

## **Integrated Responding Improves Prospective Memory Accuracy**

David Elliott<sup>1</sup>, Luke Strickland<sup>2</sup>, Shayne Loft<sup>3</sup>, Andrew Heathcote<sup>1</sup>

<sup>1</sup> The School of Psychology,  
The University of Tasmania, Australia

<sup>2</sup> The Future of Work Institute,  
Curtin University, Australia

<sup>3</sup> The School of Psychological Science,  
The University of Western Australia, Australia

### **Author note**

All data and code associated with this manuscript are located at <https://osf.io/ryd7f/>

Prospective memory (PM) supports the planning and execution of future activities, and is particularly important in applied settings. We investigate a new response method that aims to improve PM accuracy by integrating the responses to an occasional PM task and a routine ongoing lexical-decision task. Instead of the most common three-choice method where the PM response replaces the ongoing response, participants were obligated to make explicit PM (present vs. absent) and ongoing (word vs. non-word) classifications on every trial through a four-choice response. Although replacement and obligatory responses were initially similar in PM accuracy an advantage emerged with practice for the new obligatory method that was not simply due to slower responding associated with making four vs. three choices. The nature of the errors differed between methods, with obligatory responding being characterised by fast PM errors and replacement by slower errors, suggesting avenues for further potential improvements in PM accuracy.

Keywords: Prospective Memory, dual tasking, response methods.

Intro: 1386

Results: 529

Discussion: 1352

3267

Prospective Memory (PM) refers to the processes involved in the successful planning and execution of activities in the future, either at a particular time (time-based PM) or in response to a particular event (event-based PM). Prospective remembering is often necessary for successful everyday functioning (Einstein et al., 1995; Kliegel & Martin, 2003). Event-based PM tasks are particularly common in safety-critical settings such as healthcare and aviation (Dismukes, 2012; Loft, Dismukes, & Grundgeiger, 2019). The prevalence of what are often high-stakes failures to act on PM cues (i.e., *PM misses*) in those settings makes it important to identify interventions to improve PM performance.

Perhaps the most common attempted intervention method to improve PM accuracy is providing individuals with reminders, and reminders can indeed often benefit PM (e.g., Chen, Lian, Yang, Liu, & Meng, 2017; Finstad, Bink, McDaniel, & Einstein, 2006; Guynn, McDaniel, & Einstein, 1998; Gilbert, 2015a, 2015b, 2015c; Loft, Smith, & Bhaskara, 2011; Vortac, Edwards, & Manning, 1995). However, reminders are not always effective (e.g., Guynn et al. 1998; Loft et al. 2011; Vortac et al. 1995) and can have drawbacks. For example, in simulated air traffic control settings, reminders have not been found to be effective in improving PM unless they are set to flash (i.e., inducing attention capture; Jonides & Yantis, 1988) when a PM cue is present and the action needs to be performed (Loft et al., 2011; Loft 2014). Such reminders may be undesirable in some settings because visual-attention capture is distracting, potentially taking an operator's attention away from other safety-critical tasks.

It has been suggested that PM misses occur because PM responses compete for retrieval with more routine responses associated with ongoing tasks (Loft & Remington, 2010). Hence, another route to improving PM accuracy might be slowing ongoing task responding so that it does not pre-empt the PM response (Loft & Remington, 2013; Heathcote, Loft & Remington,

2015). Unfortunately, although *PM costs* – slower ongoing task responses when it is possible that PM responses may be required compared to when they are not – have been shown with computational models to reflect more cautious ongoing task responding (e.g., Heathcote et al. 2015), more recent evidence, both empirical (Anderson, Rummel & McDaniel, 2018) and from modelling (Strickland, Loft, Remington & Heathcote, 2018; Strickland, Loft & Heathcote, 2020), indicates that increased ongoing-task response caution is not effective in reducing PM misses.

In the current study we test a new way of improving PM accuracy based on the method by which ongoing and PM responses are made. Previous studies have used one of two response methods. *Replacement* instructions can be either explicit (make a PM response instead an ongoing-task response, e.g., Horn & Bayen, 2015; Strickland, Heathcote, Remington, & Loft, 2017) or implicit (make PM response when the PM target is presented, e.g., Einstein & McDaniel, 2005; Loft & Remington, 2013), which also typically results in the PM response replacing the ongoing task response. *Dual* instructions require the PM response to be made after the ongoing-task response (e.g., Hicks, Marsh, Cook, 2005; Loft & Yeo, 2007). In both cases binary ongoing task responses (e.g., in our experiment participants classified a letter string as a *word* vs. *non-word*) are made with a pair of keys and a PM detection response with a third key (e.g., indicating if the letter string contains the syllable *tor*).

Here we propose a new response method, *obligatory responding*, where participants press one of four keys to simultaneously make a combined ongoing and PM task classification (e.g., *PM word*, *PM non-word*, *non-PM word* and *non-PM non-word*). This new response method is directly applicable to safety-critical settings such as air traffic control where operators interact with a computerized interface. For example, when deciding which aircraft should adjust its flight

path when a potential aircraft conflict is detected, the required pilot instructions may differ for some aircraft. The decision and standard instructions could be issued by selecting between one pair of buttons, whereas the decision and less frequently used alternative instructions (e.g., for older model aircraft, or for when the conflict occurs at an unusually low altitude) could be made with another pair (Fothergill & Neal, 2008). The increasing use of computer interfaces that can be flexibly configured to associate buttons with different responses makes this approach widely applicable (Boehm-Davis, Durso, & Lee, 2015).

