Accepted, Psychological Review 21/03/2018.

Racing to Remember: A Theory of Decision Control in Event-Based Prospective Memory

Luke Strickland
University of Tasmania & University of Western Australia

Shayne Loft
University of Western Australia

Roger W. Remington
University of Minnesota & University of Queensland

Andrew Heathcote
University of Tasmania & University of Newcastle

Event-Based Prospective Memory (PM) requires remembering to perform intended deferred actions when particular stimuli or events are encountered in the future. We propose a detailed process theory within Braver's (2012) proactive and reactive framework of the way control is maintained over the competing demands of prospective memory decisions and decisions associated with ongoing task activities. The theory is instantiated in a quantitative "Prospective Memory Decision Control" (PMDC) architecture, which uses linear ballistic evidence accumulation (Brown & Heathcote, 2008) to model both PM and ongoing decision processes. Prospective control is exerted via decision thresholds, as in Heathcote, Loft and Remington's (2015) "Delay Theory" of the impact of PM demands on ongoing-task decisions. However, PMDC goes beyond Delay Theory by simultaneously accounting for both PM task decisions and ongoing task decisions. We use Bayesian estimation to apply PMDC to experiments manipulating PM target focality (i.e., the extent to which the ongoing task directs attention to the features of PM targets processed at encoding) and the relative importance of the PM task. As well as confirming Delay Theory's proactive control of ongoing task thresholds, the comprehensive account provided by PMDC allowed us to detect both proactive control of the PM accumulator threshold and reactive control of the relative rates of the PM and ongoing-task evidence accumulation processes. We discuss potential extensions of PMDC to account for other factors that may be prevalent in real-world PM, such as failures of memory retrieval.

Keywords: prospective memory, cognitive control, dual mechanisms framework, linear ballistic accumulator model

Introduction

Event-Based Prospective Memory (PM) requires remembering to perform intended deferred actions when particular stimuli or events are encountered in the future (Kliegel, McDaniel, & Einstein, 2008). Event-based PM task requirements are prevalent in everyday life (e.g., remembering to post a letter when you next drive past a mailbox), and can be crucial to personal safety and the safety of others (e.g., remembering to slow down when driving through a school zone; Bowden, Visser, & Loft, 2017). PM function is also essential to expertise in many safety-critical work contexts such as in aviation, medicine, and defense (Dismukes, 2012; Gawande, Studdert, Orav, Brennan, & Zinner, 2003; Loft, 2014). Given the number of actions required from experts in such work settings, even small PM error probabilities can translate into significant accident rates (Dismukes & Nowinski, 2006; Shorrock, 2005), with potentially disastrous consequences. In addition, PM is associated with difficulties with higher-level activities of daily living in both clinical samples and older adults (e.g., Woods, Weinborn, Velnoweth, Rooney, & Bucks, 2012). It is essential that the design of interventions in these aforementioned contexts is based on a thorough analysis of PM processes.

Einstein and McDaniel (1990) proposed a paradigm to study event-based PM in the laboratory, in which participants perform an ongoing task that usually requires a series of binary choices (e.g., a lexical decision making task). Prior to commencing that task, participants are also instructed to perform a PM action (e.g., press '9') if they encounter a PM target (e.g., any letter string that contains "tor") during that ongoing task. Studies using the paradigm focus on two mea-

The data and models discussed in the manuscript are available at https://osf.io/t3cqw/. The results have been discussed at the Australian Mathematical Psychology Conference, the Australasian Experimental Psychology Conference, and the International Conference on Prospective Memory. Please address correspondence about this article to Luke Strickland, School of Medicine, Division of Psychology, Private Bag 30, The University of Tasmania, Churchill Avenue, Sandy Bay 7005, Australia. E-mail: luke.strickland@utas.edu.au

sures. The first is PM accuracy, that is, the proportion of PM trials to which participants make a PM response. The second is response times (RTs) to non-PM trials; that is, ongoing task trials in which no PM target is presented, and thus no PM response is required (Smith, Hunt, McVay, & McConnell, 2007). Correct RTs are typically longer in *PM blocks*, blocks of trials in which there is a PM demand, as compared with *control blocks*, blocks of trials with no PM demand. The slow down is referred to as *PM cost* (e.g., Einstein et al., 2005; Hicks, Marsh, & Cook, 2005; Loft & Yeo, 2007; Smith, 2003). PM theories use costs to infer the PM-related processes that occur not only on non-PM trials but also those that occur on PM trials, and the associated cognitive mechanisms responsible for variation in PM accuracy (e.g., Einstein et al., 2005; Guynn, 2003; Loft & Yeo, 2007; Marsh, Hicks, Cook, Hansen, & Pallos, 2003; Smith & Bayen, 2004). Yet, the PM literature lacks a quantitative theory connecting PM accuracy and PM cost. In addition, most PM analysis ignores other potentially important features of manifest behavior, such as error RT distributions and the RT distributions observed on PM trials. In the current paper we propose a detailed process theory of the manner in which control is maintained over competing demands of PM decisions and decisions associated with ongoing task activities. We instantiate the theory in a comprehensive quantitative model that is able to accurately account for all the features of the data from the Einstein and McDaniel (1990) paradigm, as well as the effects of two benchmark manipulations (PM focality and PM importance). Our findings shed new light on the mechanisms underlying PM, and seriously challenge other current PM theories.

The Race to Remember

To our knowledge, there have been two previous attempts to quantitatively model both PM and ongoing task responding. Arnal (2008) instantiated PM and ongoing task processes as two separate diffusion processes, with the first diffusion process to complete determining the overt response. Gilbert, Hadjipavlou, and Raoelison (2013) proposed a parallel distributed processing model in which the PM and ongoing task nodes are updated non-linearly over many cycles of perceptual input until activation for one node reaches threshold and the response is made. Although both of these studies simulated a PM process, neither fit their models to data. Thus, it is not clear to what degree these models provide an accurate and comprehensive account of data from PM paradigms.

Recent research provides a more fine-grained analysis that incorporates separate fits to the precise features of each individual's data set, but focuses exclusively on non-PM trials (Ball & Aschenbrenner, 2017; Boywitt & Rummel, 2012; Heathcote, Loft, & Remington, 2015; Horn, Bayen, & Smith, 2011; Horn & Bayen, 2015; Strickland, Heathcote, Remington, & Loft, 2017). Two model architectures have been fit to non-PM trial responses: the diffusion decision model (DDM; Ratcliff, Gomez, & McKoon, 2004), and the Linear Ballistic Accumulator (LBA; Brown & Heathcote, 2008). Thus far, where both were applied to PM data sets, the LBA provided better fit (Heathcote, Loft, & Remington, 2015; Strickland et al., 2017). In addition, unlike the DDM, the LBA easily extends to more than two choices without sacrificing analytic tractability. Thus, in the current study we use the LBA as the basis to include both PM and ongoing task processes. Our architecture includes a pair of accumulators corresponding to each of the binary ongoing-task responses, and a third accumulator corresponding to the PM response. Stimulus features consistent with a particular response provide excitatory input to the corresponding accumulators, as well as possibly providing inhibitory inputs to the other accumulators. The evidence total in each accumulator increases independently of the evidence totals in other accumulators, and the first accumulator to reach its threshold determines the response made. This setup with feedforward excitation and inhibition and an independent race makes the model tractable, allowing fits to all aspects of each individual subject's data. If successful, it will enable the most thorough quantitative examination of PM data sets to date.

Figure 1 depicts the model for a lexical decision ongoing task with an additional PM task demand. There are three possible responses, indicating that the stimulus is either a word, a non-word, or a PM target. Evidence for each response accrues linearly towards threshold, starting from a point that varies independently between accumulators from trial to trial according to a uniform distribution. Correct PM responses (PM hits) occur on PM trials when the PM accumulator reaches threshold before either of the ongoing task decision accumulators. Likewise, PM misses occur when one of the ongoing task accumulators reaches threshold before the PM accumulator.

Response probabilities vary depending on the values of three classes of model parameters: start-point variability, thresholds and evidence accumulation rates. Evidence accumulates at a constant rate within a given trial, but rates differ from trial to trial according to a normal distribution. Rate parameters are usually assumed to vary as a function of stimulus differences. Thresholds are set prior to stimulus presentation, and hence are not affected by stimulus characteristics that vary unpredictably from trial to trial. Thresholds, can, however, vary over blocked manipulations (e.g., PM vs control blocks), and over accumulators (e.g., the PM response threshold as compared with the ongoing task response thresholds). The level of start-point variability is assumed not to vary over the conditions we examine here, although it might vary in other circumstances.

Decision time (i.e., the time for the winning response to accumulate to threshold) is determined by the same set of parameters as response probabilities. Total RT is determined by decision time plus non-decision time, which captures the time for stimulus encoding and motor response production. We assume non-decision time to be the same for all accumulators (and take measures to ensure this, like making manual responses for all choices equally easy), and so estimate only one associated parameter. In addition, we assumed that non-decision time would not play a role in PM cost, as it has not done in any previous LBA modeling (Heathcote, Loft, & Remington, 2015; Strickland et al., 2017).

We will use our LBA-based analysis to constrain and develop PM theory. Firstly, by evaluating the ability of the model to fit the observed PM data, we can determine whether the simple assumptions that make it tractable - a parallel independent race with feedforward excitation and inhibition and linear updating - are sufficient to model the processes driving performance in the Einstein and McDaniel (1990) paradigm. Model miss-fit will suggest that more complex (and, unfortunately, likely less computationally tractable) mechanisms are required, such as the non-linear updating assumed by Gilbert et al. (2013), or perhaps other forms of non-linearity, such as the recurrent excitation and inhibition assumed in some race models (e.g., Usher & McClelland, 2001).

Assuming a sufficiently good fit, the way in which model parameters vary to capture PM data provides detailed quantitative information about the latent cognitive processes that drive observed responding. To date, verbal PM theories have almost exclusively relied on PM accuracy and ongoing task mean RT for constraint, with less attention paid to trade-offs between accuracy and RTs, to RT distributions, and to PM RTs. In contrast, our model parameters incorporate information from the entire observed data set. Often, verbal theories of mean effects (e.g., theories of the PM cost effect) specify mechanisms that map naturally to changes in LBA model parameters, and so theorizing about these mechanisms can potentially be constrained by more data than it is in standard analysis. We now review other current PM theories and how they correspond to our model.

Capacity Sharing and Spontaneous Processes

Extant theories of PM performance argue that PM is often, or always, reliant on the allocation of cognitive capacity away from the ongoing task and towards a capacity consuming PM detection process (Einstein et al., 2005; Guynn, 2003; Nowinski & Dismukes, 2005; Smith, 2003). This capacity sharing is proposed to be ubiquitous in that it occurs on both PM and non-PM trials within PM blocks. Initial studies did not find decreased non-PM trial accuracy in PM blocks compared with control blocks (Kidder, 1999; Kidder, Park, Hertzog, & Morrell, 1997; Park, Hertzog, Kidder, Morrell, & Mayhorn, 1997; West & Craik, 1999), but Marsh and Hicks (1998) showed that concurrent tasks tapping executive control could impair PM performance, and Smith (2003) then found PM cost in terms of slowed RT to non-PM trials in PM blocks versus control blocks. Smith argued that elevated non-PM trial RTs are a more sensitive measure of reduced capacity than accuracy. Since then, PM cost has remained central to assertions that capacity sharing is present (for discussions see Einstein & McDaniel, 2010; Smith, 2010). Capacity sharing theories predict that the PM cost on non-PM trials should benefit PM accuracy on PM trials. There are at least three lines of evidence supporting this. First, PM accuracy and costs to the ongoing task are often positively correlated across subjects (Smith, 2003; Smith & Bayen, 2004). Second, RTs to non-target trials preceding successful PM target detection and response can be longer than RTs to ongoing task trials preceding PM errors (Loft & Yeo, 2007; West, Krompinger, & Bowry, 2005). Third, emphasizing the importance of PM tasks or increasing the frequency of PM targets increases both costs to the ongoing task and PM performance (e.g, Kliegel, Martin, McDaniel, & Einstein, 2004; Loft, Kearney, & Remington, 2008; Loft & Yeo, 2007).

Recent PM literature acknowledges that PM cost cannot be attributed to capacity sharing without also considering non-PM trial accuracy, because RT increases may also result from strategic processes such as a shift in the speed/accuracy trade-off (e.g., Smith, 2010). To address this, researchers have applied the LBA and DDM to non-PM trials, in order to titrate the latent variables underlying PM cost and non-PM trial accuracy. The original capacity sharing hypothesis of several authors was that PM cost would be driven by decreased accumulation rates (Boywitt & Rummel, 2012; Horn et al., 2011; Horn, Bayen, & Smith, 2013). The idea, consistent with general theories of cognitive resources (e.g., Bundesen, 1990; Gobell, Tseng, & Sperling, 2004; Kahneman, 1973; Navon & Gopher, 1979; Pashler, 1984; Welford, 1952; Wickens, 1980), is that processing speed of the ongoing task is increased in proportion to the amount of cognitive capacity it receives. In the DDM, in which accumulation towards the correct response is the same as accumulation away from the error response, reduced processing speed is reflected in evidence accumulation heading less directly towards the correct response. In the LBA, reduced cognitive capacity could cost ongoing task processing in two ways. First, the accumulation of the 'correct' accumulator (e.g., in a lexical decision task, the word accumulator when the stimulus is a word) could decrease, implying that participants become less apt at identifying true signal from the stimuli. Second, the accumulation of the 'error' accumulator (e.g., non-word accumulator for a word stimulus) could increase, implying that participants are incorrectly interpreting stimulus attributes in favor of the wrong response. Several recent computational modeling studies, examining a total of nine PM cost data sets, have attempted to detect a capacity cost to accumulation rates on

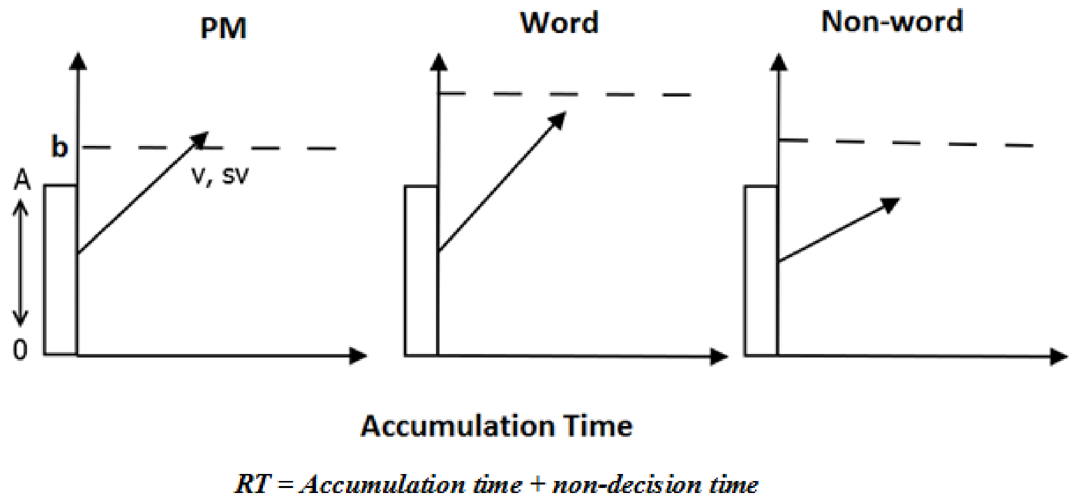

Figure 1. An LBA model of a PM task with a concurrent ongoing lexical-decision task. Evidence for each response is initially drawn from a uniform distribution on the interval [0, A]. Over time, evidence accumulates towards each response at accumulation rates drawn from normal distributions with mean v, and standard deviation sv. The first response to reach its threshold, b, is the response made. We refer in our results to B, which is $b-A$. Total RT is determined by accumulation time plus non-decision time.

non-PM trials using both the LBA and DDM (Ball & Aschenbrenner, 2017; Heathcote, Loft, & Remington, 2015; Horn & Bayen, 2015; Strickland et al., 2017). The modeling revealed that PM cost is largely due to increases in response threshold, and is not due to changes in ongoing task evidence accumulation. In the DDM and LBA thresholds are the locus of strategy, not capacity, and so these findings suggest that PM cost does not result from capacity sharing.

In contrast to PM theories that specify PM always requires ongoing task capacity (e.g., Smith, 2003), the multiprocess view of PM (Einstein et al., 2005; McDaniel & Einstein, 2000) claims that in some paradigms PM does not rely on ongoing task capacity because PM retrieval can occur 'spontaneously' on PM trials. The more recent dynamic multiprocess view (Scullin, McDaniel, & Shelton, 2013) extends the standard multiprocess view by specifying that, with complex task sets, spontaneous processes and capacity sharing can interact, for example by spontaneous retrieval processes triggering a period of capacity-consuming monitoring, but for standard laboratory paradigms the theories make similar claims. The multiprocess view proposes that spontaneous processes are more likely if PM targets are 'focal' to ongoing task performance (Einstein & McDaniel, 2010). PM targets are considered focal to ongoing task performance when there is high overlap between the information that needs to be assessed to perform the ongoing task and the information that needs to be assessed to detect the PM target; for example detecting a single PM target word may be facilitated by making lexical decisions (Einstein et al., 2005). PM targets are 'non-focal' if noticing the features requires some extra processing. For example, identifying any word within a category (e.g., make a PM response if you see a word that is an animal) during lexical decision is 'non-focal', because performing lexical decisions does not require a priori encoding of categorical concepts (Einstein & McDaniel, 2010). Many studies find high focal PM accuracy with little or no PM costs (Ellis & Milne, 1996; Harrison & Einstein, 2010; Knight et al., 2011; Marsh et al., 2003; Scullin, McDaniel, Shelton, & Lee, 2010). In contrast, the same studies find lower non-focal PM accuracy with higher PM cost.

On the basis of the reviewed effects the multiprocess view distinguishes focal and non-focal PM mechanisms, claiming that non-focal PM is reliant on PM monitoring that shares capacity with the ongoing task, whereas focal PM is not reliant on sharing ongoing task capacity (and hence instead reliant on spontaneous, cue-driven PM processing) (e.g., Einstein et al., 2005; Scullin, McDaniel, Shelton, & Lee, 2010). However, as reviewed, the recent PM cost modeling suggests that non-focal cost resulted from increased response thresholds, rather than decreased ongoing task accumulation rates, over a range of non-focal tasks (Ball & Aschenbrenner, 2017; Heathcote, Loft, & Remington, 2015; Horn & Bayen, 2015; Strickland et al., 2017). Given that, even with this lack of evidence for capacity sharing, non-focal PM accuracy was relatively high in the modeled data sets (>50%), it may be unnecessary to distinguish focal and non-focal PM tasks by their demand for ongoing task capacity. In line with this, the proposed LBA architecture represents both focal and non-focal PM in the same way, as a PM accumulator that monitors for the PM items. Although this architecture includes PM capacity, in terms of the speed of the PM accumulation process, it could potentially explain a broad range of canonical PM effects without invoking shared capacity between the PM task and ongoing task. These effects include the presence of PM errors, PM cost, the association between PM cost and PM accuracy, and also variation in cost and PM accuracy as a function of task conditions.

In our model, PM errors occur because PM and ongoing task decision processes compete in a race. Previous research has found that responses that have not been performed for a while take longer than more recently performed responses (Ruthruff, Remington, & Johnston, 2001), which suggests faster processing for frequent decisions than for infrequent decisions. Thus, the fact that PM responses are required less frequently than ongoing task responses may put them at a disadvantage in the race to response threshold, leading to substantial PM error rates. In line with this conceptualization of PM errors, Loft and Remington (2010), found that participants made more PM errors when they needed to remember to deviate from more strongly practiced ongoing task response routines. Further, Loft and Remington (2013) found that when they delayed participants' opportunity to respond by 1 second, the difference between focal and non-focal PM accuracy (long assumed to reflect different capacity-sharing requirements) was eliminated. Thus, the race between PM and ongoing task decisions can naturally, and plausibly, lead to PM errors, without failures of special-purpose monitoring mechanisms that rely on ongoing task capacity. However, a sufficient PM theory must explain many other features of PM data as well, including PM cost and the variation in PM accuracy associated with factors such as PM focality and PM importance. To fit these effects, decision process parameters would have to vary across PM conditions. As reviewed, the previous modeling indicates no variation in ongoing task capacity over PM conditions, inconsistent with previous theory. An alternative to capacity sharing is that, in response to the habitual advantage of ongoing task decisions, the cognitive system invokes "cognitive control" to aid the PM decision in completing on PM trials. Cognitive control is a general feature of the cognitive system that allows us to act in a goal-directed and flexible way, freeing us from automatic, stimulus bounded, behavior, such as the dominance of more frequent responding over rare responding (Miller & Cohen, 2001; Miyake et al., 2000). Thus, cognitive control is critical to many paradigms in which low frequency responses must overcome more frequent responses on some trials (Braver, Barch, Gray, Molfese, & Snyder, 2001), like PM responses must in the PM paradigm. Recent PM literature recognizes that cognitive control mechanisms would likely contribute to PM performance, and suggests that they should be integrated into PM theory (Bugg, McDaniel, & Einstein, 2013). We propose that cognitive control over PM and ongoing task decision processes accounts for many phenomena of interest in the PM literature, including PM cost, the effects of PM focality, and the effects of PM importance.

