Journal of Experimental Psychology: Learning, Memory, and Cognition

# Proceeding With Care for Successful Prospective Memory: Do We Delay Ongoing Responding or Actively Monitor for Cues?

Francis T. Anderson, Jan Rummel, and Mark A. McDaniel

Online First Publication, March 19, 2018. http://dx.doi.org/10.1037/xlm0000504

CITATION

Anderson, F. T., Rummel, J., & McDaniel, M. A. (2018, March 19). 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*. Advance online publication. http://dx.doi.org/10.1037/xlm0000504

Journal of Experimental Psychology:
Learning, Memory, and Cognition

© 2018 American Psychological Association
0278-7393/18/$12.00 http://dx.doi.org/10.1037/xlm0000504

# Proceeding With Care for Successful Prospective Memory: Do We Delay Ongoing Responding or Actively Monitor for Cues?

Francis T. Anderson
Washington University in St. Louis

Jan Rummel
Heidelberg University

Mark A. McDaniel
Washington University in St. Louis

In prospective memory (PM) research, costs (slowed responding to the ongoing task when a PM task is present relative to when it is not) have typically been interpreted as implicating an attentionally demanding monitoring process. To inform this interpretation, Heathcote, Loft, and Remington (2015), using an accumulator model, found that PM-related costs were associated with changes in a decision threshold parameter. This pattern was interpreted as disfavoring a monitoring process and supporting a non-capacity-consuming delayed responding strategy. The present study combined both behavioral and modeling techniques, as well as embedded parameter validation, to better illuminate the underlying processes involved in PM. We encouraged participants to use either a delayed responding or a monitoring strategy and used these conditions as anchor points for comparing a standard PM condition (with no strategy instructions). The monitoring strategy benefited PM more than did a delayed responding strategy. Most importantly, behaviors and modeling parameters associated with the standard PM instructions more closely reflected footprints of monitoring. Further, we found no individual model parameter that directly implicates monitoring behavior.

*Keywords:* delay theory, prospective memory, monitoring, costs, accumulator model

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Each day, people go through their lives performing numerous tasks necessary for independent living, such as going to get groceries, picking up their children at school, or paying their bills. Often, we form such intentions but have to delay their completion until a later point, such as forming the intention to pick up your children from school later that day when you drop them off in the morning. When an intention has been set aside for later, it introduces a unique challenge for memory; specifically, one must not only notice the appropriate event (e.g., school lets out at this time) but also retrieve the intended action from memory and coordinate its execution with your present activity (e.g., "Therefore, I need to turn my car around and go pick them up"). This phenomenon is referred to as prospective memory (PM), and differs from retrospective memory primarily in the fact that for most retrospective memory tasks, there is a direct prompt to recall (e.g., coming face-to-face with the person whose name you cannot remember). PM, by contrast, requires one to remember to remember at the appropriate moment, or the opportunity to perform the intended task could slip by unnoticed.

PM has been studied in the laboratory by having participants engage in an ongoing task coupled with an intention to perform a PM task (Einstein & McDaniel, 1990). Usually, participants are given a simple ongoing task, such as a lexical-decision task (LDT), in which they must decide if a string of letters forms a word or does not. Additionally, participants are asked to perform an intended action if a particular event occurs while they are busily engaged in the ongoing task. Participants might be asked, for example, to remember to press the *Q* key if a syllable, such as *tor*, appears within an item presented in the LDT. Using this paradigm, one can make inferences about the mechanisms underlying PM by examining both performance on the PM task and the ongoing task.

## The Monitoring View of Costs

A consistent finding is that participants' ongoing task response times (RTs) often slow down significantly when maintaining a concurrent PM task relative to a condition in which no PM task is present (Einstein & McDaniel, 2010; Jäger & Kliegel, 2008; Smith, 2003; Smith, Hunt, McVay, & McConnell, 2007). This slowing to the ongoing task (labeled *costs*) has typically been assumed to reflect a diversion of cognitive resources away from

Francis T. Anderson, Department of Psychological and Brain Sciences, Washington University in St. Louis; Jan Rummel, Psychological Institute, Heidelberg University; Mark A. McDaniel, Department of Psychological & Brain Sciences, Washington University in St. Louis.

Francis T. Anderson was supported by the National Science Foundation Graduate Research Fellowship Program under Grant DGE-1143954 during this project. Francis T. Anderson and Jan Rummel contributed equally to this work. We would like to thank our research assistant Charlie (Jiaye) Wu for his hard work collecting data, and Gil Einstein for fruitful discussions at the outset of this project. We further thank Veronika Lerche for her support with the RD modeling and Luke Strickland for his support with the LBA modeling. These data were reported in part at the 20th meeting of the European Society for Cognitive Psychology in Potsdam, Germany (September 2017).

Correspondence concerning this article should be addressed to Francis T. Anderson, Department of Psychological and Brain Sciences, Washington University, 1 Brookings Drive, Campus Box 1125, St. Louis, MO 63130. E-mail: anderson.f@wustl.edu

the ongoing task toward performing the PM task. For instance, according to some theories (e.g., the preparatory attention and memory processes theory; Smith, 2003), successful prospective memory can depend on resource-demanding monitoring processes. The idea is that when participants are given a PM task, they must engage monitoring processes in order to detect the PM target and successfully execute the intended action. Further, monitoring processes are assumed to involve limited-capacity attentional processes to search or check the environment for cues. Monitoring thus diverts resources away from the ongoing task, thereby slowing RTs. (On some views, some PM tasks can also be supported by relatively automatic spontaneous retrieval processes; McDaniel & Einstein, 2000, 2007.)

For the present study, we examined costs in a particular kind of PM task: a nonfocal PM task. A nonfocal PM task contains targets or cues that are not directly extracted in the service of performing the ongoing activity (e.g., detecting a particular syllable when making lexical decisions, such as the PM task to press the *Q* key when the syllable *tor* is present), and these nonfocal tasks are consistently associated with costs to the ongoing activity. Because the focus of the present research is determining the theoretical implications of costs in PM and because the critical cost findings have been reported with nonfocal PM tasks, we confine the scope of the present research to nonfocal PM tasks.

Most existing research supports the view that nonfocal PM tasks require the use of strategic monitoring processes and that costs index those monitoring processes. For example, research has shown that nonfocal PM performance is higher when participants exhibit greater costs to the ongoing task, which suggests that diverting more resources away from the ongoing task (and presumably toward the PM task) improves PM (Smith, 2003; Smith & Bayen, 2004). Further, people who have a higher working memory capacity generally perform better on nonfocal tasks but have a similar level of costs as people with lower working memory (Brewer, Knight, Marsh, & Unsworth, 2010; Smith, Persyn, & Butler, 2011). Neuroscientific evidence also implicates monitoring as an attentionally demanding process: Nonfocal tasks result in sustained anterior prefrontal cortex activation (Burgess, Quayle, & Frith, 2001; McDaniel, LaMontagne, Beck, Scullin, & Braver, 2013; Reynolds, West, & Braver, 2009; Simons, Schölvinck, Gilbert, Frith, & Burgess, 2006). Taken together, these findings and others implicate monitoring as a cognitively demanding process that draws on a pool of limited attentional resources or mechanisms, resulting in costs to the ongoing task. Accordingly, the consensus among researchers has been that costs to the ongoing task (when a PM task is present) are a footprint of monitoring (Einstein & McDaniel, 2010; Einstein et al., 2005; Guynn, 2003; Marsh & Hicks, 1998; Smith, 2003; Smith et al., 2007).

## Delay Theory and Costs

A recent theory proposed by Loft and Remington (2013) challenges the predominant interpretation that costs reflect monitoring processes. They proposed instead the substantially different view that costs reflect the strategic relaxation of one's decision thresholds on the ongoing task. Instead of the PM task competing for resources with the ongoing task, participants make a strategic decision to delay responding to allow more time for information related to the PM task to accumulate. In other words, participants slow down and become more careful in their responding to the ongoing task, making it more possible for PM information to accrue and thus increasing the likelihood of noticing the target.

The delay theory therefore rests on the assumption that information relating to the PM task (e.g., information as to whether or not the stimulus contains the syllable *tor*, as in *tortoise*) accumulates in parallel but at a different rate than the ongoing task (e.g., the semantic information in a LDT). Namely, when the PM task is nonfocal, PM information accrues more slowly than semantic information. On this view, if participants did not delay their responding, they would likely miss the PM target. If, however, participants were to delay their responding to the LDT by requiring more information to make a decision (thereby resulting in costs), they would, theoretically, allow more time for relevant PM information to accumulate and should thus be likely to successfully perform the intended action.

This account is primarily supported by statistical evidence from accumulator models that were applied to data from a standard laboratory PM task (Heathcote, Loft, & Remington, 2015). Accumulator models are a group of mathematical models that can generally be applied to RT and accuracy data from binary decision tasks. In recent years, this group of models has been used to investigate the cognitive processes underlying performance in a variety of cognitive tasks (see Voss, Nagler, & Lerche, 2013, for an overview). The advantage of applying these models to decision data is that they simultaneously use information from both RT and accuracy data for the estimation of the model parameters. In doing so, accumulator models provide three key parameters that can often be interpreted in a psychologically meaningful way. The first key parameter, the drift-rate parameter ($v$), is an indicator of the speed of information processing. The higher the value of this parameter, the more efficiently the information is integrated. Consequently, a high drift rate implies fast RTs alongside high accuracy. In line with this assumption, drift rates have been shown to be positively correlated with working memory capacity (Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007). The second key parameter, the threshold parameter ($a$), indicates how much information is gathered before a decision is made. Higher values of this parameter imply a more conservative approach to the task. Therefore, slower RTs that come with higher accuracy reflect a higher (i.e., more conservative) threshold parameter. The third key parameter is the so called nondecision time parameter ($t_0$). This parameter captures processes that precede and succeed the information accumulation process. These could include motor-execution processes (Voss, Rothermund, & Voss, 2004) and also mental processes that are constantly present on every trial of a task (e.g., better vocabulary knowledge results in smaller $t_0$ estimates in an LDT; Yap, Balota, Sibley, & Ratcliff, 2012).

