

Running head: Accumulating evidence about PM costs
**Accepted, Journal of Experimental Psychology: Learning, Memory,
and Cognition, 27/01/2017.**

© 2017, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors permission. The final article will be available, upon publication, via its DOI: 10.1037/xlm0000400

Accumulating evidence about what prospective memory costs actually reveal

Luke Strickland^{1, 2}, Andrew Heathcote^{2,3}, Roger W. Remington⁴
and Shayne Loft¹

¹ The School of Psychology,
The University of Western Australia, Australia

² The School of Medicine,
The University of Tasmania, Australia

³ The School of Psychology,
The University of Newcastle, Australia

⁴ The School of Psychology,
The University of Queensland, Australia

Address for Correspondence

Luke Strickland,
School of Medicine, Division of Psychology,
Private Bag 30, The University of Tasmania,
Churchill Avenue, Sandy Bay, 7005, Australia

Email: luke.strickland@utas.edu.au

Phone: 61-3-62262242

Fax: 61-3- 62262883

Abstract

Event-based prospective memory (PM) tasks require participants to substitute an atypical PM response for an ongoing task response when presented with PM targets. Responses to ongoing tasks are often slower with the addition of PM demands (“PM costs”). Prominent PM theories attribute costs to capacity-sharing between the ongoing and PM tasks, which reduces the rate of processing of the ongoing task. We modelled PM costs using the Linear Ballistic Accumulator and the Diffusion Decision Model in a lexical decision task with non-focal PM targets defined by semantic categories. Previous decision modelling, which attributed costs to changes in caution rather than rate of processing (Heathcote et al., 2015; Horn & Bayen, 2015), could be criticised on the grounds that the PM tasks included did not sufficiently promote capacity-sharing. Our semantic PM task was potentially more dependent on lexical decision resources than previous tasks (Marsh, Hicks, & Cook, 2005), yet costs were again driven by changes in threshold and not by changes in processing speed (drift rate). Costs resulting from a single target focal PM task were also driven by threshold changes. The increased thresholds underlying non-focal and focal costs were larger for word trials than non-word trials. As PM targets were always words, this suggests that threshold increases are used to extend the time available for retrieval on PM trials. Under non-focal conditions, but not focal conditions, the non-word threshold also increased. Thus, it seems that only non-focal instructions cause a global threshold increase due to greater perceived task complexity.

Keywords: prospective memory, costs, delay theory, linear ballistic accumulator model, diffusion model

In our everyday lives we often need to remember to perform an intended action when we next encounter an environmental cue (e.g., pass on a message to a co-worker), a requirement referred to as event-based prospective memory (PM). Event-based PM has traditionally been examined in the laboratory using the Einstein and McDaniel (1990) paradigm, where participants perform an ongoing task (e.g., a lexical decision task), and try to remember to make an alternative (PM) response when a target event is presented in the ongoing task (e.g., when a letter string contains a certain syllable). In these paradigms target trials, on which the stimulus is a PM target and the PM response is required, are relatively infrequent and interspersed amongst non-target trials, on which the stimulus is not a PM target and an ongoing task response is required.

Average response times (RTs) to non-target trials are often longer in blocks where participants have a PM task, versus control blocks with no PM task (Smith, 2003). This slowing is referred to as the PM cost. Costs are common for tasks in which making ongoing task decisions does not require processing information about the stimuli relevant to the PM task, referred to as *non-focal* PM tasks (Einstein & McDaniel, 2005). For example, detecting target members of a semantic category (e.g., animal) is non-focal to ongoing lexical decision making, because deciding whether a letter string is present in the mental lexicon does not require knowledge of the semantic categories to which the word belongs. An example of a focal PM task is to respond to a specific PM target word if presented during a lexical decision task (Einstein & McDaniel, 2005), because in this case the identity of the target word is part of the analysis required for the ongoing task. Focal PM costs are not always found and when they are present they are smaller than non-focal PM costs (Einstein et al., 2005).

Several theories of PM, such as the Preparatory Attentional and Memory Processes (PAM) theory (Smith, 2010), and the Multiprocess view (Einstein & McDaniel, 2005), assume that PM

costs occur because a proportion of an individual's available capacity is allocated away from the ongoing task and towards the PM task. We refer to these as *capacity-sharing* theories of PM costs to emphasize that the costs result from the division of capacity between the two tasks during stimulus processing. The empirical basis of capacity-sharing theories is slowing of mean correct RTs as a function of imposed PM task demands. The problem with this approach is that analysing mean RTs alone omits crucial information contained in the shape of the RT distributions and in error rates, which can indicate whether participants are trading speed for accuracy. This is compounded by the fact that most PM studies report near-ceiling ongoing task accuracy, where a significant change in speed could easily trade for a non-significant change in accuracy. This makes it impossible to unambiguously attribute elevated RTs for the ongoing task in the PM condition to capacity-sharing. Elevated RTs could equally result from changes in other aspects of the decision process, such as the amount of evidence required to trigger a response.

Indeed, Loft and Remington (2013) proposed an alternative to capacity-sharing theories that predicts PM costs can result from increased response thresholds. Heathcote, Loft and Remington (2015) subsequently referred to this idea as *delay theory*. Loft and Remington noted that in the Einstein and McDaniel (1990) paradigm, not only must the participant remember a new episodic PM task, but this less frequent PM task must compete for response selection with the more routine ongoing task. They reasoned that PM errors should be reduced if participants are provided with more processing time for the less frequent PM response to emerge. To test this, Loft and Remington asked participants to withhold their responses until a tone played. The tones played at varying delays after stimulus presentation. PM accuracy was improved by delays as brief as 0.2s. Heathcote et al. proposed that individuals endogenously implement something akin to this response-delay manipulation by raising their threshold to make ongoing task responses,

producing PM costs. This functional slowing could result either from a conscious strategy change, or an unconscious adaption, based on task experience.

Comparing Capacity-sharing and Delay Theories

Distinguishing delay theory from capacity-sharing theories requires a method of separating changes in processing speed from response threshold increases. Such a separation is intrinsic to *evidence accumulation models* of the decision process, which assume that evidence accumulates at an estimated rate (the drift rate) until it exceeds a certain amount (the response threshold), at which time a task response is made (Ratcliff & Smith, 2004). Capacity-sharing theories assume that PM tasks absorb capacity that could otherwise be used to speed up and/or improve the accuracy of ongoing task responses (e.g., Einstein et al., 2005; Smith, 2010). Mapping the capacity demands of one task (e.g., PM) to increased mean RT on another concurrent task (e.g., ongoing) is most consistent with the conceptualisation of capacity as the amount of work the cognitive system is able to perform per unit of time (Wenger & Townsend, 2000), which we refer to as *functional* capacity. For example, the PAM theory explicitly attributes PM costs to shared attentional resources that are functional to PM retrieval (Smith & Bayen, 2004). Theories of attention, almost universally, equate resource allocation with processing speed, regardless of whether the models posit a continuous division of resource capacity between two tasks, or a single-channel, all-or-none allocation. In resource accounts (Bundesen, 1990; Gobell, Tseng, & Sperling, 2004; Kahneman, 1973; Navon & Gopher, 1979; Wickens, 1980) attentional capacity is a processing rate multiplier. Evidence accrues in parallel for each of a set of multiple concurrent tasks at a rate directly proportional to the amount of capacity allocated to that task. In contrast, single-channel bottleneck theories treat resource sharing as iterative sampling, whereby a single resource is allocated in rapid succession between a set of current tasks (Pashler, 1984;

Welford, 1952). The rate of evidence accumulation for a given task (or channel) is determined by the frequency with which attention is allocated to it. In either case, sensory signals, decisions, and response selection are speeded in direct proportion to the attentional resources they receive. Thus, functional capacity-sharing between the PM task and the ongoing task would be reflected in decreased evidence accumulation to the ongoing task.

Drift rates agree with other measures of functional capacity, such as the nonparametric cognitive capacity coefficient provided by Systems Factorial Technology (Eidels, Donkin, Brown, & Heathcote, 2010). The capacity coefficient compares performance to a benchmark unlimited-capacity independent parallel model in specially designed experimental paradigms, and classifies participants' task performance as either exhibiting super-capacity (more capacity when there are two information streams), unlimited capacity (no interference in processing between information streams), or limited capacity (interference in processing between information streams). Eidels et al. investigated capacity requirements using a redundant target paradigm, in which the total amount of work the system must perform is manipulated by presenting either single or double targets to participants. LBA drift rates and the capacity coefficient agreed in their attribution of individual differences in capacity among participants. Donkin et al. (2014) found that LBA rates agreed with Townsend and Altieri (2012)'s accuracy adjusted capacity coefficient estimates regarding which accuracy manipulations induced increases in threshold and rates, and which selectively induced changes in threshold. Logan et al. (2014) found that increasing the array of choice alternatives in a word identification task – and hence the demand on capacity due to the need to process each alternative – reduced evidence accumulation rates. Sewell, Lilburn and Smith (2016) varied visual load by presenting different numbers of concurrent Gabor patch stimuli. They found that manipulation of the set size of

stimuli (visual load) altered drift rates. All of these results suggest that functional capacity is well measured by drift rates.