Three studies that have explicitly compared replacement with dual responding have produced mixed results. With very simple ongoing and PM tasks, Bisiacchi, Schiff, Ciccola and Kliegel (2009) and Gilbert, Hadjipavlou and Raelison (2013) reported fewer PM misses with dual responding (18% vs. 26% and 29% vs. 69%, respectively), with no effect on ongoing-task response time (RT) or accuracy. In contrast, with more difficult choices, Pereira, Albuquerque, Santos (2017) reported no significant difference in PM misses (26% vs. 28%) or ongoing-task accuracy, but slower ongoing-task RTs in the dual condition. However, these dual vs. replacement comparisons have methodological limitations. Most critically, the effect of dual responding likely depends upon the length of the inter-trial interval. In Bisiacchi et al. (2009) and in Pereira et al. (2017), PM responses were required during an approximately one second inter stimulus interval, during which participants were not occupied with any other task. Unfilled delays of this length have been shown to benefit PM in replacement paradigms (Loft & Remington, 2013; but for contrasting results see Ball, Vogel, Ellis, & Brewer, 2021), so the delay, rather than dual-responding per-se, could cause the dual improvement over replacement responding. Another limitation is that the motor response time associated with the ongoing task response in dual response conditions introduces a confound when comparing PM RTs across dual

and replacement conditions<sup>1</sup>. This is an issue for the analysis we perform here because PM RTs carry important information regarding the underlying psychological process driving PM.

Given these considerations, our experiment compares the new obligatory responding method to the replacement method. The new method makes explicit ongoing and PM task choices obligatory on every trial in the sense that although one or both choices may be wrong neither can be omitted. Unlike the traditional dual method, in the obligatory method the two choices are submitted simultaneously rather than sequentially, so PM accuracy and RT can be compared with the replacement method without confounding from differential delays between the stimulus and PM response. Loft and Remington's (2013) pre-emption perspective predicts that PM errors tend to be associated with fast responses. We use Conditional Accuracy Functions (CAFs, Thomas, 1974, see Analysis Methods for more details) to compare the speed of PM errors between the two response methods.

We hypothesise that the obligatory method might improve PM accuracy for two reasons. First, it could act as a type of implicit reminder or cue to make a PM vs. non-PM decision on every trial<sup>2</sup>, with the advantage that, compared to the explicit reminders that have been studied previously, it is integrated with the ongoing task, and thus does not involve an abrupt and attention-capturing onset. It is likely, however, that the obligatory method will detract from ongoing task performance in the sense that ongoing task responding will be slowed. That is, in line with the finding that response time slows in proportion to the logarithm of the number of response options (i.e., Hick's Law: Hick, 1952), responding should be slower for four-alternative

---

<sup>1</sup> Although we take a different approach here, a reviewer pointed out that future work might examine a version of the dual method that instructs participants to make the PM response first.

<sup>2</sup> We thank two reviewers for these suggestions.

obligatory responses than three-alternative replacement responses. To check these possibilities, we examine the impact of response method on ongoing-task RT and accuracy.

A second potential reason for improvement in PM is that obligatory responding promotes integration between the ongoing and PM tasks. Better coordination between PM intentions and ongoing task demands, particularly with respect to overlap between responses, has been found to improve performance (Marsh, Hicks, & Watson, 2002; Rummel, Wesslein, & Meiser, 2017). Also, our PM paradigm shares characteristics with dual-task paradigms, so measures that reduce dual-task costs may be beneficial. Janczyk and Kunde (2020) proposed that dual-task costs reduce when response goals are coordinated, and that practice-related declines in dual-task costs arise from the fusion of initially distinct goals to a single goal. To the degree that obligatory responding promotes task integration, improvements may also be expected due to avoiding task switching, which increases RT's and error rates relative to repeating the same task (e.g., Monsell, 2003; Kiesel et al., 2010). Given the key role of practice in these mechanisms we examined the performance of our participants over two sessions.

## **Method**

### **Participants**

A total of 36 students from the University of Newcastle, Australia, participated for partial course credit. All participants were native speakers of English. The number of participants was informed by previous research which used similar testing paradigms (e.g., Strickland et al., 2020).

## Materials

Experimental stimuli consisted of 924 words and 924 nonwords. Words and frequency counts were obtained from the Sydney Morning Herald word database (Dennis, 1995). Written frequency ranged between 2 and 6 per million, with low frequencies chosen to make the task more difficult. Difficult nonwords were created by substituting vowels of existing English words until no match was found in the word database (e.g., “chaotic” became “chaetic”). Word and nonword lists also excluded words found in the Google database of offensive English words.

Stimulus colour was used as the PM cue. To avoid confusion of PM cues across conditions, we used a unique colour palette of four stimulus colours in each condition (with the palettes randomly selected without replacement from a pool of three palettes in each session). From each palette, one randomly selected colour was used exclusively for PM trials as the PM cue, while non-PM trials used the remaining three colours, randomly selected on each trial.

## Design

The experiment consisted of 1848 trials, of which 828 were non-PM words, 828 were non-PM nonwords, 96 were PM words, and 96 were PM nonwords. Stimuli never repeated. PM cues appeared on 11% of trials.

Participants performed two sessions separated by one to two days. Each session contained four obligatory and four replacement response blocks of 113 trials. Condition order was counter-balanced between participants and across sessions<sup>3</sup>. PM cues occurred on 6 word

---

<sup>3</sup> A reviewer questioned whether, because of the within-subjects manipulation, the obligatory method interfered with performance in the replacement condition. In supplementary materials we show that in a between-subject comparison of response methods in the first half of session 1, both methods showed a similar decrease in PM misses down to ~11%. Replacement PM misses remained at this level for the remainder of the experiment, whereas obligatory PM misses continued to decrease. Hence, the results are not consistent with poorer performance in the replacement responding being due to interference from the obligatory responding.

and 6 nonword trials in each block. The first five trials in each block were non-PM word and nonword filler trials to delay the onset of the first PM trial.

Participants performed a three-minute distractor task (single-digit division math questions) after each set of response instructions and before the task commenced to further ensure a delay before the first block pertaining to these instructions.

