einstein, 2000). These authors argued that participants would only slow down to monitor in specific prospective memory situations, whereas Smith and colleagues (2007) argued in their PAM theory that effortful monitoring is always necessary for prospective memory success. This disagreement led to an intense debate regarding whether RT differences were consistently observed in prospective memory paradigms (Einstein & McDaniel, 2010; Smith, 2010), and moreover, a debate over the utility of using RTs to measure covert monitoring activity in general. Ultimately, multiprocess theory was revised (Scullin et al., 2013), with those researchers claiming that individuals drift in and out of periods of actively monitoring for the opportunity to complete their prospective memory intention. This more nuanced/dynamic multiprocess theory predicted that monitoring would only occur in brief spurts during a prospective memory task, with the relative frequency of monitoring behaviors being more frequent in prospective memory scenarios where monitoring is necessary (e.g., a non-focal prospective memory task).

In response to these debates surrounding the sufficiency of RT data, Heathcote and colleagues (2015) suggested that longer RTs during ongoing processing tasks were not indicative of monitoring at all. They proposed a delay theory of prospective memory, asserting that the prospective memory component of a task (e.g., making a special response to a specific word) took longer to respond to because of a higher response threshold (Heathcote et al., 2015; Strickland, Loft, Remington, & Heathcote, 2018). One possibility is that because the prospective memory target rarely occurred in most tasks, participants set a high response threshold, and also increased the relative response thresholds for other components of the task (e.g., determining if a string of letters is a word or non-word) to accommodate the additional prospective memory demand. As Heathcote and colleagues pointed out, this adjustment of response thresholds would have led to slower responding in a prospective memory condition, without any covert monitoring activity.

Strickland et al. (2018) suggested that both PAM and multiprocess theories of prospective memory would have predicted that at least on non-focal prospective memory tasks, RT costs should result from evidence accumulation rates being slower in a prospective memory condition compared to control. They reasoned that slower rates of evidence accumulation in a prospective memory condition would have been consistent with theories that ongoing task and prospective memory task processing shared a common resource that only allowed for parallel processing of a stimulus towards both goals (i.e., monitoring for a prospective memory target necessarily detracted from an individual’s capacity to process stimulus elements related to a response for the ongoing task). Using a linear ballistic accumulator model, this group found that longer RTs in a prospective memory condition compared to a control condition were exclusively accounted for by changes in response threshold, and not differing rates of evidence accumulation (Heathcote et al., 2015; Strickland et al., 2018). They suggested that in a prospective memory condition, compared to a control condition, an independent evidence accumulator needed to be added to the model to account for the prospective memory response, but this extra evidence accumulator did not interfere with stimulus processing relative to either response for the ongoing task. Furthermore, their model suggested that there was no quantitative reason for differentiating between focal and non-focal prospective memory tasks, because evidence accumulation parameters for the prospective memory response did not vary between them (Strickland et al., 2018).

The delay theory of prospective memory has not been adopted by many prospective memory researchers, despite claims that it accounts for prospective memory behavioral data better than PAM and multiprocess models. Only recently have other labs started to test the delay theory. Anderson, Rummel, and McDaniel (2018) sought to pit delay and monitoring based accounts of prospective memory performance against each other. Anderson et al. (2018) gave participants the typical ongoing lexical decision task, but varied their prospective memory instructions to participants. In the first two conditions, they had participants either intentionally slow down, or encouraged them to actively monitor. In each of these conditions, participants were told that the strategy they were being given would lead to an increased ability for them to remember and complete their prospective memory intention. Additionally, participants in a third prospective memory condition were given the standard prospective memory instructions, which do not emphasize slowing down or actively monitoring. In all three of the prospective memory conditions, participants were assigned the non-focal task of identifying the syllable “tor” (Einstein et al., 2005). Anderson et al. (2018) predicted that performance in the delay and monitoring conditions would differ from performance in the standard prospective memory instruction condition to the extent that their respective mechanisms actually facilitated prospective memory behaviors.

The pattern of results that emerged was generally consistent with monitoring theories of prospective memory (Scullin et al., 2013; Smith, 2003). Anderson et al. (2018) reported that the delay instructions did result in participants slowing down, relative to the standard prospective memory instructions. This finding was critical, as it demonstrated that the manipulation was effective. Interestingly, prospective memory accuracy was significantly higher in the monitoring-emphasis condition than in either the delay or standard conditions. Heathcote et al. (2015) have suggested that individuals need only to slow down to remember to remember, but this advice is inconsistent with the findings of Anderson et al. (2018). The Anderson et al. (2018) behavioral data suggests that monitoring, not adjusting response thresholds, drives prospective remembering.

In addition to the results of Anderson et al. (2018), there are other empirical patterns that delay theory currently does not account for. Perhaps the most dramatic barrier between the delay theory and other accounts of prospective memory is that there is substantial evidence distinguishing performance and costs between focal and non-focal prospective memory scenarios (e.g., Brewer, Knight, Marsh, & Unsworth, 2010; Einstein & McDaniel, 2005; Einstein & McDaniel, 2010; Einstein et al., 2005; Scullin et al., 2013; but see also Smith, 2003, 2010). Even in datasets where RTs between focal and non-focal prospective memory conditions were not found to be significantly different, rates of prospective memory success are reliably higher in focal conditions. One question then, is how the delay theory modeling work has failed to differentiate between focal and non-focal tasks. Inspection of the details of the delay theory reveals that many of the model parameters are set to be very broad/flexible (Strickland et al., 2018). It is possible that this resulted in the model being insensitive to focality effects between prospective memory task conditions. Delay theory struggles to account for other previously reported findings in prospective memory research (e.g., the usefulness of implementation intentions; Gollwitzer, 1999), but a lack of differences between focal and non-focal prospective memory task scenarios is likely the most important sticking point.

Alternative Methods of Assessing Monitoring

Most empirical studies of how people successfully fulfill their prospective memory intentions have involved interpreting differences in ongoing task RTs, but a few studies have employed alternative measures of prospective memory processes. Reynolds, West, and Braver (2009) used functional magnetic resonance imaging (fMRI) to track brain activity throughout a prospective memory task, and did so in an attempt to detect activity indicative of active monitoring and spontaneous retrieval processes, as outlined by McDaniel and Einstein (2000). Reynolds et al. (2009) used an atypical prospective memory task. They had participants perform a verbal n-back for their ongoing task (specifically 1-back). Stimuli were presented in several different ink colors on a black background, but on the majority of trials, ink color was irrelevant. For the embedded prospective memory task, participants were told to respond by pressing a special key if a stimulus was presented in a specific color, and this color occurred infrequently. Using fMRI, Reynolds et al. (2009) observed that participants had sustained activation in their anterior prefrontal cortex during the prospective memory task, and that higher levels of sustained activation in this area was correlated with faster responding on prospective memory trials. They interpreted this as neurophysiological evidence for active monitoring during a prospective memory task. Additionally, Reynolds et al. (2009) saw that when participants correctly identified a prospective memory target, it appeared to generate transient activity along an area of the right middle temporal gyrus. They had predicted such transient activation would occur when participants were using spontaneous retrieval. Collectively, they interpreted their findings as implicating both monitoring and spontaneous retrieval mechanisms (McDaniel & Einstein, 2000), suggesting that their participants were able to use both top-down attention control to facilitate monitoring, and a bottom-up process of retrieving their prospective memory intention based on features of external stimuli (Scullin et al., 2013; Shelton & Scullin, 2017).

In another study, Czernochowski, Horn, and Bayen (2012) used event-related potential (ERP) data to again explore if and when participants might monitor when given a prospective memory intention. Czernochowski et al. (2012) manipulated the prospective memory demands of their task within participants. All participants began the experiment by completing a control block of lexical decision task only, and then participants completed two additional blocks of lexical decision, each with a different prospective memory component. For both prospective memory blocks, participants were told to make a special response if they saw a string of letters beginning with the letter “G”, “H”, or “M”. The prospective memory blocks varied in the frequency with which a prospective memory target appeared. In the low-frequency block, prospective memory targets appeared on only 3% of trials, but during the high-frequency block, prospective memory targets appeared on 20% of trials. All participants completed the control block first, due to concerns about their potential inability to inhibit prospective memory responding after receiving the prospective memory intention, but the order in which participants completed the low- and high-frequency prospective memory conditions was counterbalanced between participants. Czernochowski et al. (2012) found that in both of the prospective memory conditions, participants had sustained frontal positivity relative to the control condition, and they interpreted this as measuring active monitoring. Interestingly, while participants were able to more accurately respond to the prospective memory target in the high-frequency condition, and they had faster RTs, the sustained frontal activity was not significantly different between the low- and high-frequency prospective memory conditions. Czernochowski et al. (2012) concluded that the sustained frontal ERP observed in their participants represented participants' efforts to maintain a representation of their prospective memory intention until they could act on it.

Recently, in another attempt to circumvent the issues with relying on RT observations, a novel paradigm was devised to observe participants engaging in overt monitoring processes. Using an eye-tracking computer, Shelton and Christopher (2016) had participants identify living as opposed to non-living images in collages that were presented on-screen. A small portion of the screen was separated from the collage, and had individual images that changed frequently. Some participants were told that in addition to the living versus non-living judgment task, they should make a special response if they noticed an apple image appearing in the separated portion of the display. Participants with this additional prospective memory task frequently gazed at the non-collage images to overtly monitor for the prospective memory target image (i.e., an apple). By utilizing the eye tracker, Shelton and Christopher (2016) were able to observe monitoring processes as measured by fixations in the prospective memory target area, much more directly than by measuring monitoring using RT differences.

It is worth noting that the procedure used by Shelton and Christopher (2016) uniquely required overt monitoring, whereas most investigations of prospective memory are concerned with more covert/internal monitoring processes. I propose that a novel paradigm for more directly assessing monitoring processes is needed in prospective memory experiments. The current experiments accomplished this by using eye-tracking technology to track changes in pupil dilation during prospective memory tasks.

Pupillometry

Pupil dilation has previously been used to measure changes in attentional focus and cognitive load (Beatty, 1982; Kahneman, 1973; Kahneman & Beatty, 1966; Wierda, van Rijn, Taatgen, & Martens, 2012). Pupil dilation has been theorized to provide an accurate and timely index of cognitive load because it is particularly sensitive to functioning in the locus coeruleus norepinephrine (LC-NE) system. Aston-Jones and Cohen (2005) demonstrated this relationship between LC-NE and pupillary response in a monkey. They found that pupil dilation would promptly respond to changes in activation for LC neurons. Furthermore, Ashton-Jones and Cohen (2005) observed that monkey LC was on the receiving end of direct projections from anterior cingulate and orbitofrontal cortices, two frontal areas of the brain that they proposed would facilitate monitoring for task-related utility. From these observations, Ashton-Jones and Cohen (2005) reasoned that pupil dilation would make for a temporally sensitive measure of cognitive monitoring processes.

Further studies have found convergent evidence linking the LC-NE system and pupil dilation using fMRI (Hou, Freeman, Langley, Szabadi, & Bradshaw, 2005), and by explaining observed connections between pupil size and task performance because pupil size is largely reflective of LC inputs (Gilzenrat, Nieuwenhuis, Jema, & Cohen, 2010). In addition to the link between pupil size and LC-NE, Wang and Munoz (2015) suggested that pupil dilation is also reflective of activation in the superior colliculus (SC). Specifically, Wang and Munoz (2015) outlined how SC played a central role in the orienting of visual attention, and reported that microstimulation to SC in monkeys appeared to modulate attention for stimuli in the visual environment, and that this is further tied to rapid change in pupil dilation. Wang, Brien, and Munoz (2015) further pointed out that in humans, SC has been implicated in the top-down control of attention in tasks such as anti-saccade. In the anti-saccade task, participants must inhibit a prepotent saccade towards a stimulus that would typically lead to visual orienting (e.g., a flashing red image), and instead make a voluntary saccade away from that stimulus. Successful execution of an anti-saccade depends on top-down attentional control, relative to a pro-saccade task, wherein participants are told to make a saccade towards the stimulus. During the anti-saccade trials, the participants' pupil size was larger immediately preceding the presentation of an anti-saccade stimulus when they responded correctly, compared to pupil size prior to a correct pro-saccade trial or an erroneous anti-saccade trial (Figure 1; Wang et al., 2015). Wang et al. (2015) interpreted their results as supplementing accounts of pupil dilation reflecting LC-NE activation, which facilitates performing tasks requiring the control of attention. These authors further argued that SC also plays a role in orienting attention to prepare for the execution of an intention, and that changes in SC activation are reflected in pupil dilation.

In some of the earliest work relating pupillary responses to cognitive load, Kahneman and Beatty (1966) used an intricate paradigm to track pupil dilations during a short-term memory task while they manipulated cognitive load within participants. They defined cognitive load as the amount of information being processed at a given time. Operating in an era preceding modern eye-tracking technology, Kahneman and Beatty (1966) photographed participants' eyes from behind a one-way mirror during their task. For their part, participants were asked to perform immediate recall on lists of either 3 to 7 digits, 4 digits requiring an arithmetic transformation, and 4 monosyllabic nouns. Kahneman and Beatty theorized that lists with more digits would burden participants with a greater cognitive load than lists with few digits, and similarly that lists requiring a transformation would result in greater cognitive load than the other two types of lists, which required no modification to items. The researchers photographed participants' eyes once every second, beginning 5 seconds before the presentation of a list, and continuing for 4 seconds (i.e., 4 additional photographs) after participants concluded recall of a given list. This method allowed Kahneman and Beatty to measure changes in pupil dilation from before encoding began, through encoding and reporting, and then after participants finished reporting information. Then, using the photographs, they measured the diameter of participants' pupils, and compared pupil diameters between load conditions. These authors reported marked increases in pupil size during the encoding of lists, and more importantly, statistically significant increases in pupil diameter in conditions theorized to place the heaviest burden on cognitive load (i.e., 7 digits and digit transformation lists) compared to less demanding conditions. Kahneman and Beatty (1966) interpreted their findings as evidence that changes in pupil dilation can be used to index variations in cognitive load as participants are asked to process varying amounts of information.

The concept of cognitive load has previously been closely tied to working memory load (Barrouillet, Bernardin, & Camos, 2004; Unsworth & Robison, 2015). Working memory can be defined as a system for the temporary storage and manipulation of information (Baddeley & Hitch, 1974). Indeed, Unsworth and Robison (2015) recently demonstrated that pupil dilation increased when participants were asked to hold more items in working memory during a brief delay period. Consistent with Kahneman and Beatty's (1966) conclusion, Unsworth and Robison (2015) found that when they manipulated the amount of information that participants held in working memory for a short period of time, pupil diameter responded accordingly. To test this, Unsworth and Robison presented participants with arrays of one to eight colored squares. Each array was presented for only 250 ms, then following a 4000 ms delay, the array was presented again with one of the squares circled. Participants were tasked with identifying whether or not the selected square was the same color as originally presented. Participants' pupil diameter was found to be larger during trials that included a greater number of squares in the array (e.g., pupils were more dilated when presented with 8 squares compared to 2). Unsworth and Robison interpreted their data as evidence in favor of increased pupil dilation resulting from increased demands on the working memory system (Figure 2). These researchers even observed that fluctuations in pretrial pupil dilation were predictive of accuracy on the subsequent trial, similar to the anti-saccade data reported by Wang et al. (2015). Based on that finding, they interpreted pretrial pupillary fluctuations as indicative of lapses in attention, which they concluded were the result of insufficient working memory. In a follow-up study, Unsworth and Robison (2016) had participants complete a sustained attention task, and found again that lapses of attention, which they inferred would be associated with varying LC-NE activation modes, were well predicted by pupil dilation.

Recently, Wierda and colleagues (2012) demonstrated how pupil dilation could be used to track changes in what participants were attending to with a high temporal resolution (i.e., at the millisecond level). This level of temporal resolution is particularly beneficial to research questions that hinge on participants' rapid processing of information in the environment. They used an eye-tracking computer, specifically the Eyelink 1000 Plus, to track changes in pupil dilation within participants during an attentional blink task. Participants were presented with a stream of mostly non-target digit stimuli, and told to identify between 0 and 2 target letters. Individual stimuli were only presented for 100 ms, and at the conclusion of each trial, participants were asked to recall target stimuli (i.e., letters) in the order that they were presented. Wierda and colleagues (2012) observed significant spikes in pupil dilation coinciding with the presentation of target stimuli. Typically these changes in dilation occurred within 1000 ms of the target stimulus onset, indicating that participants required some time to process the stimulus. This paradigm of identifying the rare presentation of a target stimulus within a stream of non-target stimuli is similar to the typical prospective memory task where critical prospective memory targets are embedded within a lexical decision task. Wierda and colleagues (2012) reported on pupillary responses to the presentation of a target stimulus within a fairly simple task, and a natural extension of this line of research would be to measure changes in pupil dilation throughout a prospective memory task, so as to assess the extent to which participants engage in actively monitoring for a prospective memory target.

Working Memory

Research into monitoring processes has often asked how it is that individuals actively monitored for a prospective memory target, while also maintaining adequate performance on ongoing tasks. Smith and Bayen (2005) suggested that active monitoring was likely facilitated by the working memory system. Due the working memory system's role in temporarily maintaining and manipulating information within the focus of attention, Smith and Bayen (2005) reasoned that working memory would provide the necessary cognitive infrastructure for managing ongoing task demands and the need to prepare for an eventual prospective memory opportunity. Some of the early work that shaped the way future researchers thought about the working memory system centered on individuals' abilities to simultaneously process and encode distinct units of information, as well as on research demonstrating that short-term memory appeared to be related to complex cognitive activities, such as reading comprehension (Baddeley & Hitch, 1974; Daneman & Carpenter, 1980; Engle, Cantor, & Carullo, 1992; but see also Engle, 2018).

In their first experiment, Daneman and Carpenter (1980) illustrated the relationship between individual differences in working memory and reading comprehension in the following way. They tested participants' working memory by having participants read several series of sentences, and then prompting them to recall the final words of each sentence in order, at the conclusion of each sentence (i.e., a reading span task). Each series of sentences consisted of between 2 and 6 sentences, with sets of 3 series of each length presented to all participants, until they failed to answer correctly on all series from a set. Daneman and Carpenter (1980) then calculated the correlation between their participants' working memory as measured by the reading span task, and participants' scores on two measures of reading comprehension. The correlations were quite high (for fact questions $r = .72$; and for pronoun reference questions $r = .90$), and when, in a second experiment, they compared participants' working memory scores to listening comprehension instead of reading comprehension, Daneman and Carpenter (1980) found a similar pattern of results. All of this was taken as evidence that the working memory system played a critical role in complex behaviors such as reading and listening, and that working memory did this by facilitating the efficient processing, and perhaps, more specifically, by the active chunking of idea units (Cowan, 2001; 2010).