In contrast to capacity-sharing theories, delay theory attributes PM costs to the strategic slowing of ongoing task response selection to allow more time for the PM response to reach threshold. Consequently, it predicts that PM costs should be caused by changes in response criteria, reflected in the response threshold parameters of evidence accumulation models. Thus, the delay theory and capacity-sharing theories of costs produce opposing predictions. A finding that PM demands significantly affect mean drift rates, with no significant change to threshold, provides strong evidence against delay theory. Capacity-sharing theories do not predict that costs are solely accounted for by response threshold, as this parameter reflects strategy rather than the speed of processing. A finding of threshold changes with no changes in mean drift rate provides strong evidence against capacity-sharing.

Recently (Heathcote et al., 2015; Horn & Bayen, 2015), two evidence accumulation models have been fit to PM costs: the Linear Ballistic Accumulator (LBA) and the Diffusion Decision Model (DDM). Heathcote et al. applied the LBA and DDM to three experiments that used lexical-decision ongoing tasks. The PM task was to respond to either letter strings of a specific colour (non-targets were presented in multiple other colours), or to letter strings that contained a specific syllable (e.g., “tor”). Horn and Bayen (2015) reported DDM analyses of four experiments. Participants performed a lexical-decision ongoing task. The PM task was to make a PM response when items started with specific letters (e.g., G, H or M). In both Heathcote et al. and Horn and Bayen, there was no evidence that holding a PM intention decreased average drift rates. Instead, there was strong evidence in all seven of the data sets that PM costs were caused

by increased response thresholds. These results favour the delay theory of PM costs over capacity-sharing theories.

There is, however, an alternative view of cognitive capacity. It has been proposed that another type of capacity drives an overarching executive structure that controls otherwise modular processing systems (McVay & Kane, 2012; Unsworth & Spillers, 2010). We refer to this as *executive* capacity. Under this view, the processing speed of any one system (e.g. the ongoing task) remains intact as long as the executive system is able to keep the process engaged. However, increased load (e.g. as a result of PM) will occasionally exceed the capacity of the executive system. The result is a total lapse of attention, during which processing of all tasks is halted or drastically slowed. If PM loads executive capacity, then processing speed during PM blocks would be the same as in control PM blocks on most non-target trials, but would occasionally be much slower due to increased attentional lapses. Thus, PM demands would not be expected to have a strong effect on mean PM cost. Instead PM demands would primarily affect RT variability via longer RTs during PM-induced attentional lapses (Ihle, Ghisletta, & Kliegel, 2016). Similarly, in terms of evidence accumulation parameters, executive capacity load should affect drift rate variability, rather than mean drift rate. This view was supported by McVay and Kane (2012). They found that the extent to which working memory capacity predicted performance of a simple RT task was not strongly mediated by mean drift rate, but was mediated both by drift rate variability and subjective reports of attentional lapses. Further, Hawkins et al. (2015) demonstrated that when simulated data is mixed from two LBA models, one with a drift rate of 1 (regular trials) and one with a drift rate of 0 (lapsed trials), a single LBA model captured the mixed data by an increase in the drift rate variability parameter. In one of the previously modelled experiments, Heathcote et al. (2015) did find increased variability in drift

rates with PM. However, the drift rate variability shift predicted very little of the PM costs. In addition, no such increase was found in any other experiments of either Heathcote et al. or Horn and Bayen. Thus, there is currently no strong evidence that PM costs are caused by the PM task loading executive capacity.

It has also been argued that capacity-sharing could increase the “non-decision time” parameters of evidence accumulation models (Horn & Bayen, 2015). Non-decision time reflects the processes that occur in sequence with response selection, such as stimulus encoding and motor response production. Horn and Bayen proposed that increased non-decision time may reflect the slowing effect of capacity-sharing on encoding time, or it may reflect the time taken for sequential PM target check processes (i.e., a target check that occurs either before or after the ongoing task decision). However, encoding processes are often data limited, meaning their processing rate is determined largely by the quality of the input rather than the allocation of resources (Norman & Bobrow, 1975). Further, sequential PM target checking, strictly speaking, is not an account that involves competition for cognitive capacity or resources between the ongoing task and detection of the PM cue. Rather, it interposes additional cognitive operations that occur outside of ongoing task response selection. These putative operations presumably are directed at identifying whether the stimulus is the PM cue. Waiting for these operations to complete could slow RTs whether or not the operations draw capacity from the ongoing task. Thus, to the extent that capacity-sharing theories of costs have considered PM costs as positive evidence for capacity-sharing, they predict that PM resource demand slows ongoing task processing speed and not non-decision time. Both Horn and Bayen (2015) and Heathcote et al. (2015) found evidence that a portion of non-focal PM costs were due to an increase in the non-decision time parameter. However, non-decision time accounted for a much smaller proportion

of costs than threshold. In addition, Heathcote et al. reported increased non-decision time under PM conditions for the DDM, but not for the LBA. It was only in Heathcote et al. that both the DDM and LBA models were fit, and they found that the LBA produced better fits than the DDM, primarily through changes in threshold. This is consistent with simulation showing that the DDM mimics LBA threshold increases with increases in both threshold and non-decision time (Donkin, Brown, Heathcote, & Wagenmakers, 2011).

Although existing PM costs modelling has clearly agreed on the critical role of increased thresholds, there are differences in the nature of the threshold increases that have been observed. Heathcote et al. (2015) found evidence in the data of Lourenço et al. (2013) that ongoing task thresholds increase in proportion to the response's competition with the PM response. Lourenço et al. had participants perform a lexical decision task, and some participants were informed that the PM target syllable would only appear in word trials. As PM targets were always word trials, the word ongoing task response would be much more likely to pre-empt the PM response. There was a larger increase in threshold for word responses than for non-word responses. Heathcote et al. argued that this selective threshold increase is targeted strategically to increase the probability of PM responses on PM trials. In contrast to Heathcote et al., Horn and Bayen (2015) argued that threshold increases occur because the PM instructions make the task set appear more complex to participants. This view was supported by Horn and Bayen's fourth experiment, for which they reported an equal increase in word and non-word thresholds, despite participants being informed that PM items would always be words. The task complexity view suggests a more general strategy adjustment, which may or may not be functional to PM.

The Current Study

It is possible that past failures to find decreased drift rates under PM conditions could be attributed to the nature of the non-focal PM tasks examined, which have included detecting targets of a certain colour, first letter, or containing a particular syllable. Although capacity-sharing theories have unequivocally attributed costs to capacity-sharing under these conditions on the basis of mean RT (Einstein & McDaniel, 2005; Smith, 2010), there is no assurance that these represent the kinds of PM demands that would be most likely to drive capacity-sharing during stimulus processing. Capacity-sharing theories assume that information about two or more tasks could accrue and compete for resources under the same configuration of the cognitive system. PM cues defined by a unique colour could lead to detection prior to much evidence accumulation on the lexical decision task, consistent with the proposal that colour identification occurs at an early stage of processing (Craik, 2002). Initial letters or internal syllable strings might also not be the kinds of PM cues for which information would accumulate in parallel with the lexical evaluation. In terms of visual attention, lexical processing is optimized around the centre of the stimulus (O'regan, Lévy-Schoen, Pynte, & Brugailière, 1984; Yao-N'Dré, Castet, & Vitu, 2013), whereas first letter detection is optimized at the beginning of the stimulus (Scullin, McDaniel, Shelton, & Lee, 2010), and syllable detection might require a serial scan of the stimulus. Thus, it cannot be assumed that the previous lack of PM interference with evidence accumulation will generalize to other PM paradigms. Given the long-standing and widespread agreement that non-focal costs are positive evidence of capacity-sharing (McDaniel & Einstein, 2000; Smith, 2010) it is crucial to ascertain whether the previous lack of drift-rate-driven costs will generalize to other PM tasks in which there is a stronger *a priori* case for capacity-sharing.

With this in mind, in the current study our approach was to keep the ongoing task consistent with previous studies (lexical decision), but to extend modelling to the costs from the commonly used categorical non-focal task (e.g., Dewitt, Hicks, Ball & Knight, 2012; Loft & Humphreys, 2012; Loft & Remington, 2013; Marsh, Hicks, & Cook, 2005; Marsh, Hicks, Cook, Hansen, & Pallos, 2003; Marsh, Hicks, & Watson, 2002; Meeks, Hicks, & Marsh, 2007; Meeks & Marsh, 2010), which requires detecting words belonging to target categories, for example remembering to make a PM response to any animal word. The categorical PM task requires prolonged accumulation of evidence about the meaning of the stimulus (unlike the colour PM task), and the same visual analysis as lexical decision (unlike the first letter or the syllable task). As lexical and semantic processing are likely to occur in parallel, the PM costs induced by the task may be a result of capacity-sharing. This proposition is supported by Marsh et al. (2005), who reported that categorical PM accuracy was reduced when participants were instructed to increase their lexical decision effort, but that PM accuracy was not affected for the more visual PM task of detecting palindromes. This evidence for more capacity-sharing between categorical PM and lexical decision was present despite there being a larger mean PM cost for the palindrome task. Thus, our categorical PM condition may be more conducive to mean drift rate decreases than the tasks previously modelled even if it produces smaller mean RT costs.

Similarly, the categorical PM task may be more conducive to changes in drift rate variability than the previously modelled tasks. This would map to attentional lapses induced by the PM task loading executive capacity. The categorical information required to detect PM is presumably obtained at later stage of processing than perceptual information, such as stimulus colour. It is possible that conflicts at this later stage between the concurrent management of lexical and

categorical processing may be particularly taxing for the executive system, leading to more lapses of attention and increased drift rate variability.