Dual Mechanisms of Prospective Memory Decision Control

Recently Braver (2012) proposed that there are two classes of cognitive control: proactive and reactive. This dual mechanism theory has been useful for understanding cognitive control in many paradigms, including working memory tasks (e.g., Braver, Gray, & Burgess, 2007; Burgess & Braver, 2010; Marklund & Persson, 2012), the AX- Continuous performance task (e.g., Braver, Barch, Keys, et al., 2001; Locke & Braver, 2008; van Wouwe, Band, & Ridderinkhof, 2011), the stop signal task (e.g., Boehler, Schevernels, Hopf, Stoppel, & Krebs, 2014; Stuphorn & Emeric, 2012), the Stroop task (e.g., Kalanthroff, Avnit, Henik, Davelaar, & Usher, 2015; West, Choi, & Travers, 2010), and the cued task-switching paradigm (e.g., Chevalier, Martis, Curran, & Munakata, 2015; Lucenet, Blaye, Chevalier, & Kray, 2014). Furthermore, proactive and reactive control have been found to correspond to distinct patterns of brain activity (Appelbaum, Boehler, Davis, Won, & Woldorff, 2014; Braver, 2012; Irlbacher, Kraft, Kehrer, & Brandt, 2014). Braver proposed that both control modes may be relevant to PM, and this has since been discussed in the PM literature (Ball, 2015; Bugg et al., 2013). Below we propose a way to identify both forms of cognitive control in the PM paradigm by using our model. We refer to this theoretical perspective as PM decision control (PMDC).

Proactive Control. Proactive control is deployed in order to "bias attention, perception and action systems in a goal-driven manner" (Braver, 2012, p. 2). The key distinguishing feature of proactive control is that it is active in advance of the cognitively demanding event, so that it will already be in effect when that event occurs. Thus, in the Einstein and McDaniel paradigm (1990), PM-related proactive control should be present on all PM block trials, including non-PM trials. In this way it is similar to the preparatory capacity-sharing mechanisms specified in previous PM theory (e.g., Smith, 2003), and to theories of 'attentional allocation policies' (Marsh, Hicks, & Cook, 2005; Marsh et al., 2003). It is also similar to these theories in that it reflects an ongoing adjustment to decisions due to an appreciation of increased task demands. Using a standard PM analysis, proactive control and capacity sharing are difficult to compare, because their effects on the dependent variables are qualitatively the same. Either more proactive control, or more capacity sharing from the ongoing task towards the PM task, could increase PM cost and PM accuracy. However, the PMDC model can parse proactive control from capacity sharing. As reviewed, capacity sharing should take a toll on processing speed, which corresponds to accumulation rates. In contrast, threshold changes naturally correspond to proactive control, as thresholds are the parameters of the decision process that are set prior to trial presentation, and the parameters that cause strategic trade-offs between responses and RTs. For example, Verbruggen & Logan (2009) propose that "proactive slowing" of one decision process can be implemented so that another process has more time to complete. This is consistent with the *delay theory* of PM cost, which proposes that participants slow their ongoing task decision process in order to give a parallel PM process more to reach response selection (Heathcote, Loft, & Remington, 2015; Loft & Remington, 2013).

The scope of delay theory is more limited than PMDC, because it makes few specific claims about the PM process itself. Thus, delay theory is silent on potentially important PM-related phenomena. For example, proactive control may be exerted on the PM response threshold. With a lower PM threshold, the PM accumulator is more likely to win the PM race when PM targets are presented. In our PMDC architecture, we propose that both forms of threshold adjustments are possible. That is, as in delay theory, ongoing task thresholds may be increased to increase the probability that the PM accumulator completes before the ongoing task accumulators. In addition, the PM threshold may be decreased for the same reason. It was not possible to examine this mechanism in previous modeling efforts, because they focused only on non-PM trials, but here we are able to test it because we also fit performance on PM trials.

Reactive Control. Reactive control is applied to influence responding "only as needed, in a just-in-time manner" (Braver, 2012, p. 2). In the context of the Einstein and McDaniel (1990) paradigm, reactive control would be active on PM trials, and not active on non-PM trials. Recent PM literature suggests that reactive control may sometimes facilitate PM (Bugg et al., 2013), but to our knowledge the PM literature lacks a detailed process theory of how this reactive control would operate. With PMDC we can quantitatively characterize reactive control by comparing parameters between PM trials and non-PM trials. Reactive control will not affect thresholds, as they are set prior to identifying whether the item is a PM target. Thus, in order to facilitate PM accuracy, reactive control would have to affect evidence accumulation rates on PM trials. Note that reactive control can affect accumulation rates on PM trials while leaving accumulation rates on non-PM trials intact. It can thus be distinguished from capacity sharing, which predicts that accumulation rate decreases are active for both PM trials and non-PM trials in PM blocks.

Figure 2 presents the framework for reactive accumulation rate control in the PMDC architecture, using the example of a lexical decision task in which participants have a PM task demand. The encoding process includes detectors (squares) for each possible response to the task: 'word', 'non-word', and 'PM'. The detectors receive input from stimulus features. Output from the detectors can directly contribute (solid lines) to the input to the corresponding evidence accumulator, which we refer to as excitation of the accumulator. Thus, on PM trials, detection of PM-related stimulus attributes will directly speed up PM accumulation (A1). This may allow the PM accumulator to outpace ongoing task accumulators, and thus for a PM response to reach threshold before an ongoing task response. Higher PM accumulation on PM trials is a trivial prediction of PMDC, and not unique to it, as PM stimuli contain evidence that a PM response should be made, whereas non-PM stimuli do not contain this evidence. Faster PM accumulation could also result from an increase in the capacity devoted to the PM decision process.

More specific to PMDC, activation of the PM detector can also inhibit ongoing task response production through inputs to the two ongoing task accumulators (B1 & B2). This inhibition of ongoing task accumulators distinguishes reactive control from spontaneous retrieval mechanisms, or increased PM capacity, which would both increase excitation of the PM accumulator. The LBA does not explicitly model interactions between accumulators during the accumulation process, such as competition between PM and ongoing task accumulation, or correlations between inputs to different accumulators. These assumptions enable the likelihood for a particular accumulator winning at a particular time to be written (although see Heathcote & Love, 2012), which in turn supports sufficient estimation (i.e., estimation that takes account of all information in the data). However, the LBA can use "feedforward" interactions to achieve similar effects without sacrificing analytic tractability (van Ravenzwaaij, Brown, Marley, & Heathcote, n.d.). Feedforward interactions occur when the input to an accumulator associated with one response is affected by evidence associated with alternative responses. For example, evidence for the presence of a PM target might affect the input to accumulators associated with the ongoing-task responses. PMDC allows for such feedforward effects, enabling it to account for competitive or cooperative interactions among stimulus attributes associated with different choices (see also Trueblood, Brown, & Heathcote, 2014).

PMDC's feedforward inhibition structure is illustrated in Figure 2 (dashed lines). Conceivably, input to the detectors for each response could inhibit each of the competing responses. However, such feedforward interactions are not always identifiable. For example, decreased non-word accumulation on word trials is not uniquely attributable to inhibition from the word detector. It could equally occur if lexical evidence is drawn from a single dimension (and thus word trials would necessarily contain less non-word evidence). PM trials are unique in that PM stimuli contain no less evidence towards the ongoing task responses than non-PM stimuli. For example, a PM target that is also a word

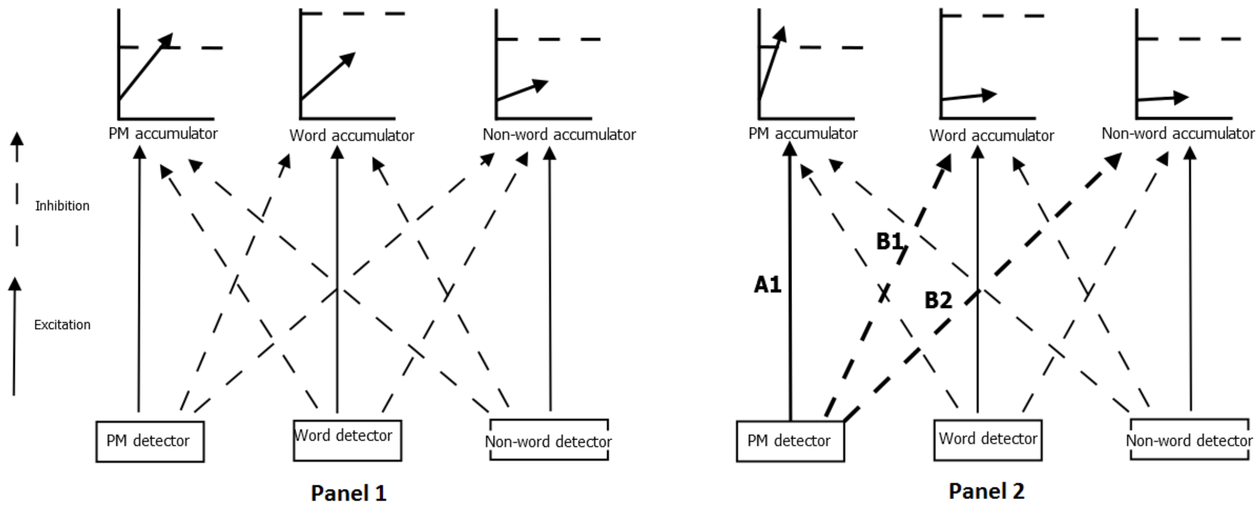

Figure 2. Conceptual model of reactive control in the PM paradigm. Panel 1: The underlying architecture. Input to each detector from stimulus encoding carries forward to decision processing to excite accumulation towards the relevant response (solid lines), and inhibit accumulation towards competing responses (dashed lines). Note we depict a threshold bias against word responding, as is often induced if PM items are always words (Heathcote, Loft, & Remington, 2015; Strickland et al., 2017). Panel 2: Reactive control from the PM detector. On PM trials, PM stimulus inputs will trigger the PM detector, potentially increasing both excitation of PM accumulation (A1), and inhibition of ongoing task accumulation (B1 and B2). Inhibition via B1 and B2 can be identified by comparing word and non-word accumulation rates on PM trials with non-PM trials.

should not have any less evidence in its stimulus features for the word accumulator than a non-PM word. Therefore, lower accumulation towards ongoing task responses on PM trials, as compared with non-PM trials, can be attributed to feedforward inhibition from the PM detector.

Inhibition of ongoing task decisions from the PM detector could explain the effects of PM "lures". There are two types of PM lure trials: 1) control block trials in which a PM target is presented despite participants being informed there is no need to perform the PM task, and 2) non-PM trials in PM blocks in which certain stimuli have some, but not all, the necessary PM features. On both types of lure trials, RTs are longer than on regular non-PM trials (Knight et al., 2011; Scullin, Einstein, & McDaniel, 2009; Scullin, McDaniel, & Einstein, 2010). Although this RT slowing on lures strongly suggests that some type of spontaneous process is triggered by PM input and interferes with ongoing task responding, the mechanism underlying this interference is unclear (Knight et al., 2011). PMDC provides a process account, under which input to the PM detector on lure trials would activate reactive inhibition of ongoing task accumulation, extending total ongoing task decision time.

One objection to the reactive control account of the PM lure effect is that ongoing task RTs on PM error trials (i.e., ongoing responses made on PM trials) are sometimes faster than ongoing task RTs on non-PM trials (Marsh, Hicks, & Watson, 2002). At first glance, this might seem to imply that PM accumulation actually speeds up ongoing task processing, rather than inhibits it. However, fast PM error RTs can emerge from our model with no such assumption. All trials on which the ongoing task accumulators are slower than the PM accumulator will lead to a PM hit, and hence be excluded from the ongoing task RT distribution for PM error trials. This would lead to quicker average ongoing task RTs on PM trials by "statistical facilitation" (Raab, 1962). Consistent with this, Gilbert et al. (2013) found that their simulation model of PM predicted fast PM miss (i.e., ongoing responses on PM trials) RTs despite their model explicitly specifying that PM stimulus inputs should lead to slower ongoing task processing. Our modeling will go further by testing whether slowed ongoing task processing on PM trials fits the actual observed RT distributions from our experiments.

Testing Prospective Memory Decision Control

We performed two studies that test the PMDC architecture. One objective of our studies is to evaluate whether the model is capable of fitting the full array of data from a PM paradigm. Adequacy of fit will guide us as to whether the assumption that PM and ongoing task decision processes compete in an independent race is reasonable. Misfit might also occur if the LBA is the wrong model of each accumulation process, but given the breadth of past successful LBA fits this seems less likely. The LBA has been successfully applied to many experimental tasks, including: lexical decision (Brown & Heathcote, 2008), brightness discrimination (Brown & Heathcote, 2008), absolute identification (Brown & Heathcote, 2008), random dot motion (Forstmann et al., 2008), redundant target detection (Eidels, Donkin, Brown, & Heathcote, 2010), the stop signal paradigm (Matzke, Love, & Heathcote, 2017), and mental rotation (Provost & Heathcote, 2015).

Assuming we find good fit, we will evaluate, based on model parameters, both the capacity sharing accounts of the PM paradigm and the PMDC account. We will do this in part by comparing trials from control blocks with non-PM trials from PM blocks (proactive control), and in part by comparing non-PM trials from PM blocks with PM trials from PM blocks (reactive control). For both experiments, the ongoing task was lexical decision. Each experiment focused on one benchmark manipulation from the PM literature. The first is a manipulation of PM target focality. Experiment 1 included two separate blocks of PM: one in which participants have a focal PM task, and one with a non-focal PM task. Focal PM demands typically produces higher PM accuracy with lower PM cost than non-focal PM demands. Experiment 2 manipulates the emphasis of the importance of the PM task. Again it includes two types of PM blocks: one where the instructions emphasize the importance of the PM task over the ongoing task, and one where the instructions emphasize the importance of the ongoing task over the PM task. Emphasizing the importance of the PM tasks typically produces higher PM accuracy, with higher PM cost (e.g., Einstein et al., 2005; Kliegel et al., 2004; Loft & Yeo, 2007; Smith & Bayen, 2004).

In terms of the standard PM analysis of mean ongoing task non-PM trial RT and mean PM accuracy, previous PM theory can already account for the effects of focality and importance. For example, capacity theories claim that focal PM has a lower capacity requirement (Einstein & McDaniel, 2005; Smith, Hunt, McVay, & McConnell, 2007), which means that high PM accuracy can be achieved with less PM costs, and that PM emphasis causes more attention to be allocated to the PM task, which results in increased cost and increased PM accuracy (Einstein et al., 2005; Smith & Bayen, 2004). However, the decision processes that capacity sharing theories would predict to underlie these effects differ from the decision process mechanisms specified by PMDC. We now outline the differing predictions of the theories.

Capacity sharing. A ubiquitous PM capacity demand on the ongoing task should cause decreased non-PM trial accumulation rates in PM blocks. As reviewed, this account has been rejected for nine previous data sets (Heathcote, Loft, & Remington, 2015; Horn & Bayen, 2015; Strickland et al., 2017, Ball & Aschenbrenner, 2017). However, the current approach to analysis differs greatly from previous work, and thus rejecting capacity sharing is not a foregone conclusion. We made several departures from previous analysis. Firstly, and most obviously, this is the first time that PM trials will be in the model, and as some parameters are shared between PM and non-PM trials, the addition of these trials could potentially affect our result. Secondly, Experiment 2 includes a condition in which PM is important, and capacity sharing is considered to be particularly likely under such conditions (Smith & Bayen, 2004). In addition, we estimate parameters using Bayesian methods, whereas all previous modeling has used optimization methods like maximum likelihood. We also used posterior inference, in which a model that can potentially produce all relevant effects is fit and conclusions are drawn based on posterior parameter distributions (see Brooks & Gelman, 1998), rather than model selection.

Proactive control. PMDC predicts that thresholds will be proactively controlled in favor of PM responding in PM blocks. Thus, in line with previous findings (Heathcote, Loft, & Remington, 2015; Strickland et al., 2017), we expect that ongoing task response thresholds will increase in PM blocks compared with control blocks. In Experiment 1, we include PM instructions that are "stimulus-specific" (Lourenço, White, & Maylor, 2013). That is, PM stimuli can only be one type of ongoing task stimulus (in our case, they could only be words in a lexical decision task). This enhances the competition between the PM response and one of the relevant ongoing task responses (e.g., more competition between PM and 'word' if PM targets are always words), and decreases the competition between the PM response and the other ongoing task response. Thus, if, as specified in PMDC, thresholds elevations are strategically targeted at buying the PM decision process more time on PM trials, we would expect to see larger threshold increases for the ongoing task accumulator relevant to PM (word) than the irrelevant ongoing task accumulator (non-word), an effect we refer to as selective delay. Note that, as participants do not know in advance what items they will be presented, and thresholds cannot be reactively adjusted, this threshold increase will be active across all trials in PM blocks - word trials, non-word trials, and PM trials.

An alternative view is that thresholds will increase equally between word and non-word responding, due to a general perception of increased task complexity (see Horn & Bayen, 2015; Strickland et al., 2017). Previous literature has produced a mixed picture regarding whether general caution increases result from stimulus-specific PM instructions (cf. Heathcote, Loft, & Remington, 2015; Horn & Bayen., 2015; Strickland et al., 2017). The presence vs. absence of general caution increases appears to depend on task conditions. For example, in Strickland et al. (2017), we found that focal PM was associated only with selective delay. In contrast, non-focal PM caused a larger selective delay than focal PM, but also a non-word (general) threshold increase. We argued that participants implement the latter general caution increase because the complexity of the non-focal task might be perceived to be higher. In Experiment 1, our task conditions are very similar to Strickland et al. (2017), and thus we expect to replicate this finding.

In Experiment 2, we include PM targets in both types of stimuli (words and non-words). This would lead to similar competition between the PM response and both ongoing task responses, and so under PMDC we expect to see substantial threshold increases for both accumulators (word and non-word). In addition, we would expect that thresholds should increase more when the importance of the PM task is emphasized. In Experiment 2, the PM task is identical in the blocks where it is important and blocks where it is unimportant, and thus increased ongoing task thresholds as a function of PM importance can be ascribed uniquely to proactive control and not to task complexity.

These predictions of PMDC could all also have been derived from the delay theory of PM cost (Heathcote, Loft, & Remington, 2015). However, PMDC goes further in specifying that the PM threshold may also vary in favor of the PM response. Prior to PMDC, variation in PM threshold had not been considered, despite being a potential source of differences in PM accuracy. It is not clear how the PM threshold will vary in non-focal PM conditions, as compared with focal PM conditions, in Experiment 1. On the one hand, the non-focal PM threshold might be lower than the focal PM threshold, in order to compensate for slower non-focal PM accumulation. On the other hand, the non-focal PM threshold might be higher because the non-focal PM rule requires more evidence. Nonetheless, it is important to examine how the PM threshold varies, and also to control for threshold variations when examining other PM-related parameters. In Experiment 2, PMDC makes a direct prediction about the PM threshold; when participants are instructed to prioritize the PM task, they will proactively lower their PM threshold so that the PM accumulator has a higher chance to reach threshold before the ongoing task accumulators.

Reactive control. PMDC also extends beyond delay theory by specifying that reactive control may affect the accumulation rates of PM trials as compared with non-PM trials. Obviously, the PM accumulation rate should be higher on PM trials, due to the processing of PM related stimulus attributes that are not present on non-PM trials. More importantly for PMDC's account, reactive control may reduce ongoing task accumulation on PM trials. For example, on a PM trial with a word stimulus (and perhaps particularly under the stimulus-specific PM conditions of Experiment 1), PMDC allows that accumulation towards the 'word' response could be reduced by an inhibitory input from the PM detector, despite their being equal evidence that the item is a word as there is on non-PM trials.