As an empirical test of delay theory, Heathcote et al. (2015) applied two different accumulator models—the Ratcliff diffusion (RD) model and their own linear ballistic accumulator (LBA) model—to data from a nonfocal PM task. By reanalyzing ongoing-task conditions from existing studies (coupled with converging evidence from their own experiments), Heathcote et al. were able to investigate how the addition of a PM task affects model parameters. Interpreting costs as reflecting monitoring processes presumes that limited cognitive resources are diverted away from the LDT and toward the PM task. Therefore, Heathcote et al. argued that one should expect to see a reduction in the drift-rate parameter (i.e., hampered accumulation of semantic information) if partici-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

pants were using a monitoring strategy (compared with a condition with no PM task). Instead, Heathcote et al. observed an increase in the decision threshold parameter when a PM task was present relative to when it was not, rather than a reduction in the drift rate (see also Strickland, Heathcote, Remington, & Loft, 2017). Based on these findings, Heathcote et al. argue that participants in a typical PM task use a delay strategy to detect the PM target (e.g., *tor*), and that the PM-induced costs observed in such PM tasks are a manifestation of this strategy. Specifically, the theoretical premise is that participants institute a more stringent criterion for their ongoing-task responding to allow more time for PM information to accrue, and therefore are successful in detecting the PM target.

Heathcote et al.'s (2015) interpretation of the modeling results relies on the assumption that the two candidate processes underlying the PM cost effect (i.e., monitoring and response delaying) can be mapped onto the drift-rate and decision threshold parameters in a one-to-one fashion. This assumption has not been uniformly embraced, however. For example, Horn and Bayen (2015) argued that monitoring processes could manifest themselves in a strategy of checking for PM targets after the information accumulation processes for the lexical decision are completed (a similar argument could be that participants check for the target *before* proceeding with the lexical decision in a serial fashion). In this case, one would expect monitoring processes to be reflected by changes in the nondecision parameter, which covers the processes subsequent to information accumulation. Because nondecision time incorporates many different processes, all unrelated to the actual decision itself, it is plausible that PM-related processes such as monitoring could be accounted for by this parameter. However, it is unlikely that this parameter, which also covers predecision perceptual and postdecision motor processes, is a process-pure reflection of monitoring. Interestingly, Heathcote et al. consistently found higher estimates of the nondecision parameter in the presence of a PM task; however, they did not consider these nondecision parameter changes an indicator of monitoring (see also Strickland et al., 2017).

In sum, whereas most behavioral evidence has been interpreted as evidence favoring a monitoring view of costs (cf. Loft & Remington, 2013), some current modeling evidence is interpreted as supportive of a delay theory account of costs and as disfavorable to a monitoring interpretation (again, note the reasoning rests on the claim that monitoring is expected to be exclusively reflected by changes in the drift-rate parameter; Heathcote et al., 2015). If, however, monitoring would be funneled, among other processes, into the nondecision parameter (Horn & Bayen, 2015), modeling evidence would then indicate that people engage in a combination of monitoring and response delaying when performing PM tasks. Accordingly, in light of formal modeling outcomes, the interpretation of PM-induced costs to the ongoing task has become equivocal. Both the monitoring and delayed responding views accommodate the existence of costs when maintaining an embedded PM task, but the processes assumed to underlie costs are substantially different in each view. Most problematic, the use of costs as a direct footprint of monitoring is presently undermined, and the corresponding interpretations made by many PM researchers operating under this assumption bear reexamination.

## The Present Paradigm

One key limitation of the existing modeling of PM-induced costs is that very little work has been conducted toward parameter validation within the PM paradigm. That is, little evidence exists that directly links a delay response strategy to changes in the decision threshold parameter or monitoring behavior to changes in the drift rate. One study has shown that participants adjust their decision threshold to expectancies regarding the likelihood of a PM target occurring, suggesting that the threshold parameter reflects a general preparedness to respond to PM targets, and that this lies under metacognitive control (Boywitt & Rummel, 2012, Experiment 1). Similarly, there is also very little evidence supporting the assumption that monitoring processes result in an isolated reduction of the drift-rate parameter. In fact, previous research employing standard PM manipulations that should affect monitoring—like increasing the PM target frequency, stressing the importance of the PM task, and making the PM targets especially salient or focal to the ongoing task—did not result in consistent effects on the drift-rate parameter, but rather on the nondecision parameter (Horn & Bayen, 2015). From these mixed results, it becomes evident that a more direct validation of the mapping of model parameters on PM strategies is required.

Therefore, a central objective of the present study was to directly inform how the model parameters map onto putative PM strategies. To do so, we used an instructional manipulation that stimulated participants to adopt a delay or a monitoring strategy. We then observed the behavioral results on the PM task and LDT performances, as well as the candidate parameters of the latent PM processes (i.e., decision threshold, drift rate, and nondecision parameters) that may characterize each strategy. In the first half of the experiment, all conditions completed a *baseline* block of trials containing only lexical decisions. In the second half of the experiment, or *active* block, we manipulated instructions designed to push participants toward engaging in a particular strategy. Then, any changes observed between the baseline and active blocks are presumed to be a result of our instructional manipulation.

In one condition, the *boundary PM* condition, we encouraged participants to be more stringent in their decision boundaries, or thresholds, by instructing them to slow down and focus on achieving perfect accuracy on the LDT. Additionally, they were asked to perform an embedded PM task. As a manipulation check, we included a *boundary control* condition to make sure that the parameters associated with a change in decision thresholds did not differ depending on whether or not participants additionally had a PM task. In this condition, participants received the same accuracy instructions but were not given a PM task. If our instructional manipulation is effective in these two boundary conditions, then we should see a corresponding speed–accuracy trade-off (slower RTs and increases in accuracy) from the baseline block to the active block. Further, if the model parameters are sensitive to these behavioral changes, both conditions should display significantly increased decision thresholds. Delay theory additionally predicts that there should be no change in the drift rate for either condition.

In another condition, the *monitoring* condition, we instructed participants to check every trial for the PM target (in the active block). In this condition, parameter estimates should change for the drift rate from the baseline block to the active block to the extent to which monitoring processes are engaged for the PM task.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

That is, a decrease in drift rate may reflect the limited attention available (because of the demands of monitoring) for extracting the semantic information relevant to the LDT. Critically, the use of such a monitoring strategy would not result in significantly better LDT accuracy. That is, in sharp contrast to the boundary PM condition, there should be no increase in LDT accuracy from the baseline block to the active block. Alternatively, one could (additionally) expect monitoring processes that people engage in after their lexical decision to be funneled into the nondecision parameter (Horn & Bayen, 2015). Finally, evidence that the boundary PM and monitoring instructional conditions produce different PM processes would be reflected in differences in PM accuracy. Specifically, obtaining differences in PM performance would suggest that delayed responding and monitoring are qualitatively different processes used in PM.

These instructional conditions, boundary PM and monitoring, were then used as anchor points against which to compare a final instructional condition. In this condition, the *standard* condition, participants were given typical PM instructions for the active block. These instructions did not mention delaying responding or monitoring. This condition resembles the standard nonfocal PM instructions reported in the majority of PM studies and was nearly identical to the one from Heathcote et al. (2015, Experiment 2). By observing which of the two anchor points (boundary PM, monitoring) the standard condition more closely approximates, we can draw inferences about the processes typically engaged when participants perform a nonfocal PM task. Specifically, as seen in Table 1, if the standard condition reflected the parameter changes and behavioral markers associated with the boundary PM condition (increase in decision thresholds and corresponding increase in LDT accuracy), then this would provide strong support for delay theory. If, however, the standard condition more closely resembled the monitoring condition (reduction in drift rate or increase in nondecision time and no change in LDT accuracy), then the predominant interpretation of costs—a diversion of attentional resources from the LDT to the PM task—would be supported.

Finally, to provide additional support for the assumptions that monitoring requires increased attentional demands (i.e., the monitoring condition) and that a delay strategy does not require such an increase (i.e., the boundary PM condition), we computed correlations between working memory capacity and PM accuracy for each condition (a similar approach to correlating costs with PM; Smith, 2003). Further, we reasoned that the correlation for the standard group would provide additional evidence regarding the PM processes people typically engage. If working memory is correlated with PM accuracy in this group, it would provide further evidence for a limited-attentional-capacity account of PM.

# Method

## Participants and Design

After obtaining approval for the study by the Washington University Institutional Review Board, 104 young adult participants ($M$ age = 19.51, $SD$ = 1.10) received course credit for participating. Eight participants were excluded either because of an inability to follow instructions or computer failure, leaving 96 participants in total. The design was a 2 × 4 mixed factorial that examined the within-subjects variables of trial block (baseline, active) and the between-subjects variable of instructional condition (boundary control, boundary PM, standard, monitoring). Each of the four conditions contained 24 participants.

## Procedure

**Baseline block.** Participants came to the lab knowing that they would be asked to engage in a series of computer-based tasks measuring working memory, PM, and their ability to identify words. They first received instructions for performing a LDT, in which they were told to press a key labeled "yes" if the string of letters presented on the screen formed a word, and to press "no" if the string of letters did not form a word. After practicing the LDT, participants completed a computerized backward digit span task, selecting the presented numbers in reverse order and submitting their response using the mouse. Number strings ranged from four to nine items. They were provided accuracy feedback on the backward digit span after each of the 10 trials.