Our non-focal PM targets were always words, and this feature of the PM task was highlighted in the PM instructions. This allowed us to compare the delay theory view of threshold shifts as strategic waiting for PM retrieval (Heathcote et al., 2015), with the view of threshold shifts as a reaction to perceived task complexity (Horn & Bayen, 2015). If the threshold shift is implemented in order to ensure that the PM response is retrieved on PM trials, then participants will become more biased against responding to a word under PM conditions (i.e., the word threshold will increase relative to the non-word threshold). In contrast, if the threshold shift is not implemented specifically to extend time on PM trials for PM retrieval, then threshold shifts may be equal for word and non-word responses.

We also modelled costs from a focal PM demand to make a PM response to a single target word. Focal PM costs can often reach statistical significance (Loft & Remington, 2013), but they tend to be much smaller in magnitude than costs caused by non-focal PM demands (Meeks & Marsh, 2010). As with non-focal costs, capacity-sharing theories of PM all propose that, when focal PM costs are observed, they occur because a proportion of an individual's available capacity (whether actually required for PM retrieval or not) is allocated away from the ongoing task and towards the PM task (McDaniel, Umanath, Einstein, & Waldum, 2015; Smith, Hunt, McVay, & McConnell, 2007). Delay theory can account for focal PM costs without capacity-sharing. If thresholds can account for the majority of the larger non-focal PM costs – where the PM retrieval is presumably more likely to interfere with evidence accumulation – and not account for the majority of the smaller focal PM costs – where the PM interference with

evidence accumulation should be minimal – that would be inconsistent with delay theory, and would bring into question conclusions drawn from the non-focal modelling.

In addition, comparison of the focal and non-focal conditions, where two identical PM response modes are employed, can reveal the effects of increasing the actual and perceived demands of PM target detection. For example, if participants are sensitive to the fact that focal PM can be supported by lexical decision processing, the focal PM task may not induce the same level of increased perceived task complexity as the non-focal PM task. If this is the case, a global threshold shift would be less likely in the focal than in the non-focal condition. Horn and Bayen (2015)'s modelling went against this proposition, as they found that focal PM costs were accounted for with a global increase in DDM thresholds. However, their focal PM task required subjects to make a PM response to any one of three PM target words. Our focal PM task required participants to remember to make a response to one single target word, a manipulation more in line with the focal PM demand traditionally imposed by researchers when using lexical decision tasks (see Einstein & McDaniel, 2010). In contrast to a one-item PM target list, which individuals can likely easily hold in focal attention, a memory retrieval operation may well be required to access the multiple-target list in order to compare it against each presented letter string (Loft, Humphreys, & Whitney, 2008; Öztekin, Davachi, & McElree, 2010). It is, therefore, possible that the retrospective memory requirement in the Horn and Bayen study increased perceived task complexity, resulting in the global threshold change.

The current study also permits us to examine the reproducibility and generality of non-decision time effects. As noted by Horn and Bayen (2015), the degree of target checking may depend on task demands. The tasks which have produced non-decision time effects were a first letter task, in which the target letter was guaranteed to be presented in a specific visual location

(the start of a word), and a syllable detection task. Both may require a visual analysis separate from the lexical decision, and thus arguably would encourage non-decision time effects. In the current PM task, where parallel accumulation between PM and ongoing responses is more plausible, non-decision effects should be less likely. Non-decision effects have to date only been found in DDM, which has provided a worse fit than the LBA to PM costs data, suggesting they may be a result of model mimicry. Thus, if we do find increased non-decision time, it will be crucial to determine whether the result is obtained from the best fitting model architecture.

Method

Participants

Thirty-five participants were tested (21 females), with ages ranging from 17-79¹ ($M = 25.77$). Participants performed three one-hour sessions on separate days. English as a first language was required to participate. Participants received either \$45 AUD or credit towards their first year psychology unit. The study was approved by the University of Western Australia's Human Research Ethics Office.

Materials

Two thousand one hundred and eighty eight low frequency (occurring 2-6 times per million) English words (of length between 4 and 10 characters) were selected from the *Sydney Morning Herald* word database (Dennis, 1995), plus 62 extra words from the MRC 2.00 database (Wilson, 1988) for a total of 2250 words. A non-word was created from each word by replacing every vowel with a random alternate vowel (e.g., *chemist* to *chamust*). In addition, 34 low frequency

¹ In response to a reviewer who noted our wide age range, we re-ran all analyses in the paper with participants older than 40 (of which there were 6) excluded. The patterns of significance in the conventional results were almost identical, as were the models selected and the patterns of significance in the parameter ANOVAs. The inclusion vs exclusion of these participants did not influence the interpretation of the results.

words from each of three categories (animal, food, human body part) were selected to be PM targets.

Participants performed 9 blocks of 500 trials: three non-focal, three focal, and three control blocks. Participants performed 1 block of each type on each day. Block order was balanced across days so that participants would not get a condition in the same position twice. For example, if on day 1 a participant had the order focal/non-focal/control then focal would not be presented first, non-focal not presented second and control not presented third on day 2 or 3. There are 12 day orders that satisfy these conditions (e.g., day 1 – focal/non-focal/control, day 2- non-focal/control/focal, day 3- control/focal/non-focal), and these orders were approximately counterbalanced over the final 35 participants.

Under non-focal conditions, participants were presented with 250 non-words, 217 non-target words and 33 target words from one of the target categories (e.g., 33 different animal words). Under control conditions, participants were presented with 250 non-words and 250 words. Under focal conditions, participants were presented with 250 non-words, 217 non-target words and 1 target word (e.g., ‘giraffe’) was presented 33 times. Each target category was in one non-focal block for each participant. For each participant, one focal target word was drawn randomly from each of the target categories (without replacement; each word was only used as a focal target for one participant), and if a word was to be presented as a target in the focal block then it was not presented in the non-focal blocks for that participant. The assignment of target category to each day’s non-focal block, and of which target category the word from the focal block was drawn, was random (without replacement) except that the focal word for a given day was never from the non-focal PM category for that day. The order of non-target stimuli was randomized. Under PM conditions, one target was presented in a random position once every 15 trials, starting from

trials 6-20. Thus, 33 trials of the 500 trials in each block were PM targets. Target trials were separated by at least 2 non-targets. To balance against any general effect of word repetition in the focal condition, one word in the control condition was repeated 33 times according to PM target position rules for approximately half the participants. The order of the PM stimuli assigned to the PM target positions in the non-focal blocks was random.

Procedure

For the lexical decision task, participants were instructed that they would be presented with letter strings and to press a key to indicate whether strings were words or non-words as quickly and as accurately as possible (the assignment of the keys ‘f’ and ‘j’ to word and non-word responses was approximately counterbalanced). Under PM conditions, participants were also instructed to press the ‘9’ key instead of the word response when they were presented targets (a specific target word for focal PM, or a member of a specific category for non-focal PM). Under control conditions, participants were instructed that they only needed to make lexical decision responses for that block. Each trial began with a fixation cross ‘+’, displayed in white on a black background for 0.5s. The fixation cross was then replaced by a blank screen for 0.25s, which was followed by a white letter string that remained on the black screen until the participant responded. Each day, participants first completed 20 practice lexical decision trials. Participants were then presented with either the focal, non-focal or control instructions. Participants then completed a three-minute distractor puzzle, after which they began their first block of trials. In between blocks participants rested for two minutes. After each block had been completed, participants were instructed to disregard the instructions from the previous block.

Results

In addition to stimulus type (word, non-word) and condition (focal, non-focal, control) the analyses included a day factor (day 1, day 2, day 3) to capture the potential effects of task repetition. The first two trials of each block were excluded, as were trials where RTs were very fast ($<0.2s$) or very slow ($>$ mean RT plus 3 times the interquartile range / 1.349, which is a robust equivalent to a mean +3 SD cutoff). This resulted in the exclusion of 5.22% of non-target trials and 3.66% of PM trials. One of the 35 participants was excluded from all analyses because they completely neglected non-word responding under PM conditions (no non-word responses over 3000 trials). We also excluded a total of eight blocks (that came from four participants) because accuracy was near chance ($<60\%$), suggesting a guessing strategy. Another block (from one of the same four participants) was excluded because the participant made a PM response to 20/32 repeats of a stimulus in a control block, suggesting they mistakenly deduced that it was a focal PM block. The exclusion of these 9 blocks of trials resulted in the removal of 2.2% of the remaining data.

We had to exclude the four participants with missing blocks from LBA and DDM fitting, which require a balanced data structure. Overall then, the LBA and DDM was fit to 30 participants out of the 35 participants tested, resulting in a 14% participant exclusion rate. This exclusion rate is higher than typical, and likely driven by the unusually large number of trials participants had to complete over three days of testing. The supplementary materials contain conventional analysis of the 30 participant data set to which we fit the DDM and LBA, which produces a similar pattern of results to analysis of the full data set.

In text we report the results of analysis on the full data set ($N = 34$), which we achieved by using random-effects models that are robust to the presence of missing cells in the data structure.

Response accuracies (both PM error rates and lexical decision error rates) were analysed using a generalized linear model with a binomial probit link function. For PM analysis, the dependent variable for the probit model was the presence versus absence of a PM response on each PM trial. For the ongoing task analysis, the dependent variable for the probit model was lexical decision accuracy. We analysed RTs using a general linear model with the dependent variable mean correct RT. The random-effects model results are tabulated in the supplementary materials, and the pattern of significant effects are reported in the body of the text below. Within-subject standard errors, both in text and in the graphs, were calculated using the Morey (2008) bias corrected method.