Our experimental manipulations may also affect reactive control. Focal PM, as compared with non-focal PM, may be more strongly activated by bottom-up retrieval processes on PM trials, increasing reactive control of ongoing task decision processing (McDaniel, LaMontagne, Beck, Scullin, & Braver, 2013). We can identify with our modeling which pathways of reactive control are increased. First, focal PM targets might increase the excitation towards the PM response (i.e., increased PM response accumulation on PM trials), without diminishing ongoing task processing (i.e, no change in ongoing task response accumulation on PM trials). Alternatively, focality may increase feedforward inhibition of ongoing task decision accumulation on PM trials (i.e., decreased ongoing task accumulation on PM trials with no change in PM accumulation rates). It is also possible that both excitation of PM accumulation and inhibition of ongoing task responding may be increased by focality.

Although previous PM theory has argued that emphasizing the importance of the PM task would modify proactive control factors, such as monitoring or capacity sharing (McDaniel & Einstein, 2000; Smith & Bayen, 2004), it is also possible that importance emphasis could affect how reactive control operates. Motivation via reward has previously been shown to increase reactive response inhibition (Boehler et al., 2014). In terms of PMDC, participants may alter the sensitivity of their reactive control structure to input from the PM detector. PM input may excite PM accumulation more when PM importance is emphasized, or it may inhibit ongoing task accumulation more. Alternatively, both pathways may be affected. However, we cannot differentiate the last option from the possibility that the detector input itself was somehow increased (e.g., more focus on the PM target), as that would also cause both increased excitation and inhibition.

Experiment 1: PM Target Focality

We designed our experiments with the goal of accurate process measurement on PM trials. Our design includes three features to this end. First, because most PM studies conducted using the traditional Einstein and McDaniel (1990) paradigm use only a small number of PM stimuli, they do not produce enough data to reliably constrain a model of PM processes. Thus, we modified the paradigm to increase the power of our model fitting by keeping the PM target trial to ongoing task trial ratio within the bounds of previous literature (1:14), but presenting participants substantially more experimental trials (~4000 trials over 3 days). Second, we instructed participants to make their PM response instead of their ongoing task response, as we have done in our previous computational modeling efforts (Heathcote et al. 2015; Strickland et al., 2017). We used this measure of PM because it is relevant to common everyday errors, in which individuals perform a routine task action instead of a less common but required deferred task action (Norman, 1981; Reason, 1990). In addition, this means we record only one RT and one response on every trial, allowing us to fit PM RTs without confounds from the production of other responses. Third, we modified the response key arrangements from the typical paradigm. Usually in the PM literature participants rest their fingers on the ongoing task keys and have to make a larger movement for the PM response. This would cause uneven non-decision time between responses (via motor response production time), which adds needless complexity to the model that is not relevant to understanding PM processes, and neglecting substantial differences in non-decision times between responses can lead to biased estimates of other model parameters (Voss, Voss, & Klauer, 2010). Hence, we asked participants to rest their fingers on both ongoing task response keys (with one hand), and on the PM response key (with the other hand), so we could assume an equal motor response time.

In Experiment 1, we used a repeated measures design in which participants completed a lexical decision ongoing task under three conditions: control (no additional PM task), focal PM (the PM task was to respond to a single target word) and non-focal PM (the PM task was to respond to any word within a PM target category). The task to make PM responses to a single target word is one of the most commonly employed focal PM demands (see Einstein & McDaniel, 2010). Relative to the single target task, the task to respond to any member of a target category is non-focal, because in order to detect PM targets participants need not only read the word (as is necessary for lexical decision and the single target task), but also map the stimulus to the relevant semantic category. In line with this, PM accuracy is found to be be higher to the focal single target task than the non-focal categorical task, with lower PM cost (e.g., Loft & Remington, 2013; Strickland et al., 2017).

Our design is very similar to that in Strickland et al. (2017), in which we modeled responding to non-PM trials and found that both focal and non-focal PM costs were attributed to increased thresholds, rather than changes in ongoing task evidence accumulation rates. We expect to replicate these previous findings. Our PM instructions were "stimulus-specific" (Lourenço et al., 2013). That is, PM targets could only be one type of ongoing task item (words), and participants were informed of that. As outlined above, PMDC predicts that under these conditions the word threshold will increase to allow the PM process more time to complete on PM trials (selective delay). Typically, non-focal PM cost is higher than focal PM cost. Thus, under PMDC, we expect to find a greater degree of ongoing task threshold increases in the non-focal PM condition than the focal PM condition.

In the current work, we directly assess the PM process by modeling the PM trials. Thus, we can examine the PM threshold, and how this differs between focal and non-focal PM tasks. It is not clear at the outset how the two thresholds would differ. For example, the non-focal PM instructions may create the impression of a globally harder task, and cause the PM threshold to increase. Alternatively, participants may reduce their non-focal PM threshold below the focal PM threshold if they expect slower evidence accrual and wish to compensate.

We also examine reactive control on PM trials, and how this differs as a function of target focality. Trivially, we expect a faster PM accumulation rate on PM trials in PM blocks than non-PM trials in PM blocks, as PM trials contain more evidence that participants should make a PM response than non-PM trials. In addition, reactive control may cause the ongoing task accumulation rates to decrease on PM trials as compared with non-PM trials. In our paradigm, detection of non-focal PM targets requires adequate semantic encoding of PM stimuli to match them to a category, whereas detection of focal PM targets does not. The simpler requirements of the focal PM task may support efficient extraction of PM signal from the PM-related stimulus attributes. Thus, in terms of the PMDC, focal PM conditions should be associated with increased input to the PM detector on PM trials. This could lead to a faster PM accumulation rate for the focal PM condition. In addition, focal PM targets may trigger more reactive inhibitory control than non-focal PM targets. This could leader to slower accumulation towards ongoing responses on PM trials in the focal PM condition.

Method

Participants

The University of Western Australia's Human Research Office approved both Experiment 1 and Experiment 2. For both experiments, all participants were students, who received course credit for their participation. For both experiments, the upper age limit for participation was 35, and English as a first language (the language spoken in the childhood home) was required. For both experiments, participants performed three one-hour sessions, with each session on a separate day. In Experiment 1, two participants were excluded (see results), with 35 participants remaining (19 females) with ages ranging from 17-34 (mean = 20.46).

Materials

Both of our experiments were programmed using E-prime (Schneider, Eschman, & Zuccolotto, 2002). In Experiment 1, one thousand nine hundred and eighty 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). A non-word was created from each word by replacing every vowel with a randomly chosen alternate vowel (e.g., chemist to chamust). All 1980 non-words were presented once to all participants, 1812 of the words (randomly selected) were presented once to all participants. Twenty nine low frequency words [1] from each of 3 categories (animal, food, part of the human body) were also selected from the TMSH database to be PM targets.

Participants performed 9 blocks of 440 trials - 3 non-focal blocks, 3 focal blocks and 3 control blocks -1 block of each type a day. Block order was balanced across days so that participants would not get a condition in the same position twice, and the 12 orders that satisfy these conditions were approximately counterbalanced across the 35 participants. In control blocks, participants were presented with 220 non-words and 220 words. In non-focal blocks, participants were presented with 220 non-words, 192 non-target words and 28 PM target words from one of the PM categories (e.g., 28 different animal words). In focal blocks, participants were presented with 220 non-words, 192 non-target words and 1 PM target word (e.g., 'giraffe') was presented 28 times. Each PM category was in one non-focal block for each participant. For each participant, one focal PM target word was drawn randomly from each of the category lists, and if a word was to be presented as a PM target in the focal PM block then it was not presented in the non-focal PM blocks. The assignment of PM category to each day's non-focal block, and of which category the word from the focal block was drawn from was random (without replacement) except that the focal word for a given day was never from the non-focal PM category of that day. Each of the 29 words in each category was used as a focal PM target for one of the first 29 participants, and 7 words were randomly redrawn (without replacement) from each category for the remaining 7 participants. The order of lexical decision stimuli was randomized across blocks. Participants were given three 1 minute long breaks within each block - one after each 110, 220 and 330 trials were completed- and PM targets were presented 28 times per 440 trial block; on trials 6-20, 21-35, 36-50, 51-65, 66-80, 81-95, 96-110 of each quarter. Target trials were separated by at least 2 lexical decision trials.

Procedure

For the lexical decision task, participants were instructed that they would be presented with letter strings and that they should press a key to indicate whether strings were words or non-words (e.g., press 's' for word, 'd' for non-word). Participants were instructed to make their responses as quickly and accurately as possible. Depending on the PM block, participants received either control, focal or non-focal PM instructions. Before control blocks, participants were instructed that they only needed to make lexical decision responses for that block. Before focal PM blocks, participants were instructed to press an alternative key instead of their word response when they saw a specific target word, for example, *"In the next block of lexical decision trials, if you see the word 'jelly' then press 'd' instead of 'j"'*. Before non-focal PM blocks, participants were instructed to press an alternative key instead of their word response when they saw any word within a category, for example, *"in the next block of lexical decision trials, if you see ANY word that is an ANIMAL then press 'd' instead of 'j"'*. Four response key assignments were counterbalanced across participants; 1) s = word, d = non-word, j = PM, 2) d = word, s = non-word, j = PM, 3) k = word, j = non-word, d = PM, and 4) j = word, k = non-word, d = PM). The key order was the same for each participant across the entire experiment. Participants were instructed before the commencement of each quarter of a block (including control blocks) to rest their fingers on their designated response keys.

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 the presentation of a white letter string which remained on the black screen until the participant pressed any key. If the participant made a correct word or non-word response (including a correct lexical decision response on a PM target trial, which is a PM miss), or a correct PM response, the next trial immediately began (next fixation cross). If the participant made an incorrect response the word 'INCORRECT' was presented in white for 1s, after which the subsequent trial would begin (next fixation cross).

Each day, participants first completed 24 practice lexical decision trials and received summative feedback on the accuracy of their responses (e.g., "87.50% correct"). Participants then proceeded to the experimental blocks and were presented with either control, focal, or non-focal PM instructions. Participants next completed a 3 minute distractor puzzle, after which they began the first block of experimental trials. After completion of each quarter of a block, participants were presented summative feedback on the accuracy of their responses. In order to avoid cueing the PM intention, both a correct ongoing task response and a correct PM response counted towards the 'correct' status of trials in this summative feedback. In addition to the breaks within blocks, participants were instructed to rest for 2 minutes between blocks.

Results

We first report conventional analysis to check whether our manipulations had the expected effects. Two participants were excluded, one who made no 'word' lexical decision response for an entire block, and one for which there was a computer error. The first two trials of each block, and the first two trials after each rest period, were excluded from further analyses. The two trials following each PM trial, and following any PM false alarms, were excluded from further analysis. We also excluded trials where participants responded with a key that was not designated to indicate their PM or lexical-decision responses (0.03% of trials), and trials with

[1] The human body category contained 3 words of frequency 1 per million and 2 of frequency 7 per million.

outlying RTs, defined as less than 0.2s or 3 times the interquartile range / 1.349 (a robust measure of standard deviation) above the mean (4.65% of responses overall). Two participants responded once each with the PM response key in the control condition; these trials were excluded from further analyses.

In addition to stimulus type (word, non-word) and PM block (focal, non-focal, control) the subsequent analyses included a day factor (day 1, day 2, day 3) to capture any effects of task repetition. In our omnibus significance testing for accuracy effects we used generalized linear mixed models with a probit link function. In our omnibus significance testing for mean correct RTs we used general linear mixed models. Significance was assessed with Wald's chi-square tests, and an alpha level of 0.05 was used in all analyses. The results of our omnibus analyses are tabulated in the supplementary materials. All standard errors reported in text and displayed in graphs were calculated using the Morey (2008) bias corrected method.

Non-PM Trials

Accuracy was higher for non-words (95.3%) than words (90.5%), and decreased slightly over days (day 1 M = 93.6%, SE = 0.71%; day 2 M = 93.0%, SE = 0.74%, day 3 M = 92.1%, SE = 0.92%). There was an interaction between PM block and stimulus type. Planned comparisons revealed that non-word accuracy was higher in both the non-focal condition (M = 95.5%, SE = 0.45%), t (34) = 2.19, p = 0.04, d = 0.37, and the focal condition (M = 95.5%, SE = 0.43%), compared to the control condition (M = 94.9%, SE = 0.48%), t (34) = 2.19, p = 0.04, d = 0.37. In contrast there was no significant difference in word accuracy between the control condition (M = 90.9%, SE = 0.58%), and the non-focal PM (M = 90.5%, SE = 0.68%), t (34) = 1.87, p = 0.07, or focal PM (M = 90.1%, SE = 0.79%), t (34) = 1.04, p = 0.31, conditions.

Mean correct ongoing task RTs are displayed in Figure 3. RTs decreased over days (day 1= 0.682s, day 2= 0.641s, day 3= 0.619s). Participants responded faster to non-words (0.640s) than words (0.654s). This effect interacted with day; correct responses to non-words sped up more over days than correct responses to words. Correct responses were fastest in the control blocks (0.633s), intermediate in the focal blocks (0.643s), and longest in the non-focal blocks (0.665s). The effects of PM block and stimulus type interacted. Planned contrasts revealed that non-word RTs increased under non-focal conditions (0.652s) compared with control (0.635s), t (34) = 5.03, p < .001, d = 0.85, but not for focal conditions compared with control (0.634s), t < 1. Word RTs were slower than control (0.631s) under focal conditions (0.652s) t (34) = 5.29, p < .001, d = 0.89, and slower than control under non-focal conditions (0.679s), t (34) = 11.14, p < .001, d = 1.88. Thus, we observed stimulus-specific PM costs in the focal PM condition, and a much larger effect size for costs to words than non-words in the non-focal condition.

PM Trials

PM responses were scored as correct if the participant pressed the PM-response key instead of a lexical decision response key on the PM target trial. PM accuracy was higher for focal PM blocks (M = 87%, SE = 3%), than non-focal blocks (M = 60%, SE = 3%), and decreased over days (day 1 M = 78%, SE = 4%; day 2 M = 74%, SE = 3%, day 3; M = 70%, SE = 4%). Thus, we observed the expected effect of target focality on PM accuracy. However, note the magnitude of this effect varied widely across participants: for some participants there was a major shift in PM accuracy, and for others much less of a shift (Figure 4). The average false alarm rate was very low, 0.11%, and differed little between participants, ranging from 0 to 0.3%. Correct PM responses were faster in focal PM blocks (M = 0.653s, SE = 0.010s) than in non-focal PM blocks (M = 0.764s, SE = 0.014s), and there was an effect of day (day 1 M = 0.724s, SE = 0.016s; day 2 M = 0.705s, SE = 0.014s; day 3 M = 0.697s, SE = 0.013s).

Non-PM trials compared with PM trials

It is possible that reactive control over ongoing task decisions could lead to slower ongoing task RTs on PM trials in PM blocks, as compared with non-PM trials in PM blocks. Thus, in order to check whether reactive control was evident without the model-based analysis, we compared correct ongoing task RTs on PM trials to correct ongoing task trials. That is, RTs to make the 'word' response on PM trials were compared with RTs to make the word response on non-PM trials (in the PM blocks). Note that these 'correct' ongoing task responses on PM trials are PM misses. We ran a linear mixed effects model including the effects of PM trial status (PM trial vs non-PM trial), PM block, and day. Word RTs were significantly faster on PM trials (0.575s) than on non-PM trials (0.665s). There was an interaction between PM trial status (PM, non-PM), and PM block. Word RTs on PM trials were faster than on non-PM trials in both focal (non-PM trial M = 0.652s, SE = 0.008s; PM trial M = 0.527s; SE = 0.019s), t (33) = 9.11, p < .001, d = 1.56, and non-focal blocks (non-PM trial M = 0.679s, SE = 0.008s; PM trial M = 0.617s; SE = 0.010s), t (34) = 9.37, p < .001, d = 1.58, but the speed-up on PM trials was greater for the focal blocks, t (33) = 5.01, p < .001, d = 0.86. Note, however, that reactive control on PM trials is confounded in raw RT by statistical facilitation from the PM response. Thus, the critical test of reactive control is on accumulation rates, as presented in the modeling section below.

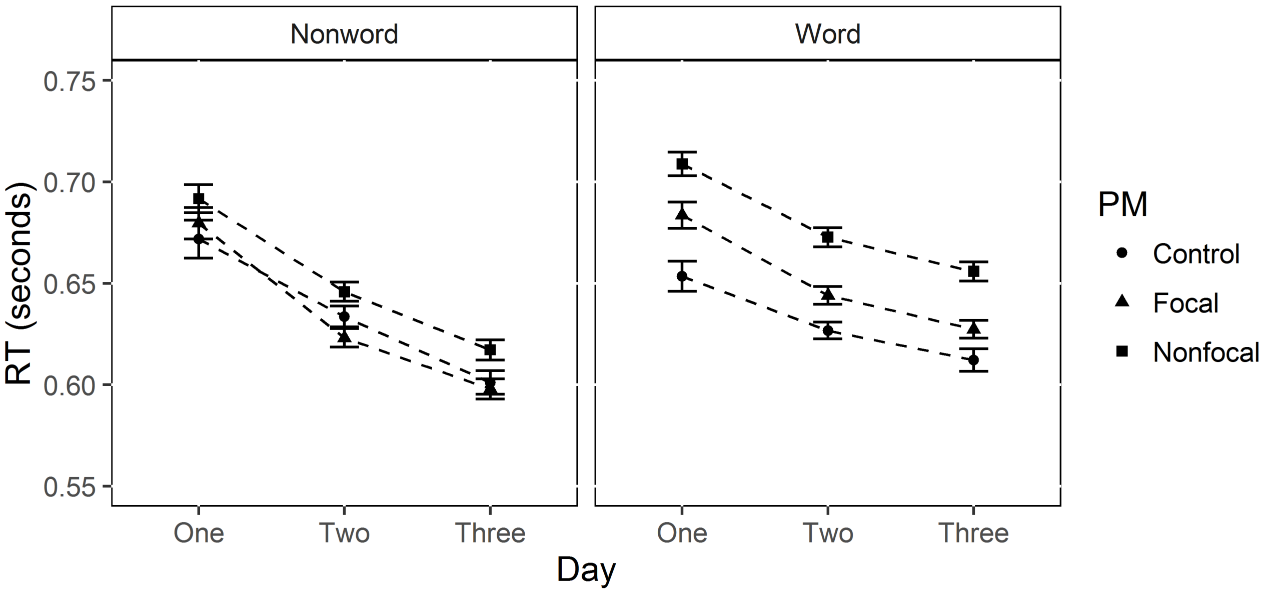

Figure 3. Experiment 1, ongoing lexical decision task mean correct RTs by stimulus type by PM block by day. The standard error bars were calculated using the Morey (2008) bias corrected method.

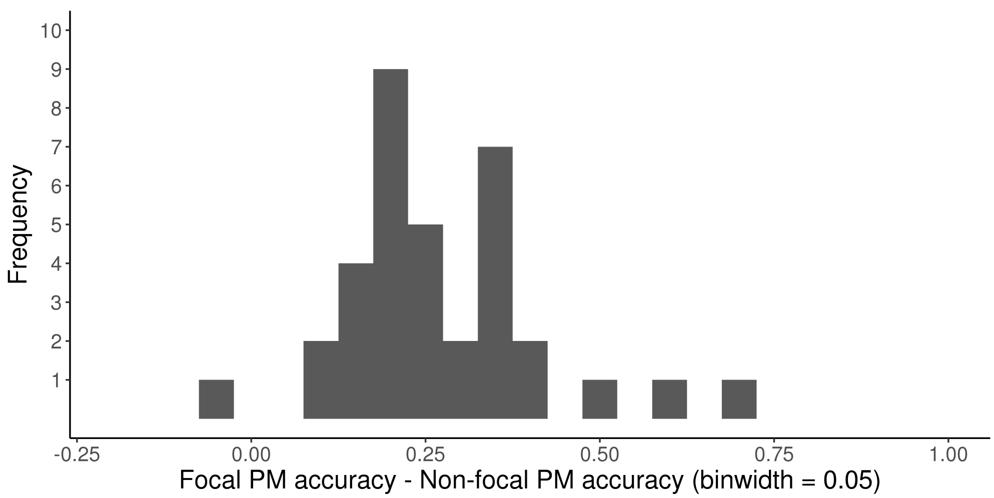

Figure 4. Frequency histogram demonstrating the spread across participants of the effect of PM focality on PM accuracy. Positive values demonstrate an advantage for focal PM accuracy. For most participants, focal PM accuracy was higher, but the magnitude of the effect varied substantially.