### **Procedure**

Participants received spoken and written instructions explaining the lexical decision and PM task. In the replacement condition, participants were instructed to indicate whether the string of letters that appeared on the screen formed a word or a nonword by pressing the 's' or 'd' keys, but on trials where the string of letters was presented in the PM target colour to instead press the 'j' key. In the obligatory response condition, participants were instructed to press one of four keys ('s', 'd', 'j', 'k') assigned to each of the four possible responses: the letter string is a word in the PM target colour; the letter string is a word not in the PM target colour; the letter string is a non-word in the PM target colour; the letter string is not a non-word not in the PM target colour. The mapping of keys to responses was counterbalanced between subjects, with the location of the ongoing task response keys and PM response keys switched between left and right, and the location of word and non-word response keys within ongoing and PM tasks (for the obligatory method) switched between left and right, as a function of participant number. Participants were instructed to respond as quickly and accurately as possible.

Participants performed a practice block of 20 non-PM lexical decision trials at the start of each session with an equal number of word and non-word stimuli.

Each trial started with a 500ms presentation of a fixation cross, followed by a 250ms blank screen. The stimulus letter string was then displayed and remained visible until the

participant pressed a response key. A 500ms interval followed each trial, during which the screen remained blank.

Participants were given self-paced breaks between each block and each condition.

### **Analysis Methods**

We excluded all data from one participant due to a pattern of excessively long RTs (up to 52s). Additionally, we excluded four blocks of trials from three other participants with lexical decision accuracy at or below chance levels (<60%). We also excluded any trial with an RT greater than 5s, and the first two trials of each block as practice trials. In total 4.5% of trials were excluded from analysis. This excluded 3.24% of all PM trials. Raw data are provided at OSF

We conducted mixed effects model analyses using the R programming language (R Core Team, 2020) and the ‘lme4’ package (Bates, Mächler, Bolker, & Walker, 2015). These models included participant as a random intercept term. We report two distinct sets of mixed models in the two subsequent sections. The first set focused on PM performance. Pressing any PM key was scored as correct on PM trials – the one “PM” key for the replacement condition, and either the “PM word” or “PM non-word” key for the obligatory condition. The models included fixed effects for response condition (Replacement, Obligatory), stimulus type (word, non-word), and testing day (one, two). A second set of models focused on lexical decision performance. In the obligatory condition, a correct lexical decision response could be submitted with either the non-PM keys or the PM keys, permitting us to examine lexical decision performance on both PM and non-PM trials. Thus, our second set of models included the same fixed effects to the first set, with one exception: a “trial-type” factor (Replacement non-PM, Obligatory non-PM, Obligatory PM) was substituted for the two-level response condition factor. Replacement non-PM includes all lexical decision trials in the replacement condition with no PM target. Obligatory non-PM

includes all lexical decision trials in the obligatory condition where PM targets are not present. Obligatory PM includes all lexical decision trials in the obligatory condition where PM targets are present.

To analyse accuracy, we fit generalised linear mixed models with a probit link function to binary accuracy data. To analyse RT, we fit linear mixed effects models to participant mean RTs. In the results we focus on using these models to test whether the obligatory response method facilitated or impeded performance in the ongoing lexical decision and PM tasks. Effect sizes are reported in terms of Cohen's *d*. Full details of all analyses are provided in the supplementary materials, with those most relevant to the aims of the paper reported in the results section. All reported effects are significant unless stated otherwise.

We used Conditional-Accuracy Functions (CAFs: Thomas, 1974) to investigate the causes of differences in PM accuracy between response methods. CAFs plot the probability of an error as a function of overall speed by ordering trials on RT, dividing them into a set of equal-sized bins, calculating the error rate for each bin, and plotting it as a function of the of the mean RT in the bin. We constructed CAFs based on PM trials, plotting PM error (i.e., PM miss) rates using 8 bins, with separate functions for replacement and obligatory responding on days one and two. Responding dominated by fast errors has a decreasing CAF and responding dominated by slow errors has an increasing CAF.

## Results

### Prospective Memory Task

The effect of response method on PM accuracy differed between sessions (Figure 1). In the first session, the proportion of PM misses did not significantly differ between response methods, whereas in the second session there were substantially fewer misses with the obligatory

method<sup>4</sup>. However, from an individual-differences perspective, 27% of participants in the first session gained an immediate benefit from using the Obligatory method (Figure 2), where “immediate benefit” was defined as having an initial proportion of PM misses as low as (or lower than) the proportion of PM misses observed in session 2. As a cohort, these participants already had fewer PM misses (4.2%) in session 1 than the group average in session 2 (7%) and substantially fewer PM misses than the other participants (19.4%) in session 1. Furthermore, despite such low PM Miss rates in session 1 using the Obligatory method, this cohort’s PM Miss rates in session 1 using the Replacement method (16%) were not significantly lower than other participants,  $t(14.8) = -0.31, p = 0.76$ . Response method had a more uniform effect on PM-task RTs (Figure 3): mean correct RT was slower for the obligatory method in both session one and session two. RTs in the obligatory condition decreased substantially more across sessions than in the replacement condition.

We also investigated the effect of response method on PM task performance using Conditional Accuracy Functions (CAFs: Thomas, 1974), which show how the probability of an error changes with overall RT. Figure 4 shows that PM misses in the obligatory condition were predominantly fast, whereas in the replacement condition they were predominantly slow, especially in session one. While both response methods led to a modest reduction in the proportion of PM misses at RTs greater than 600ms in session two, there was a much more

---

<sup>4</sup> A reviewer suggested that the observed improvement is due to consolidation of motor skills between sessions (e.g., Walker et al., 2002). However, given that 27% of participants achieved a level of performance in the first session greater than the group performance in the second session without the benefit of motor consolidation (see supplementary materials for details) it is more likely that the group differences between sessions reflect individual differences in multitasking ability and the amount of practice required to benefit from use of Obligatory response method. Hence, although consolidation may have played a role it was not the only factor. Motor-skill learning may also have contributed to the reduction in speed differences between the two response methods in the second session as Hick’s Law slowing is known to reduce with practice (Davis, Moray, & Treisman, 1961; Hale, 1968; Wifall, Hazeltine, & Mordkoff, 2016).

marked reduction in fast PM misses in the obligatory condition between sessions, indicating a reduction in faster errors.