Prospective Memory Task

PM responses were scored as correct if the participant pressed the PM response key instead of a lexical decision response key on the target trial. PM accuracy was higher in the focal condition ($M = 80\%$, $SE = 3\%$), than the non-focal condition ($M = 58\%$, $SE = 3\%$), and decreased over days (day 1 $M = 77\%$, $SE = 3\%$; day 2 $M = 70\%$, $SE = 3\%$; day 3 $M = 60\%$, $SE = 4\%$). Correct PM responses were faster in the focal condition ($M = 0.795s$, $SE = 0.021s$) than the non-focal condition ($M = 0.916s$, $SE = 0.024s$). The PM false alarm rate was 0.17% (ranging from 0 to 0.5% across participants).

Lexical Decision Task

PM target trials, false alarms and the two lexical decision trials following each target trial or false alarm were excluded. Trials in which participants were presented repeated non-targets in the control condition and the two trials following were also excluded. Lexical decision accuracy was higher for non-words (93.3%) than for words (91.7%), and decreased over days (day 1 $M = 93.9\%$, $SE = 0.9\%$; day 2 $M = 92.2\%$, $SE = 1.1\%$; day 3 $M = 91.3\%$, $SE = 1.1\%$). There was an

interaction between PM condition and stimulus type (see Figure 1). Planned comparisons suggested there was not a difference between non-focal non-word accuracy (93.4%) compared to control non-word accuracy (92.5%), $t(33) = 1.46$, $p = 0.15$, but there was a difference for focal non-word accuracy (93.8%) compared to control non-word accuracy, $t(33) = 3.07$, $p < 0.01$, $d = 0.53$. Non-focal word accuracy (91.8%) was not significantly different to control word accuracy (91.9%), $t(33) = 0.73$, $p = 0.47$, but there was a difference between focal word accuracy (91.3%) and control word accuracy, $t(33) = 2.18$, $p < 0.05$, $d = 0.37$.

Lexical decision response times decreased over days (day 1 $M = 0.695$ s, $SE = 0.015$ s; day 2 $M = 0.645$ s, $SE = 0.010$ s; day 3 $M = 0.625$ s, $SE = 0.010$ s), and this effect interacted with stimulus type. There was an effect of PM condition (non-focal $M = 0.672$ s, focal $M = 0.651$ s, control $M = 0.644$ s), which interacted with the effect of stimulus type (Figure 2). Word RTs were slower in non-focal blocks (0.682s) compared to control blocks (0.633s), $t(33) = 5.15$, $p < 0.001$, $d = 1.24$, and in focal blocks (0.653s) compared to control blocks, $t(33) = 3.24$, $p < 0.01$, $d = 0.56$. Non-word RTs were not significantly different between non-focal blocks (0.662s) and control blocks (0.654s), $t(33) = 1.14$, $p = 0.26$, or between focal blocks (0.649s) and control blocks, $t < 1$.

LBA Analysis

The LBA provided a substantially better account of the data than the DDM, both in terms of the AIC and BIC selected models. In fact, even a very simple LBA – one that accounts for all the effects of practice and PM condition in threshold (see Table 1, 16 parameters per participant) – had a lower deviance than the top DDM model (48 parameters per participant). Qualitatively, the DDM consistently under-predicted error rates to word stimuli, even for the most flexible DDM. Thus, we focus on inference from the LBA fits. The details of the DDM analysis can be found in

supplementary materials, and the implications of the DDM analysis are addressed in the discussion section.

The LBA proposes that evidence accumulates separately for each response over time, and the response made is the first to accrue evidence that reaches its threshold (Figure 2). Each accumulator begins a decision trial with a starting amount of evidence drawn from the uniform distribution $[0, A]$. Evidence increases at a speed given by the drift rate, which is drawn from a normal distribution with mean ν and standard deviation $s\nu$, until it reaches a threshold b . The first accumulator to reach the threshold decides the response, and the time taken to make the response is decided by the time taken to reach the threshold plus a non-decision time constant $t0$. The drift rate parameters ν and $s\nu$ reflect processing efficiency and capacity, as well as stimulus features, with higher rates indicating more efficient processing of response-relevant stimulus features. Better task performance (increased speed and accuracy) can be achieved with high drift rates towards the correct response and lower drift rates towards the error response. Threshold reflects the caution to make a response, with a higher threshold corresponding to a shift in bias against making a response, which is expressed by decreased response frequency and slower responses. Threshold increases for all responses yields increases in both RTs and accuracy. We report threshold in terms of B ($B = b - A$). Non-decision time ($t0$) reflects the time taken for processes such as the encoding of stimulus and response production, and thus non-decision time increases will increase RTs with no effects on response frequency.

In addition to the day factor, the PM condition factor, and the stimulus type factor reported in our conventional analyses, the LBA results described below include a response factor and correspondence factor. The response factor allows parameters to vary by the specific accumulator (e.g., participants may be biased against word responding, with a higher B for the

word accumulator), and thus can be either ‘word’ or ‘non-word’. The correspondence factor varies as a function of whether the response accumulator is ‘correct’ for a stimulus (e.g., to capture better than chance accuracy, v is higher for correct responses than incorrect responses). We sequentially fit different parameterizations of the LBA as described in Donkin, Brown and Heathcote (2011), building to a most complex “top” model, that we allowed to be highly flexible but included some reasonable restrictions for the sake of reducing computational effort. As B is assumed to be set prior to onset of the stimulus, in the LBA top model it is fixed over stimulus type but can vary over the response factor. In order to test the delay theory, B was also free to vary over condition, and in order to capture practice effects it was free to vary over days. Following previous applications of the LBA, including Heathcote et al. (2015), we only estimated one value for start point noise parameter A (i.e., it did not vary over condition, day, or response).

In order to test capacity-sharing theories, the top model allowed the mean drift rate parameter v to vary over stimulus type, condition, day and correspondence. The sv parameter could vary over the correspondence factor (correct vs incorrect response). It could also vary across PM conditions, in order to test for PM-induced attention lapses (i.e., PM load on executive capacity). The sv for the false accumulator in the control condition was fixed at 1, as a scaling parameter. The sv parameter was fixed over the day factor. Non-decision time $t0$ was free to vary over condition, stimulus and day, but fixed between responses, as is conventional when fitting the LBA. With this top model, our sequential model fitting resulted in 11,265 model fits per subject, for a total of 337,950 fits.

Model parameter values were estimated using maximum likelihood estimation (Myung, 2003). The summed log likelihood of all trials was maximised. The deviance, our measure of

quality of fit, is equal to two times the negative maximised log likelihood. Table 1 shows our model selection results. For model selection we used the AIC and BIC criteria (Myung & Pitt, 1997), which attempt to select models that provide a good trade-off between minimizing the number of parameters (p) and minimizing the deviance². AIC was calculated for each individual subject by applying the penalty term $2p$ to the deviance. The group AIC was calculated by summing together the participant AICs. The group BIC was calculated by summing the deviances across participants, then applying the penalty term $P \times \log(N)$ to the summed deviance (where N and P are the total number of observations and parameters summed over participants).

The AIC selected model retained full flexibility for B , but dropped the PM condition factor for both v and sv , and dropped both the PM condition and stimulus factors for $t0$. The BIC selected model dropped the response factor for B , and both the condition and day factors for v and $t0$. In contrast to the AIC model, the stimulus factor was retained for $t0$. The AIC model did not fit significantly worse than the top model, $\chi^2(1290) = 486, p > 0.999$, but the BIC model did, $\chi^2(1830) = 9477, p < 0.001$. Thus, the AIC model results are reported in text and details of the BIC model results are reported in supplementary materials. The AIC-selected LBA predictions are plotted in Figure 3. For our parameter analyses we report the results obtained from the AIC-selected LBA model.

AIC LBA Parameter Analysis

For estimates of B (Figure 4), a response by PM condition by day ANOVA indicated a main effect of response, $F(1, 29) = 4.62, p < .05, \eta_p^2 = .14$, with higher B for word responding (1.076) than non-word responding (1.034). This suggests a bias against responding ‘word’. There was

² We also calculated model weights (Wagenmakers & Farrell, 2004) for all of our selected models, which revealed conditional probabilities for near one for the selected models, and conditional probability near 0 for the top model.

also a main effect of PM condition, $F(2, 58) = 34.39$, $\varepsilon = .87$, $p < .001$, $\eta_p^2 = .54$ (control $B = 1.018$, focal $B = 1.047$, non-focal $B = 1.101$), and B decreased over days (day 1 = 1.174, day 2 = 1.020, day 3 = 0.972), $F(2, 58) = 17.32$, $\varepsilon = .96$, $p < .001$, $\eta_p^2 = .37$. The effect of response interacted with condition, $F(2, 58) = 29.62$, $\varepsilon = .85$, $p < .001$, $\eta_p^2 = .51$. Planned contrasts revealed differences in B estimates for the word accumulator under focal conditions (1.184) compared to control conditions (1.108), $t(29) = 4.79$, $p < .001$, $d = 0.87$, but not for the non-word accumulator (focal $B = 1.153$, control $B = 1.139$), $t < 1$. In the non-focal condition, estimates of B were higher than control for both words (1.251), $t(29) = 8.60$, $p < .001$, $d = 1.57$, and non-words (1.207), $t(29) = 3.77$, $p < .001$, $d = 0.69$. The increase in B under non-focal conditions compared to control was larger for words than for non-words, $t(29) = 6.59$, $p < .001$, $d = 1.20$. These contrasts suggest a selective threshold effect in both PM conditions; that is, bias against responding word increased compared to control. The selective threshold effect was larger in the non-focal condition; that is, the shift in bias against responding word (as compared to control) was significantly higher under non-focal conditions than under focal conditions, $t(29) = 3.50$, $p < .01$, $d = 0.64$. There was also an interaction between response type and day, $F(2, 58) = 9.25$, $\varepsilon = .83$, $p < .001$, $\eta_p^2 = .24$, with B decreasing more over days for non-words than for words, (day 1 non-word $M = 1.17$, $SE = 0.032$; day 1 word $M = 1.181$, $SE = 0.031$; day 2 non-word $M = 0.995$, $SE = 0.027$; day 2 word $M = 1.045$, $SE = 0.031$; day 3 non-word $M = 0.940$, $SE = 0.026$; day 3 word $M = 1.003$, $SE = 0.024$). This is consistent with participants learning that they can lower their non-word threshold without increasing PM errors (i.e. learning over days to be increasingly biased against word responding). However, there was no 3-way interaction between this effect and PM condition, $F(4, 116) = 1.59$, $p = .18$. There was also no interaction between day and condition ($F < 1$).