Following the backward digit span, participants moved into the baseline block of the LDT. The procedure here followed closely from that used in Heathcote et al. (2015, Experiment 2). Partici-

Table 1
*Predictions Made by Both the Delay and Monitoring Theories Regarding How the Boundary PM and Monitoring Conditions Should Align With the Standard Condition*

| Theory | DV | Predictions |
|---|---|---|
| Delay | Threshold | Boundary PM and standard conditions display increases in threshold |
| | Drift rate | Neither the boundary PM nor the standard conditions display reductions in drift rate |
| | Nondecision | Less straightforward[a] |
| | Accuracy | Boundary PM and standard conditions display increases in accuracy |
| Monitoring | Threshold | Monitoring and standard conditions display equivalent threshold values |
| | Drift rate | Both the monitoring and standard conditions display reductions in drift rate |
| | Nondecision | Less straightforward[a] |
| | Accuracy | Neither the monitoring nor the standard conditions display increases in accuracy |

*Note.* DV = dependent variable; PM = prospective memory.
[a] There is some evidence to suggest that an increase in nondecision time could reflect an increase in monitoring behavior (via a pre- or post-decision target check), but these predictions are not specifically outlined by Heathcote, Loft, and Remington's (2015) delay theory.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

pants made word/nonword judgments for 220 randomized trials, which consisted of 55 high-frequency words, 55 low-frequency words, and 110 nonwords. The stimuli were preceded by a 500-ms fixation cross and were presented in 24-point Arial Narrow white font on a black background. If participants made an incorrect decision, then the word "incorrect" was presented in blue for 1,000 ms before continuing to the next trial.

**Active block.** The active block was nearly identical to the baseline block, containing 220 trials with high- and low-frequency words as well as nonwords. However, the traditional LDT instructions were manipulated according to each participant's experimental condition. In the boundary control condition, participants were encouraged to focus on accuracy rather than speed. Specifically, they were told,

> During this block of trials, we are interested in seeing how accurate your responding to the lexical decision task can be. Toward that end, it is critical that you take your time and make sure you are correct in your answer before responding—you should try for 100% accuracy on this task.

In the boundary PM condition, participants were also encouraged to focus on their accuracy in the LDT but were additionally told to complete a concurrent PM task. For the PM task, participants were asked to press the *Q* key if they saw a word or nonword containing the syllable *tor* (or those three letters adjacent to one another anywhere). Three PM targets were high-frequency words (*history*, *story*, *senator*), three were low-frequency words (*tutorial*, *ancestor*, *tortoise*), and three were nonwords (*torpun*, *cintor*, *extoruar*).

In the boundary PM condition, participants were told,

> During the next half of the task, we are interested in seeing how accurate your responding to the lexical decision task can be. Toward that end, it is critical that you take your time and make sure you are correct in your answer before responding—you should try for 100% accuracy on this task. We would also like you to perform a prospective memory task. Specifically, if, during the lexical decision task, you see a string of letters containing the syllable *tor* we would like you to press the *Q* key instead of yes or no. For example, you would press the *Q* key to the word torrent, the nonword toruntt, or any other string of letters containing the syllable *tor*. Previous research has shown that the syllable, when it occurs, will "pop" into mind—so do not spend any time looking for it. Your focus should be on performing the LDT accurately, and it is important that you take your time and that you are careful in your judgments.

In the standard condition, participants were simply told to perform the LDT and the PM task without any additional instructions regarding how they should approach these challenges, as occurs in most PM experiments. They were told,

> We would also like you to perform a prospective memory task. Specifically, if, during the lexical decision task, you see a string of letters containing the syllable *tor* we would like you to press the *Q* key instead of yes or no. For example, you would press the *Q* key to the word torrent, the nonword torunt, or any other string of letters containing the syllable *tor*.

In the monitoring condition, participants were given instructions to perform the LDT and the PM task but were additionally encouraged to monitor for PM cues throughout this phase. They were told,

> In order to perform this prospective memory task, we would like you to make sure you **check every trial** to see if it contains the syllable *tor*. It is important that you check every trial or you might miss one of the targets.

Participants in all conditions had to correctly repeat the instructions to the experimenter before continuing.

After receiving instructions for the active block, participants again performed 10 trials of the computerized backward digit span before proceeding with the active block of the LDT. As mentioned, the active block contained high- and low-frequency words, as well as nonwords, but three items from each category were replaced with the corresponding PM targets. Targets were randomly selected and occurred on Trials 12, 39, 56, 85, 123, 130, 159, 190, and 209. Following the active block, participants completed one block of trials in a shortened version of the operation span task (Foster et al., 2015). The entire experiment lasted approximately 45 min.

The block order was not counterbalanced; therefore, all participants completed the baseline block first and then moved onward to the active block. We believe that a constant block order is most suitable for the present study for two reasons. First, our study was designed to closely follow Experiment 2 in Heathcote et al. (2015), and in that experiment, they found minimal effects of block order. Second, as shown in several PM studies (Smith, 2010; Smith et al., 2007), carryover effects from an active to a baseline block when counterbalancing in within-subjects designs are a concern, and especially so in the present study in which we instructed different groups to approach the task differently. Given that our critical analyses target the interaction between baseline and active blocks, we elected not to counterbalance the order of the baseline and the active blocks so as not to bias participants' approach to the baseline block (e.g., receiving boundary instructions in the first block could cause participants to adopt a similar strategy in the baseline block when it came second).

## Materials

Stimuli were selected using the Balota et al. (2007) norms, and divided randomly into two lists (A, B) of 220 items. All items were between four and eight characters in length ($M = 6.45$). High-frequency words ($n = 110$) had an average Log_Freq_HAL (frequency norms based on the Hyperspace Analogue to Language corpus) of 9.44 and a mean accuracy of .96. Low-frequency words ($n = 110$) had an average Log_Freq_HAL of 4.96 and a mean accuracy of .76. Nonwords ($n = 220$) had a mean accuracy of .92. PM targets were borrowed from Heathcote et al. (2015).

## Modeling Method and Selection

We fitted the full RD model with seven parameters to the ongoing task RT (trimmed 2.5 *SD*s above the mean and under 300 ms) and accuracy data using the *fast-dm-30* software (Voss, Voss, & Lerche, 2015). We focus on the RD model because it is currently the most widely used accumulator model in experimental psychology (cf. Voss et al., 2013) and has been applied to PM data numerous times (Boywitt & Rummel, 2012; Heathcote et al., 2015; Horn & Bayen,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

2015; Horn, Bayen, & Smith, 2011, 2013; Rummel, Kuhlmann, & Touron, 2013). We associated the upper threshold with correct responses and the lower threshold with incorrect responses (Heathcote et al., 2015; Horn & Bayen, 2015). In order to identify the model that best fitted our data while being as parsimonious as possible, we estimated model versions with different restrictions using a maximum likelihood optimization criterion that has been shown to provide the most reliable parameter estimates with trial numbers comparable with those of the present study (Lerche & Voss, 2017). We relied on the Bayesian information criterion (BIC) to identify the best-fitting model. For model selection, we started with a very simple baseline model and then increased model complexity step-by-step (cf. Donkin, Brown, Heathcote, & Wagenmakers, 2011).

For the baseline model, we imposed several model restrictions. We set the variability parameters of the drift-rate parameter and the starting point to zero because recent simulation studies showed that, unless trial numbers are very high, these restrictions allow for a better recovery of the other parameters and more reliable parameter estimates even when there is variability in the data (Lerche & Voss, 2016b; van Ravenzwaaij, Donkin, & Vandekerckhove, 2017). Furthermore, we set the starting point $zr$ to $a/2$ because no decision bias is expected when responses are coded as correct versus incorrect (Voss et al., 2015). As we were primarily interested in whether our manipulations differently affected the changes in mean parameter estimates from the baseline to the active task block, we started by allowing the key parameters (i.e., drift rate [$v$]), threshold [$a$], and nondecision [$t_0$]) to vary between blocks one after the other. Then, we allowed the parameters to additionally vary with stimulus type (high- and low-frequency words, and nonwords). Model-fit indices of the baseline and the more complex models are presented in Table 2. As evident from Table 2, the model with all key parameters varying between blocks and the $v$ parameter additionally varying with stimulus type resulted in the best model fit. Allowing the $a$ or $t_0$ parameters to additionally vary with stimulus type resulted in only minor fit changes. Because our primary interest was for block effects, we thus decided to focus on the simpler model. Heathcote et al. (2015) additionally allowed the starting point $zr$ to vary with block, but in our case, such a relaxation would have resulted in slightly worse model fits (see Models RD 7 and RD 8 in Table 2), and therefore we decided to use the model with fixed $zr$. Graphical model fit inspection indicated an acceptable model fit for most participants (see Figure 1). Excluding one participant for whom the graphs indicated a misfit would not have changed the present pattern of results.

As already noted, the RD model is probably the most prominent information accumulator model in cognitive psychology. One feature of this model is that it assumes that a single accumulation processes underlies both responses in binary choice tasks. Unlike the RD model, an alternative LBA model assumes independent accumulation processes for the two responses in a binary choice task. Simulation studies have shown that both models usually produce similar outcomes, at least in terms of theoretical implications regarding the cognitive underpinnings in a given task (Donkin et al., 2011). Furthermore, both RD and LBA produce comparable parameter estimates when the starting point is set to a/2, as in our best-fitting model (Donkin et al., 2011). However, when applying both models to the same PM data sets, Heathcote et al. (2015) found that the two models led to somewhat different conclusions. We therefore also report LBA modeling results in Appendix A for the interested reader.[1] Indeed, for the present data set, the best-fitting LBA model differed from the best-fitting RD model, and the manipulation effects were recovered by different parameters of the two models. In the Results section we focus on the RD results, but we consider those differences in the Discussion section.