### **Lexical Decision Task**

This section focuses purely on lexical decision performance and is not conditional on PM performance. On word trials, lexical decision accuracy (Figure 5) did not significantly differ between replacement non-PM, obligatory non-PM, and obligatory PM trials. For brevity, we do not distinguish between PM hit and PM miss trials as there was no significant difference in lexical decision performance between the two. For non-word trials, accuracy did not differ between replacement non-PM and obligatory non-PM trials but was lower for obligatory PM trials. Mean correct RTs for obligatory non-PM trials were slower than replacement non-PM trials for both non-words and words (Figure 6). Mean RT was even slower on obligatory PM trials for both non-words and words. As with PM responses, RTs in the obligatory condition decreased more across sessions than in the replacement condition.

### **Discussion**

Obligatory responding was effective in increasing PM accuracy after practice, producing a PM miss rate in the second session that was less than half that of the replacement method. This improvement was achieved even though the overall level of PM misses was low, with the average obligatory PM miss rate reduced to little more than 6% in the second session, much lower than has been reported in any earlier study comparing response methods. In the remainder of the paper, we discuss potential explanations for the reduction in PM misses and directions for future research.

The obligatory response method may provide an implicit reminder that becomes more effective with practice. Guynn et al. (1998) found that reminders are most effective when they present both the PM cue and response, with reminders involving the PM response alone being less effective and reminders of the PM cue alone being ineffective. They proposed that reminders with both cue and response are effective because they strengthen the cue-action association, so that the PM cue is more likely to lead to retrieval of the PM response (see also Vortac et al. 1995, in the context of air traffic control). In this view, the small benefit to PM accuracy from the PM-response-only reminders observed by Guynn et al. (1998) was likely mediated by the implicit rehearsal of the association of the (reminded) PM response to the PM cue. It is possible that, in learning to use a different button pair to signal a PM response under the two different within-subject conditions in the current study, participants rehearsed and hence strengthened the association between the PM response and the PM cue to a greater degree in the obligatory compared to replacement condition, because the PM response was relevant to every trial for the obligatory condition as compared to potentially every trial for replacement, and that increased the effectiveness of the PM response as a retrieval cue (reminder).

Practice with the obligatory method may also be beneficial because it enables participants to better integrate the PM and ongoing tasks task when using the obligatory response method (Janczyk & Kunde, 2020; Marsh et al., 2002; Rummel et al., 2017). This could also have the benefit of reducing the costs of switching between the different tasks (Monsell, 2003). It has been proposed that practice with “bivalent” stimuli with two attributes (each relevant to a different task) reduces task-switch costs by binding together the attributes into a single “compound cue” (Arrington & Logan, 2004; Kahneman, Treisman & Gibbs, 1992; Schumacher et al., 2018). Obligatory responding is likely to encourage the formation of compound cues

because it explicitly requires participants to associate different pairs of attributes (e.g., a particular colour and type of letter string) with each response<sup>5</sup>. Integration of tasks (potentially through compound cues) provides a more plausible mechanism than reminders to account for the observed between-sessions learning effect, as it is unclear why reminders would require such a relatively long timescale (>800 trials) to elicit a reminder effect compared to previous studies (e.g., Gilbert, 2015a). However, an important caveat is that if the learning effect does indeed reflect PM-ongoing task unification, it was observed between sessions rather than within sessions, and we did not predict this in advance as it was not clear *a priori* on what time scale learning would emerge.

A third possibility is that the slowing associated with the obligatory method having more response options is itself efficacious for improving PM accuracy through a speed/accuracy tradeoff. However, strategic slowing by increasing response-caution has not been found to improve PM accuracy with traditional response methods (Anderson et al., 2018; Strickland et al., 2020). Further, it is unlikely that slowing by itself can explain all of the reduction in PM misses, as the large improvement in obligatory PM accuracy between sessions 1 and 2 was accompanied by a substantial speeding in RT. Deeper insight into this issue is provided by considering our finding on the speed of PM misses from the perspective of evidence-accumulation models of choice accuracy and RT, which provide a detailed explanation of speed/accuracy tradeoff in

---

<sup>5</sup> A reviewer noted that integration of tasks bears some resemblance to the concept of “focality”, defined as the degree of overlap in processing requirements between the ongoing and PM tasks, in the sense that obligatory method introduced an overlap between ongoing task responding and PM task responding, and that like performance in our obligatory condition (at least in the second session) PM accuracy is often at ceiling in focal tasks. The PM task in the current study (respond to stimuli color) was non-focal to the ongoing lexical decision task, thus one interesting avenue for future research would be to investigate whether the obligatory response method can also further benefit already focal PM tasks, at least if they can be made sufficiently difficult that there is headroom for improvement.

terms of cognitive and neural processes (see Donkin & Brown, 2018, for a review) and have been applied extensively to PM paradigms (e.g., Boywitt & Rummel, 2012; Horn, Bayen & Smith, 2011; Heathcote et al., 2015; Strickland et al., 2018, 2020).

Evidence accumulation models identify two types of errors, “response-speed” errors and “evidence-quality” errors, which both occur to varying degrees in most decision tasks (Damaso, Williams, & Heathcote, 2020). Evidence-quality errors tend to be slower than correct responses, and they arise because of faulty evidence favoring the wrong response. We found that these slow errors predominated on PM trials in the replacement condition. Evidence-quality errors cannot be avoided by trading speed for accuracy, as taking longer to respond only magnifies the effect of the faulty evidence, consistent with previous findings with the replacement method (Anderson et al., 2018; Strickland et al., 2020). One potential reason for such errors is that the replacement method encourages separate task-sets for the ongoing and PM task, leading to a failure to attend to the PM cue on some trials, and hence low-quality evidence.