The model selection indicated no effect of PM condition on either mean drift rate, v , or variability in drift rates, sv . For v , a stimulus type by day by correspondence ANOVA indicated a main effect of stimulus type, $F(1, 29) = 6.56, p < .05, \eta_p^2 = .18$, with word rates ($M = 0.806, SE = 0.344$) higher than non-word rates ($M = 0.741, SE = 0.351$). As expected with better than chance accuracy, there was an effect of correspondence, $F(1, 29) = 182.79, p < .001, \eta_p^2 = .86$, with correct rates ($M = 2.010, SE = 0.130$) higher than error rates ($-0.463, SE = 0.140$). For sv , there was a significant effect of correspondence, with variability in rates lower for correct responses ($M = 0.445, SE = 0.018$) than for incorrect responses (which were fixed at 1), $F(1, 29) = 481.55, p < .001, \eta_p^2 = .94$.

The model selection indicated no effect of PM demands, or stimulus type, affecting non-decision time, $t0$. There was an effect of day, $F(2, 58) = 3.60, \varepsilon = .97, p < .05, \eta_p^2 = .11$, with $t0$ increasing on later days (day 1 $M = 0.069, SE = 0.007$; day 2 $M = 0.087, SE = 0.006$; day 3 $M = 0.092, SE = 0.006$).

In order to assess the individual differences in our observed effects, we plotted the top model parameters that can directly contribute to the magnitude of PM costs (non-decision time, mean drift rate, variability in drift rate and response threshold) on a per-subject basis (see supplementary materials). Corroborating our model selection, there was a remarkably consistent picture of null mean drift rate effects, with minimal non-decision time and drift rate variability effects. In contrast, PM threshold effects were common, particularly for the word threshold. For example, for a little less than half the participants the non-focal threshold increases were visibly larger for word thresholds than for non-word thresholds, whereas for no participants were the non-word threshold increases appreciably larger than word threshold increases.

Discussion

The aim of the current study was to investigate the latent variables that underlie costs induced by categorical non-focal PM task demands and by focal PM task demands. Costs were observed under non-focal conditions (mean RTs were 0.04s larger than control), and were present but smaller under focal PM conditions compared to control conditions (0.01s). Costs were larger for correct word responses (non-focal = 0.058s, focal = 0.024s), which were the greatest competitor with the PM response, than for correct non-word responses (non-focal = 0.022s, focal = -0.002s).

The modelling results extend Heathcote et al. (2015) and Horn and Bayen (2015) by unambiguously finding no effect of PM conditions on mean drift rates. This applied for both our single target focal task and our categorical non-focal task. To date then, there has been a failure to find drift rate driven PM costs in the modelling of eight data sets, which used four conceptually different non-focal PM tasks, and two focal PM tasks. Given the accruing consistency of the null mean drift rate effects, arguments for capacity-sharing would need to disavow any link between mean drift rates and capacity. However, this goes against the almost universally accepted link between processing speed and capacity in theories of attention (Bundesen, 1990; Gobell, Tseng, & Sperling, 2004; Kahneman, 1973; Navon & Gopher, 1979; Wickens, 1980), as well as empirical evidence supporting that link (Donkin et al., 2014; Eidels et al., 2010; Sewell et al., 2016). Moreover, if the effects of capacity-sharing were somehow split between threshold changes and drift rate, then it is likely at least one of the eight modelled studies would have found some evidence for a mean drift rate effect, but none have. It is not clear what other parameters the claims of the capacity-sharing theories could correspond to. They certainly do not predict that costs would be entirely threshold-driven, as this parameter reflects

response strategy and preference rather than resource allocation. Although non-decision time could be extended by a capacity demanding process, a non-decision time prediction is inconsistent with the proposal that costs are positive evidence for capacity-sharing, as there is no way to identify whether the processes which extend non-decision time are capacity consuming or not. Thus, the modelling evidence to date strongly ties PM costs in ongoing task RTs to increases in response thresholds underlying more cautious responding, and fails to provide any support for capacity-sharing during processing.

There are two points worth noting about what the null drift rate finding does *not* imply. First, the current findings do not rule out the existence of PM tasks that do require capacity sharing between ongoing tasks and PM tasks. The costs effects have only been modelled in paradigms in which the information required for PM retrieval is simple and available in the same general spatial location as the information required for ongoing task decisions. The previously modelled tasks include the canonical laboratory demonstrations of PM capacity sharing, but it may be that capacity sharing only occurs when there is a need to sample spatially separated information sources. One example would be an air traffic control task in which the PM task requires responding to an aircraft ‘callsign’ property that has no relevance to aircraft conflict detection, and is in a completely separate visual location from the conflict detection information, where it would be unlikely to be processed at all without the PM requirement (Loft, 2014). It seems likely in this case that switching attention between the stimuli for the two distinct tasks could affect the rate of evidence accumulation for the ongoing conflict detection task.

The second point is that the current findings do not imply that the PM task requires no cognitive capacity at all. Our model is limited to explaining PM costs, and hence can only provide insight into PM processes to the extent that changes in ongoing task response decision

processes in response to PM task demands can be mapped to PM response decision processes. PM task may require cognitive capacity that is not shared with the ongoing task. Testing this would require a model that includes a process account of the PM task decision. The PM process could, for example, be modelled in an LBA computational architecture as a third, additional, evidence accumulator, that occurs in parallel with the two ongoing task accumulators.

The null effect on ongoing task drift rates extended not just to mean drift rates, but also to drift rate variability. Thus, there was a lack of evidence for PM induced attentional lapses. To date then, model-based evidence for PM induced attentional lapsing has only been found in one experiment out of eight (Heathcote et al., 2015; Horn & Bayen, 2015). This suggests that, in the modelled ongoing tasks, the burden of PM on executive capacity has been fairly limited. This is ostensibly inconsistent with previous studies indicating that taxing the executive control system can produce PM decrements (e.g., Marsh & Hicks, 1998), and studies showing that working memory is correlated with PM accuracy (e.g. Smith & Bayen, 2005). We can think of two reasons for this apparent contradiction. Firstly, although it is often assumed that capacity-sharing processes are common to both PM target trials and non-target trials, they may not be. For example, executive capacity may only be required to perform the PM retrieval process when triggered by the PM target, or to coordinate the PM response procedure. To date we do not know of a study that can disentangle the executive requirements of non-target trials from the executive requirements of PM target trials. Our analysis included only PM costs to non-target trials, whereas previous analyses of working memory and PM accuracy did not include a parallel analysis of PM costs. Secondly, it is possible that executive capacity affects the PM related processes that occur on non-target trials without being significantly strained by those processes. For example, executive capacity may be used to adjust thresholds to meet PM task demands. In

this case, individuals with lower executive capacity (e.g. lower working memory span, imposed working memory load) may fail to increase their thresholds in PM blocks, or fail to increase their thresholds selectively when PM instructions are stimulus-specific. If executive capacity tends only to be engaged to modify thresholds when executive capacity is available, this could explain why PM tasks appear not to induce attentional lapses, which would require overloading the executive system.

We found that non-focal PM costs were driven by increased response thresholds. This supports the delay theory of PM costs, which proposes that PM costs reflect a strategy of waiting for potential PM retrieval. The best fitting model (LBA) suggested that both the word and non-word thresholds increased. This is consistent with the proposition by Horn and Bayen (2015) that PM instructions can cause a more generally cautious approach due to an increase in perceived task complexity. In contrast to the predictions of Horn and Bayen's complexity account, we also observed that our non-focal threshold increases were higher for the word response. This selective shift suggests that threshold increases may be targeted particularly at modifying the response selection on PM trials. As word accumulation is the most competitive with PM response selection (because PM items are also words), this threshold increase would be responsible for the majority of the delay on PM trials. However, note that the BIC selected LBA did not include a selective shift for word responding (for either PM condition). We reported the AIC model, because the model was no worse in fit than the top model, whereas the BIC model was statistically inferior in fit to the top model. One weakness of the AIC model is that it over-predicted error rates to word trials in non-focal blocks, whereas the BIC model did not. This could have occurred because the AIC model included a stronger bias against word responding than was actually present in the data.