Model Analysis

The PMDC model of the current experiment (Figure 1) includes several parameters for each accumulator. Each accumulator begins a decision trial with a starting amount of evidence drawn from the uniform distribution [0, *A*]. After stimulus presentation, evidence accrues linearly according to an accumulation rate that is drawn from a normal distribution with mean *v* and standard deviation *sv*. Evidence accrues until some response reaches a threshold *b*. We report threshold in terms of *B* (*B* = *b* - *A*); note that as *A* is the same across all conditions, differences in *B* across conditions reflect pure threshold effects. Finally, there is the non-decision time parameter *t0*, which captures additional RT that falls outside decision time.

The experiment includes several factors that model parameters can vary over, including latent response (i.e., word, non-word and PM accumulator), and three manifest factors, stimulus type, day and block type. The latent response factor refers to the accumulators that can lead to each response, that is, ‘word’, ‘nonword’, and ‘PM’. It is important to be clear that the latent response factor corresponds to the accumulators, and not the response that was actually observed; the observed response is predicted by, not included in, the model. The stimulus type factor has three levels: non-PM words, non-PM non-words, and PM (always words). The 'day' factor – included to capture task repetition effects such as learning to tune the response threshold – had three levels: day one, day two, day three. Finally, the blocked PM manipulation had three levels: control blocks, focal PM blocks, and non-focal PM blocks. Note that because there were no PM trials (and thus no PM responses) in control blocks of our experiments, we modeled control performance with only 2 accumulators, not three.

We applied some sensible *a priori* constraints on the model to reduce its complexity. First, following common practice with the LBA (Donkin, Brown, & Heathcote, 2011; Heathcote, Loft, & Remington, 2015), we estimated only one *A* parameter for each participant. Second, we allowed the *sv* parameter to vary over the stimulus and latent response factors, but not PM block. This is more flexible than most previous LBA modeling, which only allows *sv* to vary as a function of whether the latent accumulator matches or does not match the stimulus, because in our current model there are two types of "correct" response for PM trials (correct PM and correct lexical decision). Third, we varied only the response threshold parameter over days of the experiment. This is consistent with Strickland et al. (2017), in which we found that day's effects were expressed in shifts in the *B* parameter. Fourth, we estimated only one non-decision time parameter for each participant. We follow previous modeling in assuming that non-decision time is constant across trials (e.g., Heathcote, Loft, & Remington, 2015; Horn & Bayen, 2015; Strickland et al., 2017), and we do not allow non-decision time to vary over PM block because it has not varied in previous LBA modelling of PM paradigms (Heathcote, Loft, & Remington, 2015; Strickland et al., 2017). Finally, because there were very low numbers of PM false alarms, we pooled estimates of false alarm rate parameters (both *v* and *sv*) over all experiment factors. Furthermore, the false alarm *sv* (i.e., the *sv* for the PM accumulator on non-PM trials) was fixed at 1, as a scaling parameter. Even with these restrictions, the model had fifty-two free parameters: one non-decision time, one *A*, twenty-four *B* (PM block $\times$ latent response accumulator $\times$ day), nineteen *v* (stimulus type $\times$ PM block $\times$ latent response accumulator), and 7 *sv* (stimulus type $\times$ latent response accumulator) parameters.

Sampling

As opposed to previous studies, which relied on maximization to obtain point estimates of the parameter values (Heathcote, Loft, & Remington, 2015; Strickland et al., 2017; Horn & Bayen, 2015), we used Bayesian techniques to estimate the entire probability distribution of the parameters. We opted for Bayesian estimation because it is better equipped than previous maximum likelihood approaches to handle the relatively sparse PM trials. In terms of model selection, previous maximum likelihood efforts have penalized model complexity with a mere count of parameters. This is clearly inappropriate for the full PM data set, where adding a PM-related parameter (e.g., an extra PM accumulation rate) would contribute to flexibility of fit on far fewer trials than adding a non-PM trial related parameter (e.g., an extra ongoing task threshold). Our Bayesian selection will, in contrast, punish model complexity proportionately to the flexibility that extra parameters add to the fits. In terms of estimation, Bayesian fitting will estimate uncertainty in parameter estimates, which can then be incorporated into our tests of parameter differences. This is particularly important for PM trial parameters, for which data is relatively sparse.

We could have fit a hierarchical model that assumes common population distributions of parameters across all participants. However, as our design included thousands of trials per participant, we had enough within-participant power to rely on separate parameter estimation for each participant. We did so for three reasons. First, as mentioned in the standard analysis section, the effect of target focality on PM accuracy was substantially dispersed across participants. As this is the first time that our model has been fit, we did not have knowledge of the appropriate form of population distributions. Thus, we were concerned that inappropriate assumptions combined with hierarchical shrinkage effects could cause undesirable biases. Second, fitting a hierarchical model would complicate exploring individual differences across participants, due to the aforementioned shrinkage. Third, the Bayesian analysis was already computationally demanding for individual fits, and would be much more so in the hierarchical case. Bayesian analysis requires the researcher to specify their prior beliefs about the probabilities of the model parameters. However, note that because of our large sample sizes and use of posterior based inference the influence of the priors was small, and we selected fairly uninformative prior distributions (Table 1). The priors for control, focal and non-focal PM conditions were the same.

We estimated posterior parameter distributions using the differential evolution MCMC algorithm (Turner, Sederberg, Brown, & Steyvers, 2013), which is more effective than standard MCMC for dealing with the high level of parameter correlations in evidence accumulation models. The number of chains was three times the number of parameters (e.g., for a 52 parameter model 156 chains per parameter). The chains were 'thinned' by 20 (i.e., only one iteration in every 20 was saved). We continued to sample for each participant until a small Gelman's multivariate potential scale reduction factor (<1.1, Gelman et al., 2013), calculated with the number of chains doubled by considering the first and second halves of each as separate chains (Gelman et al., 2013), indicated convergence, stationary, and mixing. This was confirmed with visual inspection. We retained the same number of samples for each participant: each chain was 180 iterations long, and thus with 156 chains there were 28,080 posterior samples (iterations x chains) for each parameter for each participant.

Model Results

In order to evaluate fit, we sampled 100 posterior predicted data sets for each participant. As shown in Figure 5, the model provided a good fit to both ongoing task and PM accuracies, as well as to the observed ongoing task and PM RT distributions. It fitted the PM cost effect, the effect of focality on PM accuracy, and the effect of focality on PM RTs. In their simulations, Gilbert et al. (2013) explored the effect of competing PM and ongoing task decisions on the coefficient of variation of RT. Their model predicted that coefficient of variation would be highest on correct non-PM trials, intermediate on PM miss trials, and lowest on PM hit trials. For non-focal PM conditions, we found such a trend. For focal PM conditions, we did find coefficient of variation was higher for non-PM trials than PM trials. However, coefficient of variation was similar between focal PM 'hits' and PM 'misses'. In the supplementary materials we plot these trends, and show they corresponded reasonably well to the predictions of our PMDC model. In sum, overall the PMDC architecture fitted the trends in Experiment 1's data closely. We now focus on inference from the model by assessing how model mechanisms varied across PM conditions.

Model Selection. We first applied model selection to determine whether we could statistically justify constraining parameters across our blocked conditions (Control/ Focal PM/ Non-focal PM), and in order to determine the necessity, and the importance, of different parameters in capturing the effects. For this purpose we used WAIC (Watanabe, 2013), a computationally tractable approximation to cross validation. WAIC balances goodness of fit and model complexity, with the aim that the model with the lowest WAIC should be the best at fitting new data. We determined whether a difference in WAIC is sufficient to provide evidence in favor of the model with the lowest WAIC by comparing that difference with its standard error. As a rule of thumb a standardized difference (i.e., the WAIC difference divided by this standard error of the difference) larger than two indicates support for the model with the lower WAIC (Vehtari, Gelman, & Gabry, 2017)

Table 2 displays the results of model selection. We sequentially tested the importance of parameters related to PMDC, that is thresholds and PM trial accumulation rates. Then we tested the importance of the remaining parameters of the top model, which were related to differences in non-PM trial accumulation rates across blocks.

We began by comparing the overall contribution of thresholds vs accumulation rates. We tested a 'No Proactive Control' model – where accumulation rate parameters could vary freely among the PM and control conditions, but thresholds could not – against an 'Only Proactive Control' model, in which both the ongoing task and PM thresholds could vary over PM conditions, but accumulation rates could not. The WAIC for the Only Proactive Control model was much lower than the WAIC for the No Proactive Control model. Given the difference had a standard error of 125, the difference in standard units is very large, 20.66. This clearly indicates that variation in thresholds is a very important part of the account of differences among the PM and control conditions.

Next, we allowed accumulation rate of the PM response on PM trials to vary between Focal and Non-focal conditions, to allow for differences in 'excitatory' reactive control. Relative to the Only Proactive Control model, the WAIC for this model was lower by a reasonable margin, 66.90, standard error of the difference = 23.82, difference in standard units = 2.81. The improvement in fit in this case, although still substantial, was much smaller than for the previous comparison, suggesting a much more moderate influence of excitatory reactive control on our data. However, it is important to keep in mind that reactive control, by definition, can only influence performance on PM trials, and because such trials are rare reactive control can only exert a limited effect on overall fit.

Next we allowed only the ongoing task accumulation rates on PM trials to vary between Focal and Non-focal conditions, to capture differences in 'inhibitory' reactive control between the conditions. We compared this model to the previous 'excitatory' reactive control model. We found that the inhibitory only reactive control model had a WAIC significantly lower than the excitatory only model, by 245.13, with a standard error of the difference = 43.56. This difference, at 5.63 standard units, was quite substantial. This suggests that for fitting the effects of focality, it is more important to allow the ongoing task accumulation rates to vary on PM trials than it is to allow the PM accumulation rate to vary. However, a comparison of the best model so far – the proactive and reactive (inhibitory) control model – to the full PMDC model (which also allows excitatory reactive control) showed that the full PDMC model was preferred, with a WAIC difference = 87.34, standard error of the difference = 23.59, and standardized difference = 3.70, supporting the need for both types of reactive control.

Finally, we compared the PMDC model to the top model, which also allows non-PM trial accumulation rates to vary over PM condition. Relative to the PMDC model, the WAIC for the top model was lower by a solid margin, 195.11 with a standard error of 45.39. Although the standardized difference in this case is on the moderate side, 4.30, compared to, say, the difference in fit provided by allowing thresholds to vary, it is still clearly sufficient to support the presence of different rates among the control and PM conditions in the rates for non-PM trial accumulation.

Table 1
Both experiments, choice of priors for each parameter

Model Parameter	Distribution	Mean	SD	Lower	Upper
A	Truncated Normal	1	1	0	10
B	Truncated Normal	1	1	0	None
v (Correct Ongoing Task Response)	Truncated Normal	1	2	None	None
v (Incorrect Ongoing Task Response)	Truncated Normal	0	2	None	None
v (Correct PM Response)	Truncated Normal	1	2	None	None
v (PM false alarm)	Truncated Normal	-1	2	None	None
sv	Truncated Normal	1	1	0	None
t0	Uniform			0.1	1

Table 2
Experiment 1, WAIC model selection. To evaluate each model using data from the entire group of participants, we first concatenated the log-likelihoods under the model for each trial, for all participants, together into one pointwise log-likelihood matrix where points are trials. We then calculated WAIC for each model using its log likelihood matrix. Lower WAIC indicates more preference for the model.

Model	Estimated Number of Parameters	WAIC
Top	1035.5	-110467.4
Proactive Control & Reactive Control (both)	898.1	-110272.3
Proactive Control & Reactive Control (inhibitory)	886.4	-110184.9
Proactive Control & Reactive Control (excitatory)	884.4	-109939.8
Only Proactive Control	862.5	-109872.9
No Proactive Control	753.2	-107290.4

Given that the WAIC appeared to always favor more complexity, it was not diagnostic regarding our competing cognitive theories. Although the WAIC is currently the closest computationally tractable approximation available to the gold standard of leave-one-out cross validation (Vehtari et al., 2017), and is recommended to be used above other Bayesian model selection such as the deviance information criterion (Gelman et al., 2013), it remains possible that it did not select the most parsimonious model of our data. For example, Millar (2017) demonstrated poor performance of the WAIC even in correctly selecting models that are relatively simple in comparison to our fifty-two parameter LBA. In light of our results, in the next section we test the direction and magnitude of differences among conditions in the parameters of the top model. Testing directions of differences is particularity critical, as PM theories not only predict differences but also which condition should have larger parameter values. For example, capacity sharing predicts non-PM trial accumulation rates should be greater in the control condition than the PM conditions.

Model Summary

In order to summarize the central tendency of model parameters across participants, we obtained a subject-average posterior distribution, by calculating the mean of each posterior sample over all participants. Our primary interest focused on mean rate and threshold parameters, which we examine in the next sections. The other parameters had reasonable values. The non-decision time mean of the averaged posterior samples was 0.13s (posterior *SD* = 0.002). The *A* posterior mean was 0.34 (posterior *SD* = 0.013). The *sv* parameters are summarized in Table 3. As is typical in other applications of the LBA (Heathcote & Love, 2012), *sv* is lower for the accumulator that matches the stimulus (e.g., the word accumulator for a word stimulus).

Table 3
Experiment 1, Mean (SD) of the posterior distributions, averaged across participants, of the standard deviation of the evidence accumulation rates.

Accumulator	*Stimulus Type*		
	Word	Non-word	PM
Non-word	1.51 (.04)	0.65 (0.01)	1.08 (0.04)
Word	0.65 (0.01)	1.55 (0.03)	1.49 (0.04)
PM	Fixed at 1		0.39(0.01)

We tested the direction, and magnitude, of differences in top-model parameters across experimental conditions in order to evaluate support for capacity sharing, proactive control, and reactive control. To do so, we calculated the posterior distribution of the differences. For example, to test the difference between the non-PM trial accumulation rates in control and PM conditions, and hence the capacity-sharing theory prediction that the control rate should be

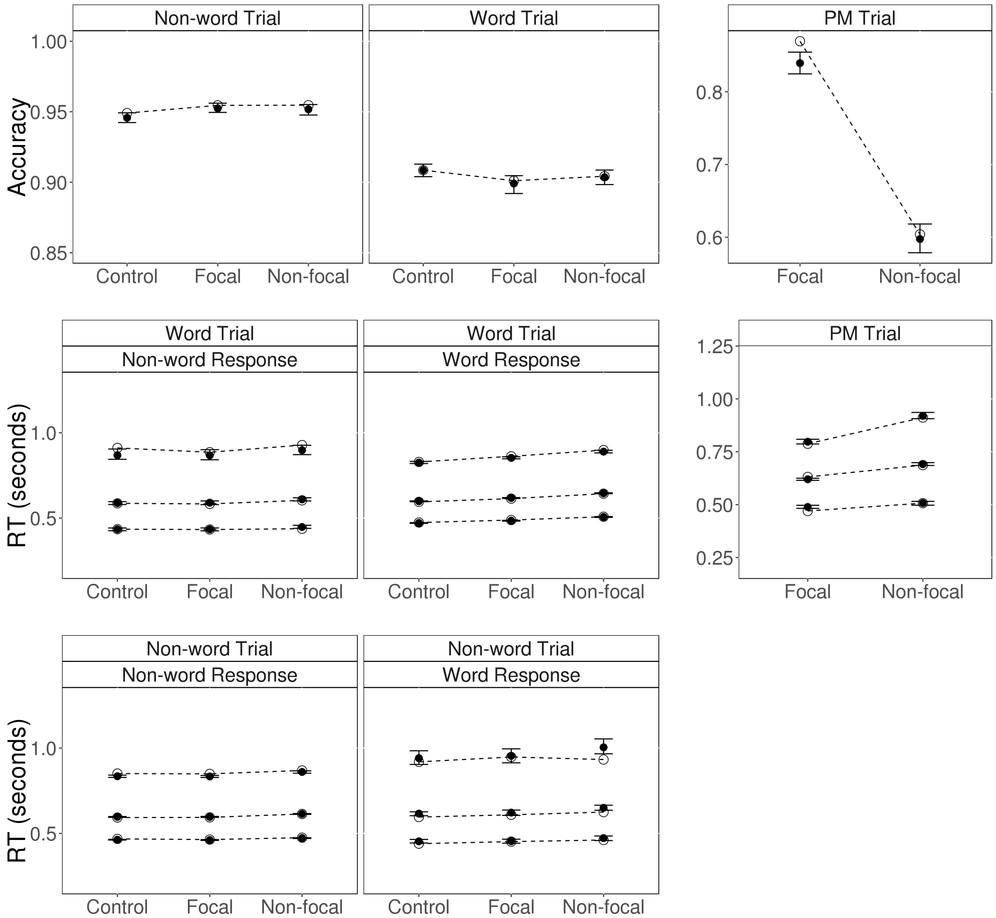

Figure 5. Fits of the LBA to the Experiment 1 data. The open circles indicate the data averaged over participants. The black dots indicate the posterior predictive mean, with 95% credible intervals. The RT distributions are summarized with three order statistics (the three lines on each RT graph): the 10^{th} percentile, which captures the leading edge of the distribution (i.e., the fastest responses), the median, and the 90^{th} percentile, which captures the tail of the distribution (i.e., the slowest responses). We calculated the plotted statistics (accuracies and quantile RTs) by concatenating the observed data from all participants into one data frame, and then calculating the statistics for that entire data frame. We used this same procedure for both the observed data and the simulated data. The alternative is to calculate statistics on a participant-by-participant basis (e.g., calculate a median RT for every participant), and then average. The approaches yielded similar results. Due to the low number of PM trials, PM RTs are collapsed over correct and error responses.

larger, for every posterior sample we subtracted the control rate from the rate in the focal PM condition, and similarly for the non-focal PM condition. We calculated these distributions independently for each participant, and then we averaged each posterior sample across participants. For each participant-average difference distribution we report a Bayesian posterior-predictive p value (Meng, 1994). These p values indicate the one-tailed probability that the difference between parameters was less than 0. Due to our powerful design, many of our observed parameter differences had p = 0 (i.e., a probability of 1 that there was an effect), yet some of the differences were much larger than others. Thus, to illustrate the magnitude of the effects we also report the standardized difference between parameters, that is, M / SD of the difference distribution. As our posterior parameter distributions are approximately normal, the interpretation of this statistic is similar to a Z score, and thus we refer to the score as Z from here on.

Capacity Sharing (Non-PM Trial Accumulation). The non-PM trial rates for the ongoing task accumulators (non-word and word) are displayed in the left two panels of Figure 6. The capacity sharing theories of PM predict lower accumulation rates towards the correct accumulator (i.e., the one that matches the ongoing stimulus) in PM than control conditions (e.g., Horn & Bayen, 2011; Boywitt & Rummel, 2012), which would lead to faster correct responses in the control condition. They also predict lower correct rates in the harder non-focal PM condition than the easier focal PM condition. Capacity sharing theories may also predict higher ongoing task rates for the error accumulator (i.e., the one that does not match the ongoing stimulus) under PM conditions, which would lead to more errors (and, similarly, a prediction of higher error accumulation rates for non-focal than focal PM).

The effect sizes and p values for comparisons between non-PM trial accumulation rates are displayed in Table 4. Inconsistent with capacity sharing predictions, correct accumulation to non-PM words was actually higher for focal PM blocks (2.94) than control blocks (2.88), and was highest for non-focal blocks (3.03). Error accumulation rates for non-PM words were similar between focal PM (-0.29) and control (-0.27) blocks, and actually lower for non-focal PM blocks (-0.49) than control blocks. Thus, accumulation rates to non-PM word trials shifted in the opposite direction to that predicted by capacity sharing theories. Similarly, correct accumulation to non-PM non-words was marginally higher in focal PM blocks (2.85) than control (2.82), but similar between non-focal PM blocks (2.81) and control. Error accumulation to non-word trials was similar for focal PM blocks (-0.97) and control blocks (-0.98), but higher for non-focal PM blocks (-0.69) than control. The latter effect (an increase in the error accumulation rate) is the only one that is consistent with capacity sharing, and so the overall pattern of results convincingly refutes the predictions of capacity sharing theories.