In contrast, response-speed errors tend to be faster than correct responses, and they arise because insufficient evidence is collected before a response is made. We found that these fast errors predominated on PM trials in the obligatory condition. Response-speed errors act in the same way as Loft and Remington’s (2013; but also see finding by Ball et al., 2021) pre-emption mechanism and can be ameliorated by trading speed for accuracy. If the advantage for obligatory responding was due to a speed/accuracy tradeoff it would be expected to have fewer fast errors than replacement responding, but we found the exact opposite. However, the predominance of fast PM errors does raise the intriguing possibility that, contrary to previous finding with traditional replacement responding, slowing may be effective in increasing obligatory PM accuracy. Indeed, the gains from this approach might be substantial, particularly in high-stakes

scenarios where any PM error might be disastrous, as we found that the slowest half of obligatory responding (greater than  $\sim 0.75$ s) PM misses consistently occurred at only 2%, four times less than the level of replacement PM misses at any speed. Although this avenue for further improvement is promising, it might not be appropriate in applications where fast ongoing-task responses are required (Loft et al., 2019). The same limitation is applicable to obligatory responding in general, in that both ongoing task RT and PM RT were slowed, although this disadvantage relative to traditional (replacement) response methods does decrease with practice. Further research might examine whether extended practice can sufficiently reduce this disadvantage of the obligatory response method. In any case, we believe that the wide applicability of obligatory responding to human-computer interfaces that can be used to flexibly automate many different courses of action based on simple key presses is likely to make it useful in at least some scenarios. Past research has shown that reminders or contextual cues set to notify individuals when a PM cue is present (or could possibly be present) can decrease and even eliminate ongoing task slowing during other times in the PM retention interval (Bowden, Smith & Loft, 2021; Loft, Smith & Remington, 2013), but as described earlier, reminders that capture visual attention could be distracting.

To better understand the cognitive processes underlying the benefits of obligatory PM responding, it would be desirable to develop an evidence-accumulation model that can directly assess the prevalence and roles of response-speed and evidence-quality errors. Strickland et al.'s (2018) Prospective Memory Decision Control (PMDC) model of the replacement PM paradigm assumes a separate accumulation process (Brown & Heathcote, 2008) for each of the three possible responses. A straightforward extension would involve four accumulators, one for each of the four possible responses in the obligatory method. However, our preliminary explorations

have suggested the need for competitive mechanisms to account for Hick's Law effects. Recent developments have shown this is possible (van Ravenzwaaij et al., 2020) but make the extension of PMDC less straightforward, and so will be left to future work. In summary, the present work has established the potential practical utility of obligatory responding and ruled out simple speed-accuracy tradeoff as the reason why it is effective. However, further research is required to determine the roles of stronger associations between PM cues and PM responses, compound cues that avoid task switch costs, a unified representation of ongoing and PM task goals, or some combination of these and other mechanisms. An extension of Strickland et al's (2018) PMDC evidence accumulation model provides a promising avenue to better understand these mechanisms.

### 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*, 1046-1050.
- Arrington, C. M., & Logan, G. D. (2004a). Episodic and semantic components of the compound-stimulus strategy in the explicit task-cuing procedure. *Memory & Cognition*, *32*, 965–978.
- Ball B. H., Vogel A., Ellis D. M., & Brewer G. A. (2021) Wait a second: Boundary conditions on delayed responding theories of prospective memory. *Journal of Experimental Psychology: Learning, Memory, Cognition*, *47*, 858-877.
- Bates D, Mächler M, Bolker B, Walker S (2015). “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01.
- Bisiacchi, P. S., Schiff, S., Ciccola, A., & Kliegel, M. (2009). The role of dual-response and task-switch in prospective memory: Behavioural data and neural correlates. *Neuropsychologia*, *47*(5), 1362–1373.
- Boehm-Davis, D. A., Durso, F.T., & Lee, J. D. (2015). *APA handbook of human systems integration*. Washington, DC: American Psychological Society.
- Bowden, V. K., Smith, R. E., & Loft, S. (2021). Improving prospective memory with contextual cueing. *Memory & Cognition*, *49*, 692–711.
- Boywitt, C. D., & Rummel, J. (2012). A diffusion model analysis of task interference effects in prospective memory. *Memory & Cognition*, *40*, 70–82.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153-178.

- Chen, Y., Lian, R., Yang, L., Liu, J., & Meng, Y. (2017). Working memory load and reminder effect on event-based prospective memory of high-and low-achieving students in math. *Journal of learning disabilities, 50*(5), 602-608.
- Damaso, K., Williams, P. & Heathcote, A. (2020). Evidence for different types of errors being associated with different types of post-error changes, *Psychonomic Bulletin & Review, 27*, 435–440.
- Davis, R., Moray, N., & Treisman, A. (1961). Imitative responses and the rate of gain of information. *Quarterly Journal of Experimental Psychology, 13*, 78-89.
- Dennis, S. (1995). The Sydney Morning Herald word database. In Noetica: Open forum (Vol. 1). Retrieved from <http://psy.uq.edu.au/CogPsych/Noetica>
- Directions in Psychological Science, 21*, 215–220.
- Dismukes, R. K. (2012). Prospective memory in workplace and everyday situations. *Current*
- Donkin, C., & Brown, S. D. (2018). Response Times and Decision-Making. In E.-J. Wagenmakers (Ed.), *The Stevens Handbook of Experimental Psychology and Cognitive Neuroscience* (4 ed., Vol. 5).
- Einstein, G. O., & McDaniel, M. A. (2005). Prospective memory: Multiple retrieval processes. *Current Directions in Psychological Science, 14*, 286–290.
- Einstein, G.O., McDaniel, M.A., Richardson, S.L., Guynn, M.J., & Cunfer, A.R. (1995). Aging and prospective memory: Examining the influences of self-initiated retrieval processes. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21*, 996–1007.
- Finstad, K., Bink, M., McDaniel, M., & Einstein, G. O. (2006). Breaks and task switches in prospective memory. *Applied Cognitive Psychology, 20*, 705–712.