Focal PM costs were also entirely threshold driven, supporting the delay theory, but they differed in two regards from the non-focal PM costs. First, in the focal blocks, the word threshold increased but the non-word threshold did not. This suggests that focal PM instructions did not increase the overall perception of task complexity. This result contrasts with the previous model of focal PM costs presented by Horn and Bayen (2015), in which both thresholds increased. However, their focal PM task was more complex than ours. Our task used a single-target focal PM requirement, whereas Horn and Bayen required participants to respond to any word from a previously encoded three-item list. Perhaps the higher retrospective memory load of the Horn and Bayen focal PM task caused an increase in perceived task complexity.

A higher retrospective memory load might also have contributed to the larger caution increase in our non-focal condition compared to the focal condition. In the single target focal task each PM cue is identical to the target and reinforces its representation in memory directly. This is not the case for our non-focal cues where there is an indirect activation of the category node by the cue. Note that both increases in threshold due to retrospective memory load and increases due to task complexity are not *a priori* features of delay theory. Thus, it seems that, although the selective delay of ongoing task responses to enable PM responses does occur, it occurs in the context of a wider range of PM task-related processes.

The second difference between the focal model and non-focal models is that the word threshold increases were smaller overall in the focal case. Assuming that focal PM retrieval is faster than non-focal retrieval, this would follow from the delay theory. A faster focal PM retrieval means that less response delay is required for the PM response to out-race the word response. This faster PM retrieval may owe to the activation of spontaneous retrieval processes, as specified in the Multiprocess view (McDaniel & Einstein, 2000).

To the extent that threshold increases were not selective to word thresholds, they cannot be attributed purely to delay theory. Although group model selection favoured a selective threshold increase in both non-focal and focal conditions, an examination of individual differences (see supplementary materials) suggests that the pattern did not hold for all participants. About half of our participants disproportionately increased their word threshold in non-focal conditions, but the remainder raised both thresholds equally. This heterogeneity in the extent of selective versus general threshold increases may have driven the aforementioned miss-fit of the AIC model to non-focal word accuracy. Perhaps the benefit of the selective delay strategy is not immediately obvious to participants, either because the stimulus-specific nature of the PM task was not salient at encoding time, or because consideration was not given to the underlying dynamics of decision-making which would occur on PM trials. Such a difference might explain why some studies find stimulus-specific instructions cause selective costs and other studies find costs to both word and non-word trial types (e.g., Cohen, Jaudas, Hirschhorn, Sobin, & Gollwitzer, 2012; Horn & Bayen, 2015).

If the selective delay strategy is indeed optimal, then extended experience with the PM task might encourage it. Consistent with this possibility, we observed that the bias against word responses increased over the days of the experiment. Although the same trend was observed in the control condition, where the word response does not compete with a PM response, this could be caused by within-subjects carryover effects (Poulton, 1982) from the PM blocks. As there were two PM blocks (focal and non-focal) for each control block, the PM delay strategy might persist into control blocks.

There were no changes in non-decision time across conditions with the LBA. However, the DDM accounted for a sizeable portion of costs with increased non-decision time: non-

decision time was equal in size to the entirety of focal PM costs (0.010s), and about half of non-focal PM costs (0.022s). One interpretation of this effect is that it reflects capacity-sharing between a PM process and stimulus encoding and/or response production processing. Another interpretation of the effect is that it reflects a sequential target check. That is, detection of the PM cue could require capacity-demanding operations that occur after response selection on the lexical decision task but prior to response execution. Consequently, if the non-decision time account of our costs is accepted then our conclusions are confined to the issue of shared capacity during processing. There is also the strong possibility that the non-decision time increases are an artefact of the particular assumptions of the DDM. This interpretation is supported by the superior fit of the LBA, and previous findings that in simulations LBA threshold increases can cause increases in both threshold and non-decision time in the DDM (Donkin, Brown, Heathcote, et al., 2011).

During the review process some concerns were raised over the fact that our PM instruction told participants to make the PM response instead of the word response. This instruction is similar to the more common instruction used in the PM literature for participants to make the PM response when the target is presented (e.g. Einstein & McDaniel, 2005; Scullin et al., 2010; Smith & Bayen, 2004). An alternative type of instruction is to explicitly tell the participant to make the ongoing task response before the PM response (Hicks, Marsh, & Cook, 2005; Loft et al., 2008; Marsh et al., 2005). The latter instruction might allow participants to remember to make their PM response after they make their ongoing task decision, rendering threshold increases less useful for PM. However, Heathcote et al. (2015) found a similar pattern of results to the current study when they modelled a task with the instruction to make the ongoing task response before the PM response: selective threshold increases accounted for PM

cost. These threshold increases may reduce PM errors because terminating stimulus processing to make an ongoing task response interrupts processing of the stimulus features required to detect the PM target (after the ongoing task response is made, the stimulus is removed from the display).

In summary, PM costs induced by both a single-target focal PM task and a categorical non-focal task were driven by changes in response thresholds, rather than mean drift rates, supporting the delay theory of PM costs over capacity-sharing theories. Both the word and non-word thresholds increased in our non-focal PM condition, but the increase in the word threshold was larger. This selective threshold increase in favour of PM responding supports the delay theory, which states that thresholds increase in order to allow time for the detection of the PM target. However, threshold increases were not entirely selective to the word accumulator for the non-focal PM task, and this suggests the involvement of additional mechanisms that are not included in delay theory. It cannot be ascertained from the current data what those mechanisms are, but possibilities include increases in the perceived task complexity of the PM instructions, and increases retrospective memory load. In contrast, our AIC selected model suggests that under focal conditions, selective word threshold increases accounted for the entirety of PM costs. Finally, we found no evidence of a change in drift rate variability that would signify that the PM demand negatively impacted executive capacity.

References

- Boywitt, C. D., & Rummel, J. (2012). A diffusion model analysis of task interference effects in prospective memory. *Memory & Cognition*, 40(1), 70–82.
<https://doi.org/10.3758/s13421-011-0128-6>
- Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, 97(4), 523.
<https://doi.org/10.1037/0033-295x.97.4.523>
- Cohen, A. L., Jaudas, A., Hirschhorn, E., Sobin, Y., & Gollwitzer, P. M. (2012). The specificity of prospective memory costs. *Memory*, 20(8), 848–864.
<https://doi.org/10.1080/09658211.2012.710637>
- Craik, F. I. (2002). Levels of processing: Past, present... and future? *Memory*, 10(5–6), 305–318.
<https://doi.org/10.1080/09658210244000135>
- Dennis, S. (1995). The sydney morning herald word database (Vol. 1). Presented at the Noetica: Open Forum. Retrieved from <http://psy.uq.edu.au/CogPsych/Noetica>
- Dewitt, M. R., Hicks, J. L., Ball, B. H., & Knight, J. B. (2012). Encountering items previously paired with prospective memory target events can serve to reactivate intentions. *Journal of Cognitive Psychology*, 24(8), 981–990. <https://doi.org/10.1080/20445911.2012.727389>
- Donkin, C., Brown, S. D., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the linear ballistic accumulator. *Journal of Mathematical Psychology*, 55(2), 140–151. <https://doi.org/10.1016/j.jmp.2010.10.001>
- Donkin, C., Brown, S. D., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review*, 18(1), 61–69.

- Donkin, C., Little, D. R., & Houpt, J. W. (2014). Assessing the speed-accuracy trade-off effect on the capacity of information processing. *Journal of Experimental Psychology: Human Perception and Performance*, 40(3), 1183.
<https://doi.org/https://doi.org/10.1037/a0035947>
- Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychonomic Bulletin & Review*, 17(6), 763–771.
<https://doi.org/10.3758/PBR.17.6.763>
- Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 717–726.
<https://doi.org/10.1037/0278-7393.16.4.717>
- Einstein, G. O., & McDaniel, M. A. (2005). Prospective memory multiple retrieval processes. *Current Directions in Psychological Science*, 14(6), 286–290.
<https://doi.org/10.1111/j.0963-7214.2005.00382.x>
- Einstein, G. O., & McDaniel, M. A. (2010). Prospective memory and what costs do not reveal about retrieval processes: A commentary on Smith, Hunt, McVay, and McConnell (2007). <https://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. <https://doi.org/10.1037/0096-3445.134.3.327>
- Gobell, J. L., Tseng, C., & Sperling, G. (2004). The spatial distribution of visual attention. *Vision Research*, 44(12), 1273–1296. <https://doi.org/10.1016/j.visres.2004.01.012>

- Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological Review*, 122(2), 376–410.
<https://doi.org/10.1037/a0038952>
- Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. *Journal of Memory and Language*, 53, 430–444. [https://doi.org/https://doi.org/10.1016/j.jml.2005.04.001](https://doi.org/10.1016/j.jml.2005.04.001)
- 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(1), 95–117. <https://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/Revue Canadienne de Psychologie Expérimentale*, 65, 69–75. <https://doi.org/10.1037/a0022808>
- Ihle, A., Ghisletta, P., & Kliegel, M. (2016). Prospective memory and intraindividual variability in ongoing task response times in an adult lifespan sample: the role of cue focality. *Memory*, 1–7. [https://doi.org/https://doi.org/10.1080/09658211.2016.1173705](https://doi.org/10.1080/09658211.2016.1173705)
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Loft, S. (2014). Applying Psychological Science to Examine Prospective Memory in Simulated Air Traffic Control. *Current Directions in Psychological Science*, 23(5), 326–331.
<https://doi.org/10.1177/0963721414545214>
- Loft, S., & Humphreys, M. S. (2012). Enhanced recognition of words previously presented in a task with nonfocal prospective memory requirements. *Psychonomic Bulletin & Review*, 19(6), 1142–1147. <https://doi.org/10.3758/s13423-012-0303-1>