Proactive Control (Thresholds). Proactive control over ongoing task decisions (as in delay theory) predicts higher ongoing task decision thresholds in PM blocks than in control blocks. The effect sizes and p values for our comparisons of thresholds are displayed in Table 5. As shown in Figure 7, word thresholds were higher in focal PM blocks (1.30) than control blocks (1.21), and higher in non-focal PM blocks (1.43) than either focal PM blocks or control blocks. Non-word thresholds differed by much less. They were not substantially higher in focal PM blocks (1.19) than control blocks (1.18). They were, however, larger in non-focal PM blocks (1.21) than both control and focal blocks. The weaker effects on non-word accumulator thresholds suggest that the proactive control was targeted, rather than a generic result of increased task difficulty (recall that non-words could never cue a PM response). There was also an effect on the PM threshold. The PM threshold was lower in focal PM blocks (1.11) than in non-focal PM blocks (1.31). All of these effects are consistent with proactive control, both of the ongoing task and PM decisions.

Reactive Control (PM vs. Non-PM Trial Accumulation). Examining reactive control requires comparing accumulation rates between non-PM trials (left two panels of Figure 6) and PM trials (rightmost panel of Figure 6). The effect sizes for comparisons relevant to reactive control are displayed in Table 6. As expected given the rarity of PM false alarms, PM accumulation rates on PM trials (right panel of Figure 6) were much faster than PM false alarm accumulation rates, that is, accumulation rates towards a PM response on non-PM trials, (M = -2.42). More importantly, PMDC predicts that stimulus characteristics, such as target focality, could increase excitation of the PM accumulator, driving up its rate. In line with this prediction, the PM accumulation rate on PM trials was higher in the focal PM block (2.48), than the non-focal PM block (2.30).

PMDC predicts that the lexical (ongoing task) accumulation rates will be reduced on PM trials, compared with non-PM trials, due to inhibitory control of the ongoing task decision from the PM detector. In line with this prediction, accumulation towards the ‘word’ response (non-PM word trials) was much lower on focal PM trials (0.49) than on non-PM word trials in focal PM blocks (2.94). The same rate suppression on PM trials also occurred in non-focal PM blocks, but to a lesser degree (PM trial word accumulation = 1.64, non-PM trial word accumulation = 3.03). The finding that non-focal reactive inhibition was weaker than focal reactive inhibition is consistent with PMDC, which predicts that a lower input to the PM detector from the non-focal stimuli would lead to weaker reactive control, and therefore higher ongoing task accumulation. The reactive inhibitory control effect was weaker for the incorrect (i.e., non-word) accumulator, but was again present to a greater degree in focal blocks. In focal PM blocks with word stimuli non-word accumulation was reduced on PM trials (-1.56) compared to non-PM trials (-0.29). There was a trend towards the same effect in non-focal PM blocks (PM = -0.57, non-PM = -0.49), however it was much weaker. This suggests that reactive control was targeted towards the word accumulator, which was most likely to cause a failure of control in the form of a word response that preempts a PM response.

Parameter Recovery

We performed a ‘parameter recovery’ study (Heathcote, Brown, & Wagenmakers, 2015) to confirm that PMDC can be treated as a ‘measurement model’, a model whose estimated parameters reliably index the processes they control. This involves simulating data from known parameter values, then estimating parameters from that simulated data to examine whether the fitted estimates match the ‘true’ data-generating values. In particular, we assessed whether we could adequately recover realistic parameter values for individual subjects given our design, priors, and sampling settings. We simulated from the top model with parameters equal to the posterior mean values reported above. We simulated an individual subject, using for each cell of the design the average number of trials (rounded up) that we obtained per participant (thus accounting for loss of power due to ex-

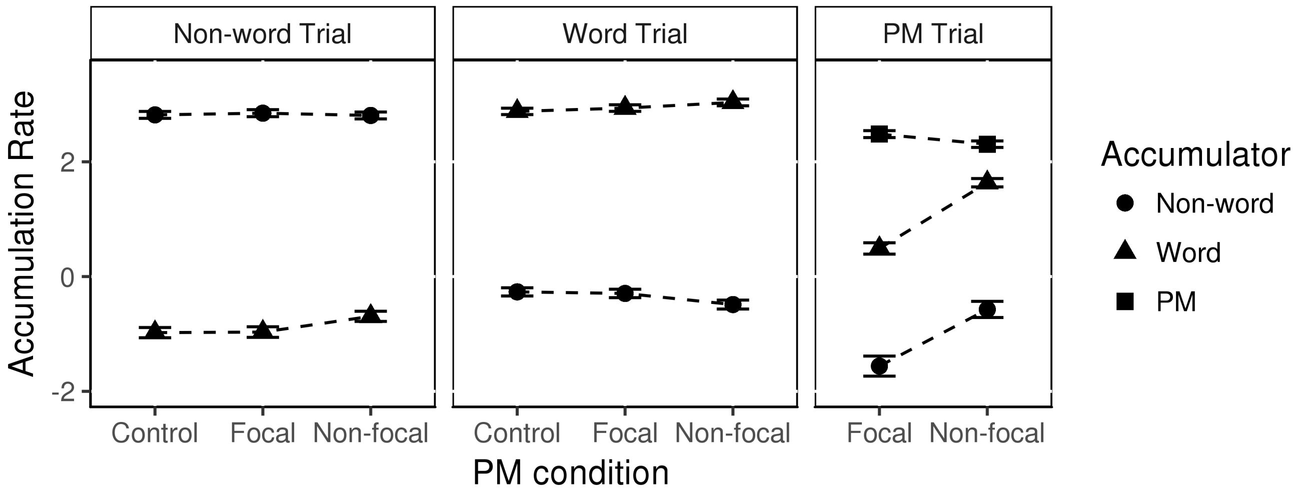

Figure 6. Experiment 1, posterior distributions of the mean accumulation rates, averaged across participants. The central symbols are the posterior means, and the bars are the mean + or - the posterior standard deviation. Plotted by stimulus type, by latent response accumulator, and by PM block. There is very little difference by PM block in the accumulation rates of the non-PM trials, hence the overlapping symbols. False alarm accumulation, that is accumulation towards the PM response on non-PM trials, not pictured (M = -2.42, SD = 0.13).

Table 4

Experiment 1 Z values (with associated posterior predictive p values in brackets) for contrasts between non-PM trial mean accumulation rates. The "correct" accumulator refers to the matching ongoing task response (e.g., word on word trials), and the "error" accumulator refers to the mismatching ongoing task response (e.g., non-word on word trials).

Contrast	*Word Trial Accumulator*		*Non-word Trial Accumulator*	
	Correct	Error	Correct	Error
Focal - Control	3.03 (.001)	-0.60 (0.27)	1.71 (0.04)	0.16 (.43)
Non-focal - Control	8.16 (0)	-4.61 (0)	-0.58 (0.28)	4.74 (0)
Non-focal - Focal	4.82 (0)	-3.93 (0)	-2.24 (0.01)	4.37 (0)

Table 5

Experiment 1 Z values (with associated posterior predictive p values in brackets) for contrasts between thresholds.

Contrast	*Accumulator*		
	Word	Non-word	PM
Focal - Control	9.47 (0)	1.14 (.13)	
Non-focal - Control	21.46 (0)	4.57 (0)	
Non-focal - Focal	12.73 (0)	3.42 (<.001)	9.35 (0)

clusion of outlier RTs, etc.) We replicated this simulation 100 times, and then fit the model back to the 100 simulated data sets. In the supplementary materials, we graph the parameter estimates that we obtained from these fits. Our results support the measurement abilities of PMDC. The estimated posterior means for each simulation clustered around the true values, other than some over-estimation of A and $t0$ (which were both fixed across conditions). In addition, our Bayesian estimates of uncertainty for each simulation correspond well to the level of uncertainty across simulations.

As our design, in terms of the number of trials per participant, was more powerful than usual, we also explored a simplification of the model suitable for less powerful designs. For the full model, the sv parameters, which measure the trial-to-trial variability in accumulation rates, were the most difficult to estimate. Thus, we performed a recovery study on a model with variability in rates fixed at 1. This model led to more consistent and more certain estimation of some model parameters, particularly mean rates. We also fit the simpler model to our actual data, and found that it resulted in virtually the same parameter inference as we reported above. Thus, for future work with fewer trials per participant, fixing sv may firm up estimation while still allowing reasonable inference.

One pertinent question related to parameter recovery is whether we would obtain support for capacity sharing if it was the true casual account of the data. That is, if our data was driven by shifts in non-PM accumulation rather than

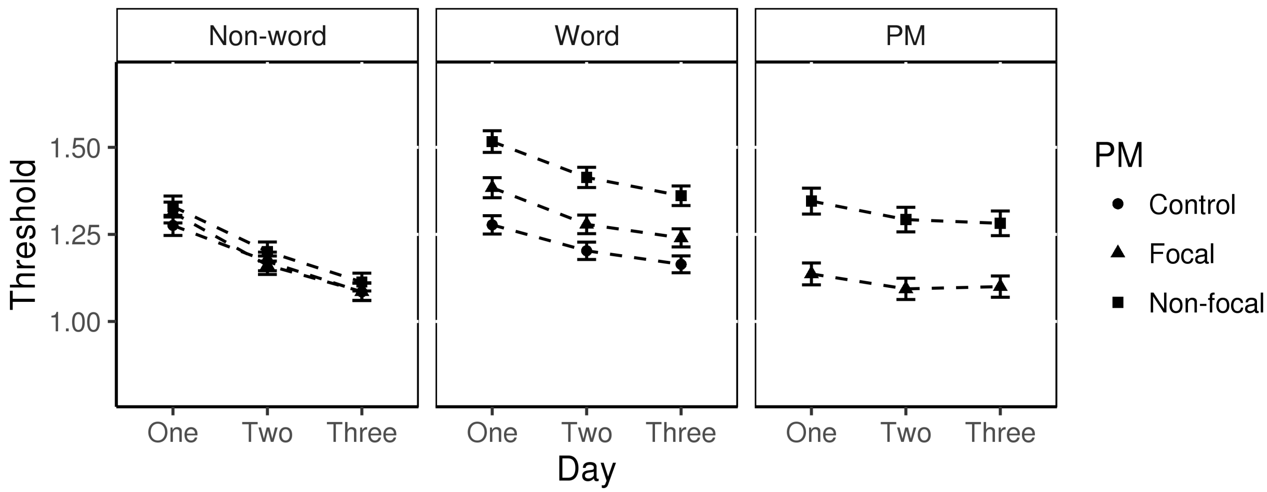

Figure 7. Experiment 1, posterior distributions of response threshold. The central dots are the posterior means, and the bars are the mean + or - the posterior standard deviation. Plotted by latent response accumulator, by PM block, by day.

Table 6
Experiment 1 Z values (with associated posterior predictive p values in brackets) for mean accumulation rate contrasts relevant to reactive control (i.e., for word stimuli only). In the first and second row we compare PM trials with non-PM trials. In the third row, we compare the difference in reactive control between focal and non-focal blocks. Thus, the third row displays the difference between PM blocks (focal - non-focal) of the differences by PM trial status (PM - non-PM).

Contrast	*Accumulator*		
	Word	Non-word	PM
Focal: PM Trial - Non-PM Trial	-23.59 (0)	-6.72 (0)	51.14 (0)
Non-focal: PM Trial - Non-PM Trial	-20.77 (0)	-0.53 (.30)	47.73 (0)
Focal difference - Non-focal difference	-13.44 (0)	-6.17 (0)	4.55 (0)

proactive control, would our fitting procedures be able to detect it? We performed another recovery simulation to address this issue. The simulation was the same as the one reported above, except that the parameters we simulated from the posterior mean parameters of the pure 'capacity-sharing' model of Experiment 1, which attributed all effects of PM condition to changes in accumulation rates. The tables in the supplementary materials detail our finding that we recovered similar non-PM rate effects to those we simulated from. We also checked whether the rate-driven data would produce spurious evidence of control over thresholds. We found this was not the case, with any changes in threshold an order of magnitude or more smaller than the actual effects we found. Thus, we conclude that the PMDC model can adequately detect accumulation-rate effects, and differentiate them from proactive control.

Exploring Model Mechanisms

In complex non-linear models, the effect of a given parameter effect on different aspects of performance can be hard to discern. Here we examine those effects by removing them, and evaluating the model mis-fit that results. To the extent that removing a parameter effect from the model causes a particular type of mis-fit, it drove the effect in the full model. We took two approaches to removing effects from the model. One was to remove control and capacity effects entirely, by setting parameters to baseline values (e.g., setting non-PM accumulation to control levels). The other was to replace selected parameters with the average across conditions (e.g., replacing estimates of the focal and non-focal PM thresholds with the average of the two). In the supplementary materials, we provided detailed descriptions of the results of these explorations. Here, we focus on PM cost and PM accuracy, and the differences between these measures under focal and non-focal conditions.

To study the effect of removing capacity sharing effects, we set all non-PM trial accumulation rate parameters equal to those from control conditions. This actually increased the predicted PM cost, suggesting that model predictions of cost were not at all driven by changes in accumulation rate. This was the case for both focal and non-focal PM conditions. To remove proactive control effects, we set all the ongoing task response thresholds (i.e., word and non-word) to the level of control conditions. This drastically reduced predictions of

PM cost for both focal and non-focal PM conditions, suggesting that control over thresholds was critical to the full-model account of cost. Removing this type of proactive control also affected predictions of non-focal PM accuracy; the full model predicted 99% of the observed non-focal PM, whereas the model with no proactive control only predicted 85%. There was far less effect on focal PM accuracy; the full model predicted 96% of focal PM, the model with proactive control removed predicted 94% of focal PM. Taken together, the two results demonstrate that the extra proactive control under non-focal conditions mitigated the advantage for focal PM accuracy.

To remove reactive control from the model, we set accumulation rates on PM trials for each PM condition equal to the accumulation rate on non-PM trials. This resulted in predicting only 52% of observed focal PM accuracy (compared with 96% in the full model), and 55% of non-focal PM accuracy (compared with 99% in the full model), suggesting a large role for reactive control in both types of PM. Removing reactive control also reduced the difference between focal and non-focal PM accuracy. Whereas the full model predicted 92% of the advantage of focal PM accuracy over non-focal, removing reactive control reduced the predicted effect to only 47% of the observed.

We also examined how differences between focal and non-focal PM thresholds and PM accumulation rates contributed to differences in PM accuracy. We did so by replacing each conditions' parameter values with the average across conditions. Both PM thresholds and PM rates contributed to the advantage of focal PM accuracy. The full model predicted 92% of the advantage of focal PM accuracy over non-focal, whereas the model with PM rates averaged predicted only 72% of the effect, and the model with PM thresholds averaged predicted 63%.

Individual Differences

Correlations. As depicted in Figure 4, the effect of focality on PM accuracy varied across participants. Thus, here we assessed how individual differences in model mechanisms related to individual differences in PM accuracy. We did so by examining correlations between model mechanisms (i.e., model parameters or differences in model parameters) and PM accuracy. We also correlated differences between focal and non-focal model parameters with the difference between focal and non-focal PM accuracy. To perform inference we estimated the distribution of plausible correlations values (Ly et al., 2017). That is, we calculated Pearson correlations across participants for each posterior sample, resulting in posterior distributions of correlations. We then applied a transformation to the sample correlations that allows for uncertainty in generalizing inference to the population, rather than sample, level (Ly, Marsman, & Wagenmakers, in press). We examined several mechanisms: changes in non-PM accumulation rates (i.e, capacity effects), changes in ongoing task thresholds (proactive control), changes in ongoing task accumulation on PM trials (reactive control), PM accumulation rates, and PM thresholds. In the supplementary materials, we graph the posterior means and credible intervals for all of the correlations. Here, we report the posterior means of the correlations that had at least a 95% probability of being different to 0.

No ongoing task parameters were correlated with focal PM accuracy, and only increases in the rate of error accumulation to word trials with PM (a form of increased ongoing task capacity) was correlated positively with non-focal PM (.42). Thus, neither decreases in shifts in ongoing task capacity, nor proactive control effects, were associated with individual differences in PM accuracy. In contrast, PM accuracy was strongly correlated with differences in reactive control of word accumulation, for both focal (.67) and non-focal PM accuracy (.73). Thus, individual differences in PM accuracy were explained by individual differences in inhibitory reactive control. The advantage of focal PM accuracy (i.e., focal PM accuracy - non-focal PM accuracy) was correlated with the advantage of the focal PM accumulation rate (.37), and with nothing else. Thus, individual differences in the effect of target focality on PM accuracy were explained by individual differences in the effect of focality on PM accumulation rate.

Model Mechanisms. We also explored the causal relation between model mechanisms and PM accuracy for individual participants, by examining individual fits to PM accuracy with the full model, and comparing them to the fit of the model with components removed. We examined five models: the full model, the model with proactive control removed, the model with reactive control removed, the model with differences between focal and non-focal PM excitation removed, and the model with differences in focal and non-focal PM thresholds removed. Supplementary materials contain plots of each model's fit to each individual's focal PM accuracy, non-focal PM accuracy, and the difference between the two.

Our results indicated fairly consistent effects of parameters across participants, similar to the exploration of model mechanisms based on averages over participants. Whereas the full model fit PM accuracies well, removing proactive control resulted in underestimation of non-focal PM accuracy across participants. In contrast, it did not consistently result in underestimation of focal PM accuracy. Thus, removing proactive control resulted in some underestimation of the advantage of focal PM accuracy. Removing reactive control resulted in large underestimation of focal and non-focal PM accuracy for all participants. It also caused caused drastic mis-fit to the difference between focal and non-focal PM accuracy, particularly for participants for whom this difference was large. Averaging PM thresholds across focal and non-focal conditions seemed to cause a fairly consistent under-estimation of the difference between focal and non-focal PM accuracy, for participants where it existed, as did averaging the PM accumulation rate.

Discussion

The LBA model provided a good fit to ongoing task accuracies and PM accuracies in Experiment 1, as well as a comprehensive account of the distribution of ongoing task and PM RTs. This included fit to the PM cost effect, and to the effects of manipulating PM target focality on PM costs, PM accuracy, and PM response RT. The behavior of the model did not suggest at all that changes in capacity sharing requirements between PM and ongoing tasks were responsible for the increased costs, or decreased PM accuracy, under non-focal compared to focal conditions, and instead suggested that both proactive and reactive control were active over PM and ongoing decision processes, as specified by PMDC.

Capacity sharing. In both focal and non-focal PM blocks, evidence accumulation actually improved for non-PM words in PM blocks (i.e., higher correct accumulation rates, lower incorrect accumulation rates), contrary to the predictions of capacity sharing theories. In contrast, evidence accumulation towards non-words was marginally worse in the non-focal condition (error accumulation rates increased). These non-PM trial effects are inconsistent, with many indicating increased capacity with PM and only one indicating decreased capacity. In addition, they are substantially weaker than the threshold effects discussed below (standardized differences less than half the size). They also have little relation to PM cost: for both PM blocks, cost was highest to word trials, and yet evidence accumulation to word trials actually sped up in PM blocks. In fact, when we averaged the accumulation rate effects out of the model, it actually predicted more PM cost. Thus, conceptually, we replicate previous studies that found no relation between PM cost and accumulation rate (Heathcote, Loft, & Remington, 2015; Horn & Bayen, 2015; Strickland et al., 2017; Ball & Aschenbrenner, 2017).

One distinction between the current findings and previous work is that here we find non-PM trial accumulation effects that were inconsistent with capacity sharing, whereas previous work has selected models that exclude non-PM trial accumulation effects entirely (e.g., Heathcote, Loft, & Remington, 2015; Strickland et al., 2017). In our WAIC model selection, the pure PMDC model, with non-PM trial accumulation effects totally removed, was selected second after the top model. However, when we examined graphically the average fit of the PMDC model (supplementary materials), it was nearly indistinguishable from the top model, with all the relevant PM-induced trends in the data accounted for. It seems likely, therefore, that our WAIC method selected an overly complex model, and more importantly, with regard to capacity sharing, one that was not psychologically plausible.

Overall then, we did not find evidence for the central tenant of previous PM theories that focal PM has a lower shared ongoing task capacity requirement than non-focal PM (Einstein & McDaniel, 2005; Smith, Hunt, McVay, & McConnell, 2007), and that this drives increased costs or decreased PM accuracy under non-focal conditions. Our findings indicated that neither focal nor non-focal PM shares capacity with the ongoing task.