## Results

For all dependent measures (other than PM accuracy), our primary interest was determining the effect of each condition's instructions on the ongoing LDT. Accordingly, our analyses target between-subjects effects of condition (boundary control, boundary PM, monitoring, standard) for the within-subjects change of task block (baseline to active). Main effects of condition, therefore, are not directly informative; consequently, we present the theoretically motivated comparisons outlined in the introduction targeting the change from baseline block to active block between specific conditions. Thus, the analyses of variance (ANOVA) models contain the within-subjects variable of block and the between-subjects variable of instructional group, with the particular groups included reflecting the particular a priori comparisons of interest. It should be noted that all omnibus tests (i.e., with all conditions analyzed concurrently) of main effects and interactions reported here were significant (see right column of Table 3). In addition, although we allowed drift rate to vary by stimulus type (high- and low-frequency words, and nonwords), this variable never resulted in a theoretically informative three-way interaction with task block and condition (all $F$s $< 1$), so nothing further will be reported regarding trial type. Finally, there were no significant differences in baseline performance for any behavioral results or parameter estimates among the four conditions. Descriptive statistics for RTs, accuracy rates, model parameters and results of the overall omnibus tests are reported in Table 3, and the difference scores for these dependent variables are in Table 4.

### Decision Threshold Manipulation Check

To determine whether our accuracy instructions produced similar effects in the boundary PM and boundary control conditions, we examined differences between the baseline and active blocks for both ongoing task accuracy, ongoing task RTs, and $a$-parameter estimates. A significant main effect showed increases in accuracy from the baseline to the active block for both boundary conditions, $F(1, 46) = 82.30$, $p < .001$, mean squared error ($MSE$) $= .001$, $\eta_p^2 = .64$ ($M$ increase $= .04$), and there were no significant differences between the two boundary conditions, $F < 1$. An ANOVA of ongoing task RTs (trimmed 2.5 *SD*s above the mean and under 300 ms) indicated that RTs slowed for the active relative to the baseline block, $F(1, 46) = 102.40$, $p < .001$, $\eta_p^2 = .69$. This effect interacted with condition, such that the boundary PM condition slowed significantly more than the boundary control condition, $F(1, 46) = 6.06$, $p < .05$, $MSE = 12774.07$, $\eta_p^2 = .12$ (boundary PM, $M$ slowing $= 290$ ms; boundary control, $M$ slowing $= 177$ ms).

[1] We are grateful to Luke Strickland for applying the LBA system to our data and providing us with the LBA parameter estimates.

Table 2
*Ratcliff Diffusion Models and Corresponding Bayesian Information Criteria*

| | Model specifications | | | | |
|---|---|---|---|---|---|
| Model | Varying with block | Varying with stimulus type | Restrictions | *K* | Σ (BIC) |
| Baseline | — | — | $zr = .5$; $szr = 0$; $sv = 0$ | 4 | 1,951 |
| RD 1 | $v$ | — | $zr = .5$; $szr = 0$; $sv = 0$ | 5 | 1,920 |
| RD 2 | $v$; $a$ | — | $zr = .5$; $szr = 0$; $sv = 0$ | 6 | 1,856 |
| RD 3 | $v$; $a$; $t0$ | — | $zr = .5$; $szr = 0$; $sv = 0$ | 7 | 1,851 |
| **RD 4** | $\boldsymbol{v; a; t0}$ | $\boldsymbol{v}$ | $\boldsymbol{zr = .5; szr = 0; sv = 0}$ | **11** | **1,809** |
| RD 5 | $v$; $a$; $t0$ | $v$; $a$ | $zr = .5$; $szr = 0$; $sv = 0$ | 15 | 1,814 |
| RD 6 | $v$; $a$; $t0$ | $v$; $t0$ | $zr = .5$; $szr = 0$; $sv = 0$ | 15 | 1,808 |
| RD 7 | $v$; $a$; $t0$ | $v$ | $szr = 0$; $sv = 0$ | 12 | 1,814 |
| RD 8 | $v$; $a$; $t0$; $zr$ | $v$ | $szr = 0$; $sv = 0$ | 13 | 1,812 |

*Note.* The *s*-parameters (*sv*, *szr*, and *st0*) are the intertrial variability parameters for the corresponding parameters. The best-fitting model is noted in bold (we favored Model 4 because it was simpler than Model 6 and had a comparable BIC value). Parameters of the seven-parameter model are $v$ = drift-rate; $a$ = threshold; $t0$ = nondecision component; and $zr$ = starting point; $K$ = number of model parameters per subject; BIC = Bayesian information criterion (with smaller values suggesting a better fit); RD = Ratcliff diffusion model version.

For *a*-parameter estimates, we obtained a significant increase in decision thresholds, $F(1, 46) = 74.23$, $p < .001$, $MSE = .10$, $\eta_p^2 = .62$, without any differences between conditions, $F(1, 46) = 2.06$, $p > .05$. These parameter changes, coupled with the corresponding increase in ongoing task accuracy, indicate that our delay instruction, as expected, resulted in more cautious responding in the two boundary conditions. Although the boundary conditions converged in terms of threshold change and ongoing task accuracy, the boundary PM condition produced significantly greater costs, and we sought to determine which parameter estimates captured this increased slowing. Greater slowing in the boundary PM condition was associated with marginally reduced drift rate for the interaction, $F(1, 46) = 3.14$, $p = .08$, $MSE = 2.71$, $\eta_p^2 = .06$, and significantly increased nondecision time, $F(1, 46) = 6.57$, $p < .05$, $MSE = .003$, $\eta_p^2 = .13$, relative to the boundary control condition.

To take stock, our delay instruction resulted in the expected speed–accuracy trade-off and a parallel increase in decision thresholds for the active blocks (relative to the baseline blocks) of the boundary control and boundary PM conditions. The additional costs observed when a PM task was added to the delay instruction (i.e., the boundary PM condition) relative to a delay-only condition (boundary control), in conjunction with the other parameter estimates (marginally reduced drift rate and significantly increased nondecision time), suggests that the boundary PM condition did not achieve a pure delay-response strategy (cf. Heathcote et al., 2015). Nevertheless, the key footprints of increased cost and *a*-parameter estimates from that condition can be considered an initial empirical benchmark of a PM delay strategy against which to compare the patterns of the other PM conditions.

## Comparison of Boundary-PM and Monitoring Conditions

We compared the boundary PM condition to the monitoring condition to reveal whether a monitoring strategy in fact produces different patterns relative to a delay strategy (boundary PM condition). There was a significant increase from baseline to active blocks in ongoing-task accuracy, $F(1, 46) = 24.57$, $p < .001$, $MSE = .001$, $\eta_p^2 = .35$. Importantly, this main effect was qualified by a significant interaction, $F(1, 46) = 17.97$, $p < .001$, $MSE = .001$, $\eta_p^2 = .28$, such that the increase was limited to the boundary PM condition (boundary PM, *M* increase = .04; monitoring, *M* increase = 0). Ongoing-task RTs also increased from baseline to active blocks, $F(1, 46) = 121.94$, $p < .001$, $MSE = 10942.07$, $\eta_p^2 = .73$. This block main effect again interacted with condition, $F(1, 46) = 6.51$, $p < .05$, $MSE = 10942.07$, $\eta_p^2 = .12$, such that the boundary PM condition displayed greater slowing (290 ms) than the monitoring condition (181 ms). Regarding PM accuracy, the boundary PM condition ($M = .55$, standard error ($SE$) = .06) was significantly worse in remembering to perform the PM task than the monitoring condition ($M = .73$, $SE = .05$), $t(46) = 2.25$, $p < .05$.

In terms of parameter estimates, we compared the boundary PM and monitoring conditions for $a$, $v$, and $t_o$ parameters. For *a*-parameter estimates, there was a significant main effect of block, $F(1, 46) = 81.18$, $p < .001$, $MSE = .06$, $\eta_p^2 = .64$, with both conditions showing an increase in decision thresholds. There was also an interaction, $F(1, 46) = 13.75$, $p < .05$, $MSE = .06$, $\eta_p^2 = .23$, indicating that the boundary PM condition had a greater threshold increase in the active block than did the monitoring condition. For *v*-parameter estimates, there was a significant decrease in the drift rate between the baseline and the active block in both conditions, $F(1, 46) = 58.03$, $p < .001$, $MSE = 2.38$, $\eta_p^2 = .56$ (and no interaction with condition, $F < 1$). In terms of $t_o$-parameters, both groups displayed increased nondecision time, $F(1, 46) = 41.96$, $p < .001$, $MSE = .003$, $\eta_p^2 = .48$, and the $t_o$ increase did not interact with condition, $F(1, 46) = 2.88$, $p = .10$, $MSE = .003$, $\eta_p^2 = .06$.

In sum, there is converging evidence from the patterns of ongoing-task responding, PM accuracy, and *a*-parameter estimates that the boundary PM and monitoring conditions reflected the use of somewhat different strategies during the active block.

## Standard Condition

We are now in a position to investigate whether the standard PM condition shows patterns of results that approximate the patterns