Subsequent research further identified a positive relationship between individual differences in working memory and a variety of other cognitive abilities, including executive attention (Engle, 2002), long-term memory (Cantor & Engle, 1993), and fluid intelligence (Kyllonen & Christal, 1990). Indeed, several theories of working memory highlighted the need for the working memory system to interact with information stored in long-term memory (e.g., Oberauer, 2009). Yet again, reading comprehension serves as a good example of the role that the working memory system plays in complex cognition. In comprehending the text in a novel, a reader must frequently incorporate newly presented ideas (i.e., what is currently being read) with previously encountered and stored ideas (i.e., information from a previous chapter in the novel). Additionally, one of the abilities most commonly connected to working memory has been the ability to inhibit goal-irrelevant information (Conway, Cowan, & Bunting, 2001; Hasher & Zacks, 1988; Unsworth & Engle, 2007). One demonstration of this was provided by Conway et al. (2001), whom observed that performance on a dichotic listening task varied as a function of working memory. In the dichotic listening task (Moray, 1959) participants are asked to shadow speech heard in one channel (e.g., the right ear), while ignoring speech in the other channel. Conway et al. (2001) presented participant's own names in the to-be-ignored channel of speech, and later asked participants if they heard their name. Participants that had scored lower on tasks measuring individual differences in working memory were more likely to report having heard their name, and this was interpreted as evidence that participants lower in working memory were less able to inhibit distracting information.

In defining the working memory system, Engle and Kane (2004) conceptualized working memory as facilitating the active maintenance of goal-relevant information within mind, particularly in spite of distractors. Brewer et al. (2010) noted that this description of working memory ability is particularly similar to what is required in many prospective memory tasks, when participants must attempt to monitor for a prospective memory target while maintaining a suitable level of performance on some ongoing task. To explore potential working memory differences in prospective memory performance, Brewer et al. (2010) grouped participants as being either high or low working memory spans. They accomplished this by standardizing (transforming to *z*-scores) and combining (averaging) participant's scores on the operation-span and reading-span tasks, and then only analyzing data from participants whose combined working memory score was in either the first or fourth quartile. They then asked participants to complete focal and non-focal prospective memory tasks. As an ongoing task, participants in the Brewer et al. (2010) experiment completed blocks of lexical decision, deciding if strings of letters were a word or a non-word.

After a baseline block of lexical decision, all participants received one of the two sets of prospective memory instructions. In the focal condition, participants were told to press a special key if they saw the word “packet”. In the non-focal condition, participants were told to press a special key if they saw the syllable “tor” (Einstein et al., 2005). The syllable “tor” only appeared on word trials, and both conditions presented a target on the 25^{th}, 50^{th}, 75^{th}, and 100^{th} trials. This allowed Brewer et al. (2010) to compare prospective memory performance between working memory span groups and between prospective memory target focalities. They observed statistically significant span differences on the non-focal prospective memory task, such that individuals higher in working memory appeared to be better able to detect the critical prospective memory target. However, on the focal prospective memory task where processing of the ongoing lexical decision task was hypothesized to facilitate detecting the prospective memory target, they failed to identify a statistically significant relationship between working memory and prospective memory success. In addition, although non-focal lexical decision task RTs were slower than the focal and control conditions, there was no significant difference between working memory span groups. Brewer et al. (2010) concluded that working memory was only related to prospective memory performance when the prospective memory task necessitated active monitoring.

Christopher, Fansher, and Redick (in prep) conducted a relatively large experiment in which participants (n = 299) completed a lexical decision task as their ongoing task, and were given one of two prospective memory intentions. Participants in the non-salient condition were told that in addition to the lexical decision task, they should make an alternative keypress if they ever noticed a word describing a color (e.g., the word “blue”). In the salient condition, participants were instructed to make an alternative keypress if they ever noticed a word presented in red ink (Smith et al., 2007), as opposed to the black ink color all other stimuli were presented in. In Christopher et al. (2019), working memory was measured by combining scores on the operation-span and symmetry-span tasks, similar to the method used by Brewer et al. (2010), though it is important to note that participants from across the entire spectrum of working memory ability were included for analysis (i.e., participants were not grouped, as in Brewer et al., 2010). Christopher et al. (2019) further extended the work of Brewer et al. (2010) by using Bayesian analyses. Given that Brewer et al. (2010) reported that individual differences in working memory were not significantly related to prospective memory accuracy, it seemed necessary to use Bayesian methods, which are uniquely capable of finding and quantifying evidence supporting a null hypothesis.

We (Christopher et al., 2019) found that when prospective memory targets were particularly salient, individual differences in working memory were unrelated to successful prospective remembering. It was only when the prospective memory target was not salient, and therefore the need to monitor was likely greater, as with non-focal targets, that individual differences in working memory were positively correlated with successful prospective memory recall (Figure 3). Based on previous literature describing ongoing task RT costs as the result of active monitoring (Smith et al., 2007), and our belief that the working memory system facilitated monitoring, we had hypothesized that individual differences in working memory would be correlated with ongoing task RTs. Instead, we found evidence favoring a null hypothesis in both the salient and non-salient prospective memory conditions. Using ongoing task RTs as a measure of in-task monitoring for the prospective memory target, we were unable to detect a more direct relationship between participants' monitoring and individual differences in working memory. This was despite the relationship between individual differences in working memory and prospective memory accuracy in the non-salient condition, which to us indicated, albeit less directly, that monitoring was differentially utilized in the non-salient condition more than in the salient condition.

Furthermore, in our data, there was no difference in ongoing task RTs between the salient and non-salient conditions. A multiprocess view of prospective remembering would have predicted that participants in the salient condition would have been faster than those in the non-salient condition; however, a lack of RT differences is actually consistent with Smith et al. (2007), which demonstrated that costs still occurred within a salient prospective-memory-target condition. Smith et al. (2007) concluded that prospective memory tasks in which the prospective memory target cue is salient will still lead participants to actively monitor in a fashion that is no different than what would be expected if participants needed to use a non-salient target to cue their prospective memory intention. Based on our finding (Christopher et al., 2019) that individual differences in working memory were only correlated with accurate prospective remembering in the non-salient prospective memory condition, it would appear that participants do perform prospective memory tasks differently based on target-cue characteristics. Specifically, it may be that monitoring, and by extension working memory, are much more important in some prospective memory situations than others. To the extent that this is true, it would appear from the Christopher et al. (2019) data that ongoing task RTs are not a particularly sensitive measure for detecting these differences in prospective memory tasks (Einstein & McDaniel, 2010; Smith, 2010).

While previous studies had investigated the relationship between individual differences in working memory and prospective memory (Brewer et al., 2010; Smith & Bayen, 2005; West, Bowry, & Krompinger, 2006), there are many advantages of our approach versus other research. First, our study used a large sample size better suited for individual differences research, in contrast to most previous studies. We also used multiple working memory span tasks to measure individual differences in working memory, removing the influence of task-specific measurement differences when using one working memory task (Conway et al., 2005). We used the full distribution of working memory scores, instead of creating high/low working memory groups based on medians or quartiles. Finally, our study was the first to use Bayesian analysis techniques. The strength of this Bayesian approach is that it uniquely allowed us to find evidence in favor of a null hypothesis, such as the predicted lack of a relationship between working memory and focal/salient prospective memory. This analytical approach has been missing from previous prospective memory experiments, despite a priori predictions of null results (Einstein et al., 2005).

The conditional relationship between individual differences in working memory and prospective memory performance (Brewer et al., 2010; Christopher et al., 2019) would not have been predicted by PAM theory (Smith, 2010; Smith et al. 2007; Smith & Bayen, 2005), but was more consistent with a multiprocess model of prospective memory (McDaniel & Einstein, 2000). As we reasoned in Christopher et al. (2019), the PAM theory asserts that regardless of the situation, the only effective mechanism for completing a prospective memory intention would be active monitoring, which we argued would be facilitated by the working memory system. Conversely, a multiprocess model of prospective memory would selectively emphasize the need for monitoring process in situations where the prospective memory target was not focal/salient. It is less clear what the delay theory of prospective memory would have predicted regarding the relationship between working memory and prospective memory, because delay theory is less directly concerned with mechanistic explanations of prospective memory behavior.

Delay theorists have however been explicit that their model did not distinguish between focal and non-focal prospective memory (Strickland et al., 2018), and this would likely have extended to a predicted lack of differences between attention-demanding and non-attention-demanding prospective memory tasks (Christopher et al., 2019; Smith et al., 2007). Based on this, the delay theory would predict that any relationship between individual differences in working memory and prospective memory success would be consistent between prospective memory tasks. Furthermore, it seems that delay theory would have predicted that it was unnecessary for prospective memory processes to recruit the working memory system to complete any prospective memory intentions. This prediction is less obvious than the former, but given that delay theory suggested prospective memory performance did not rely on any shared cognitive resource to process ongoing task and prospective memory task elements of a given stimulus (Heathcote et al., 2015; Strickland et al., 2018), it is unclear why the working memory system would be particularly burdened by a prospective memory scenario. Importantly, neither of these predictions was consistent with our results. Our observation, that working memory was only related to prospective memory success when participants were tasked with responding to a target that was not salient, was taken as evidence favoring a multiprocess approach (Brewer et al., 2010; McDaniel & Einstein, 2000; Scullin et al., 2013).

As previously noted, individual differences in working memory have been shown to be predictive of performance on tasks in a number of different cognitive domains beyond prospective memory (Engle, 2002). Motivated out of concern that individuals scoring high on measures of working memory might simply be giving more effort generally, which could have been artificially creating or inflating the relationship between working memory and other higher-order cognitive abilities, Heitz, Schrock, Payne, and Engle (2008) manipulated incentives within working memory groups, and then measured mental effort using pupil size. Heitz et al. (2008) initially sorted participants into the first and forth quartile of working memory ability based on their performance on the operation-span task. In a subsequent session, participants returned to the lab to complete a reading span task, while their pupil size was recorded using an eye-tracking computer. On the reading span task, participants were presented with series of sentences and letter pairs. When presented with a sentence, participants were to read it aloud, and the following letter was also to be read aloud, and remembered for later recall. Scores on the reading span task were based on a participant’s ability to recall letters in the order that they were presented. Performance on the operation-span and reading-span tasks has been shown to be positively correlated (e.g., $r = .61$ in Heitz et al., 2008), so it was expected that participants in the top quartile would perform better on the reading-span task than participants in the bottom quartile. Importantly though, if typical working memory effects were really being driven by differences in mental effort, then increasing incentives for good performance should have produced an interaction, such that participants previously low in working memory would perform better than expected on the reading-span task. Conversely, participants high in working memory would not have been expected to respond very much when given high incentives, if they were already displaying high levels of mental effort.

Heitz et al. (2008) found that when they offered participants greater monetary incentives for improved performance on the reading-span task, pupil size and performance increased for both high- and low-working memory groups. Participants that initially reached higher scores on the operation-span task typically scored higher on the reading-span task, and participants in both working memory ability groups scored higher when incentives were increased. Their conclusions were later extended by Tsukahara, Harrison, and Engle (2016), whom found that mental effort could not explain the relationship between pupil size and working memory. Importantly, Tsukahara et al. (2016) reported that participants high in working memory ability and participants low in working ability showed equivalent increases in pupil size under increased cognitive load (Figure 4). Furthermore, Unsworth and Robison (2017) proposed an LC-NE account describing the relationship between individual differences in working memory and attention control, using pupil dilation as a proxy for LC-NE functioning.

Current Research

As I reviewed, there have been many studies designed to provide evidence for the different accounts of prospective memory, but few experiments have come up with conclusive evidence for or against the three main theories (PAM, multiprocess, delay). Part of the challenge is the difficulty in the use of ongoing task RTs as indicators of monitoring processes. Given the connection between pupil dilation and cognitive load, the current research was designed to use this signal as an ongoing indicator of monitoring processes. In addition, to further explore the connection between individual differences in working memory and prospective memory, I collected data from a large sample of young adults using tasks commonly used in both literatures. Finally, given the central importance of null relations for specific predictions and theories, I used Bayesian analyses to assess what model the evidence was most consistent with.

I reasoned that employing an active monitoring mechanism to prepare for prospective memory targets during an ongoing task (e.g., a lexical decision task), would lead to increases in the amount of information being processed in working memory (Brewer et al., 2010; Christopher et al., 2019; Smith & Bayen, 2005). I endeavored to measure this additional working memory load created by monitoring processes through changes in pupil dilation. My hypothesis was that when the parameters of the prospective memory task encouraged monitoring, pupil dilation would increase during the task (Kahneman & Beatty, 1966), and prospective remembering would be positively correlated with individual differences in working memory in these scenarios (Christopher et al., 2019). More specifically, I predicted that monitoring as measured by pupil dilation would moderate the relationship between individual differences in working memory and prospective memory performance. Furthermore, even on prospective memory tasks where participants were most likely to monitor (e.g., when a non-focal prospective memory target was used), pupil dilation was expected to fluctuate between ongoing task trials, with dilation in this high cognitive load condition sometimes mirroring dilation during a control/no-prospective memory instruction condition. This pattern of results would be consistent with the dynamic multiprocess view of prospective memory (see Figure 5 for an example of the pupil dilation patterns predicted by various theories of prospective memory), which allows for lapses in monitoring activity during a prospective memory task (Scullin et al., 2013). To summarize, the specific research questions that were investigated, and what each of the three major theories of prospective memory would predict, are as follows:

1. Does pupil dilation vary between prospective memory and control conditions?
    a. PAM: Yes, pupil dilation in any and all prospective memory conditions should be greater than in a control condition because of active monitoring.
    b. Multiprocess: Sometimes, depending on the nature of the prospective memory task. For example, pupil dilation would be similar in focal prospective memory and control conditions.
    c. Delay: Yes, pupil dilation in any and all prospective memory conditions should be greater than in a control condition because of a greater number of potential responses.
2. Does pupil dilation vary between focal and non-focal prospective memory conditions?
    a. PAM: No, because monitoring levels should not vary between prospective memory conditions based on focality.
    b. Multiprocess: Yes, because active monitoring will be mostly unnecessary during focal prospective memory tasks.

    c. Delay: No, there is no reason to distinguish between focal and non-focal prospective memory tasks.

3. Is pupil dilation correlated with individual differences in working memory?
    a. PAM: Yes, because working memory facilitates monitoring, and active monitoring is reflected in the pupillary response.
    b. Multiprocess: Sometimes, because working memory facilitates monitoring, and active monitoring is reflected in the pupillary response, individual differences in working memory will be correlated with pupil dilation when the prospective memory task requires participants to engage in active monitoring.
    c. Delay: No, because monitoring is unnecessary.
4. Is pupil dilation correlated with prospective memory accuracy?
    a. PAM: Yes, in all prospective memory tasks there will be a positive correlation.
    b. Multiprocess: Sometimes, there will be a positive correlation when the prospective memory task requires active monitoring (e.g., a non-focal prospective memory task).
    c. Delay: No, pupil dilation will be unrelated to prospective memory accuracy because participants do not need to engage in active monitoring.

5. Is pupil dilation correlated with ongoing task RTs?
    a. PAM: Yes, because both are an index of active monitoring.
    b. Multiprocess: Yes, because both are an index of active monitoring.
    c. Delay: No, because slowed responding during a prospective memory task has nothing to do with active monitoring.

EXPERIMENT 1

To test these predictions, I utilized eye-tracking technology to measure covert monitoring processes that were expected to occur during a commonly used prospective memory task paradigm, the lexical decision task. In addition to extending the literature on prospective memory by using eye-tracking technology, I also used larger samples and more prospective memory targets than is typical, and analyzed the data using Bayesian methods of analysis. Collectively, these elements allowed me to measure prospective memory processes more precisely than what has previously been possible. Elements of the prospective memory task were manipulated between participants.

Method

Apparatus

Participants were asked to complete a lexical decision task on the EyeLink 1000 Plus, which is the eye-tracking computer Wierda and colleagues (2012) reported had sufficient temporal resolution (SR Research Ltd, 2014). During all experiments, the EyeLink was monocularly recording at 1000 Hz. In addition to accuracy and RT data, pupil dilation was measured throughout the task. On the EyeLink 1000 Plus, pupil size is measured in arbitrary units, and is accurate to within 0.01 mm. Because the units are arbitrary, participants' pupil data were considered in terms of change relative to pupil size during an initial baseline period (SR Research Ltd, 2014). Furthermore, as the measurement of pupil size is sensitive to head orientation, all participants were required to use a chin rest when on the EyeLink 1000 Plus. The chin rest ensured a visual angle of 161°, and kept participants 36 in. from the 27 in. display monitor. The monitor had a resolution of 1920 x 1080. According to the EyeLink 1000 Plus user manual (SR

Research Ltd, 2014), the use of a chin rest and a central fixation point between trials (as described in the experimental methods below) is sufficient to ensure accurate measurement of pupil size.

Participants

Undergraduate students from Purdue's Introductory Psychology course (n = 244) were recruited to participate in Experiment 1, in exchange for credits going towards partial completion of the course. From this larger sample, data from participants was excluded for a variety of reasons, prior to analyzing the final sample's data (n = 182). Due to the verbal nature of the lexical-decision task, participants' data were excluded when they self-identified as non-native English speakers (n = 20). Participants' data were further excluded if they failed to achieve 50% accuracy on the lexical decision task (n = 1), or on the processing portion of either of the complex span tasks (n = 1). Finally, to ensure that participants encoded the prospective memory intention, and that prospective memory failures were not due to lapses in retrospective memory, participants' data were excluded if they failed to recall what the prospective memory intention was at the end of the lexical decision task (n = 40). Specifically, participants were asked what they were supposed to remember to do in addition to performing the word vs. non-word task, and data were only included from those participants that indicated recalling being instructed to press the space bar when presented with certain targets. In the final sample, participants primarily identified as female (66%), and were between 18 and 24 years of age (M = 18.79, SD = 1.08).

Procedure

Participants completed the experiment individually, and in the lab. All participants were randomly assigned to one of three conditions. In a control condition, participants completed the lexical decision task without any prospective memory instructions. Participants in the other two conditions received either focal or non-focal prospective memory instructions as a part of the lexical decision task. Following the lexical decision task, all participants answered some demographic questions and completed two complex span tasks designed to measure individual differences in working memory.

The lexical decision task (Figure 6) was similar to tasks used in previous prospective memory experiments (Brewer et al., 2010; Christopher et al., 2019; Smith et al., 2007), but with the critical difference that it was administered on an eye-tracking computer. Participants were presented with 228 words and non-words, all of which were between 4 and 6 characters. The selected stimuli were correctly identified as either words or non-words with 95% accuracy during a lexical decision task as part of the English Lexicon Project (Balota, et al., 2007). Participants were instructed to respond to non-words by pressing the “F” key with their left index finger, and to words by pressing the “J” key with their right index finger using the keyboard in front of them. After receiving these initial task instructions, participants practiced the lexical decision task for 20 trials (10 words and 10 non-words randomly intermixed). Pupil data from the practice trials was used as the baseline pupil measurements, and later pupil size measurements were considered relative to pupil size during the practice trials. Because pupil diameter is recorded in arbitrary units on the EyeLink 1000 Plus, this correction was necessary to facilitate analyzing the pupil data.

Each trial began with the presentation of a central fixation cross, which remained onscreen for 500 ms. Following fixation cross offset, the stimulus was immediately presented, and remained onscreen until a response was made. During practice trials, accuracy feedback was provided to participants for 500 ms following each response. Participants in either of the prospective memory conditions were given the prospective memory instructions after completing the lexical decision task practice trials. After completing the practice trials, all participants completed 36 addition and subtraction problems, which typically takes participants approximately 4 minutes (Christopher et al., 2019). This arithmetic phase served as a distractor task. Following the distractor arithmetic problems, participants began the real lexical decision task trials. Trials on the lexical decision task were presented randomly, but with the following constraints: a given stimulus type (e.g., word) could only be presented four consecutive times at most, and no stimulus could be presented to a participant more than once.