One reviewer pointed out that, due to 'carryover' effects (Poulton, 1982) in our within-subjects design, participants might have devoted ongoing task capacity to PM monitoring in control blocks that followed PM blocks. If this was the case, capacity sharing would affect our control baseline for testing ongoing task capacity, and thus weaken our test of diminished ongoing task capacity in PM blocks. To explore this, we fit a re-parameterized PMDC model, allowing both thresholds and non-PM accumulation rates to vary depending upon whether a condition was performed first, second, or third within a session of the experiment (a factor we refer to as 'block'), rather than across day order [2]. Resulting parameters are summarized in supplementary materials. We found that the previously reported patterns of non-PM accumulation, which did not support capacity sharing, held for all levels of the block factor. Critically, there were no ongoing-task capacity costs with PM even for block 1. For a given day, participants were not assigned any PM task prior to block 1, and thus it seems highly unlikely that comparisons for this block would be confounded by unnecessary capacity sharing in control conditions. Thus, it appears that capacity sharing was not masked by carryover effects in our data. We did find some variation in control over ongoing task thresholds by within-session order. We found a decrease in control of non-focal word thresholds for block 2, but this decrease was not sustained for block 3. Thus, carryover effects might alter PM-related ongoing task threshold settings, but in this experiment they appeared not to do so in a systematic way.

Proactive Control. Replicating Heathcote et al. (2015), Horn and Bayen (2015), Strickland et al. (2017), and Ball and Aschenbrenner (2017), we found large increases in ongoing task thresholds in PM blocks compared with control blocks. As PMDC predicts should be the case with stimulus-specific instructions, we found that threshold increases were largely selective, with the word threshold increasing more than the non-word threshold. Thus, participants applied proactive control specifically to target the threshold parameter that would affect PM accuracy. However, we did find

[2]Note that we did not include this factor in our primary analyses because we could not model within-day effects, across-day effects, and PM condition effects simultaneously. Each participant performed each condition in the same within-day position only once. For example, if a participant performed the control condition on block 1 of day 1, they would not perform control on block 1 of later days.

evidence of significant (but smaller) increases in non-word thresholds in the non-focal condition (also replicating Strickland et al., 2017). Thus, non-focal PM appears to cause participants' general response caution to increase, perhaps due to increased perceived task complexity (Horn & Bayen, 2015). The effect sizes of the non-word threshold increases were much smaller than the word threshold increases. Thus, our data demonstrate a largely selective delay strategy, with a smaller role for general caution increases.

Replicating Strickland et al. (2017), the ongoing task threshold increases were greater under non-focal conditions than focal conditions. This suggests that participants applied more proactive control for the non-focal PM task, perhaps because a larger degree of selective delay is required to wait for the slower PM accumulator. Our posterior exploration revealed that this threshold strategy was indeed beneficial to non-focal PM - when we removed ongoing task threshold increases from the model, the advantage of focal PM accuracy was increased beyond that observed in the data.

Despite higher ongoing task decision thresholds in the non-focal PM blocks, focal PM accuracy was greater overall. This owed to differences between focal and non-focal blocks in the parameters of the PM accumulator. In focal blocks, there was a lower threshold to respond PM. This threshold effect was not a strong prediction of PMDC, which makes no claim either way about whether PM thresholds should vary in response to stimulus characteristics. Nonetheless, it would be interesting to identify the reasons underlying this shifted threshold. One possibility is that participants set a higher non-focal threshold because the non-focal task requires a more difficult target-nontarget discrimination process, in which they are less confident relative to the focal task.

Reactive control. Reactive control was evident on PM trials in both focal and non-focal PM blocks. As expected, PM excitation was evident on PM trials, with accumulation towards the PM response being much higher than on non-PM trials. Consistent with PMDC, we also found feedforward inhibition of accumulation towards the correct ongoing task decision on PM trials (i.e., decreased word accumulation). We were unable to detect this with mean RTs, as word responses to PM trials were actually faster than word responses to non-PM trials, likely due to statistical facilitation (i.e., when the correct ongoing task accumulator is slow it is not observed because a PM response is made). The accumulation rate estimates provided by our model can unmask the feedforward inhibition effect because they titrate processing speed from statistical facilitation.

Although still present, evidence for feedforward inhibition of incorrect ongoing task accumulation (i.e., non-word accumulation on PM trials) was substantially weaker than for correct ongoing task accumulation. This is consistent with selective reactive inhibition (i.e., mainly inhibiting the ongoing task response associated with PM stimuli), which might, for example, arise through associative learning of the relevant contingency over the course of the experiment. However, the parameter for the non-word accumulation rate on PM trials is constrained by relatively little data – there are only 84 PM trials for each PM block (focal and non-focal), and participants made very few incorrect ongoing task responses on these trials (~2% of responses on the 84 trials). The lack of data for these response cells would allow for strong influence from the prior, which specified no difference. Thus, the data does not allow us to conclusively determine whether inhibitory reactive control was selective, or applied to all ongoing task decision processes.

Reactive control was larger in focal PM blocks than in non-focal PM blocks. There was an increase in excitation of PM accumulation. That is, the PM accumulation rate was higher under focal conditions than under non-focal conditions. This is consistent with greater bottom-up activation from focal PM stimuli than non-focal PM stimuli. There was also more feedforward inhibition of ongoing task decisions under focal conditions. That is, accumulation towards ongoing task responses was reduced more on PM trials under focal conditions than under non-focal conditions. The increase in inhibitory reactive control was larger than the increase in PM excitation, and responsible for more of the PM accuracy advantage observed in the focal condition. Arguably, this is inconsistent with the multiprocess account of focality effects, in which spontaneous retrieval brings the PM response immediately to mind (e.g., Einstein & McDaniel, 2000), without the need to inhibit the ongoing task decision process. However, it converges with more recent theoretical work (Bugg et al., 2013), and with neurological data (McDaniel et al., 2013) that implicates reactive control on focal PM trials.

Summary. The modeling of Experiment 1 was consistent with dual mechanisms of PM decision control. The behavior of our model suggested no role of capacity sharing in PM cost. Instead, there was proactive control over decision thresholds. We found that our stimulus-specific PM task, in which PM targets were always words, was associated with larger increases in word thresholds than non-word thresholds. This suggests that decision control via delay can be targeted selectively at allowing time for the PM response. There were greater increases in proactive control of ongoing task decisions under non-focal conditions, and this increased non-focal PM accuracy. Focal PM accuracy was higher than non-focal overall, due to PM-related mechanisms not indexed by PM cost. A unique contribution of our modeling is to provide a quantitative process account of these mechanisms. We found that increased focal PM accuracy owed to a lower focal PM threshold, more PM focal excitation, and, most critically, increased inhibitory reactive control of ongoing task decision processes on focal PM trials.

Experiment 2: PM Importance

In Experiment 1, we found that PMDC could provide a full distributional account of performance in the PM paradigm. PMDC accounted for non-focal PM task performance, which produced moderate PM accuracy with sizeable PM cost, while simultaneously accounting for performance on focal PM tasks, which produce high PM accuracy with lower PM cost. In Experiment 2, we instead manipulated PM accuracy by varying the importance of the PM task. Like focality, PM importance increases PM accuracy, but unlike focality, PM importance increases PM cost. In Experiment 2 we test whether the model can fit these effects, and, if it can fit them, examine how it does so. We use a within-subjects design including control blocks of trials (no PM task), PM-important blocks (participants instructed that the PM task was most important), and PM-unimportant blocks (participants instructed that the ongoing task was most important).

In Experiment 1, our modeling revealed a lack of support for capacity sharing on non-PM trials. In Experiment 2 we again compare non-PM trial accumulation between PM blocks and control blocks, allowing us to once again test for capacity sharing effects. Current PM theories claim that PM importance emphasis causes more capacity to be allocated to the PM task, resulting in increased cost and increased PM accuracy (e.g., Einstein et al. 2005; Smith & Bayen, 2004). It is possible that, even if not supported by our model of typical PM cost, capacity sharing may be uniquely supported in the case of PM-important cost. If this is the case, we should see systematic reductions in non-PM trial accumulation rates under PM-important conditions. Further, as the cognitive capacity usurped from ongoing task processing is assumed to be allocated to to monitoring ongoing task stimuli for PM features, increased capacity sharing with importance should also be associated with increased PM excitation on PM trials (faster accumulation towards the PM response) for PM-important blocks as compared with PM-unimportant blocks. There is also the possibility we will find evidence of increased *non-shared* capacity for the PM-important task. Participants may have some separate capacity pool that does not affect ongoing task decisions, which they draw more capacity from when PM is important. If this is the case, PM-important instructions will lead to an increase in PM accumulation, without any corresponding decreases in non-PM trial accumulation.

In contrast to capacity sharing theories, PMDC predicts PM cost will be reflected in increased ongoing task decision thresholds. In Experiment 2 we changed our PM task to detecting a target syllable within lexical decision items (e.g., press a different key if any letter string contains 'tor'). We swapped to this task so that we could include both PM word targets and PM non-word targets. In Experiment 1, we observed that the non-word threshold increased a relatively small amount under PM conditions. According to delay theory, this is because the non-word accumulator was unlikely to reach threshold before the PM accumulator on PM trials: as PM trials were always words, the non-word accumulation rate to PM trials was always very low. With the inclusion of non-word PM targets, the non-word threshold should be more relevant to PM accuracy, and as such we expect to find more substantial threshold shifts for the non-word threshold than in Experiment 1.

Under PMDC, increased PM cost with importance would be reflected in larger increases in ongoing task thresholds. The presence vs. absence of larger threshold increases in PM-important conditions is diagnostic as to whether ongoing decision threshold increases are implemented specifically to improve PM accuracy (Heathcote, Loft, & Remington, 2015; Strickland et al., 2017), or whether they reflect a global increase in caution due to the PM instructions increasing the overall perception of task complexity (Horn and Bayen., 2015). The perceived complexity of the PM-important and PM-unimportant tasks should be identical, and thus any ongoing task decision threshold increase in the PM-important blocks beyond the PM-unimportant blocks can be attributed to PM-specific processes, rather than general impressions of the task. Increased ongoing decision thresholds could lead to higher PM accuracy, by allowing more time for the PM accumulator to reach threshold on PM trials. Proactive control could also benefit PM-important accuracy via control of the PM threshold. Participants could lower their PM threshold in the PM-important block, so that the PM accumulator is more likely to win the race with the competing ongoing task accumulators.

Emphasizing the importance of the PM task may also alter reactive control. That is, under PM-important conditions, the reactive control architecture may react differently to each unit of PM input that it receives. This could produce differences in PM accuracies despite identical monitoring processes or thresholds. Specifically, participants may increase inhibitory reactive control when PM importance is emphasized, in line with findings that reward increases reactive response inhibition (Boehler et al., 2014). In this case, accumulation of ongoing task decision processes should decrease more on PM trials when PM is important than when PM is unimportant. In terms of our PMDC framework for reactive control (Figure 2), this would indicate that pathways B1 and B2 become more sensitive to input from the PM detector when PM is important. However, note that if we find that PM excitation (A1) also increases, we cannot identify whether the reactive control structure was modified. In this case changes in accumulation could equally owe to an increase in input to the PM detector, for example as a result of more capacity devoted to the PM-important task.

Method

Participants

Five participants were excluded from analysis (see results), which left our final data set with 36 participants (30 females) aging from 17-26 (average = 18.56 years).

Materials

All word stimuli (both PM targets and non-targets) in the experiment were low frequency English words (occurring 1-7 times per million) drawn from the TMSH database (Dennis, 1995) of length between, and including, 5 to 10 characters. All non-word stimuli were created using the Wuggy algorithm (Keuleers & Brysbaert, 2010), which replaces subsyllabic segments of words with other subsyllabic segments (i.e., that are also legal in the same position in the language of choice). Wuggy was set to replace two out of three subsyllabic segments and to match both the subsyllabic segment lengths and transition frequencies of its output non-words with the input words. One thousand eight hundred and ninety six words were selected to be non-PM word stimuli. Each word was input to Wuggy to produce a corresponding non-PM non-word stimulus. PM targets contained the substring *ver*, the substring *tor*, or the substring *per*. To create the PM target words for each substring, 28 new words were drawn from the TMSH database that contained that substring, for example *tortoise* was one of the 28 words drawn that contain the substring *tor*. To create the PM target non-words for each substring, 28 new words were drawn from the TMSH database and input to Wuggy, with filters so that only non-words that contained each of the target syllables were generated. All stimuli were presented once each to all participants.

The total number of trials per block, and the counterbalancing of the three different PM conditions, followed the same pattern as Experiment 1. For each day of the experiment for each participant, one substring (*tor*, *ver*, *per*) was assigned as a PM target for the PM-important block and a different substring was assigned to the PM-unimportant block. The assignment of substrings to PM blocks was balanced for each participant so that, over the three days, each substring was used once in the PM-important and once in the PM-unimportant block. Other than satisfying the previous two conditions, the assignment of substrings to PM blocks was random. In control blocks, participants were presented with 220 non-words and 220 words. In PM blocks (both PM-important and PM-unimportant), participants were presented with 206 non-target non-words and 206 non-target words, as well as 14 PM target non-words and 14 PM target words. For each participant, the 14 PM target words and 14 non-words used in the PM-important block were drawn randomly, without replacement, from the 28 of each available for each substring. The other 14 were used for the PM-unimportant block. The order of PM target presentation for each block was randomized. For each participant the order of lexical decision stimuli was randomized across blocks. The structure of each block, in terms of breaks, and position of PM trials, was the same as Experiment 1, except that PM target trials were separated by at least 4 lexical decision trials rather than 2.

Procedure

The procedure was identical to Experiment 1, except that the instructions were modified to match the new PM task and PM importance manipulation. In PM blocks, participants were additionally instructed to press an alternative key instead of their word or non-word response when they encountered items containing a target substring, for example, *"In the next block of lexical decision trials, if you see ANY item that contains 'tor' then press 'j' INSTEAD of 's' or 'd'. For example, if you see 'indicator' then press 'j' instead of 's' or if you see 'botoraty' then press 'j' instead of 'd"'*. For the PM-important blocks, participants were asked to prioritize remembering to perform their alternative response to PM targets; *"Please make all your responses as quickly and accurately as possible, however, concentrate on remembering to make a special response if you see items containing 'per'. That is, for this block of trials, remembering to make a special response to items containing 'per' is more important than discriminating between words and non-words"*. For the PM-unimportant blocks, participants were asked to prioritize their performance on the lexical decision task; *"Please make all your responses as quickly and accurately as possible, however, concentrate on the lexical decision task. That is, for this block of trials, discriminating between words and non-words is more important than remembering to make a special response if you see items containing 'tor"'*. The instructions also asked the participant to speak to the experimenter, so they could ask any questions they had about the task. For PM blocks, the next screen asked to indicate whether the PM or LD task was more important, e.g., *"Please indicate which of your tasks is more important. Press 'g' if responding to items containing tor is more important. Press 'h' if discriminating between words and non-words is more important. If you answer correctly, the experiment will continue. If you answer incorrectly, the experiment will return to the instructions screen for review"*. If participants made the wrong response they returned to their instructions screen for review. They were then asked the question again, and if they made the incorrect response would again be returned to their instructions. Once they correctly indicated which task was more important the experiment proceeded.

Results

Five participants were excluded from analysis: two because an entire block of their ongoing task accuracy was near chance (< 65%), two because they didn't identify a single PM target in the PM-unimportant blocks over all 84 PM presentations, apparently because they misinterpreted the instructions and decided to completely disregard the PM, and one because the participant, after completing the experiment, indicated in a conversation with the experimenter that they did not understand the importance emphasis instructions. This exclusion rate of participants is higher than typical, but not surprising given that for each participant we required valid data from nine blocks of trials over three days of testing. The first two trials of each block and after each rest period were excluded from the analyses, as were the two trials following each PM trial, and following any PM false alarms. We also excluded trials where participants responded with a key which was not designated to indicate their PM or LD responses (0.03% of trials), and trials with outlying RTs, defined as less than 0.2s or 3 times the interquartile range above the mean (4.47% of the remaining responses). One participant responded once with the PM response key in the control condition, this response was excluded from the analyses.

Prior to reporting modeling results we present standard analysis of ongoing task accuracy, PM cost, PM accuracy, and PM RT. As with Experiment 1, we analyzed accuracies with binomial probit models and mean correct RTs with general linear models (supplementary materials), and standard errors were calculated as suggested by Morey (2008). In addition to stimulus type (word, non-word) and PM block (control, PM-important, PM-unimportant), all analyses included a day order factor (day 1, day 2, day 3) to capture effects of task repetition.

Non-PM Trials

In addition to PM trials, false alarm trials (non-PM trials on which a PM response was made) and the two lexical decision trials following both PM trials and false alarm trials were excluded. Accuracy was higher for non-words (92.7%) than words (89.2%), and there was an effect of day (day 1 = 91.8%, day 2= 90.8%, day 3= 90.3%). These effects were qualified by an interaction between stimulus type and day. Non-word accuracy was relatively stable over the 3 days (in order of day: 93.1%, 92.4% and 92.7%), whereas word accuracy decreased over the 3 days (90.5%, 89.3% and 87.8%). There was no effect of PM block (PM-important M = 91.0%, SE = 0.84%; PM-unimportant M = 90.9%, SE = 0.87%; Control M = 90.9%, SE =0.75%).

Correct ongoing task RTs decreased over days (day 1= 0.809s, day 2= 0.740s, day 3= 0.709s), and were slower to non-words (0.774s) than words (0.731s). We found costs to correct ongoing task RT in PM-unimportant blocks (0.761s) as compared to control (0.712s), t (35) = 5.91, $p < .001$, d = 0.99. Costs were larger in the PM-important blocks (0.785s) than in the PM-unimportant blocks, t (35) = 4.74, $p < .001$, d = 0.79. Thus, the experiment produced the expected PM costs, and the expected increase in PM costs with PM importance. The effect of PM block interacted with the effect of day order (Figure 8), with the speed increase over day being largest in the PM-important condition and smallest in the control condition (PM-important mean RT in order of day = 0.853, 0.773s, 0.730s; PM-unimportant = 0.824s, 0.744s, 0.715s; control = 0.750s, 0.704s, 0.682s). The effects of stimulus type and PM block did not interact.

PM trials

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 decreased over days (day 1 M = 62%; SE = 5%; day 2 M = 60%; SE = 4%; day 3 M = 56%; SE = 4%), and was higher for non-word PM targets (M = 61%, SE = 5%) than for word PM targets (M = 57%, SE = 5%). We found the typical effect of importance emphasis; PM accuracy was higher in the PM-important blocks (M = 70%, SE = 4%) than in the PM-unimportant blocks (M = 49%, SE = 5%). However, the magnitude of this effect was widely distributed across participants: for some participants there was a major shift in PM accuracy, and for some none at all (Figure 9). PM response false alarms were very rare, ranging from 0 to 0.7% of trials.

Correct PM responses were faster in the PM-important blocks (M = 0.839s, SE = 0.018s) than in the PM-unimportant blocks (M = 0.884s, SE = 0.023s), and there was an effect of day order (day 1 M = 0.917s, SE = 0.018s; day 2 M = 0.849s, SE = 0.016s; day 3 M = 0.811s, SE = 0.015s). PM responses took longer for non-word PM targets (M = 0.870s, SE = 0.022s) than for word targets (M = 0.850s, SE = 0.021s).

Non-PM trials compared with PM trials

We again tested for reactive control by examining the effect of PM trial status (PM trial vs. non-PM trial) on ongoing task RT. We ran a linear mixed effects model including the effects of PM trial status, PM block, stimulus type, and day. Recall that a ‘correct’ ongoing task response on a PM trial is in fact a missed PM target. Correct ongoing task RTs were not significantly different on non-PM trials (M = 0.778s) than PM trials (M = 0.762s). There was an interaction between PM trial status, and lexical stimulus type. There was a trend for correct ongoing task responses to non-words to be slower on PM trials (M = 0.801s; SE =0.032s) than on non-PM trials (M = 0.792s; SE = 0.015s), t (35) = 1.93, p = .06, d = 0.79, whereas there was a trend for correct ongoing task responses to words to be faster on PM trials (M = 0.725s; SE = 0.027s) than on non-PM trials (M = 0.754s; SE =0.016s), t (35) = 1.92, p =.06, d = 0.32. There was no interaction between PM trial status and PM block. Although these RT effects do not strongly suggest reactive control, note again that the critical analyses are of accumulation rates (next section), and not raw RT.