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

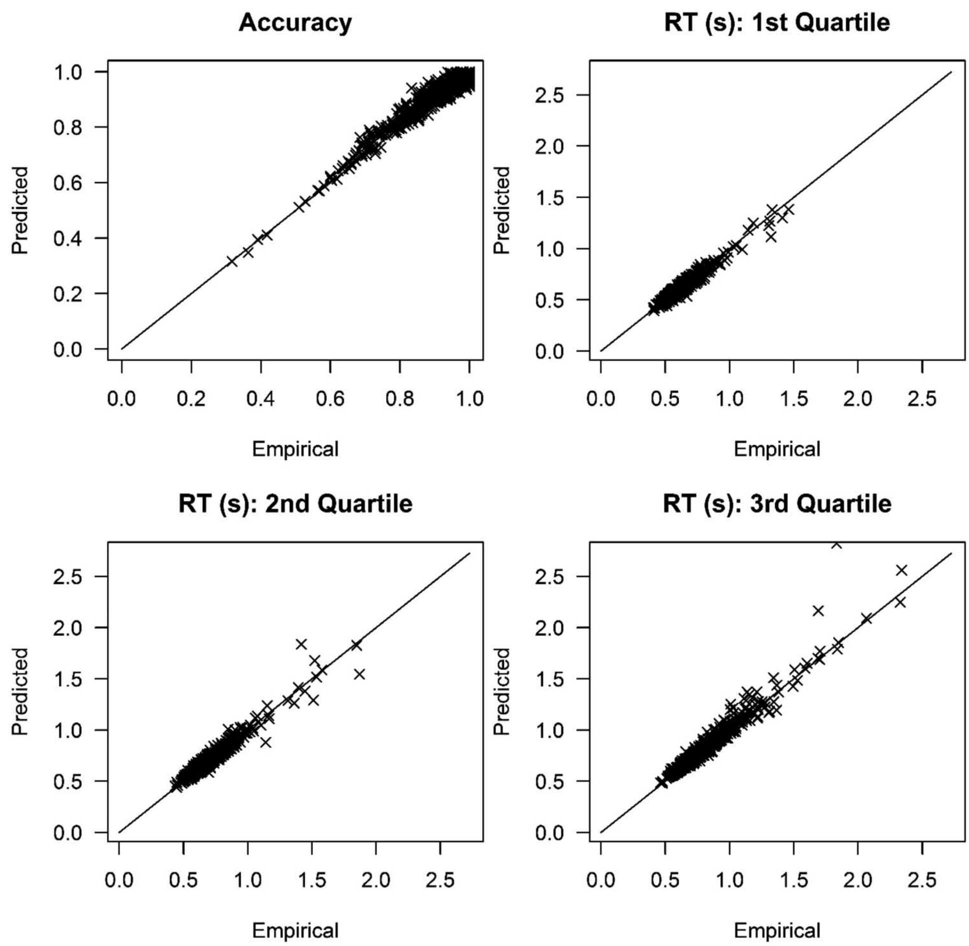


*Figure 1.* The graphs show the relation between the empirical and model predicted accuracies and RTs (separated by the 1st, 2nd, and 3rd RT-quartile) for each participant. Observations closer to the diagonal indicate a better fit. The graphs suggest acceptable individual model fits, except for the one participant with a predicted value > 2.5 in the 3rd quartile. When that participant was excluded, none of the statistical conclusions changed. RT = response time.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

found when a delay strategy is used versus when a monitoring strategy is used. That is, to illuminate the underlying processes involved in a typical PM task, of interest is which of the two anchor points the standard condition more closely resembles (refer to Table 1 for a complete list of the predictions made by both the delay and monitoring theories). To do so, in separate 2 (block) × 2 (condition) ANOVAs we computed pairwise comparisons of the standard condition with the monitoring condition and with the boundary PM condition. In terms of ongoing task accuracy, the standard condition was identical to the monitoring condition, in that neither increased in accuracy across blocks (*M* increase = 0; $F < 1$). By contrast, the boundary PM condition increased in accuracy, resulting in a significant difference relative to the standard condition, $F(1, 46) = 26.13$, $p < .001$, $MSE = .0003$, $\eta_p^2 = .36$, for the interaction. In terms of slowing (from baseline to active blocks), the standard condition slowed less (*M* slowing = 124 ms) than either the monitoring condition, $F(1, 46) = 7.34$, $p < .05$, $MSE = 2743.01$, $\eta_p^2 = .14$, or the boundary PM condition, $F(1, 46) = 16.33$, $p < .001$, $MSE = 10231.04$, $\eta_p^2 = .26$.

Finally, we examined PM accuracy to determine whether or not the use of differential strategies affected PM performance. The monitoring condition obtained greater PM ($M = .73$, $SE = .05$) than the standard condition ($M = .59$, $SE = .05$), $F(46) = 4.12$, $p < .05$, $\eta_p^2 = .08$. There was no significant difference between the standard and boundary PM conditions ($F < 1$).

Next we compared parameter estimates from the RD model. The boundary PM condition had a greater decision threshold change than the standard condition, $F(1, 46) = 21.16$, $p < .001$, $MSE = .06$, $\eta_p^2 = .32$, for the interaction, whereas there was no significant difference for threshold changes between the standard and the monitoring conditions, $F(1, 46) = 2.01$, $p > .05$, $MSE = .02$. There were no significant differences in drift-rate reductions between the standard condition and the monitoring or boundary PM conditions (*F*s < 1). For $t_O$-parameters, both the monitoring condition, $F(1, 46) = 4.24$, $p < .05$, $MSE = .001$, $\eta_p^2 = .08$, and the boundary PM condition, $F(1, 46) = 9.91$, $p < .05$, $MSE = .002$, $\eta_p^2 = .18$, had significantly increased nondecision time relative to the standard condition.

Table 3
*Descriptive Statistics and ANOVA Analyses for Behavioral and Modeling Results*

| | Condition | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Boundary control | | Boundary PM | | Monitoring | | Standard | | | |
| DV | Baseline | Active | Baseline | Active | Baseline | Active | Baseline | Active | ME (Block) | IA (Block × Condition) |
| Acc | .89 (.012) | .94 (.008) | .90 (.008) | .94 (.006) | .90 (.015) | .91 (.017) | .90 (.009) | .90 (.010) | ✓ | ✓ |
| RT (ms) | 656 (11) | 832 (28) | 618 (13) | 908 (47) | 629 (15) | 810 (25) | 632 (16) | 756 (25) | ✓ | ✓ |
| *a* | 1.10 (.04) | 1.55 (.09) | .96 (.02) | 1.60 (.10) | 1.10 (.06) | 1.36 (.07) | 1.03 (.05) | 1.22 (.06) | ✓ | ✓ |
| *v*(HF) | 5.23 (.37) | 3.85 (.22) | 5.21 (.35) | 3.17 (.19) | 4.79 (.30) | 2.89 (.19) | 5.47 (.35) | 3.34 (.19) | ✓ | — |
| *v*(LF) | 1.38 (.12) | 1.55 (.14) | 1.38 (.17) | 1.55 (.15) | 1.56 (.12) | 1.47 (.11) | 1.27 (.14) | 1.46 (.17) | ✓ | — |
| *v*(NW) | 2.19 (.12) | 2.18 (.12) | 2.82 (.16) | 2.27 (.13) | 2.35 (.16) | 1.96 (.15) | 2.56 (.15) | 2.17 (.16) | ✓ | — |
| *v*(tot) | 8.80 (.51) | 7.57 (.39) | 9.41 (.54) | 6.99 (.40) | 8.70 (.49) | 6.32 (.35) | 9.30 (.47) | 6.97 (.39) | ✓ | — |
| *t0* | .49 (.009) | .52 (.014) | .49 (.011) | .57 (.021) | .47 (.010) | .52 (.010) | .48 (.009) | .51 (.011) | ✓ | ✓ |
| PM | — | — | — | .55 (.06) | — | .73 (.05) | — | .59 (.05) | [✓] | — |

*Note.* Data are means and standard errors. Drift rate is separated by stimulus type (high- and low-frequency words, and nonwords) because the best-fitting model estimated drift rate separately, rather than combined across stimulus type. Significant results of the 2 × 4 omnibus ANOVA ($p < .05$) are marked by a check. The one check in brackets indicates a $p$ value of .05. PM = prospective memory; ME = presence of a condition main effect; IA = presence of an interaction; DV = dependent variable; Acc = ongoing task accuracy; RT = response time; $a$ = threshold; $v$ = drift rate (separated by stimulus type and summed); HF = high frequency words; LF = low frequency words; NW = nonwords; tot = total; $t0$ = nondecision component.

In sum, the standard condition closely resembled the monitoring condition for $a$-parameter estimates but showed lower PM accuracy. Conversely, the standard condition showed lower $a$ estimates than the boundary PM condition but comparable PM accuracy. We will consider these mixed patterns in the Discussion section.

Finally, we $z$-scored our working memory measures (backward digit span and OSPAN), created a composite score, and examined whether PM accuracy was correlated with working memory capacity. Because our conditions received instructions that we believed should affect the influence of working memory capacity on PM accuracy (i.e., the use of a delay strategy should not rely heavily on working memory, but monitoring would), we examined the correlation separately for each condition. In the boundary PM condition, working memory capacity was not correlated with PM accuracy, $r(24) = .09$, $p > .05$. By contrast, in both the monitoring condition, $r(24) = .49$, $p < .05$, and the standard condition, $r(24) = .41$, $p < .05$, working memory and PM accuracy were significantly correlated.

Table 4
*Difference Scores of the Statistics Presented in Table 3*

| | Condition | | | |
|---|---|---|---|---|
| DV | Boundary control | Boundary PM | Monitoring | Standard |
| Acc | .044 (.007) | .042 (.006) | .003 (.007) | .002 (.005) |
| RT (ms) | 177 (24) | 290 (39) | 181 (17) | 123 (13) |
| *a* | .46 (.09) | .64 (.09) | .27 (.04) | .19 (.04) |
| *v*(HF) | −1.38 (.37) | −2.04 (.37) | −1.90 (.30) | −2.13 (.31) |
| *v*(LF) | .17 (.11) | .17 (.12) | −.09 (.10) | .20 (.10) |
| *v*(NW) | −.01 (.07) | −.55 (.15) | −.39 (.08) | −.39 (.11) |
| *v*(tot) | −1.23 (.44) | −2.42 (.50) | −2.38 (.38) | −2.32 (.34) |
| *t0* | .029 (.012) | .084 (.018) | .049 (.010) | .024 (.007) |
| PM | — | .55 (.06) | .73 (.05) | .59 (.05) |

*Note.* Data are means and standard errors. Difference scores were computed by subtracting the baseline block from the active block. Drift rate is separated by stimulus type (high- and low-frequency words, and nonwords) because the best-fitting model estimated drift rate separately, rather than combined across stimulus type. DV = dependent variable; PM = prospective memory; Acc = ongoing task accuracy; RT = response time; $a$ = threshold; $v$ = drift rate (separated by stimulus type and summed); HF = high frequency words; LF = low frequency words; NW = nonwords; tot = total; $t0$ = nondecision component.