Participants in the prospective memory conditions received an additional prospective memory task to be completed during the lexical decision task. Einstein and McDaniel (2005) suggested that asking participants to identify words fitting a general category during a lexical decision task (e.g., color words) would only facilitate non-focal processing during the task. Accordingly, participants in the present non-focal condition were asked to respond by pressing the space bar whenever they were presented with a color word (i.e., the words “black”, “blue”, “green”, “pink”, “purple”, and “white”). Einstein and McDaniel (2005) also suggested that asking participants to identify specific words (e.g., the word “mind”) would facilitate focal processing during the ongoing task. In the current experiment, participants in the focal prospective memory condition were given a list of six unrelated words to identify (i.e., the words “code”, “mind”, “handy”, “wine”, “clone”, and “music”), matching the number of prospective memory targets in the non-focal condition. Furthermore, these focal target words were matched to the non-focal target words on length, mean RT and accuracy during lexical decision, and frequency within the English language (Balota et al., 2007). All prospective memory task instructions can be found in Appendix C. On these prospective memory trials, participants did not need to press the “F” or “J” keys to indicate whether or not the letters form a word; instead, they were instructed to just press the space bar. The prospective memory targets were evenly spaced out across the 228 lexical decision task trials, such that prospective memory targets appear on trials 33, 71, 109, 147, 185, and 223. Which prospective memory target were presented on a given trial was randomized without replacement. To equate the number of trials presented to participants across conditions, participants in the control condition also received the target words from the focal prospective memory condition, though in the control condition these words would not have possessed any special significance for participants.

After completing the lexical decision task, participants completed the operation span and symmetry span tasks (Figure 6). These automated complex-span tasks were designed to measure individual differences in working memory (Redick et al., 2012). Both tasks were administered on a standard (i.e., non-eye-tracking) computer. During the operation-span task (Unsworth, Heitz, Schrock, & Engle, 2005), participants made true/false judgments about possible answers provided for math problems, while encoding a to-be-remembered letter presented in between each math problem. Initially, participants were given an opportunity to practice the math and letter tasks independently. Following these practice trials, participants were presented with trials where the math and letter stimuli were interleaved. After being presented with 3-7 letters and math problems, participants were prompted to recall the letters in the order that they were presented. Participants were instructed to ensure that they complete at least 85% of the math problems correctly. Overall math accuracy was displayed in the top right corner of the screen throughout the task. A participant's operation-span score was determined by the number of letters correctly recalled. In total, participants were presented with 75 letters, over 15 trials, making the maximum possible score on the operation-span task 75.

On the symmetry-span task (Unsworth, Redick, Heitz, Broadway, & Engle, 2009), participants were required to make yes/no decisions about the vertical symmetry of black and white images, and to encode the locations of suddenly red squares within a white 4 x 4 grid. As in the operation-span task, participants first practiced each of these tasks independently. After the practice trials, the red square presentations and symmetry decisions were interleaved. Each trial included the presentation of 2-5 red square locations to be remembered. A participant's score was calculated as the sum of the red-square locations correctly remembered out of a possible 42 locations over 12 trials. However, participants were also instructed to maintain at least 85% accuracy on the symmetry decisions. In the symmetry-span task, overall symmetry accuracy was always displayed in the top right corner of the screen, similar to the presentation of math accuracy in the operation-span task.

Analytical Approach

In addition to some more typical frequentist analyses, data were analyzed using Bayesian modeling and model comparison. This approach facilitated a direct comparison of differing theoretical accounts of monitoring processes in prospective memory (e.g., hypothetical pupil dilation patterns for a non-focal prospective memory task, depicted in Figure 5) based on how well they fit the posterior data. A Bayesian analytical approach was expected to be particularly essential to testing theories of prospective memory, given that some theories, such as the delay theory, predicted null results. Bayesian analysis methods allowed me to first specify priors corresponding to model parameters for task performance. I was able to modify the priors for parameters to reflect different predictions from theories of prospective memory. The priors provided the initial structure for the models, and then the data were analyzed by running 300,000 simulations for each model. These simulations used a Markov chain-Monte Carlo (MCMC) sampling method. The MCMC sampling allows the analysis to sample from multi-dimensional space in an attempt to find the most likely values for the model parameters. The MCMC sampling starts in a given location based on the priors, but then moves through multi-dimensional space in an attempt to fit the observed data. The sampled locations from this process make up the posterior distributions. The posterior distributions provided a distribution of the likely values for each parameter in each model. Having distributions of possible values for parameters, rather than a single-point estimate for the most likely value of a parameter, allowed me to ask probabilistic questions of the data. For example, instead of just asking if pupil size is larger in the non-focal prospective memory condition than in the control condition, I was able to ask how likely it is that there was a difference.

The model fit was assessed using the Leave-One-Out (LOO) method. LOO facilitates model comparison by simulating how well each model accounts for posterior data, systematically leaving out one posterior observation at a time. The results are arbitrary LOO units, which are not directly interpretable by themselves. However, the LOO values can be compared to each other, when using different models for the same data sets. The lowest LOO value represents the best model fit. Additionally, the LOO values can be used to generate probabilities of how likely a given model is to account for future data sets, relative to the other models being considered. The probability, though not the LOO value itself, allows researchers to assess how much better one model is relative to the others.

In the models, individual differences in working memory were included as a predictor of performance on the prospective memory tasks. To calculate a single working memory score for each participant, performance on the operation- and symmetry-span tasks was standardized, and the resulting *z*-scores were averaged together. Pupil data were corrected relative to baseline pupil size. Baseline pupil size was measured during the lexical decision task practice trials. Data from the lexical decision task was trimmed in the following ways. RTs were trimmed if they were faster than 150 ms, or if they were slower than 5000 ms (< 1% of total trials). Additionally, RTs were only analyzed for accurate trials.

In a series of nested models (Table 1), pupil dilation during ongoing lexical decision trials was the primary dependent variable of interest. Comparing nested models allowed me to answer research questions 1-3, which specifically pertained to how much participants should be expected to actively monitor during different prospective memory tasks. Model 1.1 stipulated that pupil dilation varied only as a function of the trial, and that there was some common intercept and sigma (i.e., variance unaccounted for). This first model was akin to a null model. Conversely, Model 1.2 represented predictions that would have been generated under a PAM model of prospective memory. Intercepts for the prospective memory conditions were expected to be equivalent, but higher than in the control condition, and working memory was expected to be positively related to pupil size in both prospective memory conditions. Model 1.3 was designed to reflect multiprocess theory predictions. Within Model 1.3, pupil size was allowed to vary between prospective memory conditions, as was the effect of individual differences in working memory. Finally, Model 1.4 was designed to reflect the predictions of delay theory. This model was similar to Model 1.2, but did not stipulate any relationship between individual differences in working memory and pupil size.

Additional models (Table 1) were generated for other dependent variables of interest. Rather than using model comparison, I used sufficiently broad priors to allow me to freely estimate the most likely values for model parameters for Models 1.5-1.7. Model 1.5 was used to estimate the predictive utility of pupil dilation during the 5 trials preceding a prospective memory target, and individual differences in working memory for prospective memory accuracy between the two prospective memory conditions. To the extent that active monitoring actually facilitates remembering to remember, both of these parameters would be predictive of prospective memory accuracy. Notably, delay theory predicts that neither parameter would be related to prospective memory accuracy in either condition. The results of Model 1.5 would allow me to answer research question 4. In Model 1.6, I modeled ongoing task RTs as a function of pupil dilation and individual differences in working memory. Again, delay theory predicts that neither parameter would be related to RTs, because slowing observed during a prospective memory task should not be indicative of costly monitoring processes. Model 1.6 facilitated answering research question 5. Finally, in Model 1.7, I estimated the roles of prospective memory accuracy, individual differences in working memory, and condition in determining pupillary response on the 5 trials immediately following the presentation of a prospective memory target. If recognizing a prospective memory target leads to increased pupil dilation in one or both of the prospective memory conditions, it would suggest that monitoring levels fluctuate during a prospective memory task, and can be elicited by an external stimulus. Only dynamic multiprocess theory would have predicted that this would be the case. Furthermore, given the previous evidence that the working memory system facilitates active monitoring (e.g., Christopher et al., 2019), I reasoned that the pupillary response to a prospective memory target might vary as a function of individual differences in working memory. Model 1.7 does not directly map onto the five research questions posed previously, but it does represent an opportunity to leverage the pupil data to better understand the potentially dynamic way that participants perform prospective memory tasks, possibly adjusting their levels of monitoring in response to external stimuli (Scullin et al., 2013).

Different types of distributions were used to fit the dependent variable being modeled. Models 1.1-1.4 and Model 1.7 used Gaussian (i.e., normal) distributions, which reflected my expectation that the pupil data would be normally distributed. Model 1.5 used a binomial distribution, because the dependent variable, prospective memory accuracy, was recorded dichotomously as either a 0 or 1 depending on whether or not participants responded accurately on a given trial. Model 1.6 used an ex-Gaussian distribution, which is similar to a Gaussian distribution in that its shape is relatively normal. Importantly, the ex-Gaussian distribution has wider tails to reflect the nature of RT data. The priors used for Models 1.1-1.7 are provided in Table 2.

Results

Descriptive Statistics

Performance on the operation and symmetry span tasks (Table 3) was positively correlated, $r = .38$, $p < .001$, and consistent with normative data (Redick et al., 2012). Participants were highly accurate on the lexical decision task, which is consistent with previous prospective memory studies using this task (e.g., Scullin et al., 2013). Additionally, participants correctly responded to the prospective memory target at a rate similar to what previous researchers have observed (e.g., Einstein & McDaniel, 1990). Somewhat surprisingly, focal and non-focal prospective memory accuracy rates were very similar. Furthermore, ongoing task RTs were longest in the focal prospective memory condition. Typically, participants in these paradigms have been much more accurate and faster on the focal prospective memory task than on the non-focal prospective memory task (e.g., Einstein et al., 2005; but see Smith, 2010). While the finding that levels of prospective memory accuracy were extremely similar and ongoing task RTs were slowest in the focal condition was surprising, a possible explanation of this pattern of results is examined in the following discussion section.

Traditional Frequentist Analyses

A 3-way ANOVA revealed that pupil size varied between conditions $F(2,179) = 406.56$, $p < .001$. A follow-up Tukey post-hoc test revealed that pupil size was significantly larger in the focal prospective memory condition than in either the control or non-focal conditions (*p* < .001 for both comparisons). Differences between the control and non-focal prospective memory conditions were not statistically significant (*p* = .054). A similar 3-way ANOVA was conducted to explore RT differences between conditions, as has traditionally been done. An ANOVA of RTs by condition was also statistically significant $F(2,179) = 793.39$, $p < .001$. Tukey post-hoc analyses revealed that all comparisons reached statistical significance (*p* < .001 for all). RTs were longest in the focal prospective memory condition, and shortest in the control condition.

Bayesian Modeling

The results of nested model comparison between Models 1.1-1.4 suggested that Model 1.3 was very likely (~100%) to be the best fitting among the models compared (Table 4). Examining the posterior distributions of Model 1.3 (Table 5) resulted in a number of informative observations. First and foremost, intercepts clearly differed between the three conditions (Figure 7). Posterior distributions indicated that it was almost certain (~100% in for both comparisons) that relative to control, pupil size increased under focal instructions and non-focal instructions by an estimated 5% and 3%, respectively. Furthermore, while pupil size appeared to decline across trials in all conditions, it was estimated to be very likely that this decline (i.e., the negative slope associated with the effect of trials) was less severe in the focal prospective memory condition (99%; Figure 8). Comparing the change in pupil size across trials between the non-focal and control conditions, it was not at all likely that a similar effect could be ascribed to the non-focal prospective memory condition (~0%). Individual differences in working memory were unlikely to be positively associated with pupil size in both the control and non-focal conditions (~0% and 3%, respectively). However, such a positive relationship was very likely in the focal prospective memory condition (~100%).

Models for other dependent variables were similarly informative. Posterior distributions from Model 1.5 for the effect of prior pupil size (Table 6) suggested, most notably, that increased pupil dilation immediately preceding the presentation of a prospective memory target was very likely to predict more accurate recognition of that prospective memory target (~100%). At least a couple of informative observations can be drawn from the posteriors of Model 1.6 (Table 7). First, by comparing the intercepts for each condition, I observed that ongoing task RTs were slower in the focal condition than in either the control or the non-focal conditions (each comparison ~100%). Second, in each condition, pupil dilation was positively associated with RTs, such that larger pupil size appeared to correspond to slower RTs (~100%). From the posterior distributions for the effect of prospective memory accuracy in each prospective memory condition in Model 1.7 (Table 8), it appeared that pupil size was very likely to increase following the accurate recognition of a prospective memory target (99%).

Discussion

Experiment 1 allowed me to answer each of the five research questions under investigation. First, pupil dilation varied between prospective memory and control conditions. Second, pupil dilation further varied between the focal and non-focal prospective memory conditions, though the direction of this difference was surprising. Third, pupil dilation was positively correlated with prospective memory accuracy. Fourth, pupil dilation was also positively correlated with ongoing task RTs. Fifth, pupil dilation was positively correlated with individual differences in working memory, though only in the focal prospective memory condition.

The present pattern of results is entirely consistent with the hypothesis that pupil dilation reflects active monitoring processes. Pupil dilation increased when participants were instructed to perform a prospective memory task (Figure 7), and pupil dilation was positively correlated with ongoing task RTs. Given that RTs have traditionally been used to demonstrate that individuals actively monitor for the chance to fulfill their prospective memory intention (Smith, 2003), it was expected that RTs and pupil dilation would be related. Additionally, pupil dilation was predictive of successful prospective remembering, which is consistent with the PAM and multiprocess theories' suggestions that active monitoring facilitates prospective memory (Einstein & McDaniel, 2005; Smith et al., 2007). Furthermore, pupil size appeared to increase following the recognition of a prospective memory target, suggesting that participants were engaging in increased active monitoring (Figure 9), as predicted by the dynamic multiprocess view (Scullin et al., 2013).

The most surprising finding was that pupil dilation was increased more in the focal prospective memory condition than in the non-focal prospective memory condition. The data here suggest then that monitoring was more important to performance in the focal condition than in the non-focal condition. Moreover, prospective memory accuracy in the focal condition was not particularly higher than accuracy in the non-focal condition, and ongoing task RTs were slowest in the focal prospective memory condition. This is inconsistent with any existing theory of prospective memory performance. A related finding is that individual differences in working memory were positively related to pupil dilation in only the focal condition. This was also inconsistent with previous studies (Christopher et al., 2019; Smith & Bayen, 2005). It could be that focal prospective memory task scenarios really do require active monitoring above and beyond the levels of monitoring that are necessary in non-focal scenarios, but this seems unlikely given the prior research on this topic (e.g., Einstein et al., 2005). In several experiments, Einstein et al. (2005) reported that participants were able to successfully carry out a prospective memory intention without incurring any monitoring costs when the target was focal, but they consistently found significant monitoring costs when the prospective memory target was non-focal in nature. Beyond prospective memory target focality, Einstein et al. (2005) reported that prospective memory costs, and by extension, monitoring, varied by a number of characteristics other than focality.

Importantly, the prospective memory conditions in Experiment 1 differed in another important way. In the focal task, participants had to encode 6 prospective memory targets, while participants in the non-focal task may have only been encoding (and monitoring for) a single target. Participants in the non-focal condition could have chosen to monitor for multiple exemplars of the color-words category (e.g., “blue”, “red”, etc.). It seems likely though, that at least some participants would have only been monitoring for “color words” as a single item. Both prospective memory conditions were presented with the same number of targets, but the number of things to monitor for was likely different in the minds of many participants. Providing participants in a focal prospective memory condition with 6 specific target words is consistent with previous work in this area (Otani, Landau, Libkuman, St. Louis, & Kazen, 1997; Smith, 2003). Furthermore, when Czernochowski et al. (2012) increased the frequency with which prospective memory targets were presented to participants, participants did not appear to decrease how much they monitored, based on ERP data. Given the Czernochowski et al. (2012) data, it would seem unlikely that the number of prospective memory target presentations led participants performing the non-focal prospective memory task in the current experiment to monitor less than participants in those previous experiments, which used fewer prospective memory targets (McDaniel & Einstein, 2007). However, it appeared that in the present experiment, focality and cognitive load were confounded. The instructions given to participants in the focal and non-focal prospective memory conditions may have caused participants to encode and monitor for their prospective memory intention differently. On-the-one-hand, it was clear following Experiment 1 that pupil dilation can be used to track active monitoring processes during a prospective memory task. On-the-other-hand, it was unclear in Experiment 1 whether active monitoring, prospective memory accuracy, and ongoing task RTs were driven by focality or load.

EXPERIMENT 2

Following the surprising pattern of results observed in Experiment 1, Experiment 2 was conducted to manipulate both focality and load factors directly. Because Experiment 1 already provided substantial evidence that monitoring as measured by pupil size is increased in prospective memory conditions relative to a control condition, and because the primary question of interest is a matter of distinguishing between types of prospective memory task scenarios, no control condition was utilized in Experiment 2. Furthermore, tasks measuring working memory were not included in the second experiment. This was because in Experiment 1, the results involving individual differences in working memory were largely consistent with previous findings that have suggested that the working memory system is selectively engaged to facilitate monitoring when the prospective memory task characteristics most suggest that active monitoring would be necessary. In Experiment 2, I manipulated focality and the cognitive load imposed by the prospective memory instructions between participants. To be clear, Experiment 2 was designed to test whether participants truly do monitor more during a focal prospective memory task relative to a non-focal task, or if the degree to which participants engage in active monitoring is driven by the inherent cognitive load placed on participants as a function of the prospective memory task instructions. In Experiment 1 there were five main research questions under investigation, but Experiment 2 was motivated by a single overarching question: is the extent to which participants in a prospective memory scenario actively monitor driven by prospective memory target focality, or is it primarily driven by prospective memory load? I answered this guiding question via five specific research questions. None of the three theories of prospective memory were able to perfectly account for the results of Experiment 1, so for the research questions under investigation in Experiment 2, I generated my own *a priori* hypotheses instead of relying on predictions from those theories. The new specific research questions and my *a priori* hypotheses are as follows:

1. Does pupil size differ between the high-load focal condition and the low-load focal condition?
    a. Hypothesis: Yes, I anticipated that pupil size would be larger in the high-load focal condition.
2. Does pupil size differ between the non-focal condition and the low-load focal condition?
    a. Hypothesis: Yes, based on previous research (Brewer et al., 2010; Einstein et al., 2005), I expected that pupil size would be larger in the non-focal condition, but that the effect of focality would be smaller than the effect of load.
3. Is prospective memory accuracy affected by monitoring, as measured by pupil size, equivalently between prospective memory conditions?
    a. Hypothesis: Yes, as in Experiment 1, to the extent that participants do monitor, monitoring should facilitate successful prospective remembering.

4. Will the relationship between RTs and pupil size vary depending on the focality or load of the prospective memory intention?
    a. Hypothesis: No, as in Experiment 1, while the degree to which participants monitor may vary between conditions, both RTs and pupil data should still reflect active monitoring.
5. Will pupil size increase following the accurate recognition of a prospective memory target in all conditions?
    a. Hypothesis: Yes, upon recognizing a prospective memory target, participants were expected to engage in active monitoring, regardless of focality or load (Scullin et al., 2013).