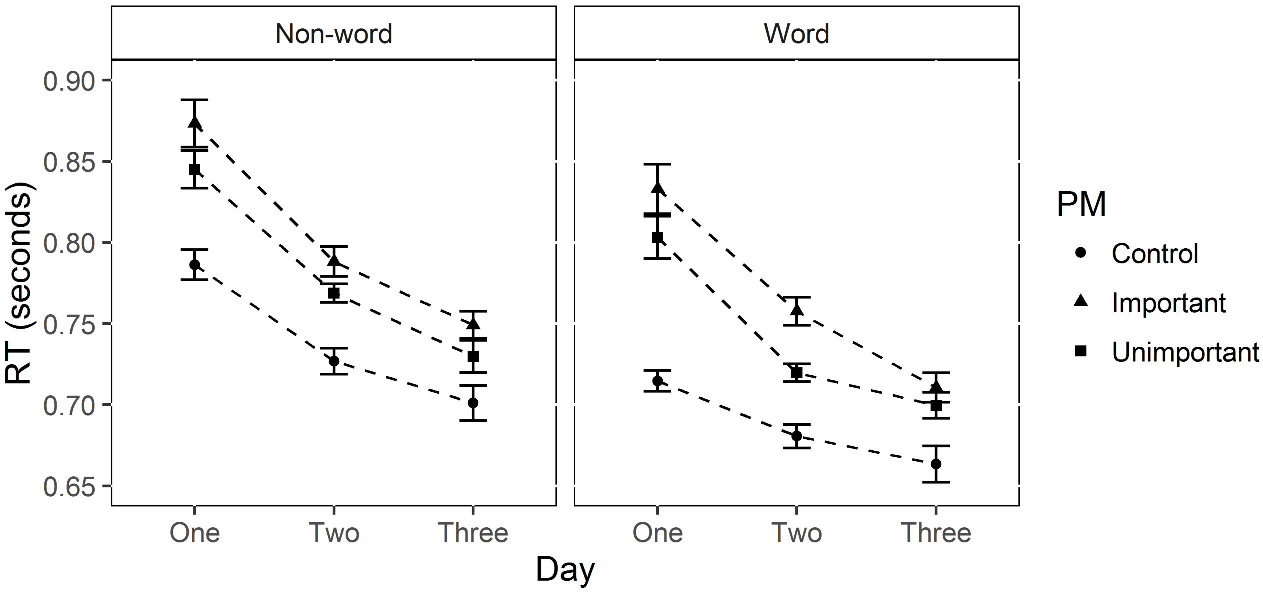

Figure 8. Experiment 2, ongoing task RTs by PM block by day.The standard error bars were calculated using the Morey (2008) bias corrected method.

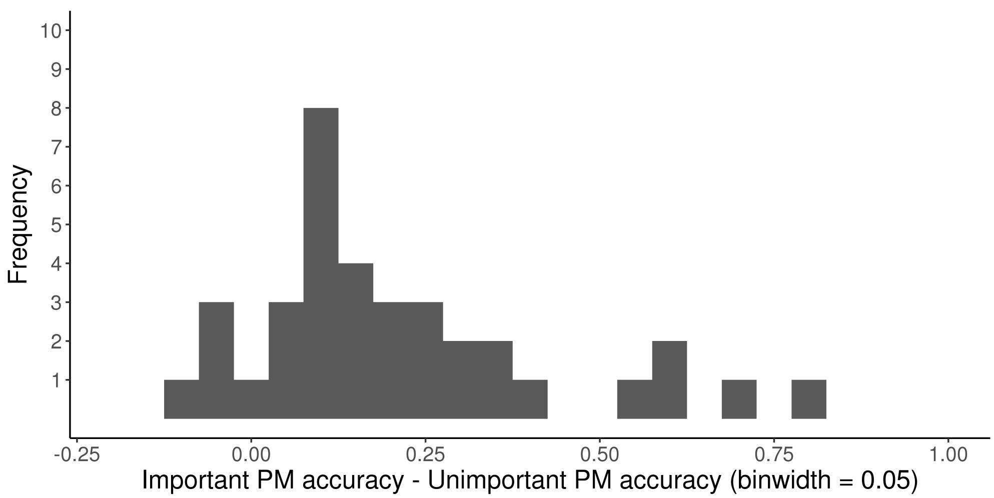

Figure 9. Frequency histogram demonstrating the spread across participants of the effect of PM importance emphasis on PM accuracy. Positive values demonstrate an advantage for the PM-important blocks. For most participants, PM-important PM accuracy was higher, but the magnitude of the effect varied substantially.

Model Analysis

Our factor structure was the same as Experiment 1, except that our stimulus type factor had four levels: word, non-word, PM word, and PM non-word. We applied the same modeling approach, parameter restrictions, priors, and sampling settings as in Experiment 1. Figure 10 plots the fits of the posterior predictions, averaged over participants, to the Experiment 2 data. Again the model provided a good fit to both non-PM trials and PM trials, including the PM cost effect, PM accuracies, PM RTs, and the effects of PM importance. We also include fits to the coefficient of variation in non-PM trial RTs, PM miss RTs, and PM hit RTs, in the supplementary materials. Again we found similar patterns to those reported by Gilbert et al. (2013), with non-PM trial coefficient of variation larger than PM trial coefficient of variation, and the PMDC model provided a reasonable fit to these trends.

We again used WAIC model selection (Table 7) to assess whether any further constraint on our model across PM blocks could be statistically justified. We used the differences in WAICS, and standard error of the differences, to

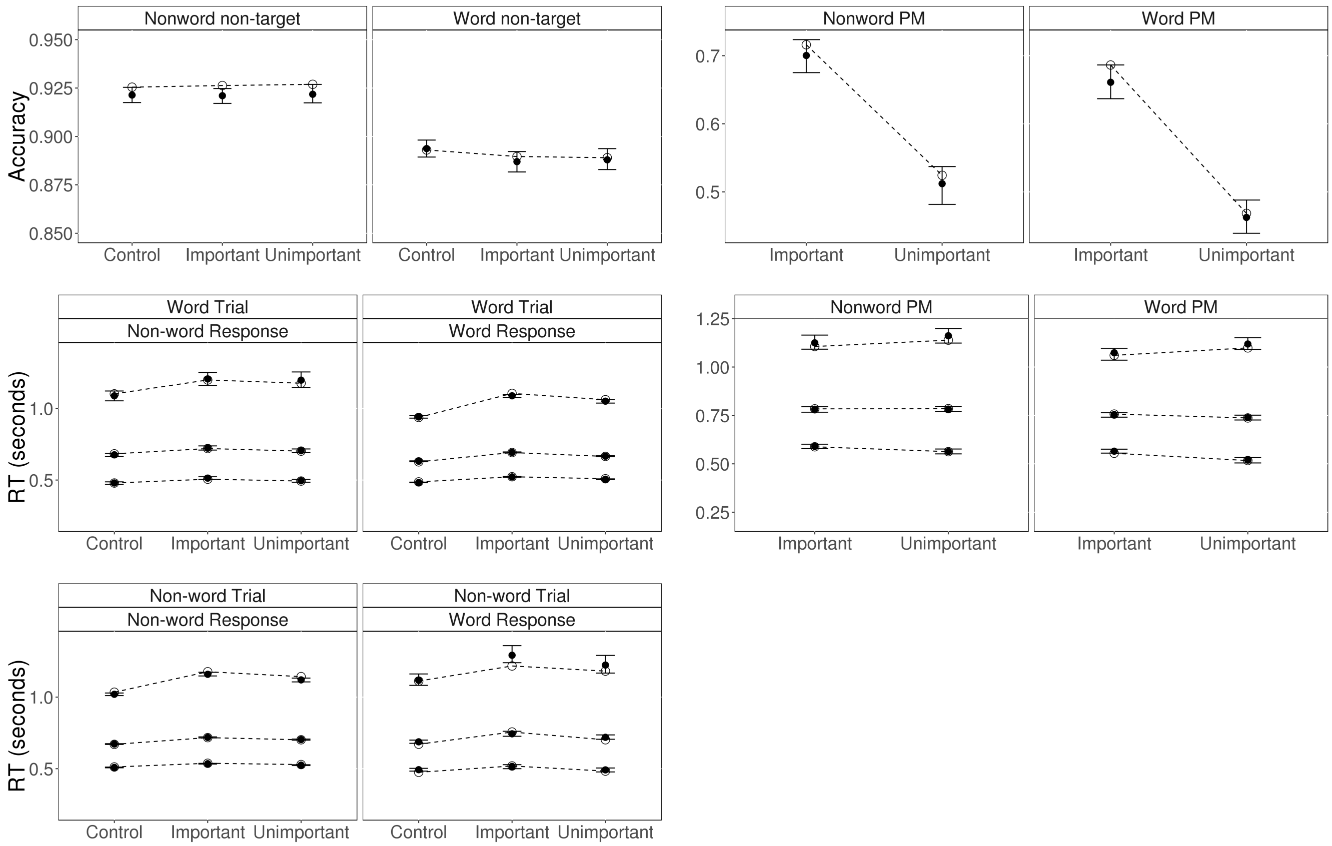

Figure 10. Fits of the LBA to the Experiment 2 data. The white dots indicate the observed data average. The black dots indicate the posterior prediction. The error bars are the 95% credible interval of the posterior prediction. The credible intervals tend to be narrow due to our large number of trials. The RT distributions are summarized with three order statistics (the three lines on each RT graph): the 0.1 quantile, which captures the leading edge of the distribution (i.e., the fastest responses), the median value, and the 0.9 quantile which captures the tail of the distribution (i.e., the slowest responses). We calculated the plotted statistics (accuracies and quantile RTs) by concatenating the observed data from all participants into one data frame, and then calculating the statistics for that entire data frame. We used this same procedure for both the observed data and the simulated data. The alternative is to calculate statistics on a participant-by-participant basis (e.g., calculate a median RT for every participant), and then average. The approaches yielded similar results. Due to the lower number of PM trials, PM RTs are collapsed over correct and error responses.

assess the relative importance of differences in accumulation rates and thresholds across PM conditions (control/ PM-important/ PM-unimportant). As with Experiment 1, we found more support for an 'Only Proactive Control' model, in which both the ongoing task and PM thresholds could vary over PM conditions, but accumulation rates could not, than for a 'No Proactive Control' model, in which accumulation rate parameters could vary freely among the PM and control conditions, but thresholds could not, WAIC difference = 4103.61, standard error of the difference = 159.54, standardized difference = 25.72.

We compared the Only Proactive Control model to a model that allowed both proactive control and also variation in PM excitation across PM block (i.e., different accumulation rates towards the PM response for PM-important and PM-unimportant conditions), and although WAIC for the latter was slightly lower, the difference was not substantial, 12.25, standard error of the difference = 15.92, difference in standard units = 0.77. We found a larger difference when we compared the model with proactive control and differences in PM excitation to a model with proactive control and differences in PM inhibition (differences in the ongoing task accumulation rates on PM trials between blocks). WAIC favored latter, with an advantage of 279.37, standard error of the difference = 40.35, standardized difference = 6.92. Consistent with the weak effect of PM excitation, a comparison of the previous model (variation proactive control plus inhibitory reactive control) to the full PMDC model (variation in proactive control, excitatory control and inhibitory control) indicated only a slight WAIC preference for the PDMC model, WAIC difference = 7.29, standard error of the difference = 16.67, and standardized difference = 0.44. Thus, in terms of the importance effects on PM trials, variation in ongoing task accumulation rates on PM trials as a function of PM importance led to big improvements in fit, whereas variation in accumulation towards the PM response led to smaller improvements in fit.

Finally, we compared the PMDC model to the top model, which also allows non-PM trial accumulation rates to vary over PM condition. As with Experiment 1, we found that the top model was the most supported of all, with a WAIC 301.06 lower than the full PMDC model, standard error of the difference = 62.03 , difference in standard units = 4.85. Thus, again it appears that non-PM trial accumulation rate flexibility increases the predictive ability of the model. However, as we demonstrate below with posterior inference from the chosen, most flexible model, we again found that the direction and magnitude of non-PM trial accumulation differences does not suggest capacity sharing.

Model Summary

As with Experiment 1, we summarize our results by averaging the posterior samples across participants. The non-decision time posterior mean was 0.16 (posterior SD = 0.002). The A posterior mean was 0.35 (posterior SD = 0.01). The sv parameters are summarized in Table 8. As with Experiment 1, all of these values are reasonable and consistent with past results. We now examine variation in mean accumulation rates and thresholds to determine the relative support for capacity sharing and PMDC mechanisms. We test differences between parameters using posterior predictive p values, and also report the standardized effect size of the difference distribution (Z).

Capacity sharing (Non-PM Trial Accumulation). The non-PM trial accumulation rates are displayed in the left two panels of Figure 11. Table 9 contains effect sizes and posterior p values for comparisons of non-PM trial rates. Capacity sharing theories predict lower correct non-PM trial accumulation in PM blocks, and higher incorrect non-PM trial accumulation in PM blocks. Inconsistent with capacity sharing, correct accumulation to non-PM words was marginally higher for PM-important blocks (2.71) than control (2.68). The difference between control and PM-unimportant (2.70) blocks was negligible. Also inconsistent with capacity sharing, error accumulation to non-PM words was lower under PM-unimportant conditions (0.13) than control (0.26), and marginally lower still in PM-important blocks (0.09) than PM-unimportant blocks. Similarly, error accumulation to non-PM, non-word trials was similar under PM-unimportant (-1.19) and control conditions (-1.20), as well as control and PM-important conditions (-1.17).

We found one non-PM trial accumulation effect in line with capacity sharing. Correct accumulation to non-PM non-words was lower under PM-unimportant conditions (2.65) than control (2.72), and similar between PM-unimportant and PM-important conditions (2.64). Although this effect could be taken as evidence for capacity sharing, it is not convincing in the context of the other effects pointing in the opposite direction.

Proactive Control (Thresholds). Proactive control to delay ongoing task decision processes should be reflected in higher ongoing task decision thresholds in PM blocks, and more proactive control should cause higher thresholds in PM-important blocks compared with PM-unimportant blocks. Figure 12, which plots the thresholds, demonstrates that this is what we found. Table 10 shows effect sizes and posterior p values for comparisons of thresholds. Word response thresholds were higher in PM-unimportant blocks (1.27) than control blocks (1.14), and higher in PM important blocks (1.36) than unimportant blocks. Non-word thresholds were higher in PM-unimportant blocks (1.33) than control blocks (1.26), and higher in PM-important (1.38) blocks than PM-unimportant blocks. We also found support for proactive control over the PM threshold: it was lower (i.e., set to favor making the PM decision) in PM-important blocks (1.12) compared with PM-unimportant blocks (1.35).

Reactive Control (PM vs. non-PM trial accumulation). We examined reactive control by comparing the non-PM trial accumulation rates (left two panels of Figure 11) with the PM trial accumulation rates (right two panels of Figure 11). The effect sizes and p values relevant to these comparisons are in Table 11. As expected, we found evidence in both PM-important and PM-unimportant blocks for excitation of the PM accumulator on PM trials, in that the PM accumulation rates on PM trials were much higher than the 'false alarm' PM accumulation rate on non-PM trials (M = -2.71).

We compared PM excitation (accumulation towards PM on PM trials) between PM-important and PM-unimportant blocks. Inconsistent with gains in PM excitation, PM accumulation to PM non-words was a little lower in PM-important blocks (1.84) than in PM-unimportant blocks (1.89). However, more consistent with increased excitation, for PM words PM accumulation was higher in PM-important blocks (1.89) than in PM-unimportant blocks (1.84). Given there opposing directions, these effects do not indicate that PM excitation increased overall in PM-important blocks as compared with PM-unimportant blocks.

We again test for reactive inhibitory control by comparing correct ongoing task accumulation rates on PM trials with correct ongoing task accumulation rates on non-PM trials. In PM-unimportant blocks, word accumulation was much lower on PM word trials (1.70), than on non-PM word trials (2.70). Similarly, in PM-unimportant blocks non-word accumulation was lower on PM non-word trials (1.63) than on non-

Table 7

Experiment 2, WAIC model selection. Lower WAIC indicates more preference for the model. To evaluate each model using data from the entire group of participants, we first concatenated the log-likelihoods under the model for each trial, for all participants, together into one pointwise log-likelihood matrix where points are trials. We then calculated WAIC for each model using its log-likelihood matrix.

Model	Estimated Number of Parameters	WAIC
Top	1171.1	-47977.3
Proactive Control & Reactive Control (both)	1013.1	-47676.3
Proactive Control & Reactive Control (inhibitory)	991.8	-47669.0
Proactive Control & Reactive Control (excitatory)	985.4	-47389.6
Proactive Control	962.3	-47377.4
No Proactive Control	877.4	-43273.8

Table 8

Experiment 2, M (SD) of the posterior distributions, averaged across participants, of the standard deviation of the evidence accumulation rates.

Accumulator	*Stimulus Type*			
	Word	Non-word	Word PM	Non-word PM
Non-word	1.36 (0.03)	0.66 (0.012)	0.98 (0.04)	1.06 (0.03)
Word	0.71 (0.01)	1.53 (0.03)	1.26 (0.04)	1.09 (0.04)
PM	Fixed at 1		0.40 (0.01)	0.43 (0.01)

PM non-word trials (2.65). In PM-important blocks, ongoing task accumulation rates were also reduced on PM trials, and the reduction in rate was larger than in PM-unimportant blocks for both word stimuli (non-PM word = 2.71 , PM word = 1.13) and non-word stimuli (non-PM non-word = 2.64, PM non-word = 1.15). Thus, in sum, we found strong evidence of reactive inhibition towards the ‘correct response’ for all PM trial types, and also strong evidence of greater reactive inhibition towards the ‘correct’ response in PM-important conditions.

In contrast to the correct rates, incorrect ongoing task accumulation rates were not consistently lower on PM trials. Although non-word accumulation was reduced on word PM trials (PM-unimportant = -0.57, PM-important = -0.57) as compared with non-PM word trials (PM-unimportant = 0.13, PM-important = 0.09), word accumulation on non-word PM trials (PM-unimportant = -0.45, PM-important = -0.83) was actually higher than non-PM trials (PM-unimportant = -1.19, PM-important = -1.17). However, these incorrect ongoing task accumulation rates were again very data poor parameters, with very few incorrect ongoing task responses ever being observed on PM trials, and so this was probably due to influence from the prior.

Model Mechanisms

We took the same approach as with Experiment 1 to exploring how parameter differences accounted for the effects in the data. That is, we simulated from models where differences between parameters were averaged out, or parameters were set equal to baseline values, and evaluated predictive mis-fit as compared with the full model. In the supplementary materials, we include detailed graphs of our findings. In text, we restrict our discussion to the benchmark PM effects (PM cost, PM accuracy, and increased PM cost and PM accuracy with importance). Again, for simplicity, we discuss the adequacy of fit to PM cost effects, and the percentage of actual PM accuracy and PM accuracy effects predicted (see supplementary materials for more detailed graphical summaries).

Again we simulated a model with all non-PM trial accumulation rate parameters equal to those from control conditions (i.e., no capacity sharing). This did not appreciably affect the model’s predictions of costs, again suggesting a dissociation between capacity sharing and PM cost. In contrast, setting all ongoing task response thresholds to the level of control conditions (i.e., no proactive control) removed the model’s predictions of PM cost, suggesting that thresholds were critical to the effect in the full model. Ongoing task thresholds also affected PM accuracy. The full model predicted nearly all of both PM-important accuracy (96% of word PM, 98% of non-word PM) and PM-unimportant accuracy (99% of word PM, 98% of non-word PM), whereas the model with ongoing task thresholds set to control levels predicted less of both PM-important accuracy (85% of word PM, 91% of non-word PM) and PM-unimportant accuracy (88% of word PM, 91% of non-word PM). Differences in ongoing task thresholds also contributed to the advantage of important PM accuracy: the full model fit most of the advantage (89% word, 96% non-word), whereas the model with

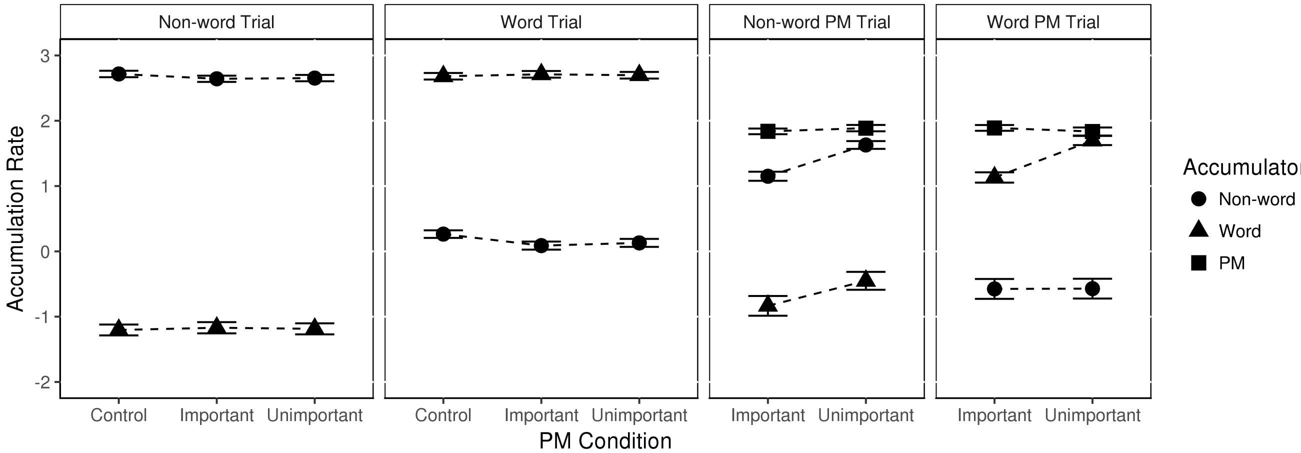

Figure 11. Experiment 2, posterior distributions of the mean accumulation rates, averaged across participants. The central symbols are the posterior means, and the bars are the mean + or - the posterior standard deviation. Plotted by stimulus type, by latent response accumulator, and by PM block. There is very little difference by PM block in the accumulation rates of the lexical decision stimuli, hence the overlapping symbols. False alarm accumulation (M = -2.71, SD = 0.11) not pictured.