- Fothergill S, Neal A. (2008). The effect of workload on conflict decision making strategies in air traffic control. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52, 39-43.
- Gilbert, S. J. (2015a, 2015-05-01). Strategic offloading of delayed intentions into the external environment. *Quarterly Journal of Experimental Psychology*, 68(5), 971-992.  
<https://doi.org/10.1080/17470218.2014.972963>
- Gilbert, S. J. (2015b, 2015-05-01). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. *Consciousness and Cognition*, 33, 245-260. <https://doi.org/10.1016/j.concog.2015.01.006>
- Gilbert, S. J. (2015c, May). Strategic use of reminders: Influence of both domain-general and task specific metacognitive confidence, independent of objective memory ability. *Conscious Cogn*, 33, 245-260. <https://doi.org/10.1016/j.concog.2015.01.006>
- Gilbert, S. J., Hadjipavlou, N., & Raelison, M. (2013). Automaticity and control in prospective memory: A computational model. *PloS one*, 8(3), e59852.  
<https://doi.org/10.1371/journal.pone.0059852>
- Guynn, M. J., McDaniel, M. A., & Einstein, G. O. (1998). Prospective memory: when reminders fail. *Memory & Cognition*, 26, 287–298. <http://doi.org/10.3758/bf0320114>.
- Hale, J. K. (1968). Dynamical systems and stability. *Journal of Mathematical Analysis and Applications*, 26, 39-59
- Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological Review*, 122, 376-410.
- Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4, 11–26.

- Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. *Journal of Memory and Language*, *53*, 430–444.
- Horn, S. S., & Bayen, U. J. (2015). Modeling criterion shifts and target checking in prospective memory monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 95–117.
- Horn, S. S., Bayen, U. J., & Smith, R. E. (2011). What can the diffusion model tell us about prospective memory? *Canadian Journal of Experimental Psychology*, *65*, 69–75.  
<https://doi.org/10.1177/0963721412447621>
- Janczyk, M., & Kunde, W. (2020). Dual tasking from a goal perspective. *Psychological Review*, *127*, 1079-1096.
- Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual onset in capturing attention. *Perception & Psychophysics*, *43*, 346–354.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology*, *24*, 175-219.
- Kiesel, A., Steinhauser, M., Wendt, M., Falkenstein, M., Jost, K., Philipp, A.M. & Koch, I. (2010) Control and interference in task switching—A review. *Psychological Bulletin*, *136*, 849-874. <http://dx.doi.org/10.1037/a0019842>
- Kliegel, M., Martin, M., (2003). Prospective memory research: Why is it relevant? *International Journal of Psychology*, *38*, 193–194. <http://dx.doi.org/10.1080/00207590244000205>.
- Loft, S. (2014). Applying psychological science to examine prospective memory in simulated air traffic control. *Current Directions in Psychological Science*, *23*, 326-331.

- Loft, S., & Remington, R. W. (2010). Prospective memory and task interference in a continuous monitoring dynamic display task. *Journal of Experimental Psychology: Applied*, *16*, 145-167.
- Loft, S., & Remington, R. W. (2013). Wait a second: Brief delays in responding reduce focality effects in event-based prospective memory. *Quarterly Journal of Experimental Psychology*, *66*, 1432–1447.
- Loft, S., & Yeo, G. (2007). An investigation into the resource requirements of event-based prospective memory. *Memory and Cognition*, *35*, 263-274.
- Loft, S., Dismukes, R. K., & Grundgeiger, T. (2019). Prospective memory in safety-critical work contexts. In Rummel, Jan (Ed) & McDaniel. Mark (Ed). *Current Issues in Memory: Prospective Memory*. (pp. 170-185). Taylor and Francis.
- Loft, S., Smith, R. E., & Bhaskara, A. (2011). Prospective memory in an air traffic control simulation: external aids that signal when to act. *Journal of Experimental Psychology: Applied*, *17*, 60–70. <http://doi.org/10.1037/a0022845>
- Loft, S., Smith, R. E., & Remington, R. W. (2013). Minimizing the disruptive effects of prospective memory in simulated air traffic control. *Journal of Experimental Psychology: Applied*, *19*, 254–265. <http://doi.org/10.1037/a0034141>
- Marsh, R. L., Hicks, J. L., & Watson, V. (2002). The dynamics of intention retrieval and coordination of action in event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 652–659.
- Monsell, S. (2003). Task Switching. *Trends in Cognitive Sciences*, *7*, 134-140.  
[https://doi.org/10.1016/S1364-6613\(03\)00028-7](https://doi.org/10.1016/S1364-6613(03)00028-7)

- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason, 4*, 61-64.
- Pereira, D. R., Albuquerque, P. B., & Santos, F. H. (2018). Event-based prospective remembering in task switching conditions: Exploring the effects of immediate and postponed responses in cue detection. *Australian Journal of Psychology, 70*, 149-157.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>
- Rummel, J., Wesslein, A.-K., & Meiser, T. (2017). The role of action coordination for prospective memory: Task-interruption demands affect intention realization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*, 717–735.
- Schumacher, E. H., Cookson, S. L., Smith, D. M., Nguyen, T. V., Sultan, Z., Reuben, K. E., & Hazeltine, E. (2018). Dual-task processing with identical stimulus and response sets: Assessing the importance of task representation in dual-task interference. *Frontiers in Psychology, 9*, 1031.
- Strickland, L., Heathcote, A., Remington, R. W., & Loft, S. (2017). Accumulating evidence about what prospective memory costs actually reveal. *Journal of Experimental Psychology: Learning, Memory and Cognition, 43*, 1616-1629.
- 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*, 851-877.
- Strickland, L., Loft, S. & Heathcote, A. (2020). Investigating the effects of ongoing-task bias on prospective memory. *Quarterly Journal of Experimental Psychology, 73*, 1495-1513.
- Thomas, E. A. C. (1974). The selectivity of preparation. *Psychological Review, 81*, 442–464.

van Ravenzwaaij, D., Brown, S. D., Marley, A. J., & Heathcote, A. (2020). Accumulating advantages: A new approach to multialternative forced choice tasks. *Psychological Review*, *127*, 186–215.