- Loft, S., Humphreys, M. S., & Whitney, S. J. (2008). Control of access to memory: The use of task interference as a behavioral probe. *Journal of Memory and Language*, 58(2), 465–479. <https://doi.org/10.1016/j.jml.2007.04.002>
- 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*, 66(7), 1432–1447. <https://doi.org/10.1080/17470218.2012.750677>
- Logan, G. D., Van Zandt, T., Verbruggen, F., & Wagenmakers, E.-J. (2014). On the ability to inhibit thought and action: General and special theories of an act of control. *Psychological Review*, 121(1), 66.
- Lourenço, J. S., White, K., & Maylor, E. A. (2013). Target context specification can reduce costs in nonfocal prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39, 1757–1764. <https://doi.org/10.1037/a0033702>
- 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(2), 336. <https://doi.org/10.1037/0278-7393.24.2.336>
- Marsh, R. L., Hicks, J. L., & Cook, G. I. (2005). On the Relationship Between Effort Toward an Ongoing Task and Cue Detection in Event-Based Prospective Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 68–75. <https://doi.org/10.1037/0278-7393.31.1.68>
- Marsh, R. L., Hicks, J. L., Cook, G. I., Hansen, J. S., & Pallos, A. L. (2003). Interference to ongoing activities covaries with the characteristics of an event-based intention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 861–870. <https://doi.org/10.1037/0278-7393.29.5.861>

- Marsh, R. L., Hicks, J. L., & Watson, V. (2002). The dynamics of intention retrieval and coordination of action in event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(4), 652–659.
<https://doi.org/10.1037//0278-7393.28.4.652>
- McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology*, 14, S127–S144. <https://doi.org/10.1002/acp.775>
- McDaniel, M. A., Umanath, S., Einstein, G. O., & Waldum, E. R. (2015). Dual pathways to prospective remembering. *Frontiers in Human Neuroscience*, 9.
<https://doi.org/10.3389/fnhum.2015.00392>
- McVay, J. C., & Kane, M. J. (2012). Drifting from slow to “d’oh!”: Working memory capacity and mind wandering predict extreme reaction times and executive control errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3), 525–549.
<https://doi.org/10.1037/a0025896>
- Meeks, J. T., Hicks, J. L., & Marsh, R. L. (2007). Metacognitive awareness of event-based prospective memory. *Consciousness and Cognition*, 16(4), 997–1004.
<https://doi.org/10.1016/j.concog.2006.09.005>
- Meeks, J. T., & Marsh, R. L. (2010). Implementation intentions about nonfocal event-based prospective memory tasks. *Psychological Research*, 74(1), 82–89.
<https://doi.org/10.1007/s00426-008-0223-x>
- Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Reason*, 4(2), 61–64.

- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1), 90–100.
- Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, 4(1), 79–95.
<https://doi.org/10.3758/BF03210778>
- Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86(3), 214. <https://doi.org/10.1037/0033-295x.86.3.214>
- Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, 7(1), 44–64. [https://doi.org/10.1016/0010-0285\(75\)90004-3](https://doi.org/10.1016/0010-0285(75)90004-3)
- O'regan, J., Lévy-Schoen, A., Pynte, J., & Brugailière, B. é. (1984). Convenient fixation location within isolated words of different length and structure. *Journal of Experimental Psychology: Human Perception and Performance*, 10(2), 250.
<https://doi.org/10.1037/0096-1523.10.2.250>
- Öztekin, I., Davachi, L., & McElree, B. (2010). Are representations in working memory distinct from representations in long-term memory? Neural evidence in support of a single store. *Psychological Science*. <https://doi.org/10.1177/0956797610376651>
- Pashler, H. (1984). Processing stages in overlapping tasks: evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, 10(3), 358.
<https://doi.org/10.1037/0096-1523.10.3.358>
- Poulton, E. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin*, 91(3), 673–690.
<https://doi.org/10.1037/0033-2909.91.3.673>

- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2), 333. <https://doi.org/10.1037/0033-295x.111.2.333>
- Scullin, M. K., McDaniel, M. A., Shelton, J. T., & Lee, J. H. (2010). Focal/nonfocal cue effects in prospective memory: Monitoring difficulty or different retrieval processes? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 736–749. <https://doi.org/10.1037/a0018971>
- Sewell, D. K., Lilburn, S. D., & Smith, P. L. (2016). Object Selection Costs in Visual Working Memory: A Diffusion Model Analysis of the Focus of Attention. <https://doi.org/10.1037/a0040213>
- 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. <https://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. <https://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. <https://doi.org/https://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(4), 243. <https://doi.org/10.1027/1618-3169.52.4.243>

- Smith, R. E., Hunt, R. R., McVay, J. C., & McConnell, M. D. (2007). The cost of event-based prospective memory: Salient target events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 734–746. <https://doi.org/10.1037/0278-7393.33.4.734>
- Townsend, J. T., & Altieri, N. (2012). An accuracy–response time capacity assessment function that measures performance against standard parallel predictions. *Psychological Review*, 119(3), 500. <https://doi.org/10.1037/e502412013-221>
- Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, 62(4), 392–406. <https://doi.org/10.1016/j.jml.2010.02.001>
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, 11(1), 192–196. <https://doi.org/10.3758/bf03206482>
- Welford, A. T. (1952). The “psychological refractory period” and the timing of high-speed performance—a review and a theory. *British Journal of Psychology. General Section*, 43(1), 2–19. <https://doi.org/10.1111/j.2044-8295.1952.tb00322.x>
- Wenger, M. J., & Townsend, J. T. (2000). Basic response time tools for studying general processing capacity in attention, perception, and cognition. *The Journal of General Psychology*, 127(1), 67–99. <https://doi.org/10.1080/00221300009598571>
- Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–257). Hillsdale, NJ: Erlbaum.
- Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, 20(1), 6–10. <https://doi.org/10.3758/BF03202594>

Yao-N'Dré, M., Castet, E., & Vitu, F. (2013). The optimal viewing position effect in the lower visual field. *Vision Research*, 76, 114–123. <https://doi.org/10.1016/j.visres.2012.10.018>

Table 1. LBA model selection. Model factors include condition (PM), stimulus type (S), day (D), response (R) and correspondence (C).

Model	B	ν	s_ν	$t0$	p	Deviance	AIC	BIC
Top Model	R, PM, D	S, PM, D, C	PM, C	S, PM, D	2340	-92812	-88132	-65744
AIC Model	R, PM, D	S, D, C	C	D	1050	-92328	-90228	-80182
BIC Model	PM, D	S, C	C	S	510	-89234	-88214	-83335
Threshold LBA	PM, D	S, C	C	1	480	-88750	-87790	-83198

Figure 1. Panel 1. Mean accuracy by stimulus type by condition. Panel 2. Mean correct RT by stimulus type condition. Error bars calculated using the Morey (2008) bias-corrected method.