Method

Many of the data collection procedures in Experiment 2 were the same as those used in Experiment 1. Purdue undergraduates were recruited and then randomly assigned to one of three prospective memory conditions. All details pertaining to the eye-tracking apparatus were identical to those in Experiment 1. Data were collected using the EyeLink 1000 Plus, with the same recording settings as in the first experiment. The most significant departures from Experiment 1 were that in Experiment 2, participants did not complete any working memory complex span tasks, and no participants were assigned to a control condition. In Experiment 2, participants were randomly assigned to either a nonfocal, focal high-load, or focal low-load condition.

Participants

Much like the sample for the first experiment, undergraduate students from Purdue's Intro Psychology course (n = 269) were recruited to participate in Experiment 2, in exchange for credits going towards partial completion of the course. From this larger sample, exclusion criteria were applied similarly to what was done with the data in Experiment 1, prior to analyzing the final sample's data (*n* = 196). Due to the verbal nature of the lexical-decision task, participants' data were excluded when they self-identified as non-native English speakers (*n* = 24). Participants' data were further excluded if they failed to achieve 50% accuracy on the lexical decision task (*n* = 2). Finally, to ensure that prospective memory failures were not due to lapses in retrospective memory, participants' data were excluded if they failed to recall what the prospective memory intention was at the end of the lexical decision task (*n* = 46). Within the final sample, participants primarily identified as female (69%), and were between 18 and 28 years of age (*M* = 18.74, *SD* = 1.20).

Procedure

The procedure used in Experiment 2 was similar to the procedure followed in Experiment 1. As before, participants completed the experiment individually, and in the lab. All participants were randomly assigned to one of three conditions. In the non-focal prospective memory condition, the instructions mirrored those of the non-focal condition from Experiment 1 exactly (i.e., respond if you see color words). Participants in the two focal prospective memory conditions were instructed to respond to either 2 or 6 specific target words. In the low-load focal prospective memory condition, participants were asked to indicate if they ever saw the words “code” or “mind”. Alternatively, in the high-load focal prospective memory condition, participants were given the instructions used in the focal prospective memory condition in Experiment 1. Importantly, in all three prospective memory conditions, participants were only presented with two prospective memory targets during the lexical decision task. This means that in the focal conditions, the only difference was in the number of targets participants potentially needed to monitor for. Participants in the low-load focal condition simply received a subset of the list of targets that participants in the high-load focal condition were told about. In Experiment 2, the prospective memory target words were presented to participants on trials 149 and 223 in all conditions. Following the lexical decision task, all participants answered some demographic questions. All other task procedures were consistent with the methods and procedures outlined for Experiment 1.

Analytical Approach

Once again, data were first analyzed using more typical frequentist tests, and then followed-up with more nuanced Bayesian modeling techniques (Table 9). Rather than using model comparison techniques again, all dependent variables were forecasted using single models that facilitated generating posterior distributions for the effects of interest (e.g., random intercepts between conditions for a model of pupil size on ongoing task trials). Data trimming procedures were the same as in Experiment 1.

In the first model, Model 2.1, pupil dilation on lexical decision trials was predicted by prospective memory condition, and change over trials was allowed to vary between conditions. Exploring the posterior distributions resulting from Model 2.1 allowed me to test whether focality or load primarily affected pupil dilation (particularly research questions 1 and 2), because the posteriors provide information about the relative likelihood of different values for a given parameter (e.g., likely intercept values for pupil size in the low- and high-load focal conditions). Model 2.1 was intentionally designed to mirror the structure and parameters of Model 1.3, which had the same dependent variable in Experiment 1. Prospective memory accuracy was the dependent variable investigated in Model 2.2, and it was predicted by condition and pupil dilation on the 5 trials immediately before the presentation of a prospective memory target. In Experiment 1, and specifically Model 1.5, accuracy was also predicted by condition and pupil dilation on previous trials, and Model 2.2 allowed me to investigate whether or not this would vary as a function of either load or focality (research question 3). In Model 2.3, RTs on lexical decision trials were forecasted as a function of condition and pupil dilation. Model 2.3 was comparable to Model 1.6, and was intended to further test the extent to which ongoing task RTs would be related to pupil size in the different prospective memory conditions (research question 4). Model 2.4 was used to forecast pupil dilation on the 5 trials immediately following the presentation of a prospective memory target. Condition and successful recognition of the prospective memory target (i.e., prospective memory accuracy) were used as predictors. This model allowed me to explore the extent to which participants in different prospective memory conditions engaged in active monitoring, particularly immediately after a prospective memory target was presented (research question 5), which may have caused participants to spontaneously retrieve their prospective memory intention and increase monitoring activity (McDaniel & Einstein, 2000; Scullin et al., 2013). Model 1.7 from Experiment 1 provided evidence that participants in both a focal and a non-focal prospective memory condition monitor more following the recognition of a prospective memory target.

As in Experiment 1, the models used different distributions to fit the dependent variable being modeled. Models 2.1 and 2.4 used Gaussian distributions, Model 2.2 used a binomial distribution, and Model 2.3 used an ex-Gaussian distribution. The rationale for using different distributions based on the dependent variable was the same in Experiment 2 as what was described in Experiment 1. Gaussian distributions were used to model pupil size data, as those data were expected to follow a normal distribution. Prospective memory accuracy was recorded as 1s and 0s, making a binomial distribution appropriate for Model 2.2. Furthermore, I again used an ex-Gaussian distribution for the RT data in Model 2.3, because wider tails on the distribution were expected to fit the RT data effectively. The priors that were used to generate Models 2.1-2.4 are provided in Table 10.

Results

Descriptive Statistics

From Experiment 1 to Experiment 2, prospective memory accuracy appeared to decline marginally (Table 11). The level of performance was still generally in line with what previous researchers have observed. Moreover, it is not entirely surprising that overall prospective memory accuracy declined, given that there were fewer opportunities to correctly respond to the prospective memory target in Experiment 2. When Czernochowski et al. (2012) and Wilson et al. (2013) manipulated prospective memory frequency, they also reported that prospective memory accuracy was much higher when participants were presented with more prospective memory targets. Moreover, if recognizing a target boosts monitoring, and monitoring facilitates successful prospective remembering, both of which were observed in Experiment 1, then it stands to reason that participants in Experiments 2 would suffer from having only one opportunity to recognize a target and boost monitoring (before being presented with the final prospective memory target). Importantly, prospective memory accuracy appeared to be highest in the focal low-load prospective memory condition, and lowest in the focal high-load condition (non-focal prospective memory was in the middle). RTs on ongoing task trials were noticeably slower in the focal high-load prospective memory condition than in the other two prospective memory conditions. Collectively, these results provided an initial clue that the surprising focality results in Experiment 1 were chiefly a matter of load, and not focality.

Traditional Frequentist Analyses

A 3-way ANOVA suggested that pupil size varied between conditions $F(2,196) = 201.16$, $p < .001$. Subsequent Tukey post-hoc tests revealed that pupil size was largest in the focal high-load prospective memory condition, and smallest in the non-focal condition, with all comparisons reaching statistical significance ($p < .001$ for all comparisons). An additional ANOVA of RTs by condition was also statistically significant $F(2,196) = 125.41$, $p < .001$. Tukey post-hoc analyses revealed that all comparisons reached statistical significance ($p < .01$ for all). RTs were longest in the focal high-load prospective memory condition, and shortest in the focal low-load condition.

Bayesian Modeling

Using Bayesian modeling allowed for an analysis of relevant posterior distributions. Based on the condition intercept posteriors of Model 2.1 (Table 12), it appeared almost certain (~100%) that pupil dilation was greater in the focal high-load condition than in either the non-focal or focal low-load conditions (estimated 4% increase in the intercept for both comparisons; Figure 10). Model 2.1 intercepts further suggested that pupil dilation did not vary between the non-focal and focal low-load conditions

(Figure 10). The change in pupil dilation across trials showed the least decline in the focal high-load condition (~100%; Figure 11). When modeling prospective memory accuracy in Model 2.2, I observed that in all three conditions, higher pupil dilation on the 5 trials immediately prior to the presentation of a prospective memory target was indicative of a greater likelihood of successfully identifying the prospective memory target (~100%; Table 13). In examining the condition intercepts for Model 2.3 (Table 14), it appeared very likely (~100%) that participants in the focal high-load condition would respond more slowly than participants in the other two conditions. Also, in all three conditions, pupil dilation was positively related to ongoing task RTs, such that larger pupils appeared to co-occur with slower RTs (~100%). According to the posteriors for the effect of prospective memory accuracy in Model 2.4 (Table 15), pupil dilation in all three prospective memory conditions would be very likely (~100%) to increase following the successful recognition of a prospective memory target.

Discussion

Experiment 2 replicated and extended the key findings from Experiment 1, allowing me to answer each of my five new research questions. First, consistent with my hypothesis, I observed larger pupil size in the high-load focal prospective memory condition than in the low-load focal prospective memory condition. Second, I did not find evidence suggesting that pupil size, and by extension monitoring, was higher in the non-focal condition than in the low-load focal condition. Third, I observed that pupil size was positively related to prospective memory accuracy in all three prospective memory conditions, suggesting that monitoring facilitated successful prospective memory retrieval in each of the prospective memory tasks being studied. Fourth, there was substantial evidence for a positive relationship between pupil size and RTs in all three conditions, such that when participants' pupils were larger they responded more slowly. Fifth, pupil size did increase following an accurate response to a prospective memory target, suggesting that external stimuli could elicit active monitoring (Figure 12).

Notably, all of the data suggested that load, and not focality, is responsible for eliciting active monitoring during a prospective memory task, as indexed by pupil dilation. When load was similar, but focality differed (i.e., non-focal condition vs. focal low-load condition), there was virtually no evidence for any differences in pupillary response. However, when load was manipulated within the focal prospective memory task scenario, participants with the high-load instructions appeared to rely on active monitoring more, even though the number of targets presented matched what was presented in the low-load focal condition. This outcome is not consistent with either the PAM or multiprocess theories, which suggested that either monitoring would be equally necessary between focal and non-focal tasks, or more necessary during non-focal prospective memory tasks respectively (McDaniel & Einstein, 2000; Smith et al., 2007). The other findings of Experiment 2 bolstered the conclusions drawn from Experiment 1. Pupil dilation certainly appeared to be a reasonable measure of active monitoring, given that it positively predicted prospective memory accuracy, and was positively correlated with ongoing task RTs, the traditional measure of monitoring.

GENERAL DISCUSSION

Using pupil dilation to measure monitoring processes at the trial level could have a significant impact on theoretical understandings of prospective memory. In the current project, the results of two highly powered experiments demonstrated the feasibility of using pupillometry as a much more temporally-sensitive measure of active monitoring than the traditional RT-cost paradigm. Due to disagreements surrounding how best to interpret RT effects in prospective memory experiments, pupillometry has the potential to yield more theoretically useful results (Einstein & McDaniel, 2010; Heathcote et al., 2015; Smith, 2010). Furthermore, because of this more nuanced measure of monitoring activity, I was able to test competing theories of prospective memory in novel ways. From each prospective memory theory, several specific predictions were generated (see Table 16 for a summary of different predictions made by each theory related to pupil data), and after two experiments, no theory had perfectly accounted for the results.

The PAM model proposed by Smith and colleagues (2007; Smith, 2003) stipulated that individuals must continuously monitor for successful prospective remembering. Their theory predicted that participants in a prospective memory condition, regardless of focality, would show consistently different dilation patterns throughout the entire task, compared to participants in a control (no prospective memory) condition. Moreover, pupil dilations were expected to be positively correlated with prospective memory success, because they are being used as a measure of active monitoring (Unsworth & Robison, 2015). These predictions were consistent with the data from the current experiments. Pupil dilation was consistently larger in prospective memory conditions relative to control, and Experiment 2 in particular revealed that focality was not a determining factor of whether or not participants would monitor. However, the PAM theory would have also predicted that individual differences in working memory would be positively correlated with prospective memory performance in all conditions (Smith & Bayen, 2005). This was not the case in Experiment 1, where individual differences in working memory were unrelated to pupil dilation on the non-focal prospective memory task.

The dynamic multiprocess model that Scullin and colleagues (2013; Shelton & Scullin, 2017) described provides a more nuanced or context-specific interpretation of active monitoring behavior. Scullin and colleagues suggested that participants alternated between periods of monitoring and not monitoring over time, and that certain events, such as identifying a prospective memory target, would be likely to instantiate monitoring again for a period of time. This prediction certainly seemed to fit the observed data, as recognizing a prospective memory target lead to a temporary increase in pupil size. According to the dynamic multiprocess model, participants in a prospective memory condition should have periods where their pupil dilation changes to be significantly different from that of control participants as they monitor for the prospective memory target. However, this model also predicted that during the task, participants would periodically revert to not monitoring for a time. Under the dynamic multiprocess view, individual differences in working memory should be related to successful prospective memory performance only when the need for active monitoring is sufficient. Scullin and colleagues might have suggested a non-focal prospective memory condition would be when individual differences in working memory would be most related to prospective memory performance (Brewer et al., 2010), but in light of the effect of load relative to focality (i.e., Experiment 2), the predicted situational relevance of individual differences in working memory still seems fairly similar to the data observed here.

Finally, the delay theory of prospective memory predicted that pupil dilation would not vary at all between prospective memory conditions as a function of cue focality, given that costly monitoring processes are supposed to be unnecessary for prospective remembering (Heathcote et al., 2015). Delay theory holds up in this instance, given that Experiment 2 of the current manuscript makes the case that focality is not of central importance to performance in prospective memory tasks. From delay theory's assertion that an additional, though critically, independent, evidence accumulator must be incorporated to allow for the prospective memory response in addition to the ongoing task responses, I have inferred for the sake of the current manuscript that pupil dilation would be predicted to increase in a prospective memory scenario, compared to a control condition (Strickland et al., 2018). In this way, the pupillary data predictions were similar between PAM and delay theories. Critically, PAM theory predicted that pupil dilation will be correlated with prospective memory success rates, since the degree of active monitoring a participant was engaged in was essential to success in that theory (Smith, 2003). Conversely, under delay theory, pupil dilation in a prospective memory condition is only related to the addition of another potential response path, and relative dilation patterns should not be correlated with prospective memory success because they are not indicative of changes in evidence accumulation rates, as PAM theory supposes.

Strangely, the evidence accumulators suggested by delay theory are the closest thing to focusing on cognitive load, rather than focality, which exists in these main theories of prospective memory. Delay theory correctly deemphasizes target-focality as a driver of prospective memory behavior. However, delay theory also stipulates that evidence accumulators will be added for each additional response option (Strickland et al., 2018), not each possible target that could be responded to. Therefore, delay theory would have predicted differences in pupil size between control and prospective memory conditions, but not between prospective memory conditions as a function of how many targets participants needed to respond to (e.g., between focal low-load and focal high-load prospective memory conditions). Also, within delay theory, it would not have been possible to predict the observed relationship been how much a participant's pupil size increased and how likely they were to successfully respond to the prospective memory target. Moreover, the relationship between individual differences in working memory and prospective memory success, which other researchers argued is the result of active monitoring (Brewer et al., 2010; Christopher et al., 2019; Smith & Bayen, 2005), would not have been predicted by delay theory. Here again, delay theory would have failed to predict components of the observed data. Testing delay theory was one of the benefits of measuring active monitoring through pupillometric data, rather than RTs. While it has previously been difficult to interpret RT differences between prospective memory and control conditions (Einstein & McDaniel, 2010; Heathcote et al., 2015; Smith, 2010), changes in pupil diameter following the inclusion of a prospective memory target more clearly indicate a change in cognitive load. RTs from the current experiments did in fact vary across trials in a pattern visually similar to the pupillometric data (Figures 13 and 14), with notably longer RTs occurring immediately following the presentation of a prospective memory target. However, continuing to test theories of prospective memory will likely require that researchers move beyond RT-cost paradigms (Smith, 2010), as was done in the current experiments.

Interestingly, each of the three theories of prospective memory that were highlighted here failed to predict some element of the data reported in the current manuscript. PAM and delay theories both correctly reject or at least downplay the importance of focality, and this de-emphasis of focality was supported by the current data. However, only the dynamic multiprocess view appropriately predicted a situational relationship between prospective memory performance and individual differences in working memory.

In Experiment 1, I found further evidence that working memory facilitates monitoring for prospective memory intentions (Brewer et al., 2010; Christopher et al., 2019). Previous research has often focused on whether or not the relationship between individual differences in working memory and prospective memory success varied depending on prospective memory target focality (Brewer et al., 2010; Smith & Bayen, 2005). In Christopher et al. (2019) we considered whether the working memory-prospective memory relationship varied as a function of target salience, not focality, and tried to reorient the conversation around whether or not the task was attentionally demanding. Based on the findings of Experiment 2, it appears that attentional load may be the more relevant element with respect to how much the working memory system will be recruited to facilitate prospective remembering, and not target focality. Future researchers interested in how the characteristics of a prospective memory intention will affect performance would be wise to place more emphasis on load, rather than focality, which has traditionally been the focus of much of the prospective memory research (Einstein et al., 2005; McDaniel & Einstein, 2000; Smith, 2010).

A critical reader of the current experiments might argue that focality was not successfully manipulated. Asking participants to respond to specific words (focal) or words that fit within a category (non-focal) is a focality manipulation that had previously produced the more typical effect, wherein participants given the non-focal instructions showed greater costs than participants with a focal prospective memory intention (Einstein & McDaniel, 2005). Einstein and McDaniel (2005) argued that judging whether a string of letters represented a word or a non-word would also facilitate processing if that word was a specific target-word being searched for, but not whether or not the word potentially fit within a category. However, it is possible that determining category membership during lexical decision is still somewhat focal relative to other focality manipulations. How focal a task is has not previously been measured, and there is not currently a method for determining the relative focality of a task. Smith (2010) rightly pointed out just how difficult it would be to determine how focal a task is, and as a result, Smith (2010) further suggested that in some situations it is not really clear if a task is focal or non-focal in nature. This distinction likely becomes even less clear in prospective memory scenarios outside of the lab. In addition to the present finding that a previously established manipulation of focality was ineffective at altering active monitoring during a prospective memory task, Smith's (2010) compelling argument that focality is currently not well defined should encourage future researchers to focus on other elements of prospective memory intentions that could impact how participants perform the task.

Based on the results of Experiments 1 and 2, pupillometry offers a promising method for studying the underlying cognitive processes that facilitate performance in prospective memory contexts. An area of interest for cognitive psychologists has been improving prospective remembering for individuals outside of the lab. One promising line of research would involve assessing how it is that previously established behavioral strategies work to improve the likelihood of successfully fulfilling one's prospective memory intentions. In particular, there are a number of published studies which reported that when participants used implementation intentions, they were much more likely to fulfill their prospective memory intention, and that implementation intentions reduced or perhaps even eliminated the need for active monitoring (Gollwitzer, 1999). Implementation intentions simply involve creating explicit statements about how one will respond upon encountering a critical prospective memory target. For example, an individual might say, "When I make my coffee in the morning, I will take my medication". Gollwitzer (1999) reasoned that this, "When X, then Y" formula leads participants to be much more likely to successfully complete their prospective memory task, because it formalized a mental pairing between stimulus and response.