## Discussion

The goal of this experiment was to combine the use of both behavioral and modeling methods to better illuminate the underlying processes involved in PM costs and, by extension, the processes supporting typical laboratory PM tasks. The large body of PM literature favors a monitoring view of costs, but this interpretation is limited by the fact that costs are an indirect indicator of cognitive processes. Accordingly, accumulator models have been applied to ongoing task accuracy and RT patterns to derive parameter estimates of the latent processes assumed to underlie the decisions made in a typical PM task. Recent modeling evidence favors a delay account of costs (Heathcote et al., 2015), but is limited by the lack of parameter validation of the underlying processes in dual-task settings (e.g., the concurrent demands of both an ongoing LDT and a PM task). The present study attempted to combine behavioral and modeling techniques to illuminate the processes reflected by costs. As an important component of this effort, we reported the first experiment to establish the particular signatures of the model parameters for two distinct PM strategies (i.e., response delaying vs. monitoring).

We first sought to determine whether parameter estimates capture the lexical decision-making processes similarly when a PM task is concurrent to the LDT relative to when there is no PM task. Therefore, we included a boundary control condition with which to compare our boundary PM condition. There was strong convergence between the two conditions in terms of patterns theoretically implicated by the use of a delay strategy. Both the boundary control and boundary PM conditions displayed the expected speed–accuracy trade-off, resulting in both slower RTs (though greater in the boundary PM condition) and increased accuracy in the active block relative to the baseline block. Further, these patterns were reflected in $a$-parameter estimates, showing that for

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

both conditions, participants' decision thresholds became more conservative in the active block. These results establish that our delaying instructional condition was successful and that the model is sensitive to these behavioral changes. It should be noted, however, that the boundary PM condition did not appear to be using a pure delay strategy, as the increased slowing in this condition was also accompanied by marginally reduced drift rate and significantly increased nondecision time relative to the boundary control condition.

Next, our first central objective was to determine whether the use of such a delay strategy resulted in patterns of behaviors and parameter estimates that diverged from those that emerged when participants were implementing a monitoring strategy for the PM task. We reasoned that if monitoring is distinctly different from delayed responding, then the difference might be reflected by a divergence in both behavioral measures and parameter estimates. There is initial support for this general approach based on previous similar manipulations of task importance (Kliegel, Martin, McDaniel, & Einstein, 2004; Loft, Kearney, & Remington, 2008; Smith & Bayen, 2004), in which emphasizing the importance of the (nonfocal) PM task improved PM and resulted in more costs to the ongoing task. As an example, our boundary PM condition resembled Smith and Bayen's (2004; Experiment 1) group in which the importance of the ongoing task was emphasized,[2] and our monitoring condition was similar to their group in which the importance of the PM task was emphasized. They found that both groups slowed relative to a pure control but that the PM emphasis group slowed more and obtained better PM.

On most measures and parameters, the two conditions did show different patterns. The boundary PM condition exhibited more costs, higher accuracy changes, and more pronounced decision threshold changes than the monitoring condition (when comparing the active block with the baseline block). Drift-rate and nondecision parameter results were less divergent, with drift rate significantly reducing and nondecision time increasing from the baseline to the active block for both groups. To the extent that drift rate and nondecision time are reflective of monitoring (Heathcote et al., 2015; Horn & Bayen, 2015), as noted above, the implication is that even in the boundary PM condition, some participants engaged in monitoring. Still, the differences in ongoing task performance and in the decision threshold parameter indicate that participants in each of the two conditions were generally using somewhat different strategies to complete their PM tasks in the active block, in line with their delayed responding or monitoring instructions. Further, the monitoring condition obtained significantly greater PM than the boundary PM condition (.73 and .55, respectively), supporting the claim that these differences have functional implications for PM.

## PM Strategies Under Standard Instructions

Our second central objective was to use the above anchor points to gauge the PM processes adopted by participants under typical PM instructions. According to Heathcote et al. (2015), costs in the typical laboratory PM task reflect the use of a delay strategy to allow more time for relevant PM information to accumulate, and with more delay (observed with greater costs and indexed by the threshold parameter), one should generally expect higher PM (e.g., Smith, 2003; Smith & Bayen, 2004). If the delay theory has merit, then we expected that the costs and parameter estimates from the standard condition should resemble those from the boundary PM condition. Further, if for some reason one condition delayed more than the other (i.e., sets a higher decision threshold), then that condition should display somewhat higher PM. By contrast, the attentional monitoring view (Smith, 2003) would be supported if the standard condition parameter estimates were similar to the monitoring condition. Again, if one condition exhibited more costs than the other, then that condition should obtain somewhat higher PM (because of a decrease in drift rate or increase in nondecision time).[3]

A first important finding was that the ongoing task performances diverged between the standard PM condition and the boundary PM condition. The standard condition slowed significantly less than the boundary PM condition, did not show a similar increase in accuracy from the baseline block to the active block (as one would expect under a non-capacity-consuming delay process), and had significantly lower decision thresholds than the boundary PM condition. By contrast, the standard condition's ongoing task indices were identical in most respects—ongoing-task accuracy, drift rates, and decision thresholds (though not in costs or nondecision time)—to the monitoring condition. In general, then, these indices suggest that the standard PM group was recruiting processes more in line with monitoring than with a delay-response strategy.

Another interesting feature of the results is that the monitoring condition *did* obtain significantly greater PM accuracy than the standard condition—that is, higher costs were associated with better PM performance in the monitoring condition relative to the standard condition. By contrast, in the boundary PM condition, higher costs (and higher threshold estimates) did not translate into higher PM performance relative to the standard condition. On the one hand, this pattern could be interpreted as indicating that the processes underlying cost were different in the standard condition relative to the boundary PM condition. Otherwise, the greater costs along with the higher decision threshold change for the boundary condition should have translated into higher PM than for the standard condition (if the standard condition were also using a delay strategy). Therefore, in conjunction with the above findings, it appears as though the costs observed in a standard PM condition, at least in our experiment, more closely resemble the behaviors associated with monitoring.

On the other hand, one might argue that if participants in the standard condition were monitoring, then they should have obtained significantly greater PM accuracy than the boundary PM condition (given the robust advantage in PM accuracy for the monitoring condition). However, participants in the boundary PM condition appeared to not only be delaying their responding but also monitoring to some degree. This interpretation is based on several observations. First, the boundary PM condition showed significantly greater slowing in the active block than the boundary control condition. If participants in the boundary PM condition

[2] But note that a delay strategy was not specified.

[3] Note that many of these predictions require the interpretation of null results. Therefore, power analyses were conducted using G*Power software for within–between interactions. Using our total number of subjects, assuming a medium-sized effect, and computing the within-subjects correlations between baseline and active blocks, all analyses of parameter values, RT, and accuracy obtained a power of .87 or greater.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

were simply using a delay strategy as instructed (the instructions emphasized very strongly that participants' primary goal was to obtain high accuracy on the ongoing task), the boundary PM and boundary control conditions would presumably have displayed equivalent slowing and decision threshold increases. Second, the drift rate decreased more (marginally) and nondecision time increased more (significantly) from the baseline to the active block in the boundary PM condition relative to the boundary control condition. Thus, the implication is that participants in the boundary PM condition were not solely engaged in a delay-response strategy. If this conclusion has currency, then it suggests that participants find it difficult or are disinclined to use a pure delay-response strategy in the typical laboratory PM task (cf. Heathcote et al., 2015; Strickland et al., 2017), even when instructed to do so.

An alternative interpretation of the just-discussed slowing and decision threshold patterns is that the boundary PM condition did implement a relatively pure delay-response strategy, but did so in an even more conservative fashion than did the boundary control condition. This is plausible on the assumption that participants in the boundary PM condition recognized that the PM-related information might take some time to accrue and adjusted their boundary for the LDT to an even more conservative threshold (than in the boundary control condition). This interpretation, however, also disfavors the idea that in the *standard* PM condition, participants were adopting a delay strategy (Heathcote et al., 2015; Loft & Remington, 2013). Were that the case, then the standard PM condition should have evidenced the degree of slowing and the increase in threshold estimates displayed by the boundary PM condition, given their very similar PM performance. This pattern clearly was not obtained, however.

Further evidence that an attentionally demanding (monitoring) process is involved in typical nonfocal PM tasks was that working memory capacity was significantly correlated with PM accuracy in both the monitoring and standard conditions but not the boundary PM condition. There is no apparent reason why working memory should be associated with a non-capacity-consuming delay process; in contrast, from a monitoring view, working memory capacity should be critical. Our results support this interpretation, showing that participants higher in working memory capacity performed better in the condition in which monitoring was instructed and in the standard condition in which monitoring was implicated by modeling parameters and behavioral markers. Additionally, this pattern was not obtained in the condition instructed to adopt a delay strategy.

In sum, in light of the constellation of ongoing task indices, modeling results and PM accuracy patterns, a plausible interpretation of the present findings is that the monitoring condition followed instructions and monitored to a high degree. The standard condition, without explicit instructions to monitor, did not engage in as much monitoring as would be most beneficial. Consistent with this idea, participants in this condition exhibited the lowest costs, implying that their monitoring was limited or inconsistent (cf. Einstein & McDaniel, 2010; West & Craik, 1999). However, even with much lower costs, the standard condition still evidenced numerically greater PM accuracy than the boundary PM condition, suggesting that the standard condition was not relying on a delay strategy. It remains possible that in the standard condition monitoring processes were perhaps additionally supported by a delay-response strategy (Heathcote et al., 2015; Horn & Bayen, 2015). The key point, however, is that the cost in the standard PM condition likely did not reflect that participants were relying extensively, if at all, on a delay strategy.