Vortac, O. U., Edwards, M. B., & Manning, C. A. (1995). Functions of external cues in prospective memory. *Memory*, *3*, 201–219. <http://doi.org/10.1080/09658219508258966>

Walker, M. P., Brakefield, T., Morgan, A., Hobson, J. A., & Stickgold, R. (2002). Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron*, *35*, 205–211.

Wifall, T., Hazeltine, E., & Mordkoff, J. T. (2016). The roles of stimulus and response uncertainty in forced-choice performance: an amendment to Hick/Hyman Law. *Psychological Research*, *80*, 555-565.

## Figures

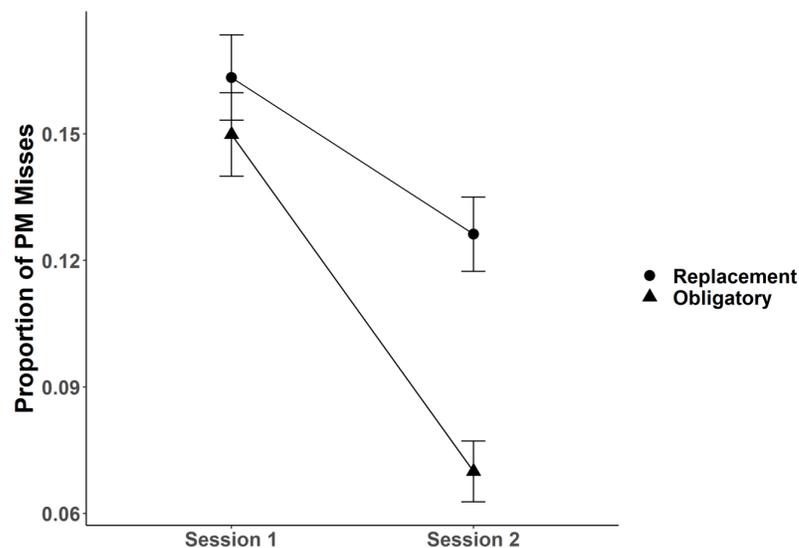


Figure 1. Two-way interaction between response method (Replacement, Obligatory) and Session (1,2) for proportion of PM misses (failure to respond to a PM event). Error bars represent within-subject standard error calculated using the Morey (2008) bias-corrected method.

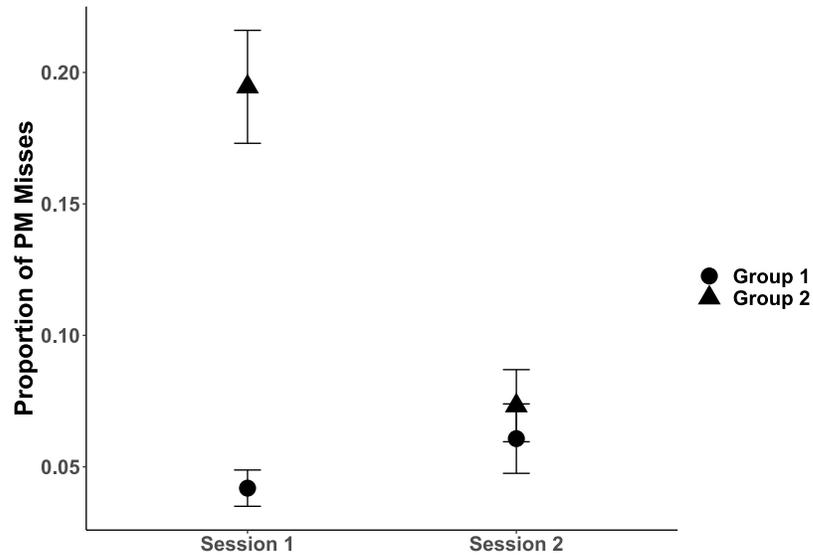


Figure 2. Proportion of Obligatory PM misses for participants who received an immediate benefit from the Obligatory condition (group 1) and the remaining participants (group 2). “Immediate benefit” was defined as an initial proportion of PM misses being at least as low as the proportion of PM misses observed in session 2. Error bars represent within-subject standard error calculated using the Morey (2008) bias-corrected method.

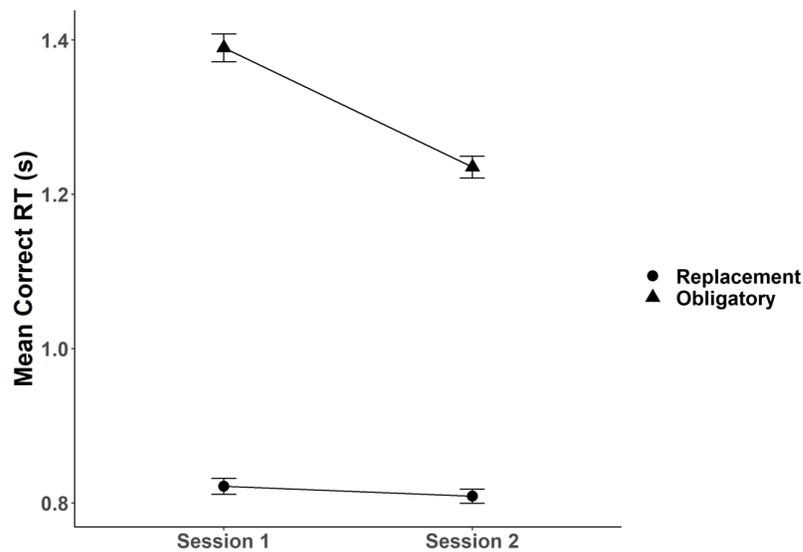


Figure 3. Two-way interaction between response method (Replacement, Obligatory) and Session (1,2) for PM mean correct RT. Error bars represent within-subject standard error calculated using the Morey (2008) bias-corrected method.