Gollwitzer (1999) argued that implementation intentions facilitated prospective remembering by eliminating, or substantially reducing, the need to actively monitor for one's prospective memory target. Implementation intentions are thought to generally automate prospective memory responding. To demonstrate this, Gollwitzer (1996) instructed participants that their prospective memory task was to make a special response whenever they were presented with one of a few specific words. Later in the experimental session, participants completed a dichotic listening task, and during the task, the prospective memory target words were presented in the to-be-ignored channel of speech. Participants that had been instructed to use implementation intentions were found to be much more likely to successfully complete the prospective memory intention compared to participants given standard prospective memory instructions. Based on this result, Gollwitzer (1996, 1999) deduced that implementation intentions caused target stimuli to become highly salient and difficult for participants to ignore, given that participants were typically successful in not processing the content of the speech in the to-be-ignored channel (Moray, 1959).

More recently, some researchers have pushed back against Gollwitzer's (1999) explanation of the benefits of implementation conditions. McDaniel and Scullin (2010) reasoned that if implementation intentions automate prospective memory responding, as Gollwitzer (1999) claimed, then prospective memory performance under implementation intentions should be immune to manipulations increasing cognitive load. Across three similar experiments, McDaniel and Scullin (2010) had participants complete a category decision task, which involved quickly determining whether or not a word on the right-hand portion of the screen would fit into a category provided on the left-hand portion of the screen. In addition to the ongoing category decision task, all participants were given a prospective memory task, to make an alternative response if they ever noticed one of a few specific words (e.g., corn). The prospective memory instructional method varied between participants, with half of participants receiving an implementation intention instruction (e.g., say aloud three times, "When I see the word corn, I will press the space bar"), and the other participants receiving standard prospective memory instructions (e.g., when you see the word corn, press the space bar). Cognitive load was manipulated within participants, such that one block of trials simply involved the category decision task as already described, while another block introduced an additional component to the category decision task, which required participants to generate a random sequence of numbers during the task. These random numbers were spoken out loud by participants at a rate of one number per second, and were believed to interfere with the category decision component of the task, thus increasing demands on cognitive load. Gollwitzer's (1999) account of implementation intentions would have predicted that participants in the implementation intention condition would complete their prospective memory intention more frequently than participants in the standard instruction condition, regardless of how demanding the cognitive load was in a given block.

What McDaniel and Scullin (2010) found was not consistent with this prediction. They observed that implementation intentions did lead to improved prospective remembering in the lower cognitive load block of trials. Yet during the high cognitive load block, implementation intentions did not appear to yield improved prospective memory performance (prospective memory accuracy was numerically higher, though this difference was not statistically significant). This pattern of results would seem to contradict Gollwitzer's (1999) theory that implementation intentions facilitate prospective remembering by automating prospective memory responding, and minimizing the need for active monitoring processes. A future experiment could have participants complete prospective memory tasks under implementation intentions, and compare pupil size levels to a standard-instructions prospective memory condition. Gollwitzer (1999) would have predicted that pupil size and prospective memory accuracy would be unrelated, because implementation intentions should make active monitoring unnecessary. Conversely, the prediction of McDaniel and Scullin (2010) would be that while participants under implementation intention instructions might monitor less overall, monitoring, as measured by pupil size, should still be related to prospective memory accuracy. Pupillometry would provide convincing and direct evidence to adjudicate the currently unsettled debate over how implementation intentions work to improve prospective memory success rates.

Future studies of prospective memory should also consider how prospective memory behaviors might vary when the time between intention formation and the opportunity to fulfill one's prospective memory intention is significantly longer than what has been typical in lab-based prospective memory studies. Often in the lab, the time between intention formation and execution would best be represented in minutes, but outside of the lab, many prospective memory scenarios likely involve intervals of hours, if not days. In Experiment 1, I found evidence that the working memory system supports active monitoring for a prospective memory intention, which is consistent with previous studies establishing a conditional relationship between individual differences in working memory and prospective memory performance (Christopher et al., 2019; Brewer et al., 2010). The working memory system is likely most relevant to prospective remembering when the interval between prospective memory intention formation and execution is short (Engle & Kane, 2004), as has been the case in these lab-based studies. Even in the current paradigm, participants appeared to monitor less in-between prospective memory targets, compared to when a target had recently brought the prospective memory intention back into the focus of attention. More research like naturalistic study conducted by Kvavilashvili and Fisher (2007) is needed to test the relative role of active monitoring and working memory when participants must remember to fulfill an intention much later. Recent advances in portable eye-tracking technology might allow researchers to measure pupil size as participants go about their day outside of the lab. It would be illuminating to test whether relationships between individual differences in working memory, pupil size, and prospective memory accuracy would be observed when participants attempted to fulfill prospective memory intentions over hours or even days.

In a recent article that was published after the current data collection began, Moyes, Sari-Sarraf, and Gilbert (2019) used a pupillometry procedure similar to the current experiments to investigate prospective memory processes. Across two experiments, participants completed lexical decision for their ongoing task. Participants in the prospective memory conditions were told to press a special key if they saw either a specific word (e.g., “tower”) or a word belonging to a specific category (e.g., “a metal”). Moyes et al. (2019) reported a pattern of results consistent with what was found in the present data, most notably that pupil dilation (a) increased when participants were given a prospective memory intention; (b) was predictive of prospective memory accuracy; and (c) increased following the accurate recognition of a prospective memory target. Moyes et al. (2019) concluded that pupil dilation would be a promising way of measuring monitoring processes during a prospective memory task.

The two experiments reported in Moyes et al. (2019) differed from the current experiments in several important ways. In both experiments in Moyes et al., sample sizes (n = 36 in each) were approximately 15% of the final sample sizes in the current experiments. Also, in both experiments there was an inter-trial interval of 3 seconds. Some prospective memory tasks do not use an inter-trial interval, and those that do typically set the interval at approximately 500 ms (my experiments reported here did not have an inter-trial interval). Moyes et al. (2019) acknowledged that this may have been important, because such a long interval may have encouraged participants to monitor for their prospective memory intention between trials. The eye-tracking computer used in Moyes et al. (2019) was an inexpensive machine, that has been shown to provide pupillometry measurement with accuracy nearly as good as the EyeLink 1000 Plus, for a much lower price (Dalmaijer, 2014). Like the EyeLink, the EyeTribe that Moyes et al. (2019) used produces pupil size measurements in arbitrary units that cannot be readily interpreted. Moyes et al. (2019) acknowledged the arbitrary pupil units, and attempted to transform them relative to baseline, but then Moyes et al. (2019) reported that this transformation "increased noise rather than reducing it" (Moyes et al., 2019, p. 86). Moyes et al. (2019) never elaborated on how they determined that transforming the pupil data increased the noisiness of the data. Not transforming the pupil data relative to baseline makes it impossible to interpret their pupil results. Moyes et al. (2019) cite Dalmaijer (2014) as evidence for the utility of the EyeTribe, but Dalmaijer (2014) is clear that pupil data from the EyeTribe can only be interpreted if it is corrected relative to baseline, and if participants are using a chin rest. Moyes et al. (2019) only used a chin rest in their second experiment. Given the oversights with how the pupil data were processed and analyzed, it is difficult to interpret any of the results of Moyes et al. (2019). While pupillometry offers some clear advantages for researchers wanting to study prospective remembering, it will also require that future researchers are particularly careful in setting up their experimental procedures, and that they establish thoughtful plans for how data will be treated and analyzed.

Conclusion

Given that one theory did not clearly perform better than the others, it is likely useful to summarize the findings. On several fronts, pupil dilation appeared to be an effective means of measuring active monitoring during a prospective memory task. Participants did engage in active monitoring, more so when they had to remember a greater number of items (i.e., when cognitive load was higher), and not based on whether or not the task was focal. The working memory system appeared to facilitate active monitoring, which supported prospective remembering, though this may only be the case when cognitive load is sufficiently high/taxing. The extent to which individuals monitored for their prospective memory target(s) appeared to fluctuate over time, and monitoring immediately before the presentation of such a target was likely to lead to a prospective memory success. Using eye-tracking technology to measure pupil dilation during a prospective memory task provided the most convincing evidence to date for active monitoring as a mechanism that supports prospective remembering, and perhaps more importantly, it highlighted strengths and weaknesses for leading theories of prospective memory.

LIST OF REFERENCES

Anderson, F. T., Rummel, J., & McDaniel, M. A. (2018). Proceeding with care for successful prospective memory: Do we delay ongoing responding or actively monitor for cues? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(7), 1036-1050. http://dx.doi.org/10.1037/xlm0000504

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annual Review of Neuroscience, 28,* 403-450. doi: 10.1146/annurev.neuro.28.061604.135709

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Vol. Ed.), *The psychology of learning and motivation: Vol. 8* (pp. 47-89). New York, NY: Academic Press.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445-459. https://doi.org/10.3758/BF03193014

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*(1), 83-100. doi: 10.1037/0096-3445.133.1.83

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin, 91*(2), 276-292. doi: 10.1037/0033-2909.91.2.276

Brewer, G. A., Knight, J. B., Marsh, R. L., & Unsworth, N. (2010). Individual differences in event-based prospective memory: Evidence for multiple processes supporting cue detection. *Memory & Cognition, 38*(3), 304-311. doi: 10.3758/MC.38.3.304

Cantor, J., & Engle, R. W. (1993). Working-memory capacity as long-term memory activation: An individual-differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1101-1114. doi: 10.1037/0278-7393.19.5.1101.

Christopher, E. A., Fansher, M., & Redick, T. S. (2019). *The working memory system supports prospective remembering in attention-demanding situations*. Manuscript in preparation.

Conway, A. A., Cowan, N., & Bunting, M. F. (2001). The cocktail party phenomenon revisited: The importance of working memory capacity. *Psychonomic Bulletin & Review*, *8*(2), 331-335. http://dx.doi.org/10.3758/BF03196169

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 97-185. doi: 10.1017/S0140525X01003922

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51-57. doi: 10.1177/0963721409359277

Czernochowski, D., Horn, S., & Bayen, U. J. (2012). Does frequency matter? ERP and behavioral correlates of monitoring for rare and frequent prospective memory targets. *Neuropsychologia*, *50*(1), 67-76. https://doi.org/10.1016/j.neuropsychologia.2011.10.023

Dalmaijer, E. S. (2014). Is the low-cost EyeTribe eye tracker any good for research? *PeerJ Preprints, 4*, 1-35. https://doi.org/10.7287/peerj.preprints.141v2.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450-466. doi: 10.1016/S0022-5371(80)90312-6

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 717-726. doi: 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. doi: 10.1111/j.0963-7214.2005.00382.x

Einstein, G. O., & McDaniel, M. A. (2010). Prospective memory and what costs do not reveal about retrieval processes: A commentary on Smith, Hunt, McVay, and McConnell (2007). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 1082-1088. doi: 10.1037/a0019184

Einstein, G. O., McDaniel, M. A., Thomas, R., Mayfield, S., Shank, H., Morrisette, N., & Breneiser, J. (2005). Multiple processes in prospective memory retrieval: Factors determining monitoring versus spontaneous retrieval. *Journal of Experimental Psychology: General*, *134*(3), 327-342. doi: 10.1037/0096-3445.134.3.327

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19-23. doi: 10.1111/1467-8721.00160

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science*, *13*(2), 190-193. https://doi.org/10.1177/1745691617720478

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Vol. Ed.), *The psychology of learning and motivation: Vol. 44* (pp. 145-199). New York, NY: Elsevier.

Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience, 10*(2), 252-269. doi: 10.3758/CABN.10.2.252

Gollwitzer, P. M. (1996). The volitional benefits of planning. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 287-312). New York, NY: Guilford.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*(7), 493-503. doi: 10.1037/0003-066X.54.7.493

Harris, J. E. (1984). Remembering to do things: A forgotten topic. In J. E. Harris & P. E. Morris (Eds.), *Everyday memory, actions, and absentmindedness*, (pp. 71-92). New York, NY: Academic Press.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Vol. 22* (pp. 193-225). New York, NY: Academic Press.

Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological Review*, *122*(2), 376-410. doi: 10.1037/a0038952

Heitz, R. P., Schrock, J. C., Payne, T. W., & Engle, R. W. (2008). Effects of incentive on working memory capacity: Behavioral and pupillometric data. *Psychophysiology*, *45*(1), 119-129. https://doi.org/10.1111/j.1469-8986.2007.00605.x

Henry, J. D., MacLeod, M. S., Phillips, L. H., & Crawford, J. R. (2004). A meta-analytic review of prospective memory and aging. *Psychology and Aging*, *19*(1), 27-39. doi: 10.1037/0882-7974.19.1.27

Hou, R. H., Freeman, C., Langley, R. W., Szabadi, E., Bradshaw, C. M. (2005). Does modafinil activate the locus coeruleus in man? Comparison of modafinil and clonidine on arousal and autonomic functions in human volunteers. *Psychopharmacology, 181*(3), 537-549. doi: 10.1007/s00213-005-0013-8

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, *154*(3756), 1583-1585. doi: 10.1126/science.154.3756.1583

Kvavilashvili, L., & Fisher, L. (2007). Is time-based prospective remembering mediated by self-initiated rehearsals? Role of incidental cues, ongoing activity, age, and motivation. *Journal of Experimental Psychology: General*, *136*(1), 112-132. doi: 10.1037/0096-3445.136.1.112

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*, 389-433. doi: 10.1016/S0160-2896(05)80012-1

Maylor, E. A. (1990). Age and prospective memory. *The Quarterly Journal of Experimental Psychology Section A, 42*(3), 471-493. doi: 10.1080/14640749008401233

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology*, 14, S127-S144. doi: 10.1002/acp.775

McDaniel, M. A., & Einstein, G. O. (2007). *Prospective memory: An overview and synthesis of an emerging field*. Thousand Oaks, CA: Sage Publications.

McDaniel, M. A., & Scullin, M. K. (2010). Implementation intention encoding does not automatize prospective memory responding. *Memory & Cognition*, *38*(2), 221-232. doi: 10.3758/MC.38.2.221

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*, 56-60. doi: 10.1080/17470215908416289

Moyes, J., Sari-Sarraf, N., & Gilbert, S. J. (2019). Characterizing monitoring processes in event-based prospective memory: Evidence from pupillometry. *Cognition*, *184*, 83-95. https://doi.org/10.1016/j.cognition.2018.12.007

Otani, H., Landau, J., Libkuman, T. M., St. Louis, J. P., & Kazen, J. K. (1997). Prospective memory and divided attention. *Memory, 5*, 343-360. doi: 10.1080/741941393

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164-171. doi:10.1027/1015-5759/a000123

Reynolds, J. R., West, R., & Braver, T. (2008). Distinct neural circuits support transient and sustained processes in prospective memory and working memory. *Cerebral Cortex*, *19*(5), 1208-1221. https://doi.org/10.1093/cercor/bhn164

Scullin, M. K., McDaniel, M. A., & Shelton, J. T. (2013). The dynamic multiprocess framework: Evidence from prospective memory with contextual variability. *Cognitive Psychology*, *67*, 55-71. doi:10.1016/j.cogpsych.2013.07.001

Shelton, J. T., & Christopher, E. A. (2016). A fresh pair of eyes on prospective memory monitoring. *Memory & Cognition*, *44*(6), 837-845. doi: 10.3758/s13421-016-0601-3

Shelton, J. T., & Scullin, M. K. (2017). The dynamic interplay between bottom-up and top-down processes supporting prospective remembering. *Current Directions in Psychological Science*, *26*(4), 352-358. https://doi.org/10.1177/0963721417700504

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. doi: 10.1037/0278-7393.29.3.347

Smith, R. E. (2010). What costs do reveal and moving beyond the cost debate: Reply to Einstein and McDaniel (2010). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(4), 1089-1095. doi:10.1037/a0019183

Smith, R. E., & Bayen, U. J. (2005). The effects of working memory resource availability on prospective memory: A formal modeling approach. *Experimental Psychology*, *52*(4), 243-256. doi:10.1027/1618-3169.52.4.243

Smith, R. E., Hunt, R. R., McVay, J. C., & McConnell, M. D. (2007). The cost of event-based prospective memory: Salient target events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 734-746. doi: 10.1037/0278-7393.33.4.734

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.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114*, 104-132. doi: 10.1037/0033-295X.114.1.104

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498-505. doi: 10.3758/BF03192720

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*, 635-654. doi: 10.1080/09658210902998047

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. doi: 10.3758/s13423-014-0747-6

Unsworth, N., & Robison, M. K. (2016). Pupillary correlates of lapses of sustained attention. *Cognitive, Affective, & Behavioral Neuroscience*, *16*(4), 601-615. doi: 10.3758/s13415-016-0417-4

Unsworth, N., & Robison, M. K. (2017). The importance of arousal for variation in working memory capacity and attention control: A latent variable pupillometry study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(12), 1962-1987. http://dx.doi.org/10.1037/xlm0000421

Wang, C. A., & Munoz, D. P. (2015). A circuit for pupil orienting responses: Implications for cognitive modulation of pupil size. *Current Opinion in Neurobiology*, *33*, 134-140. https://doi.org/10.1016/j.conb.2015.03.018

Wang, C. A., Brien, D. C., Munoz, D. P. (2015). Pupil size reveals preparatory processes in the generation of pro- and anti-saccades. *European Journal of Neuroscience*, *41*(8), 1102-1110. http://dx.doi.org/10.1111/ejn.12883

West, R., Bowry, R., & Krompinger, J. (2006). The effects of working memory demands on the neural correlates of prospective memory. *Neuropsychologia*, *44*(2), 197-207. https://doi.org/10.1016/j.neuropsychologia.2005.05.003

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, *109*(22), 8456-8460. doi: 10.1073/pnas.1201858109

Wilson, J., Cutmore, T. R., Wang, Y., Chan, R. C., & Shum, D. H. (2013). Effects of cue frequency and repetition on prospective memory: An ERP investigation. *International Journal of Psychophysiology*, *90*(2), 250-257. https://doi.org/10.1016/j.ijpsycho.2013.08.003

APPENDIX A

Table 1

Models Computed in Experiment 1

	DV	Summary of Model Parameters
Model 1.1	Pupil Size	Pupil Size varies across trials
Model 1.2	Pupil Size	Prospective memory conditions higher than control & working memory is positively related to pupil size
Model 1.3	Pupil Size	Same as Model 1.2, but effects allowed to vary between focal and non-focal prospective memory conditions
Model 1.4	Pupil Size	Prospective memory conditions higher than control
Model 1.5	PM Acc.	Accuracy differs between conditions & is positively related to working memory and pupil size on previous trials
Model 1.6	RTs	Varies between conditions and as a function of working memory and pupil size
Model 1.7	Pupil Size	Pupil size elevated following accurate response to a prospective memory target, though the effect differs between conditions and as a function of individual differences in working memory

Note. Models 1.1-1.4 were used for nested model comparison. DV = Dependent Variable; PM Acc = Prospective Memory Accuracy; RTs = Response Times.