One reviewer noted the possibility that our monitoring instruction might not only have increased the frequency of monitoring (as we argue) but could also have changed the cognitive nature of monitoring, specifically by lowering the PM threshold with a more liberal bias toward making a PM response. The low number of PM observations makes it impossible to directly model the PM-related decision parameters; however, we appeal to postexperimental qualitative data to better inform this possibility. Specifically, when participants were asked what strategy they used to help them execute the PM task, 11 participants in the monitoring condition and eight in the standard condition were coded as reporting active rehearsal of the PM task or target-searching behavior. A representative response from the monitoring condition was "I looked for *tor* in every word, before seeing if it was a word or not," and a representative response in the standard condition was, "I looked for *tor* first, then decided if it was a word or nonword." By and large, across both conditions, when participants reported monitoring, they reported engaging in a serial search for *tor* before or after making a lexical decision. Although we cannot directly test whether participants adjust their PM decision threshold, both conditions report quite similar subjective experiences of monitoring.

Our results also underscore an ambiguity in the literature regarding inconsistencies in the drift-rate parameter across PM studies using accumulator models. The standard condition in our study was patterned precisely on that used by Heathcote et al. (2015; Experiment 2), yet we obtained significant reductions in the drift rate from the baseline block to the active block, whereas Heathcote et al. did not. These conflicting patterns, even within identical paradigms, raise questions regarding the reliability of the drift-rate parameter; the obtained estimate might depend on advanced aspects of the model fitting procedures. Certain model specifications, such as allowing the starting point for the drift process to vary instead of fixating it to a/2, may affect drift rates in a way that render them either indicative or not of PM processes (Heathcote et al., 2015). Note, however, that we still found significant drift-rate changes when the starting point was allowed to vary with block (RD 8; Table 2).

Further, a more sequential engagement in monitoring before or after the lexical decision should affect nondecision time rather than drift rates. Thus, it may be premature for researchers to rely on drift rate as a decisive indicator of limited-capacity attentional processes in PM. Other studies as well have obtained mixed results regarding the drift-rate parameter. Some show a significant reduction (Boywitt & Rummel, 2012; Horn et al., 2011, 2013; Rummel et al., 2013), and some do not (Heathcote et al., 2015; Horn & Bayen, 2015; Strickland et al., 2017). However, it should also be noted (as Heathcote et al., 2015, noted) that some studies (e.g., Boywitt & Rummel, 2012; Horn et al., 2011) contain methodological biases such as low error rates and practice effects. When Heathcote et al. (2015) reanalyzed these studies, there was no evidence for a reliable reduction in drift rate. Importantly, all three studies that did not find significant drift-rate reductions did still find significant PM-induced increases in nondecision time when applying the RD model to their data.

It is noteworthy that the present interpretations rely on the results from one particular type of accumulator model: the RD model. As is evident from Appendixes A and B, an alternative LBA modeling approach, using a parameter specification that yielded the best fit for our data according to the BIC, would have indicated somewhat different conclusions. Effects on the boundary parameter were comparable between the two approaches, but no PM-associated effects on nondecision time were observed with the LBA approach. Instead, the monitoring condition showed increased drift-rate variability, in LBA terms. In a recent project, researchers from different laboratories were asked to apply their favored accumulator model to a variety of data sets while being blind to the nature of the manipulation reflected in the data (Dutilh et al., 2017). Results suggested that the RD model is generally more likely to allocate systematic variance to the nondecision parameter than the LBA model. The present differences between the results from the two model types are in line with this observation.

It remains an open question at this point, however, which model provides more valid indicators of PM processes. In the present study, we favored the RD model because it has been frequently applied to PM data. Moreover, the current LBA modeling captured the PM-related changes in ongoing task behavior across the PM conditions primarily with the threshold parameter, and was unable to distinguish between conditions that we know used qualitatively different PM strategies. Accordingly, it is uncertain whether previous LBA modeling would have signaled a monitoring strategy even had it been engaged by participants (e.g., Heathcote et al., 2015). However, the RD model also failed to pick up the increased monitoring in the monitoring condition relative to the standard condition. In short, PM researchers using accumulator models in the future should keep in mind that (a) the choice of a certain modeling approach may significantly affect conclusions regarding underlying PM processes, and (b) regardless of the modeling approach, individual parameter changes do not appear to uniquely reflect particular PM processes. Consequently, anchor conditions (for which particular strategies are instructed) like those used herein may be required when using such models to infer PM processes.

## Converging Evidence for a Capacity-Consuming (Monitoring) Process in Nonfocal PM

In closing, we note that there is converging evidence, from both our study and many other behavioral experiments, that at least some PM strategies require the engagement of attentional processes (which have been generally described as monitoring). There is support for this claim from several individual differences effects. For example, working memory has been frequently shown to mediate the relationship between costs and PM performance (Brewer et al., 2010; Smith & Bayen, 2005; Smith et al., 2011). Although participants who show greater costs typically have higher PM, those with high working memory capacity obtain better PM without any corresponding increase in costs. Further, older adults frequently perform worse on nonfocal PM tasks than younger ones (i.e., those resulting in costs; Kliegel, Jäger, & Phillips, 2008; Smith & Bayen, 2006), and when forced to divide their attention (Einstein, Smith, McDaniel, & Shaw, 1997).

Additionally, neuroimaging studies consistently implicate sustained anterior prefrontal cortex (aPFC) activation in nonfocal PM, which counters what one would expect based on delay theory. Sustained aPFC activation is viewed as implicating an active monitoring strategy, and is correlated with PM performance (Beck, Ruge, Walser, & Goschke, 2014; Benoit, Gilbert, Frith, & Burgess, 2012; Burgess, Gonen-Yaacovi, & Volle, 2011; Simons et al., 2006). By contrast, it is uncertain what role sustained aPFC activation would reflect for a strategy that simply delays responding (delay theory). It is possible that initiating a delay strategy requires the use of the aPFC to allocate attentional resources (see Strickland et al., 2017), but it is unclear why this strategic approach would require sustained cognitive control throughout the task. Prestimulus neural states in frontal, temporal, and parietal regions (as measured by EEG), too, predict successful PM for abstractly specified targets (e.g., animal words), implicating their role in an active search process (Knight, Marsh, Brewer, & Clementz, 2012). This is problematic because delay theory appears to support a more uniform response bias (i.e., higher threshold), rather than a fluctuating cognitive state, leading to successful PM.

Finally, the use of ex-Gaussian modeling of RTs indicates that entire distribution shifts (which reflects continuous monitoring), rather than greater positive skew (which has been assumed to reflect intermittent monitoring processes), results in more successful PM performance (Ball, Brewer, Loft, & Bowden, 2015; Rummel, Smeekens, & Kane, 2017). There is little reason to interpret these shifts as reflecting intermittent or continuous engagement of a delay strategy under a non-capacity-consuming process, because there would be no reason *not* to adopt a delay strategy. If, however, the processes underlying costs are cognitively demanding, then there are clear benefits to showing costs intermittently (i.e., monitoring is exhausting). Interestingly, the LBA-model-based analyses of the present data attributed monitoring effects to an increase in the variability of the drift rate (see Appendix A), which may also suggest a transient engagement in monitoring.

Finally, we would like to emphasize our agreement that the observation of PM-induced costs to ongoing tasks per se does not directly implicate monitoring processes (Einstein & McDaniel, 2010), and that accumulator models offer an exciting and fruitful way to better understand the cognitive processes involved in a PM task. The present study offers a converging approach to illuminate the processes underlying costs by jointly using behavioral data, modeling techniques, and embedded parameter validation. The results are most in line with the conclusion that PM tasks (when they are not focal to the ongoing task) require the use of attentionally demanding processes. This has been typically conceptualized as monitoring, but regardless of the exact qualities of the cognitive processes, the present study does *not* support the view that a non-capacity-consuming delay process is the primary approach that people rely on in standard nonfocal PM laboratory tasks.

## References

Ball, B. H., Brewer, G. A., Loft, S., & Bowden, V. (2015). Uncovering continuous and transient monitoring profiles in event-based prospective memory. *Psychonomic Bulletin & Review, 22,* 492–499. http://dx.doi.org/10.3758/s13423-014-0700-8

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., . . . Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39,* 445–459. http://dx.doi.org/10.3758/BF03193014

Beck, S. M., Ruge, H., Walser, M., & Goschke, T. (2014). The functional neuroanatomy of spontaneous retrieval and strategic monitoring of delayed

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

intentions. *Neuropsychologia, 52,* 37–50. http://dx.doi.org/10.1016/j.neuropsychologia.2013.10.020

Benoit, R. G., Gilbert, S. J., Frith, C. D., & Burgess, P. W. (2012). Rostral prefrontal cortex and the focus of attention in prospective memory. *Cerebral Cortex, 22,* 1876–1886.

Boywitt, C. D., & Rummel, J. (2012). A diffusion model analysis of task interference effects in prospective memory. *Memory & Cognition, 40,* 70–82. http://dx.doi.org/10.3758/s13421-011-0128-6

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,* 304–311. http://dx.doi.org/10.3758/MC.38.3.304

Burgess, P. W., Gonen-Yaacovi, G., & Volle, E. (2011). Functional neuroimaging studies of prospective memory: What have we learnt so far? *Neuropsychologia, 49,* 2246–2257.

Burgess, P. W., Quayle, A., & Frith, C. D. (2001). Brain regions involved in prospective memory as determined by positron emission tomography. *Neuropsychologia, 39,* 545–555. http://dx.doi.org/10.1016/S0028-3932(00)00149-4

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review, 18,* 61–69. http://dx.doi.org/10.3758/s13423-010-0022-4

Dutilh, G., Annis, J., Brown, S. D., Cassy, P., Evens, N. J., Grasman, R. P. P. P., . . . Donkin, C. (2017). The quality of response time data inference: A blinded, collaborative assessment of the validity of cognitive models. Advance online publication. http://dx.doi.org/10.17605/OSF.IO/S2X32

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

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. http://dx.doi.org/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,* 327–342. http://dx.doi.org/10.1037/0096-3445.134.3.327

Einstein, G. O., Smith, R. E., McDaniel, M. A., & Shaw, P. (1997). Aging and prospective memory: The influence of increased task demands at encoding and retrieval. *Psychology and Aging, 12,* 479–488. http://dx.doi.org/10.1037/0882-7974.12.3.479

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition, 43,* 226–236.