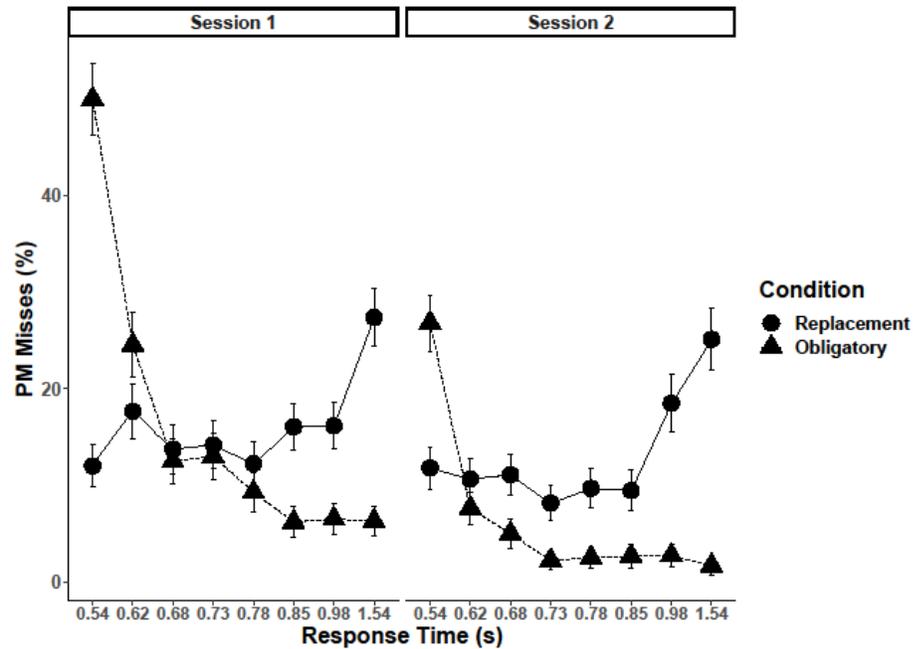


Figure 4. Conditional-accuracy functions for PM misses as a function of RT for each session and response method. Error-bars represent within-subject standard errors, calculated with the Morey (2008) bias-corrected method.

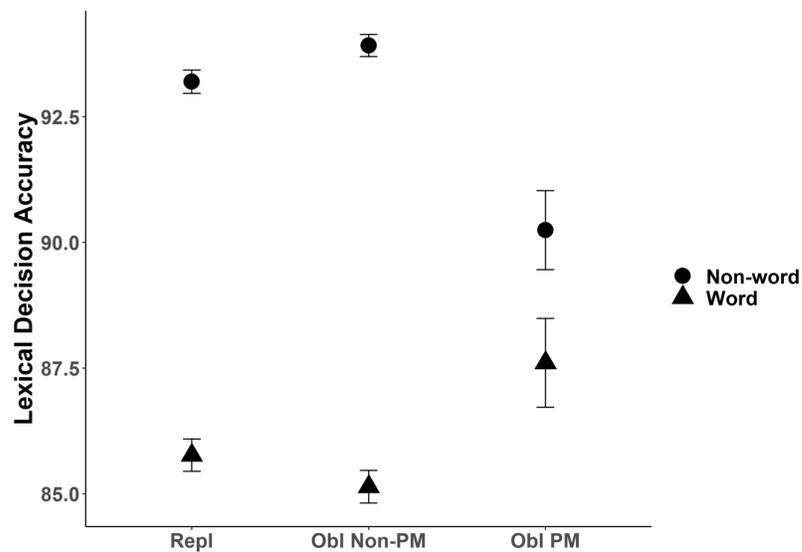


Figure 5. Two-way interaction between Trial Type (Replacement non-PM, Obligatory non-PM, Obligatory PM) and Stimulus (Word, Non-Word) for lexical decision accuracy. Error bars represent within-subject standard error calculated using the Morey (2008) bias-corrected method.

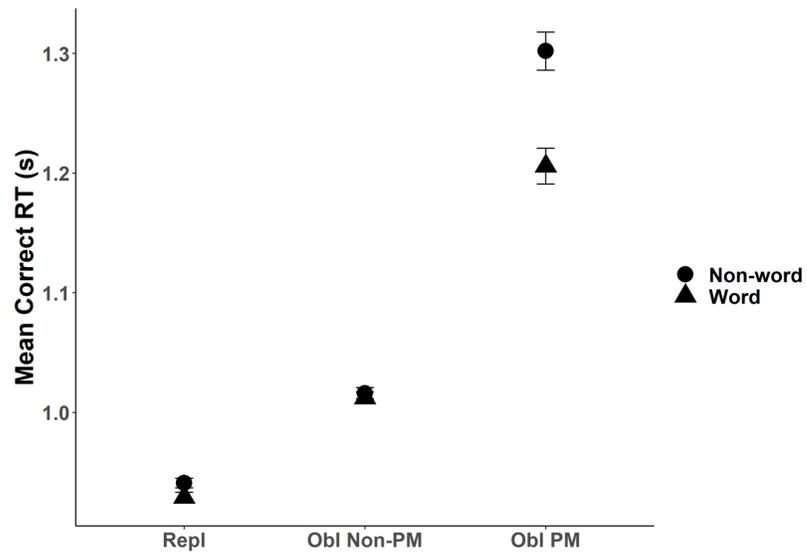


Figure 6. Two-way interaction between Trial Type (Replacement non-PM, Obligatory non-PM, Obligatory PM) and Stimulus (Word, Non-Word) for lexical mean correct RTs. Error bars represent within-subject standard error calculated using the Morey (2008) bias-corrected method.