Table 2

Priors Used for Models 1.1-1.7

	M 1.1	M 1.2	M 1.3	M 1.4	M 1.5	M 1.6	M 1.7
Intercept	1 (.1)						
Trial	-.01 (.01)						
Control Intercept		1 (.1)	1 (.1)	1 (.1)		600 (100)	1 (.1)
PM Intercept		1.2 (.1)		1.2 (.1)			
FPM Intercept			1.1 (.1)		.7 (.1)	650 (100)	1.1 (.1)
NFPM Intercept			1.2 (.1)		.4 (.1)	700 (100)	1.2 (.1)
Trial*Control		-.01 (.01)	-.01 (.01)				
Trial*PM		0 (.01)					
Trial*FPM			0 (.01)				
Trial*NFPM			0 (.01)				
WM*Control		0 (.1)	0 (.1)			0 (10)	0 (.1)

(*table continues*)

	M 1.1	M 1.2	M 1.3	M 1.4	M 1.5	M 1.6	M 1.7
WM*PM		.2 (.1)					
WM*FPM			0 (.1)		.1 (.1)	10 (10)	0 (.1)
WM*NFPM			.2 (.1)		.2 (.1)	20 (10)	.2 (.1)
PM Acc*FPM							.25 (.1)
PM Acc*NFPM							.25 (.1)
Prior Pupil Size*FPM					.01 (.1)		
Prior Pupil Size*NFPM					.01 (.1)		
Pupil Size*FPM						50 (25)	
Pupil Size*NFPM						100 (25)	
Sigma	.3 (.1)	.3 (.1)	.3 (.1)	.3 (.1)		300 (100)	.3 (.1)

Note. Parameters for the different models are identified in the left-most column. No model included all of the parameters. Priors are presented in columns as ‘Mean (SD)’. M = Model; PM = Prospective Memory; WM = Working Memory; PM Acc = Prospective Memory Accuracy; FPM = Focal Prospective Memory; NFPM = Non-focal Prospective Memory.

Table 3

Descriptive Task Statistics for Experiment 1

	Mean (*SD*)	Min, Max	Skewness	Kurtosis
Operation Span	60.56 (10.04)	[11, 75]	-1.18	2.59
Symmetry Span	30.06 (7.29)	[4, 42]	-0.86	0.60
LDT RT (Control)	717.01 (265.60)	[546.71, 1118.93]	3.83	27.08
LDT RT (FPM)	873.00 (379.84)	[625.99, 1809.75]	3.47	18.94
LDT RT (NFPM)	779.40 (288.27)	[651.48, 1035.64]	3.34	19.48
LDT Acc (Control)	0.95 (0.22)	[0.84, 1.00]	-4.15	15.24
LDT Acc (FPM)	0.96 (0.20)	[0.72, 1.00]	-4.76	20.64
LDT Acc (NFPM)	0.98 (0.15)	[0.91, 1.00]	-6.54	40.72
FPM Accuracy	0.49 (0.50)	[0.00, 1.00]	0.03	-2.00
NFPM Accuracy	0.47 (0.50)	[0.00, 1.00]	0.11	-1.99

Note. LDT RT = Mean lexical decision task response times on correct trials; FPM = Focal Prospective memory; NFPM = Non-focal Prospective Memory; LDT Acc = Lexical decision task accuracy.

Table 4

Results of Model Comparison for Pupil Dilation During Ongoing Task Trials in Experiment 1

	Proportional 'Weight'	LOO
Model 1.3	1.00	654.50
Model 1.2	0.00	927.30
Model 1.4	0.00	1001.60
Model 1.1	0.00	2538.90

Table 5

Posterior Estimates From Model 1.3 for Pupil Dilation During Ongoing Task Trials

	Mean (*SD*)	95% Credible Interval	Rhat
Control Intercept	0.89 (0.00)	[0.89, 0.90]	1.00
FPM Intercept	0.94 (0.00)	[0.93, 0.94]	1.00
NFPM Intercept	0.92 (0.00)	[0.91, 0.92]	1.00
Trial*Control	-0.00 (0.00)	[-0.00, -0.00]	1.00
Trial*FPM	-0.00 (0.00)	[-0.00, -0.00]	1.00
Trial*NFPM	-0.00 (0.00)	[-0.00, -0.00]	1.00
WM*Control	-0.01 (0.00)	[-0.01, -0.01]	1.00
WM*FPM	0.02 (0.00)	[0.02, 0.02]	1.00
WM*NFPM	-0.00 (0.00)	[-0.01, 0.00]	1.00
Sigma	0.15 (0.00)	[0.15, 0.15]	1.00

Note. Mean = Mean for a given parameter value from the posterior distribution; SD = Mean of the standard deviation for a given parameter value resulting from the MCMC sampling procedure; FPM = Focal Prospective Memory; NFPM = Non-focal Prospective Memory; WM = Working Memory.

Table 6

Posterior Estimates From Model 1.5 for Prospective Memory Accuracy

	Mean (*SD*)	95% Credible Interval	Rhat
FPM Intercept	0.63 (0.16)	[0.33, 0.94]	1.00
NFPM Intercept	0.62 (0.16)	[0.31, 0.93]	1.00
WM*FPM	0.25 (0.10)	[0.06, 0.45]	1.00
WM*NFPM	-0.09 (0.10)	[-0.28, 0.12]	1.00
Prior Pupil Size*FPM	0.60 (0.17)	[0.28, 0.93]	1.00
Prior Pupil Size*NFPM	0.56 (0.18)	[0.22, 0.91]	1.00

Note. Mean = Mean for a given parameter value from the posterior distribution; SD = Mean of the standard deviation for a given parameter value resulting from the MCMC sampling procedure; FPM = Focal Prospective Memory; NFPM = Non-focal Prospective Memory; WM = Working Memory.

Table 7

Posterior Estimates From Model 1.6 for RTs During Correct Ongoing Task Trials

	Mean (*SD*)	95% Credible Interval	Rhat
Control Intercept	544.44 (7.33)	[530.04, 558.73]	1.00
FPM Intercept	686.64 (3.90)	[678.87, 694.36]	1.00
NFPM Intercept	604.91 (3.93)	[596.67, 612.02]	1.00
WM*Control	-2.56 (2.30)	[-3.95, 0.21]	1.00
WM*FPM	13.74 (1.82)	[9.41, 16.30]	1.00
WM*NFPM	7.60 (1.63)	[4.50, 10.83]	1.00
Pupil Size*FPM	244.15 (9.38)	[221.76, 273.59]	1.00
Pupil Size*NFPM	198.69 (8.70)	[183.15, 206.58]	1.00
Sigma	313.36 (1.12)	[311.13, 315.59]	1.00

Note. Mean = Mean for a given parameter value from the posterior distribution; SD = Mean of the standard deviation for a given parameter value resulting from the MCMC sampling procedure; FPM = Focal Prospective Memory; NFPM = Non-focal Prospective Memory; WM = Working Memory.

Table 8

Posterior Estimates From Model 1.7 for Pupil Dilation on Trials Immediately Following the Presentation of a Prospective Memory Target

	Mean (*SD*)	95% Credible Interval	Rhat
Control Intercept	0.75 (0.01)	[0.74, 0.77]	1.00
FPM Intercept	0.84 (0.01)	[0.82, 0.87]	1.00
NFPM Intercept	0.79 (0.01)	[0.76, 0.81]	1.00
WM*Control	-0.01 (0.01)	[-0.02, 0.00]	1.00
WM*FPM	0.01 (0.01)	[0.01, 0.03]	1.00
WM*NFPM	0.01 (0.01)	[-0.01, 0.03]	1.00
PM Acc*FPM	0.04 (0.02)	[0.01, 0.07]	1.00
PM Acc*NFPM	0.04 (0.02)	[0.01, 0.07]	1.00
Sigma	0.15 (0.00)	[0.14, 0.16]	1.00

Note. Mean = Mean for a given parameter value from the posterior distribution; SD = Mean of the standard deviation for a given parameter value resulting from the MCMC sampling procedure; FPM = Focal Prospective Memory; NFPM = Non-focal Prospective Memory; WM = Working Memory; PM Acc = Prospective Memory Accuracy.

Table 9

Models Computed in Experiment 2

	DV	Summary of Model Parameters
Model 2.1	Pupil Size	Initial pupil size and change in pupil size across trials varies as a function of condition
Model 2.2	PM Acc.	Accuracy differs between conditions & is positively related to pupil size on previous trials
Model 2.3	RTs	RTs vary between conditions and as a function of pupil size
Model 2.4	Pupil Size	Pupil size elevated following an accurate response to a prospective memory target, though the effect differs between conditions

Note. DV = Dependent Variable; PM Acc = Prospective Memory Accuracy; RTs = Response Times.

Table 10

Priors Used for Models 2.1-2.4

	M 2.1	M 2.2	M 2.3	M 2.4
F_{LL}PM Intercept	1 (.1)	.4 (.1)	600 (100)	1 (.1)
F_{HL}PM Intercept	1.2 (.1)	.7 (.1)	700 (100)	1.2 (.1)
NFPM Intercept	1 (.1)	.4 (.1)	600 (100)	1 (.1)
Trial*F_{LL}PM	-.01 (.01)			
Trial*F_{HL}PM	0 (.01)			
Trial*NFPM	-.01 (.01)			
PM Acc*F_{LL}PM				.25 (.1)
PM Acc*F_{HL}PM				.25 (.1)
PM Acc*NFPM				.25 (.1)
Prior Pupil Size*F_{LL}PM		.01 (.1)		
Prior Pupil Size*F_{HL}PM		.01 (.1)		
Prior Pupil Size*NFPM		.01 (.1)		
Pupil Size*F_{LL}PM			50 (25)	
Pupil Size*F_{HL}PM			100 (25)	
Pupil Size*NFPM			50 (25)	
Sigma	.3 (.1)		300 (100)	.3 (.1)

Note. Parameters for the different models are identified in the left-most column. No model included all of the parameters. Priors are presented in columns as ‘Mean (SD)’. M = Model; F_{LL}PM = Focal Low-load Prospective Memory; F_{HL}PM = Focal High-load Prospective Memory; NFPM = Non-focal Prospective Memory; PM Acc = Prospective Memory Accuracy.

Table 11

Descriptive Task Statistics for Experiment 2

	Mean (*SD*)	Min, Max	Skewness	Kurtosis
LDT RT (F_{LL}PM)	765.60 (290.73)	[610.46, 1185.99]	3.29	17.45
LDT RT (F_{HL}PM)	822.20 (343.04)	[588.82, 1356.93]	3.33	18.69
LDT RT (NFPM)	776.60 (297.76)	[568.26, 1129.90]	3.31	19.42
LDT Acc (F_{LL}PM)	0.97 (0.17)	[0.90, 1.00]	-5.71	30.92
LDT Acc (F_{HL}PM)	0.97 (0.18)	[0.77, 1.00]	-6.95	52.79
LDT Acc (NFPM)	0.97 (0.16)	[0.93, 1.00]	-5.99	34.29
F_{LL}PM Accuracy	0.44 (0.50)	[0.00, 1.00]	0.23	-1.96
F_{HL}PM Accuracy	0.18 (0.38)	[0.00, 1.00]	1.68	0.84
NFPM Accuracy	0.27 (0.44)	[0.00, 1.00]	1.04	-0.93

Note. LDT RT = Mean lexical decision task response times on correct trials; F_{LL}PM = Focal Low-load Prospective Memory; F_{HL}PM = Focal High-load Prospective Memory; NFPM = Non-focal Prospective Memory; LDT Acc = Lexical decision task accuracy.

Table 12

Posterior Estimates From Model 2.1 for Pupil Dilation During Ongoing Task Trials

	Mean (*SD*)	95% Credible Interval	Rhat
F_{LL}PM Intercept	0.91 (0.00)	[0.91, 0.92]	1.00
F_{HL}PM Intercept	0.95 (0.00)	[0.94, 0.95]	1.00
NFPM Intercept	0.91 (0.00)	[0.91, 0.92]	1.00
Trial*F_{LL}PM	-0.00 (0.00)	[-0.00, -0.00]	1.00
Trial*F_{HL}PM	-0.00 (0.00)	[-0.00, -0.00]	1.00
Trial*NFPM	-0.00 (0.00)	[-0.00, -0.00]	1.00
Sigma	0.15 (0.00)	[0.15, 0.15]	1.00

Note. Mean = Mean for a given parameter value from the posterior distribution; SD = Mean of the standard deviation for a given parameter value resulting from the MCMC sampling procedure; F_{LL}PM = Focal Low-load Prospective Memory; F_{HL}PM = Focal High-load Prospective Memory; NFPM = Non-focal Prospective Memory.

Table 13

Posterior Estimates From Model 2.2 for Prospective Memory Accuracy

	Mean (*SD*)	95% Credible Interval	Rhat
F_{LL}PM Intercept	0.35 (0.59)	[0.21, 0.46]	1.00
F_{HL}PM Intercept	0.25 (0.57)	[0.15, 0.37]	1.00
NFPM Intercept	0.24 (0.59)	[0.14, 0.37]	1.00
Prior Pupil Size*F_{LL}PM	0.58 (0.60)	[0.40, 0.76]	1.00
Prior Pupil Size*F_{HL}PM	0.37 (0.60)	[0.19, 0.51]	1.00
Prior Pupil Size*NFPM	0.52 (0.60)	[0.37, 0.74]	1.00

Note. Mean = Mean for a given parameter value from the posterior distribution; SD = Mean of the standard deviation for a given parameter value resulting from the MCMC sampling procedure; F_{LL}PM = Focal Low-load Prospective Memory; F_{HL}PM = Focal High-load Prospective Memory; NFPM = Non-focal Prospective Memory.

Table 14

Posterior Estimates From Model 2.3 for RTs During Correct Ongoing Task Trials

	Mean (*SD*)	95% Credible Interval	Rhat
F_{LL}PM Intercept	747.81 (2.24)	[743.91, 753.96]	1.00
F_{HL}PM Intercept	762.23 (3.92)	[754.68, 770.14]	1.00
NFPM Intercept	747.67 (1.94)	[744.03, 752.05]	1.00
Pupil Size*F_{LL}PM	35.59 (3.93)	[26.71, 43.58]	1.00
Pupil Size*F_{HL}PM	56.32 (5.24)	[48.77, 67.21]	1.00
Pupil Size*NFPM	36.14 (3.87)	[29.12, 43.38]	1.00
Sigma	314.60 (2.19)	[310.31, 318.97]	1.00

Note. Mean = Mean for a given parameter value from the posterior distribution; SD = Mean of the standard deviation for a given parameter value resulting from the MCMC sampling procedure; F_{LL}PM = Focal Low-load Prospective Memory; F_{HL}PM = Focal High-load Prospective Memory; NFPM = Non-focal Prospective Memory.

Table 15

Posterior Estimates From Model 2.4 for Pupil Dilation on Trials Immediately Following the Presentation of a Prospective Memory Target

	Mean (*SD*)	95% Credible Interval	Rhat
F_{LL}PM Intercept	0.82 (0.02)	[0.77, 0.86]	1.00
F_{HL}PM Intercept	0.83 (0.01)	[0.80, 0.86]	1.00
NFPM Intercept	0.80 (0.02)	[0.76, 0.85]	1.00
PM Acc*F_{LL}PM	0.04 (0.02)	[-0.00, 0.09]	1.00
PM Acc*F_{HL}PM	0.06 (0.03)	[-0.01, 0.12]	1.00
PM Acc*NFPM	0.04 (0.03)	[-0.12, 0.09]	1.00
Sigma	0.15 (0.01)	[0.14, 0.16]	1.00

Note. Mean = Mean for a given parameter value from the posterior distribution; SD = Mean of the standard deviation for a given parameter value resulting from the MCMC sampling procedure; F_{LL}PM = Focal Low-load Prospective Memory; F_{HL}PM = Focal High-load Prospective Memory; NFPM = Non-focal Prospective Memory; PM Acc = Prospective Memory Accuracy.

Table 16

Predictions for Pupil Dilation Data Generated by Each Prospective Memory Theory

Theory	Different Focal vs. Non-Focal	Different Non-Focal vs. Control	Correlated with WM	Correlated with Non-Focal PM
PAM	N (✓)	Y (X)	Y (X)	Y (✓)
Multiprocess	Y (X)	S (✓)	S (✓)	Y (✓)
Delay	N (✓)	Y (X)	N (X)	N (X)

Note. N = No; Y = Yes; S = Sometimes; WM = Working memory score. PM = Prospective memory accuracy; (✓) = the prediction was supported by the current data; (X) = the prediction was not supported by the current data.

APPENDIX B

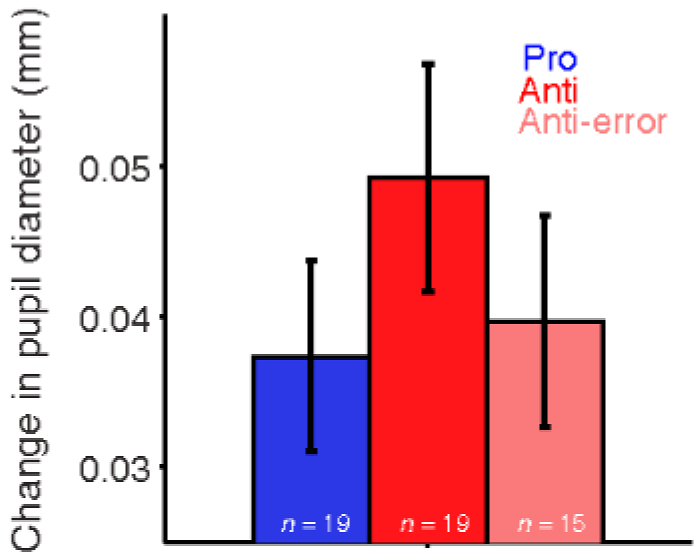

Figure 1. Adapted from Wang, Brien, and Munoz (2015). Pupil diameter is plotted between trials where participants performed pro-saccade, correctly performed anti-saccade, or erroneously performed anti-saccade. Pupil diameter was largest prior to an accurate anti-saccade trial.

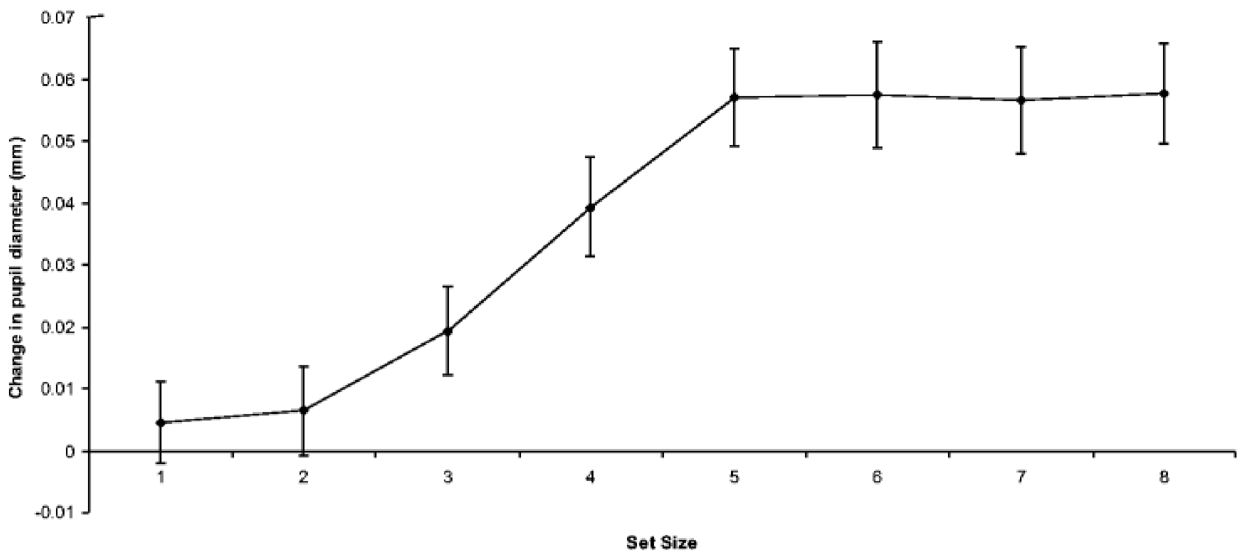

Figure 2. Adapted from Unsworth and Robison (2015). Pupil diameter increased as a function of the load placed on working memory.