Guynn, M. J. (2003). A two-process model of strategic monitoring in event-based prospective memory: Activation/retrieval mode and checking. *International Journal of Psychology, 38,* 245–256. http://dx.doi.org/10.1080/00207590344000178

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. http://dx.doi.org/10.1037/a0038952

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. http://dx.doi.org/10.1037/a0037676

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. http://dx.doi.org/10.1037/a0022808

Horn, S. S., Bayen, U. J., & Smith, R. E. (2013). Adult age differences in interference from a prospective-memory task: A diffusion model analysis. *Psychonomic Bulletin & Review, 20,* 1266–1273. http://dx.doi.org/10.3758/s13423-013-0451-y

Jäger, T., & Kliegel, M. (2008). Time-based and event-based prospective memory across adulthood: Underlying mechanisms and differential costs on the ongoing task. *Journal of General Psychology, 135,* 4–22. http://dx.doi.org/10.3200/GENP.135.1.4-22

Kliegel, M., Jäger, T., & Phillips, L. H. (2008). Adult age differences in event-based prospective memory: A meta-analysis on the role of focal versus nonfocal cues. *Psychology and Aging, 23,* 203–208. http://dx.doi.org/10.1037/0882-7974.23.1.203

Kliegel, M., Martin, M., McDaniel, M. A., & Einstein, G. O. (2004). Importance effects on performance in event-based prospective memory tasks. *Memory, 12,* 553–561. http://dx.doi.org/10.1080/09658210344000099

Knight, J. B., Marsh, R. L., Brewer, G. A., & Clementz, B. A. (2012). Preparatory distributed cortical synchronization determines execution of some, but not all, future intentions. *Psychophysiology, 49,* 1155–1167. http://dx.doi.org/10.1111/j.1469-8986.2012.01400.x

Lerche, V., & Voss, A. (2017). Retest reliability of the parameters of the Ratcliff diffusion model. *Psychological Research, 81,* 1–24. http://dx.doi.org/10.1007/s00426-016-0770-5

Lerche, V., & Voss, A. (2017). Model complexity in diffusion modeling: Benefits of making the model more parsimonious. *Frontiers in Psychology, 7,* 1324. http://dx.doi.org/10.3389/fpsyg.2016.01324

Loft, S., Kearney, R., & Remington, R. (2008). Is task interference in event-based prospective memory dependent on cue presentation? *Memory & Cognition, 36,* 139–148. http://dx.doi.org/10.3758/MC.36.1.139

Loft, S., & Remington, R. W. (2013). Wait a second: Brief delays in responding reduce focality effects in event-based prospective memory. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 66,* 1432–1447. http://dx.doi.org/10.1080/17470218.2012.750677

Marsh, R. L., & Hicks, J. L. (1998). Event-based prospective memory and executive control of working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 336–349. http://dx.doi.org/10.1037/0278-7393.24.2.336

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

McDaniel, M. A., LaMontagne, P., Beck, S. M., Scullin, M. K., & Braver, T. S. (2013). Dissociable neural routes to successful prospective memory. *Psychological Science, 24,* 1791–1800. http://dx.doi.org/10.1177/0956797613481233

Reynolds, J. R., West, R., & Braver, T. (2009). Distinct neural circuits support transient and sustained processes in prospective memory and working memory. *Cerebral Cortex, 19,* 1208–1221. http://dx.doi.org/10.1093/cercor/bhn164

Rummel, J., Kuhlmann, B. G., & Touron, D. R. (2013). Performance predictions affect attentional processes of event-based prospective memory. *Consciousness and Cognition: An International Journal, 22,* 729–741. http://dx.doi.org/10.1016/j.concog.2013.04.012

Rummel, J., Smeekens, B. A., & Kane, M. J. (2017). Dealing with prospective memory demands while performing an ongoing task: Shared processing, increased on-task focus, or both? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43,* 1047–1062. http://dx.doi.org/10.1037/xlm0000359

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General, 136,* 414–429. http://dx.doi.org/10.1037/0096-3445.136.3.414

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Simons, J. S., Schölvinck, M. L., Gilbert, S. J., Frith, C. D., & Burgess, P. W. (2006). Differential components of prospective memory? Evidence from fMRI. *Neuropsychologia, 44,* 1388–1397. http://dx.doi.org/10.1016/j.neuropsychologia.2006.01.005

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,* 347–361. http://dx.doi.org/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,* 1089–1095. http://dx.doi.org/10.1037/a0019183

Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 756–777. http://dx.doi.org/10.1037/0278-7393.30.4.756

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

Smith, R. E., & Bayen, U. J. (2006). The source of adult age differences in event-based prospective memory: A multinomial modeling approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 623–635. http://dx.doi.org/10.1037/0278-7393.32.3.623

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,* 734–746. http://dx.doi.org/10.1037/0278-7393.33.4.734

Smith, R. E., Persyn, D., & Butler, P. (2011). Prospective memory, personality, and working memory: A formal modeling approach. *Zeitschrift fur Psychologie, 219,* 108–116. http://dx.doi.org/10.1027/2151-2604/a000055

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. http://dx.doi.org/10.1037/xlm0000400

van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple empirical effects. *Psychonomic Bulletin & Review, 24,* 547–556. http://dx.doi.org/10.3758/s13423-016-1081-y

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology, 60,* 385–402. http://dx.doi.org/10.1027/1618-3169/a000218

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32,* 1206–1220. http://dx.doi.org/10.3758/BF03196893

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: A tutorial based on fast-dm-30. *Frontiers in Psychology, 6,* 336. http://dx.doi.org/10.3389/fpsyg.2015.00336

West, R., & Craik, F. I. (1999). Age-related decline in prospective memory: The roles of cue accessibility and cue sensitivity. *Psychology and Aging, 14,* 264–272. http://dx.doi.org/10.1037/0882-7974.14.2.264

Yap, M. J., Balota, D. A., Sibley, D. E., & Ratcliff, R. (2012). Individual differences in visual word recognition: Insights from the English Lexicon Project. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 53–79. http://dx.doi.org/10.1037/a0024177

## Appendix A

### Descriptive Statistics and ANOVA Analyses for the BIC-Selected LBA Model

| | Condition | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Boundary control | | Boundary PM | | Monitoring | | Standard | | | |
| DV | Baseline | Active | Baseline | Active | Baseline | Active | Baseline | Active | ME (Block) | IA (Block × Condition) |
| $A$ | .45 (.05) | | .44 (.08) | | .55 (.09) | | .35 (.08) | | — | — |
| $B$ | 1.25 (.10) | 1.65 (.15) | 1.13 (.12) | 1.86 (.22) | 1.41 (.12) | 1.84 (.14) | 1.54 (.11) | 1.87 (.13) | ✓ | ✓ |
| $v_{true}$(HF) | 3.00 (.14) | | 3.03 (.21) | | 3.13 (.19) | | 3.43 (.17) | | — | — |
| $v_{true}$(LF) | 2.41 (.12) | | 2.45 (.16) | | 2.63 (.17) | | 2.78 (.13) | | — | — |
| $v_{true}$(NW) | 2.50 (.11) | | 2.63 (.17) | | 2.69 (.16) | | 2.93 (.14) | | — | — |
| $v_{false}$(HF) | –.59 (.52) | | –.38 (.32) | | .01 (.32) | | .17 (.37) | | — | — |
| $v_{false}$(LF) | 1.29 (.13) | | 1.33 (.19) | | 1.42 (.19) | | 1.77 (.19) | | — | — |
| $v_{false}$(NW) | .49 (.16) | | .25 (.25) | | .70 (.24) | | .96 (.19) | | — | — |
| $sv_{true}$ | .38 (.02) | .56 (.02) | .36 (.03) | .52 (.05) | .34 (.02) | .62 (.04) | .41 (.02) | .65 (.04) | ✓ | — |
| $t_0$ | .09 (.02) | | .12 (.02) | | .04 (.01) | | .07 (.01) | | — | — |

*Note.* Data are means and standard errors. Significant results ($p < .05$) are marked by a check. As is standard in the LBA model, different accumulation rates are estimated for word and nonword responses and because the best-fitting model estimated drift rate separately for each stimulus type (high- and low-frequency words, and nonwords) true responses are "word" ("nonword") responses on word (nonword) trials and false responses are "nonword" ("word") responses on word (nonword) trials. For the best-fitting model, only $B$ and $sv$ estimates varied with PM block (baseline, active). BIC = Bayesian information criterion; LBA = linear ballistic accumulator; PM = prospective memory; DV = dependent variable; ME = presence of a condition main effect; IA = presence of an interaction; $A$ = maximum starting point noise; $B$ = threshold to A; $v$ = mean accumulation rates for word and nonword responses (separated by stimulus type); HF = high frequency words; LF = low frequency words; NW = nonwords; $sv$ = accumulation rate standard deviation; $t0$ = non-decision component.

(*Appendices continue*)

## Appendix B

## Follow-Up Tests on the Significant Interaction Effects on the Boundary Parameter in the BIC-Selected LBA Model

As evident from Appendix A, the only significant interaction between PM block and condition was found for the *B* parameter. Follow-up tests indicated that the increase in *B* was significantly higher in the boundary PM condition ($M = .72$; $SE = .07$) than in the boundary control ($M = .40$; $SE = .13$), the PM monitoring ($M = .43$; $SE = .05$), or the PM standard conditions ($M = .32$; $SE = .04$), all $ps < .01$. The other three conditions did not differ, all $ps > .330$. Thus, in line with Heathcote et al. (2015), the BIC-selected LBA model suggests that PM-induced RT costs were driven by threshold changes only. However, the threshold changes did not parallel the differences in PM accuracy among conditions. Additionally, the BIC-selected LBA model did not pick-up on the increased attentional monitoring in the monitoring condition compared with the standard condition (although this was also true for the RD model).

Received March 27, 2017
Revision received September 15, 2017
Accepted September 15, 2017 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.