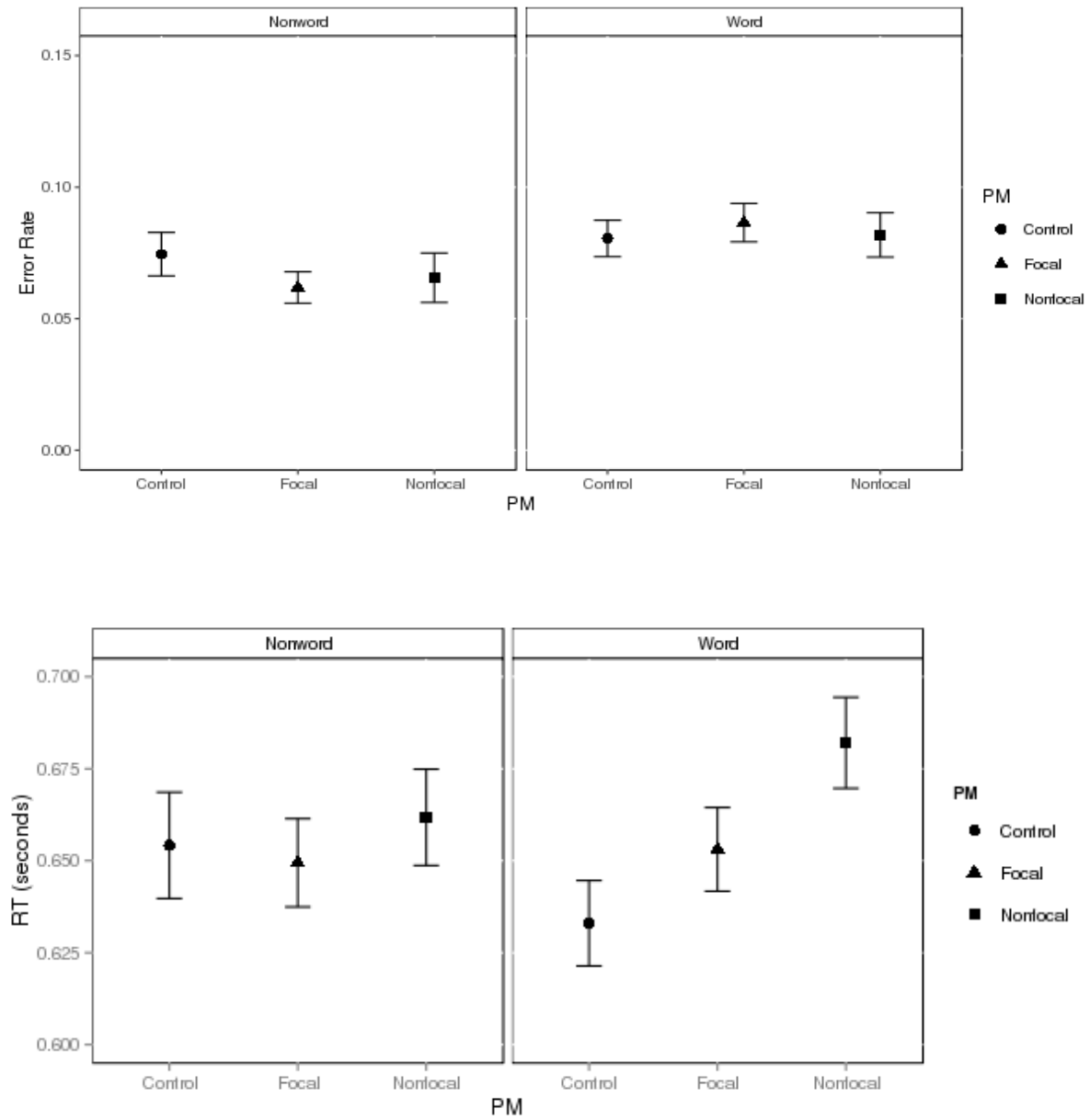
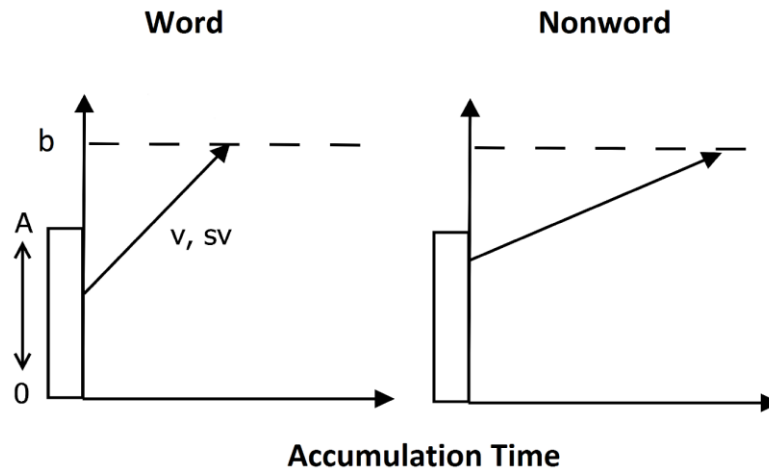
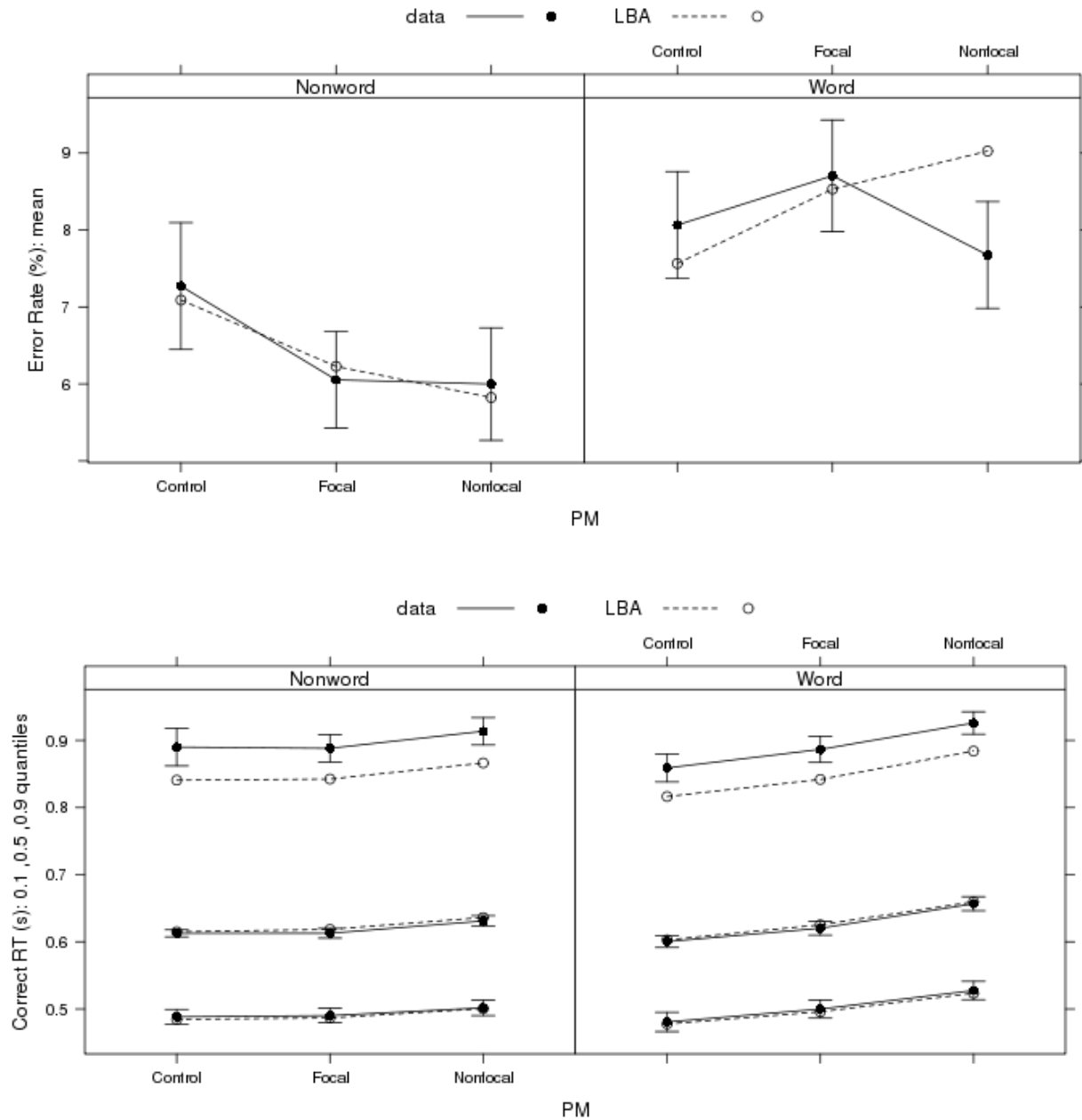


Figure 2. Illustration of the LBA (Brown and Heathcote, 2008).



$$RT = \text{Accumulation time} + t_0$$

Figure 3. LBA AIC selected model fits to error rates (left column), RT distribution for correct responses (10th, 50th, and 90th percentiles, middle column) and the 50th percentile (median) for error responses. Error bars calculated using the Morey (2008) bias-corrected method.



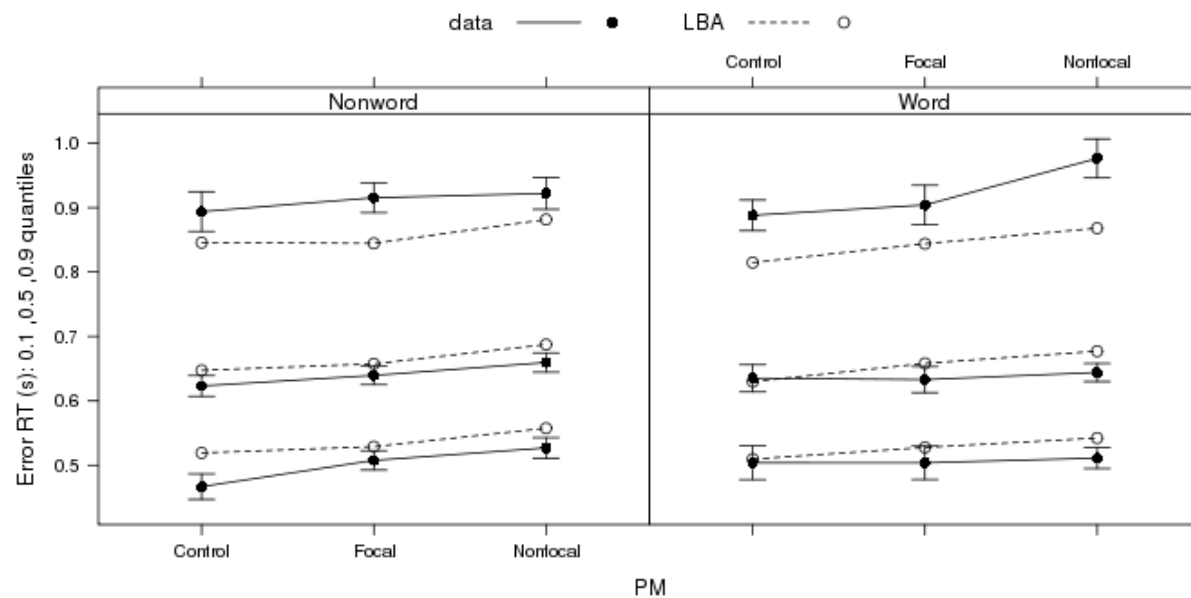


Figure 4. Mean B by stimulus type condition. Error bars calculated using the Morey (2008) bias-corrected method.