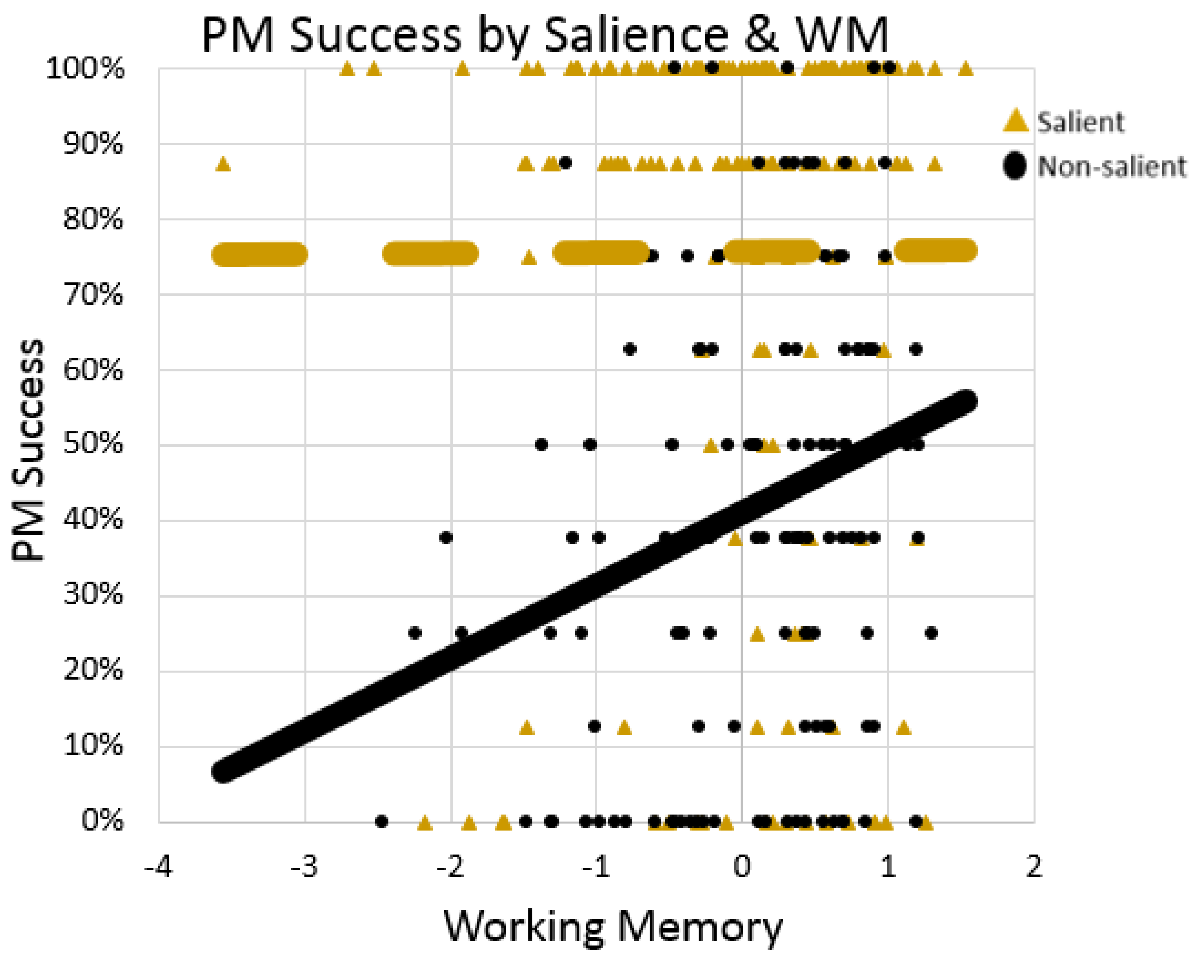

Figure 3. Adapted from Christopher, Fansher, and Redick (2019). The relationship between individual differences in working memory and prospective memory as a function of cue saliency.

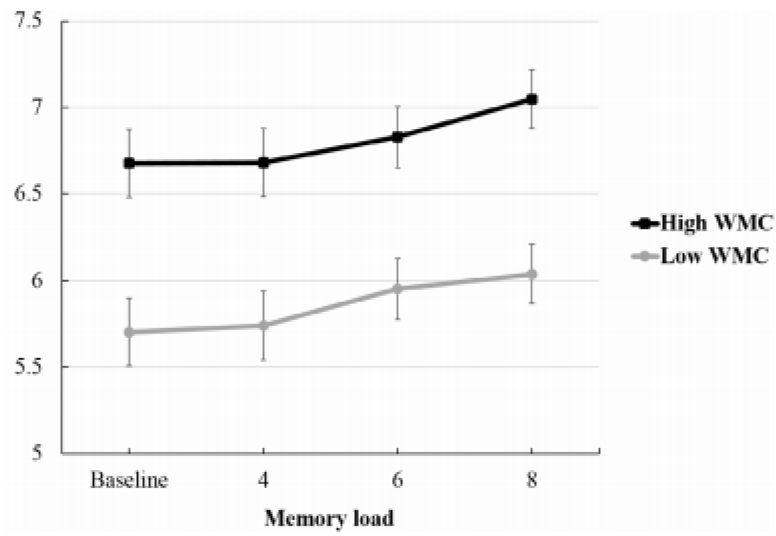

Figure 4. Adapted from Tsukahara, Harrison, and Engle (2016). Pupil diameter in millimeters is plotted vertically along the y-axis. Pupil diameter increased as a function of load, and was larger in high working memory participants than in low working memory participants. WMC = Working memory capacity.

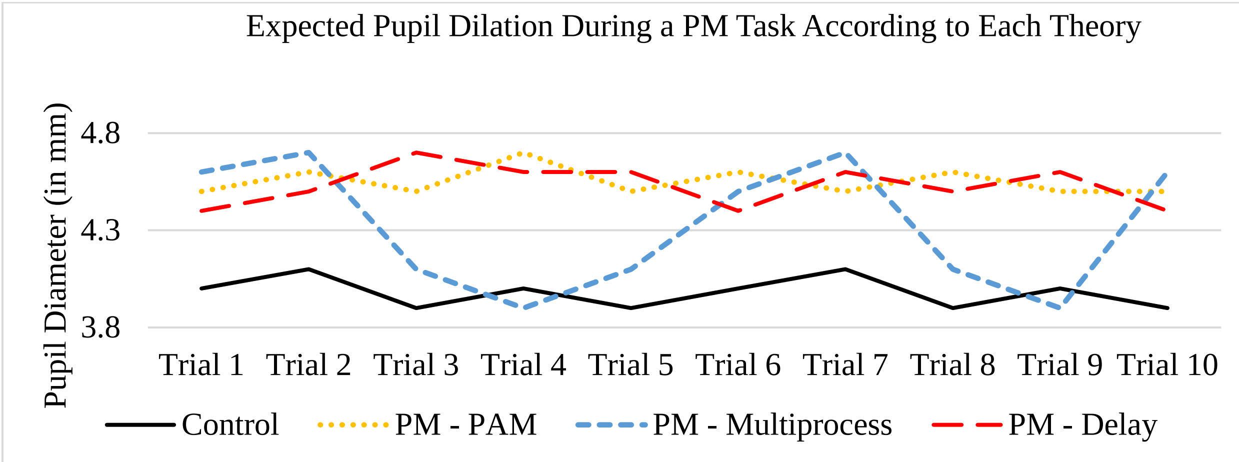

Figure 5. Divergent predictions for pupil dilation between theories of prospective memory on a non-focal task. All of the theories of prospective memory that are discussed in the current manuscript would predict that pupil dilation in a control condition would be consistently low throughout the task. PAM theory would predict consistently larger pupil dilations, compared to a control condition, across a non-focal prospective memory task. Multiprocess theory would predict that compared to a control condition, participant's pupils would be more dilated during portions of a non-focal prospective memory task, but that pupil dilation would also decrease to control levels during other portions of the task. Finally, delay theory is somewhat ambiguous about cognitive load during prospective memory tasks. A possible prediction that one could reasonably generate from delay theory is that pupil dilation in a non-focal prospective memory task would be elevated relative to control because of an additional evidence accumulator to facilitate the prospective memory response. PM = Prospective memory.

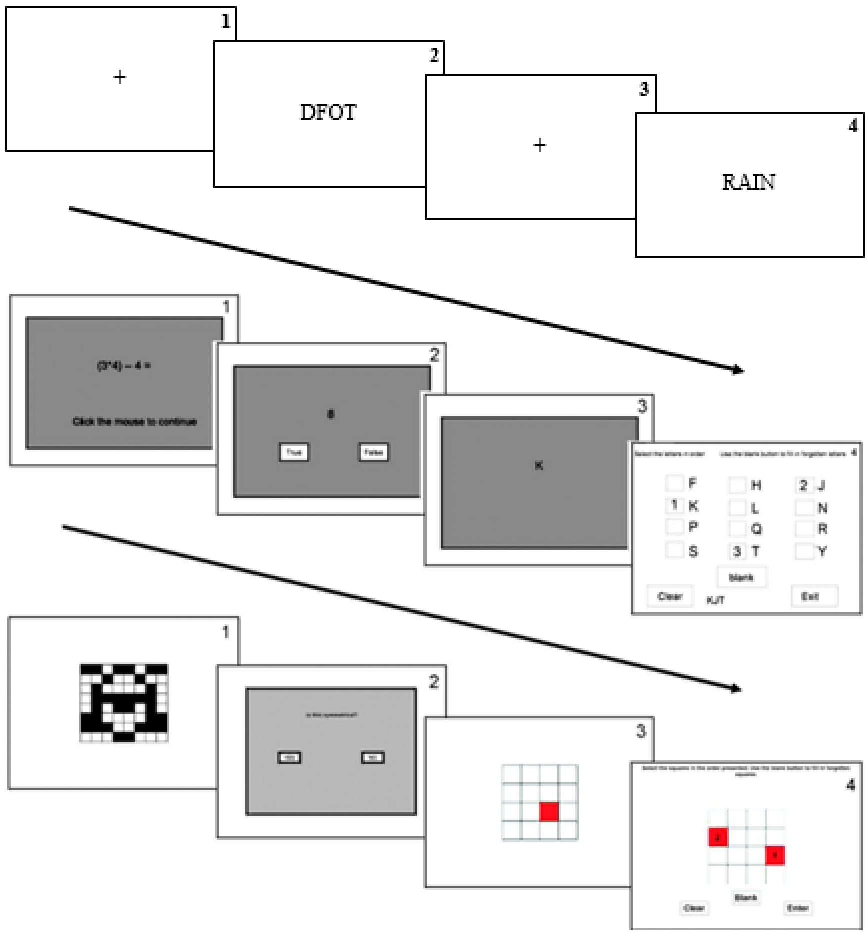

Figure 6. The progression of example trials from each task. Top: Prospective memory/lexical decision task, Middle: Operation span task, Bottom: Symmetry span task.

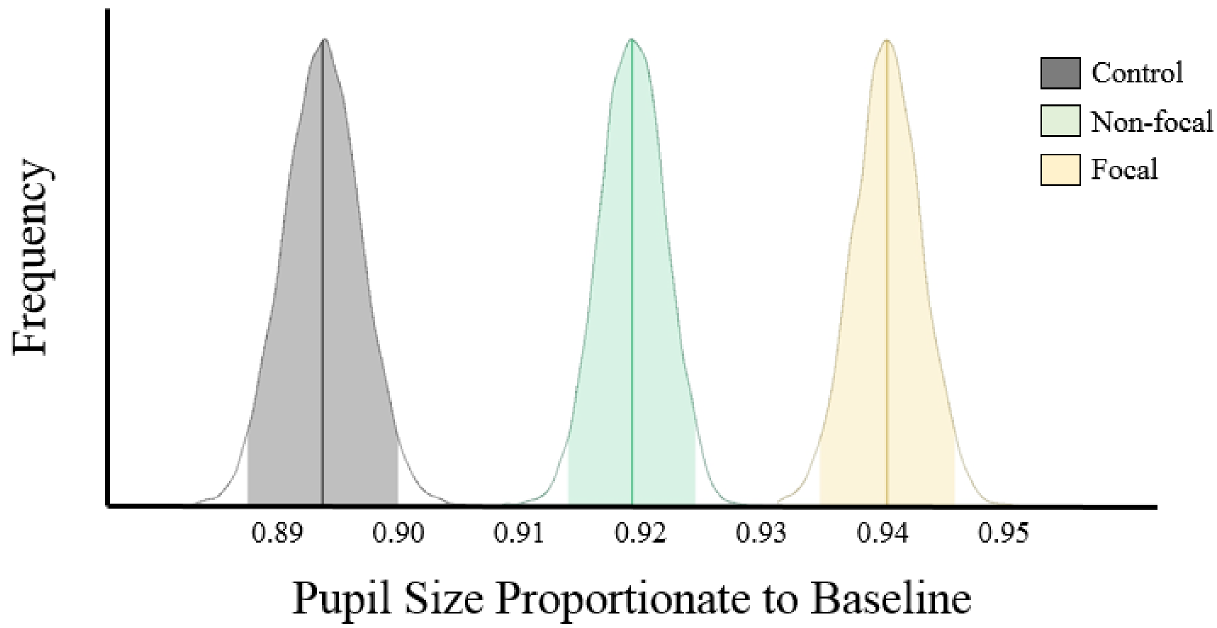

Figure 7. Posterior distributions of intercepts for each condition. Based on posterior data from Model 1.3.

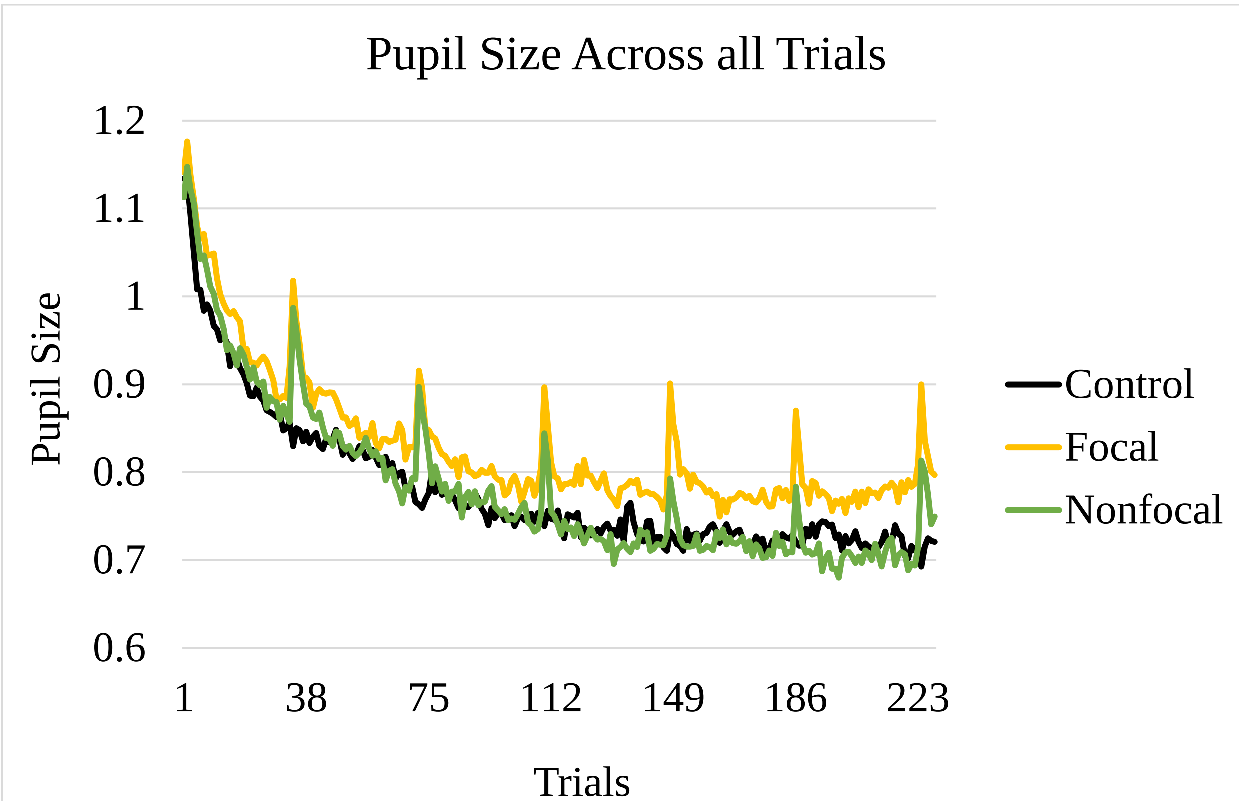

Figure 8. Observed pupil dilation data across all trials from Experiment 1.

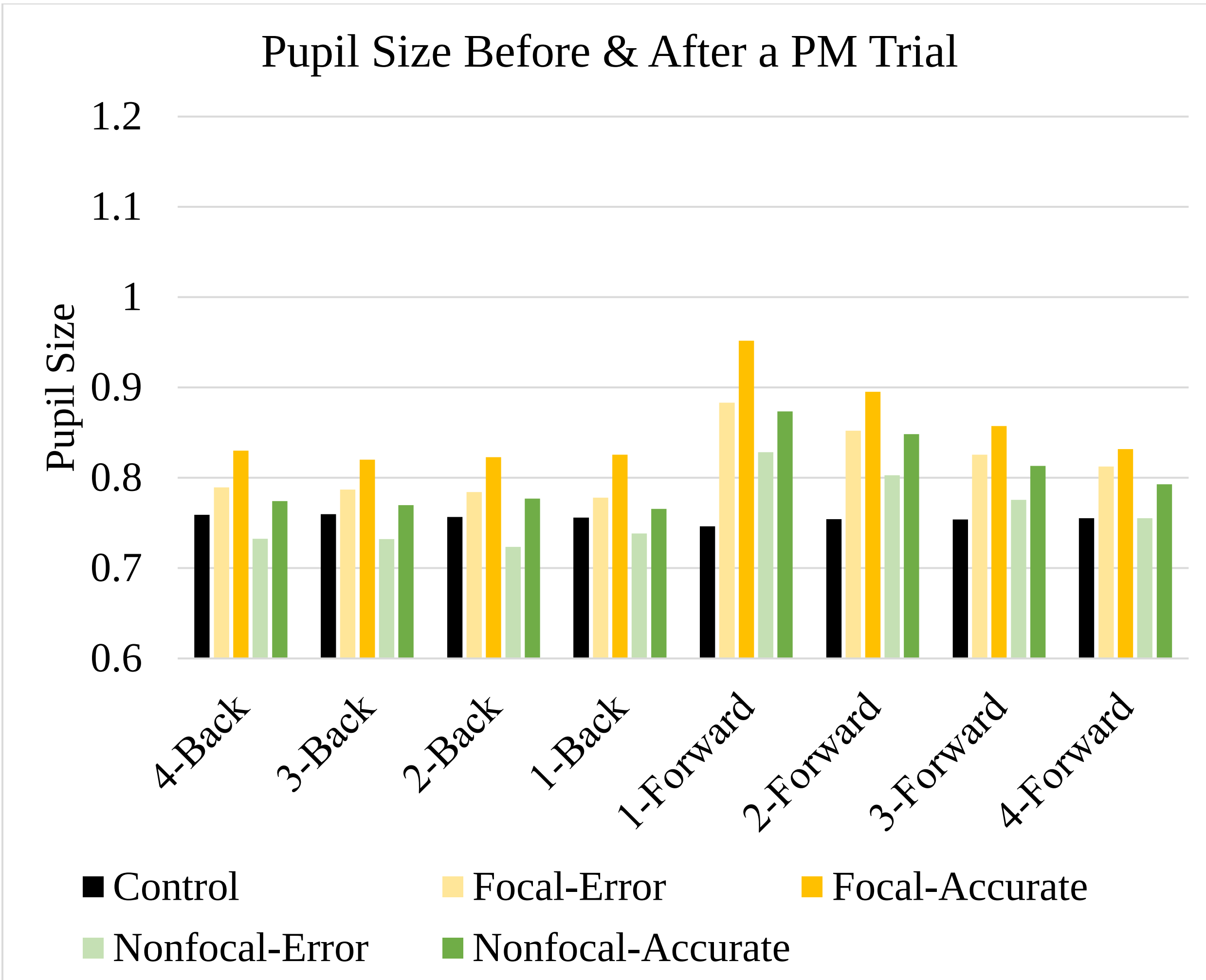

Figure 9. Pupil size is plotted by condition and whether or not the prospective memory trial was accurately responded to. Pupil size is further plotted as either some number back (i.e., before the target), or some number forward (i.e., after the target). Pupil data is from Experiment 1. PM = Prospective memory.

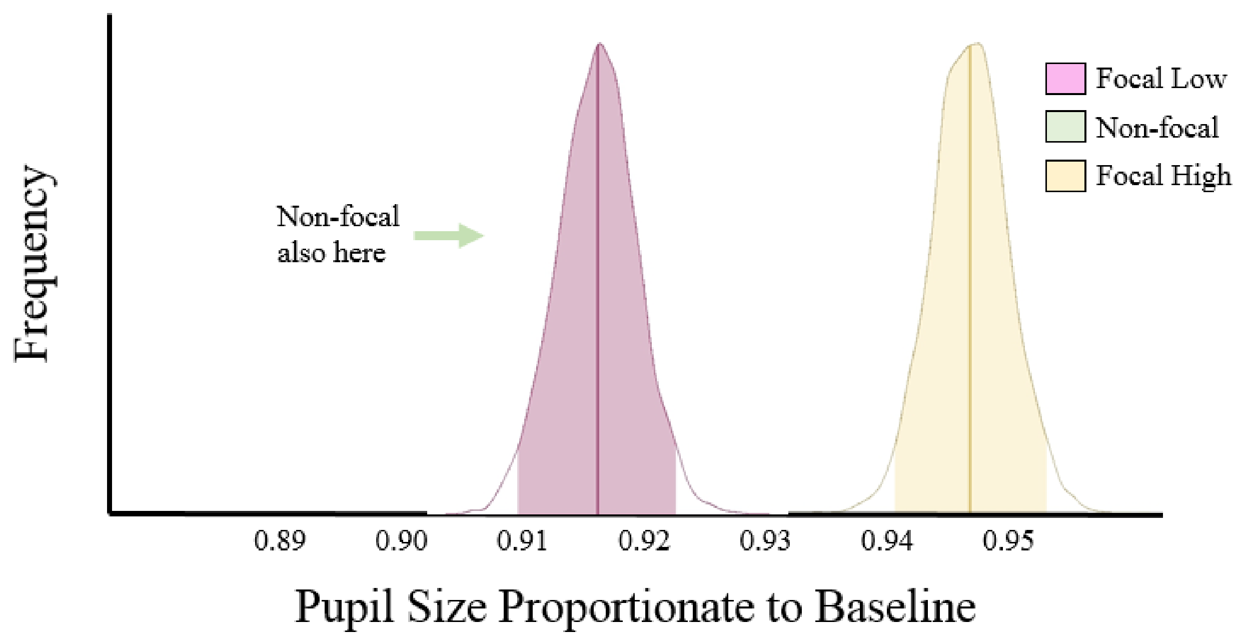

Figure 10. Posterior distributions of intercepts for each condition. Based on posterior data from Model 2.1. The posterior distributions for the focal low-load and non-focal conditions occupy the same space visually. They do not actually overlap perfectly, but are similar enough to be visually indistinct.

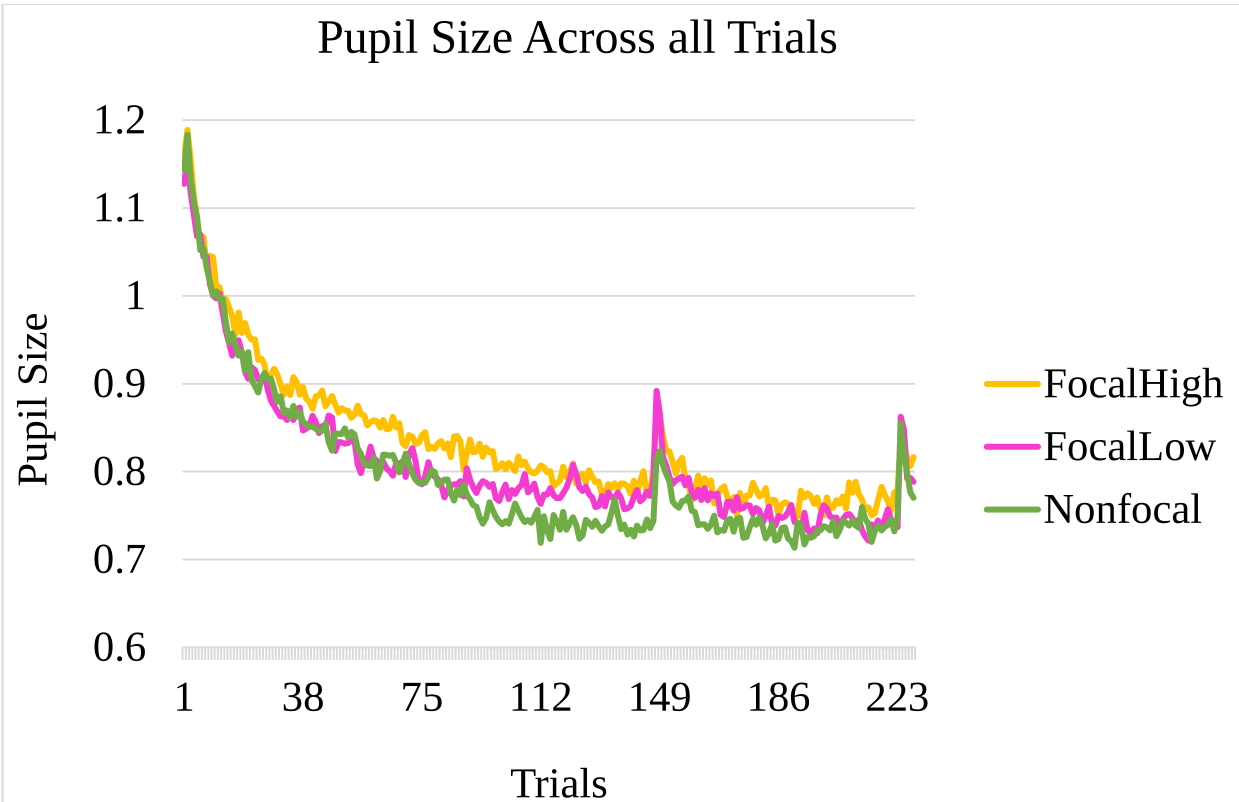

Figure 11. Observed pupil dilation data across all trials from Experiment 2.

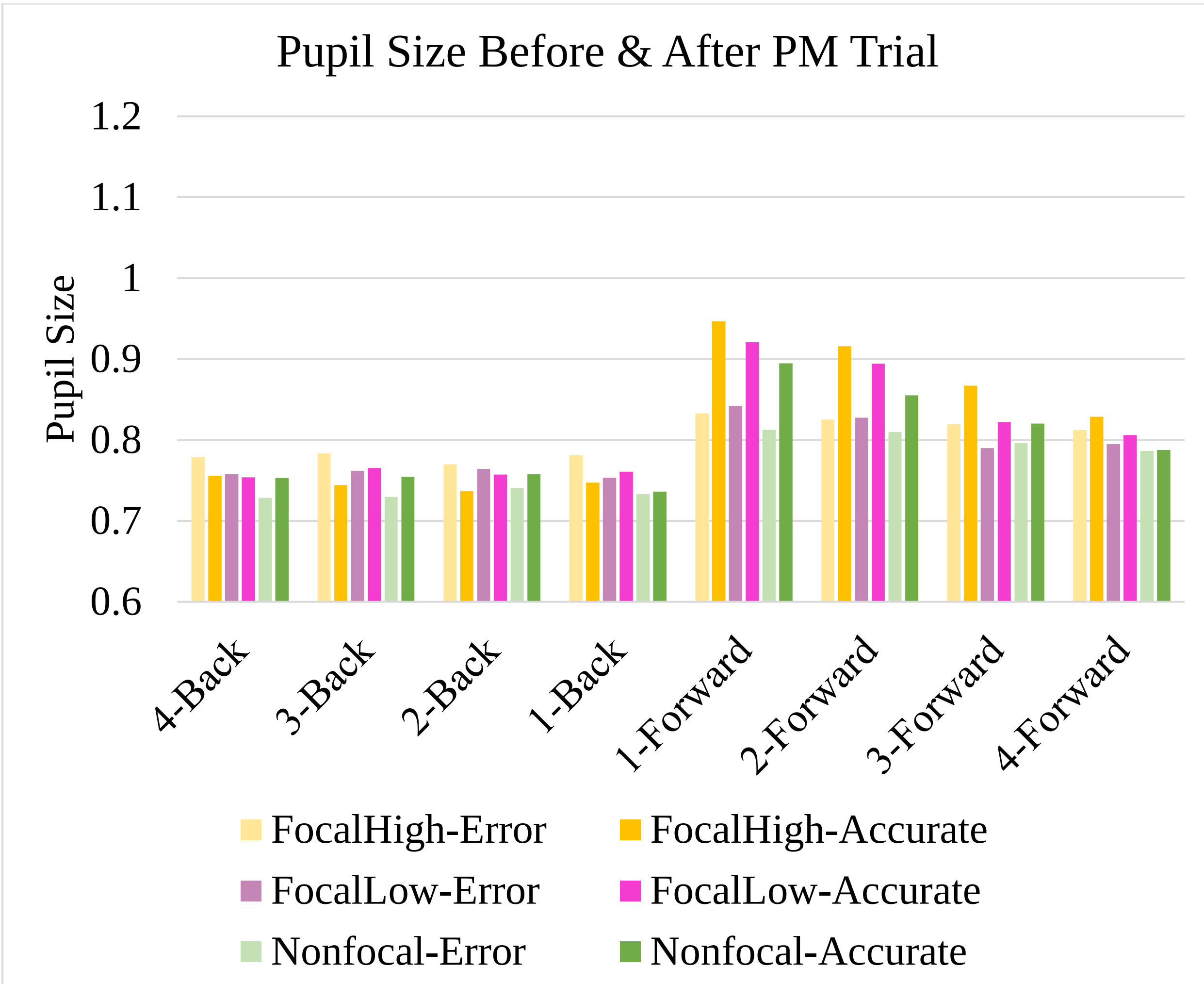

Figure 12. Pupil size is plotted by condition and whether or not the prospective memory trial was accurately responded to. Pupil size is further plotted as either some number back (i.e., before the target), or some number forward (i.e., after the target). Pupil data is from Experiment 2. PM = Prospective memory.

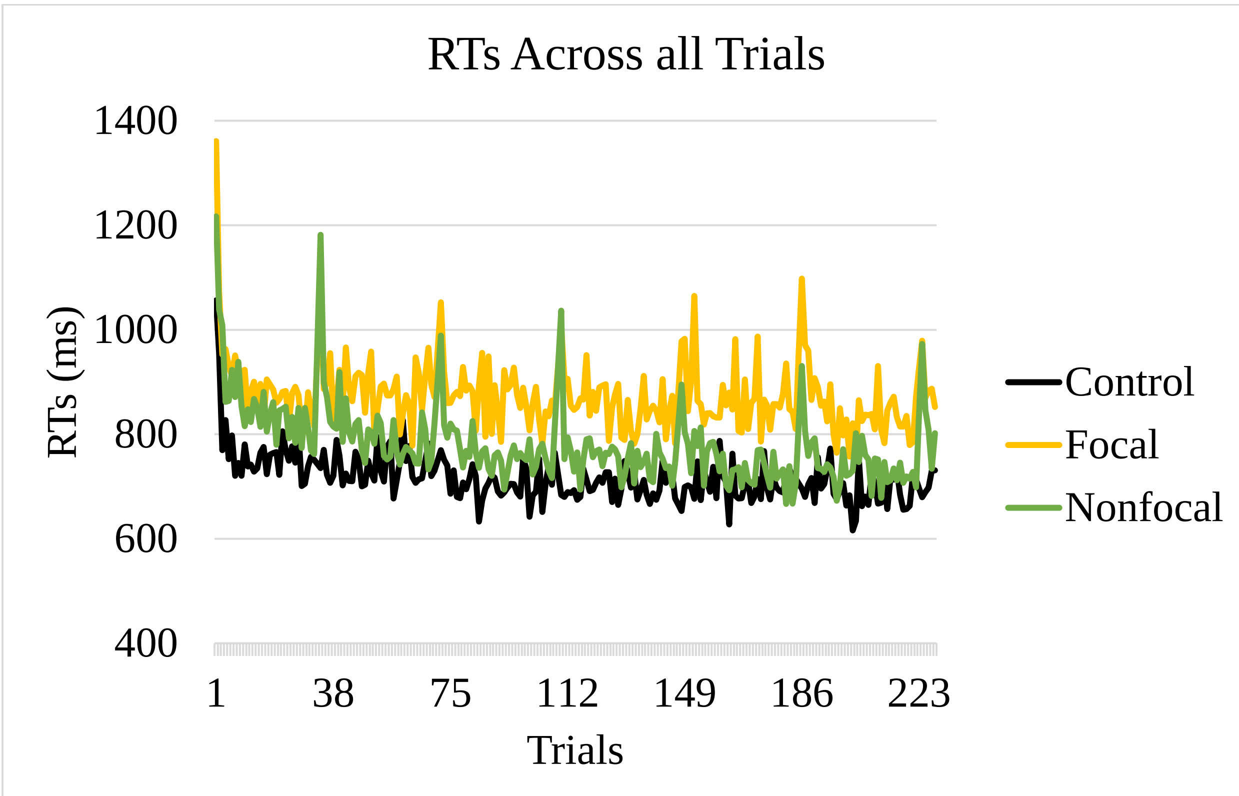

Figure 13. Observed RT data across all trials from Experiment 1.

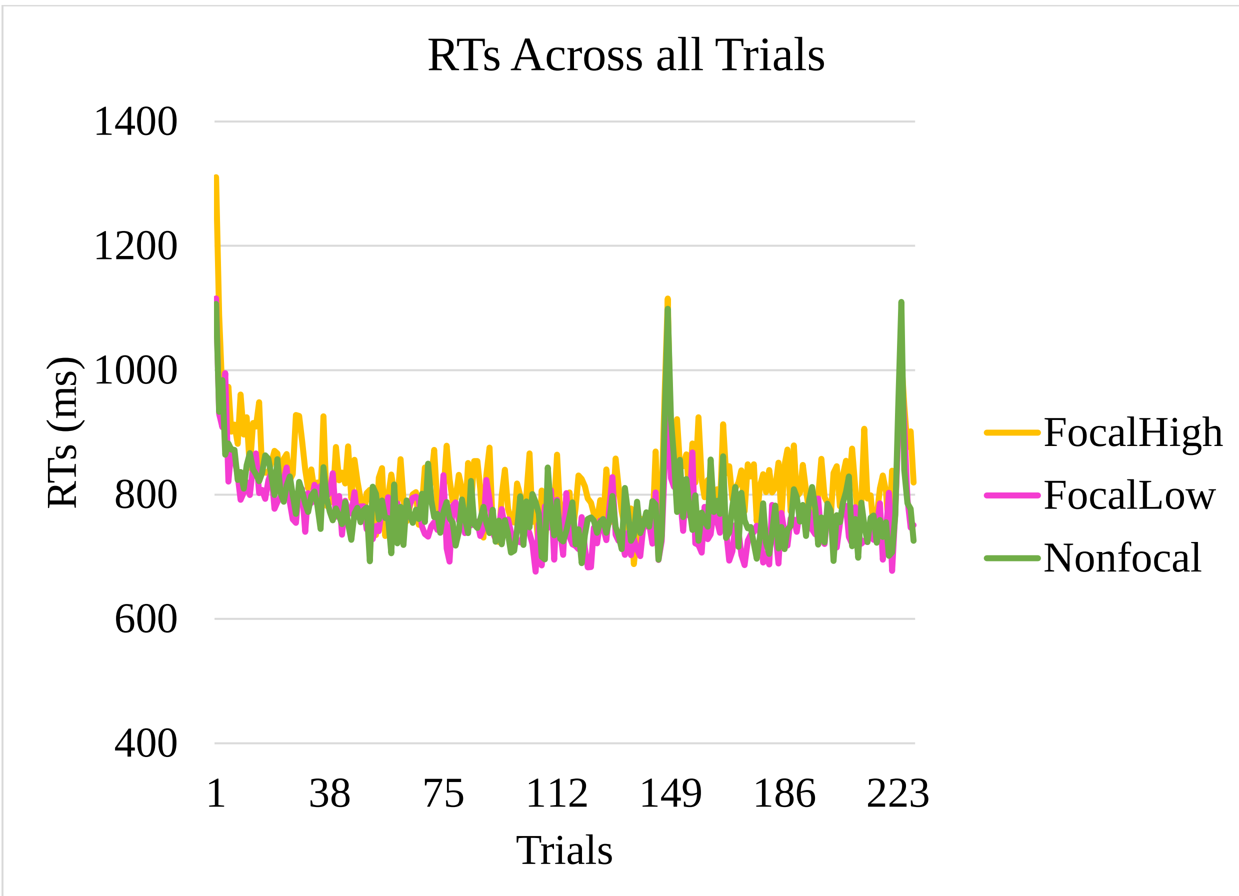

Figure 14. Observed RT data across all trials from Experiment 2.

APPENDIX C

Non-focal. Instead of pressing the "J" key for words and the "F" key for non-words, you will also be required to press the SPACE key for any word that describes a color.

Pressing the SPACE key right after pressing another key when a color name appears will still count as an accurate response.

Focal/High-load. Instead of pressing the "J" key for words and the "F" key for non-words, you will also be required to press the SPACE key for any one of the following words: code, mind, handy, wine, clone, and music.

Pressing the SPACE key right after pressing another key when one of these words appears will still count as an accurate response.

Low-load. Instead of pressing the "J" key for words and the "F" key for non-words, you will also be required to press the SPACE key for any one of the following words: code and mind.

Pressing the SPACE key right after pressing another key when one of these words appears will still count as an accurate response.