Table 9
Experiment 2 Z values (with associated posterior predictive p values in brackets) for contrasts between non-PM trial mean accumulation rates.

Contrast	*Word Trial Accumulator*		*Non-word Trial Accumulator*	
	Correct	Error	Correct	Error
Unimportant - Control	0.88 (.19)	-3.41 (<.001)	-3.89 (<.001)	0.32 (.37)
Important - Control	1.68 (.05)	-4.31 (<.001)	-4.42 (0)	0.62 (.27)
Important - Unimportant	0.77 (.22)	-1.00 (.16)	-0.64 (.26)	0.29 (.39)

ongoing task thresholds set to control levels moderately underestimated it (77% word, 91% non-word).

We simulated a model with reactive control removed, by setting ongoing task accumulation rates on PM trials to non-PM trial levels. Removing reactive control resulted in the under-prediction of both PM-important accuracy (49% of word PM, 51% of non-word PM) and PM-unimportant accuracy (53% of word PM, 54% of non-word PM). Further, whereas the full model predicted most of the effect of PM importance on PM (89% word, 96% non-word), the model with reactive control removed did not (41% word, 43% non-word).

We also examined how differences across conditions in PM thresholds and PM accumulation rates contributed to the differences in important and unimportant PM. The PM-important accuracy advantage was facilitated by the PM threshold: with the PM thresholds averaged, the percentage of PM accuracy effect predicted was down from 89% to 52% for words and from 96% to 50% for non-words. In contrast, the effect of changes in PM rate on the PM-important advantage to accuracy were inconsistent: removing PM rate differences reduced the prediction of the effect from 89% to 86% for word PM targets, but actually increased the effect predicted from 96% to 103% for non-word targets.

Individual Differences

Correlations. As with Experiment 1, we used plausible-value correlations with population corrections to explore individual differences in PM accuracies, and the strength of our PM accuracy manipulation. We correlated PM accuracy with PM accumulation, PM thresholds, and also combinations of model parameters that index PM mechanisms: decreases in non-PM accumulation as compared with control conditions to index capacity sharing, increases in ongoing task thresholds as compared with control conditions to index proactive control, and decreases in accumulation towards ongoing task responses on PM trials as compared with non-PM trials to index reactive control. In contrast to Experiment 1, here there were two types of PM trials, word and non-word. We calculated separate correlations for PM accuracies on each type of trial.

Neither PM-important accuracy, nor PM-unimportant accuracy, nor the difference in the two, correlated with decreases in non-PM accumulation rates. In contrast, there were many correlations between PM accuracy and proactive control. For word PM targets, PM accuracy under important conditions was correlated with increases in word thresholds (.33) and non-word thresholds (.40), as was PM accu-

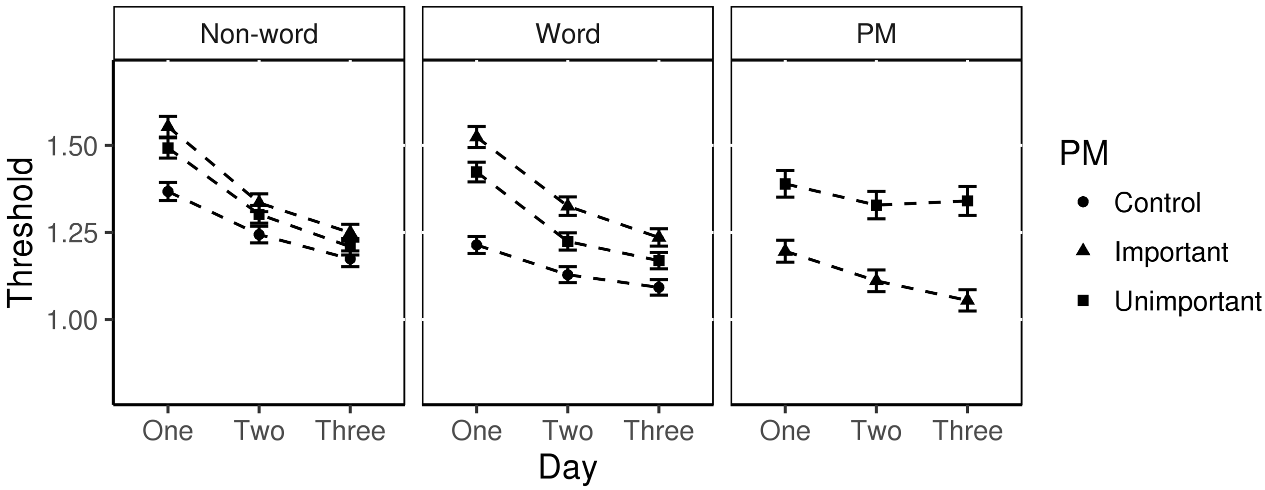

Figure 12. Experiment 2, response thresholds. The central dots are the means of the average posterior samples, and the bars are the mean + or - the standard deviation. Plotted by latent response accumulator, by PM block, by day.

Table 10
Experiment 2 Z values (with associated posterior predictive p values in brackets) for contrasts between thresholds.

Contrast	*Accumulator*		
	Word	Non-word	PM
Unimportant - Control	14.54 (0)	8.44 (0)	
Important - Control	22.29 (0)	13.18 (0)	
Important - Unimportant	9.30 (0)	4.88 (0)	-8.40 (0)

racy under unimportant conditions (word threshold correlation = .43; non-word = .38). For non-word PM targets, PM accuracy under important conditions was correlated with increases in non-word thresholds (.42) as compared with control, but not word thresholds. PM accuracy to non-words under unimportant conditions correlated with both increases in both word (.47) and non-word (.43) thresholds. We also examined how increases in ongoing task thresholds from PM-unimportant conditions to PM-important conditions correlated with the advantage of important PM accuracy. We only found one correlation, that between increases in proactive control over non-word and increased PM accuracy to non-words (.36). In sum, individual differences in PM accuracy were to some degree explained by individual differences in proactive control over ongoing task thresholds. Individual differences in the advantage of PM-important accuracy were explained by increased proactive control over non-word decisions for non-word PM targets, but not by increased proactive control over word decisions for word targets.

We next considered PM accumulator parameters. We found no correlation between PM rates and PM accuracies, and no correlations between changes in PM rate and the difference in important and unimportant PM. We found no correlation between PM thresholds and PM accuracy for important conditions, but for unimportant conditions we found negative correlations between PM threshold and both word PM accuracy (-0.36) and non-word PM accuracy (-0.38). Further, we found strong correlations between differences in PM thresholds over important and unimportant conditions and differences in PM accuracy for both word (-0.56) and non-word PM targets (-0.61). Thus, individual differences in PM thresholds explained a large amount of the individual differences in the effect of importance on PM accuracy.

Finally, we examined reactive control. For word PM targets, reactive control over word accumulation was strongly correlated with both important (0.66) and unimportant PM accuracy (0.73), whereas reactive control of non-word accumulation was not. For non-word PM targets, reactive control over non-word accumulation was strongly correlated with both important (0.52) and unimportant PM accuracy (0.56), whereas reactive control over word accumulation was not. Thus, reactive control over the 'correct' ongoing task accumulator for a given PM target explained a large degree of individual differences in PM accuracy to that target. We found no correlation between differences in reactive control across PM conditions and differences in PM accuracy. Thus, differences between PM-important and PM-unimportant reactive control were not associated with differences in the effect of importance on PM accuracy.

Table 11

Experiment 2, mean accumulation rate contrasts relevant to reactive control. We report Z (p), as defined in text. We separate the word trials (top half) and non-word trials (bottom half). In first and second row of each half we compare PM trials with non-PM trials. In the third row, we compare the difference in reactive control between PM-important and PM-unimportant blocks. Thus, the third row of each half displays the difference between PM blocks (Important- Unimportant) of the differences by PM trial status (PM - non-PM).

Contrast	*Accumulator on Word Trials*		
	Word	Non-word	PM
PM-unimportant: PM trials - Non-PM trials	-16.21 (0)	-4.27 (0)	47.25 (0)
PM-important: PM trials - Non-PM trials	-20.11 (0)	-4.03 (0)	57.99 (0)
Important difference - Unimportant difference	-8.75 (0)	0.21 (.42)	0.97 (.17)
Contrast	*Accumulator on Non-word Trials*		
	Word	Non-word	PM
PM-unimportant: PM trials - Non-PM trials	4.56 (0)	-19.13 (0)	52.33 (0)
PM-important: PM trials - Non-PM trials	1.93 (.03)	-21.40 (0)	56.84 (0)
Important difference - Unimportant difference	-2.40 (.008)	-8.17 (0)	-1.19 (.12)

Model Mechanisms. As with Experiment 1, we examined how different model mechanisms accounted for PM accuracy across individual participants by removing mechanisms from the model and examining the resulting PM accuracies for each individual. Supplementary materials contain plots of each model's fit to each individual's important PM accuracy, unimportant PM accuracy, and the difference between the two.

Our model exploration of individual participants was fairly consistent with our averaged model exploration. The contribution of the different pieces of PMDC (i.e., proactive ongoing task threshold control, proactive PM threshold control, reactive control) to PM-important and PM-unimportant accuracies were similar across participants. Removing proactive ongoing task threshold control resulted in consistent moderate underestimation of both important and unimportant PM accuracy across participants. Removing reactive control resulted in consistent large underestimation of PM accuracies.

As previously discussed, there was a large spread across participants in the effect of PM importance on PM accuracy, perhaps because participants varied in their sensitivity to our instructions to prioritize PM. Our exploration here revealed that for the participants where the advantage of important PM was substantial, all the PMDC mechanisms contributed. Removing proactive control of ongoing task thresholds produced a small underestimation of the effect, whereas removing reactive control produced a large underestimation, as did averaging PM thresholds. Averaging PM accumulation rates between important and unimportant conditions did not lead to consistent underestimation of the differences.

Discussion

The LBA model provided a good fit to the ongoing task accuracies and PM accuracies in Experiment 2, as well as a comprehensive distributional account of ongoing task and PM RTs. This included fit to the PM cost effect, and to the effects of manipulating PM importance on PM costs, PM accuracy, and PM RTs. We did not find evidence that capacity sharing between PM and ongoing task processes was responsible for the increased costs or increased PM accuracy in our PM-important condition. Instead, we found evidence that both proactive and reactive control were active over PM and ongoing decision processes, as specified by PMDC.

Capacity sharing. For both PM blocks, the shifts in non-PM trial accumulation rates as compared with control were inconsistent: non-PM non-word trials might be presented as evidence of capacity-sharing cost with PM (lower correct ongoing task accumulation), but non-PM word trials actually indicated evidence for higher ongoing task cognitive capacity with PM (lower error accumulation with PM). These accumulation effects were similar between the PM-unimportant blocks and PM-important blocks, despite the latter being argued to be particularly conducive to capacity sharing. In addition, the effects of PM condition on non-PM trial accumulation rates were substantially weaker than the increases in ongoing task thresholds, and were not associated with prediction of PM cost. Thus, our findings are consistent with Experiment 1, and previous work, in that they do not support the capacity sharing account of PM cost. We also did not find consistent increases in PM accumulation rates in the PM-important condition compared to the PM-unimportant condition. For word trials, there was a trend towards slightly faster PM accumulation with important PM, but for non-word trials, the PM accumulation rate was actually lower with important PM. Thus, not only was there a lack of evidence for capacity sharing, there was a lack of evidence for increased non-shared capacity when PM importance was emphasized.

Proactive Control. We found large increases in ongoing task response thresholds in PM blocks compared with control blocks, suggesting proactive control in favor of PM responding. We found that Experiment 2's PM task, in which PM items could be words and non-words, resulted in much larger threshold increases to non-words than in Experiment 1, where PM items could not be non-words. This is consistent with delay theory's proposition that ongoing task threshold increases are targeted specifically to slow down decision processes that might pre-empt the PM decision on PM trials (Heathcote, Loft, & Remington, 2015; Loft & Remington, 2013; Strickland et al., 2017). Furthermore, ongoing task threshold increases were larger in PM-important blocks. This suggests that when PM is important, participants applied greater proactive control over their ongoing task decisions.

One interesting finding is that, although we observed substantial overall PM cost, the increased PM cost we observed with PM importance was smaller than previous studies (only 0.024s). Perhaps this is due to our use of a within-subjects design, whereas, to our knowledge, all previous PM manipulations have used a between subjects design (Einstein et al., 2005; Kliegel, Martin, McDaniel, & Einstein, 2001; Kliegel et al., 2004; Loft et al., 2008). Proactive control settings may be subject to within-subjects carryover effects from the control blocks and PM-unimportant blocks (Poulton, 1982). We explored this possibility in the same way we explored carryover effects for Experiment 1: by re-parameterizing PMDC using within-day order ('block') as a factor, rather than across day order. The resulting parameters are presented in supplementary materials. In contrast to Experiment 1, here we found that differences in ongoing task thresholds across PM conditions did decrease for blocks performed later within a session, suggesting that indeed, within-subjects carryover reduced the strength of proactive control over ongoing task thresholds. However, we replicated Experiment 1's findings of no support for capacity sharing in non-PM accumulation even for the blocks that were performed first each day, suggesting that carryover effects did not mask capacity sharing.

Despite the small impact on PM cost, our importance manipulation was not weak *per se*, as the increase in PM accuracy with importance was quite substantial (21%). The fact that a large increase in PM accuracy could be achieved with only a small increase in PM cost suggests that PM accuracy and cost do not map entirely to a change in a single cognitive mechanism. Consistent with this, our model exploration revealed that although our cost parameters (the ongoing task threshold increases) did contribute to increased PM accuracy with importance, they contributed much less to the increase in PM-important PM accuracy than the PM accumulator parameters. For example, we found that PM importance caused proactive control over the PM threshold. In the PM-important condition, the PM threshold was lower, biasing responding in favor of PM. This caused around half of the PM accuracy effect. This effect on PM threshold was larger than the effect of focality on PM threshold from Experiment 1, consistent with the strategic nature of the importance manipulation.

Reactive Control. We found evidence for PM-induced reactive control in both PM-important and PM-unimportant blocks. As expected, in both PM blocks the PM accumulation rate was much higher on PM target trials than on non-PM trials. As in Experiment 1, accumulation towards the correct ongoing task response was greatly reduced on PM trials compared with non-PM trials, suggesting that processing of PM-related attributes inhibited accumulation towards this response. In contrast, there was not strong evidence that accumulation towards the incorrect ongoing task response was reduced on PM trials. However, the incorrect ongoing task accumulation rates for PM trials are very data-poor parameters (as we observed very few incorrect ongoing task responses on PM trials), and thus may have been subject to substantial influence from the prior, which specified no difference. Consistent with this, incorrect ongoing task accumulation rates for PM trials appeared lower than non-PM trial rates when the non-PM trial rates were above the prior value, but appeared higher than non-PM trial rates when the non-PM trial rates were far below the prior value.

We found that reactive control on PM trials was stronger when the PM task was important. In particular, when the importance of the PM task was emphasized, feedforward inhibition of the competing ongoing task response increased (i.e., accumulation towards the correct ongoing task response to PM items decreased). This feedforward inhibition accounted for around half of the advantage to PM accuracy associated with important PM. In contrast to Experiment 1's finding that PM accumulation was faster under focal conditions, we did not find that PM accumulation was substantially faster under PM-important conditions. This is consistent with reactive excitation being a stimulus-driven rather than a control-related effect. Furthermore, the lack of increase in PM accumulation with PM importance suggests there is not an increase in the capacity devoted to extracting PM-related information from the PM stimuli. Instead there appears to be an additional strategic adjustment (beyond proactive threshold adjustment) specific to reactive inhibition. It seems that under PM-important conditions participants alter their reactive control architecture to inhibit ongoing task accumulation more strongly when they encounter PM signal (in Figure 2, increasing the sensitivity of pathways B1 and B2).

Summary. The model analysis of Experiment 2 was consistent with dual mechanisms of PM decision control. We did not find evidence for capacity sharing between PM and ongoing task processes: we found no systematic difference in non-PM trial accumulation rates between PM and control conditions, even with important PM, and we found no increase in PM accumulation with important PM. Instead, PM cost was due to proactive control over ongoing task thresholds. There was more threshold control when PM importance was emphasized as compared with when ongoing task importance was emphasized, despite otherwise identical instructions. This suggests ongoing task thresholds can be modified for PM-specific reasons. Higher PM accuracy in PM-important blocks was caused in small part by the aforementioned increase in ongoing task thresholds, but in larger part by a decreased threshold to respond PM, and by increased inhibitory reactive control on PM trials in PM-important blocks.

General Discussion

We found that a three accumulator LBA provided a good fit to the entire array of observed PM data from two experiments. This included focal PM cost, non-focal PM cost, PM-important cost, and PM-unimportant cost, as well as cost from both stimulus-specific PM instructions (in which PM items could only be words) and non-specific PM instructions (in which PM items could be both types of ongoing task item). The model was able to account for the accuracy and RT of PM responding to all three of the PM tasks we used: the non-focal categorical task, the single-target focal PM task and the non-focal syllable detection task, and was also able to account for the effects of PM emphasis on PM accuracy. Thus, it appears that the three-accumulator LBA is sufficient to measure PM processes in the Einstein and McDaniel (1990) paradigm, and to predict many benchmark effects from the PM literature. The model indicated that both proactive and reactive control, as specified in our PMDC architecture, accounted for the range of observed effects.

Proactive Control. All four of our PM conditions replicated the nine previous data sets that found strong evidence for increased ongoing task response thresholds in PM blocks (Heathcote, Loft, & Remington, 2015; Horn & Bayen, 2015; Strickland et al., 2017; Ball & Aschenbrenner, 2017). This included the single target focal task (Strickland et al., 2017) as well as the two non-focal tasks (category and syllable detection)(Heathcote, Loft, & Remington, 2015; Strickland et al., 2017). Furthermore, the effect occurred under both PM-unimportant and PM-important conditions. In terms of PMDC, this demonstrates proactive control of ongoing task decision processes in favor of the PM decision. In Experiment 1, which used focal and non-focal PM tasks that were stimulus-specific (i.e., PM targets were always words), the threshold of the word accumulator increased substantially more under PM conditions than the non-word threshold. In addition, the non-word threshold did not substantially increase under focal PM conditions, in which the PM decision process was relatively fast. In Experiment 2, when the PM task was non-specific (i.e., PM targets were words and non-words) there were substantial increases in both word and non-word thresholds. Further, in Experiment 2 when we instructed participants that the PM task was more important than the ongoing task, their ongoing task thresholds increased more than we instructed them the ongoing task was more important, enabling higher PM accuracy. This combination of results suggests that ongoing task threshold increases are specifically implemented to allow time for PM accumulation on PM trials. This form of proactive decision control is consistent with the delay theory of PM cost (Loft & Remington, 2013; Heathcote, Loft, & Remington, 2015).

Although our stimulus-specific instructions in Experiment 1 led to a large word threshold increase, we did also observe a smaller increase in non-word thresholds in the non-focal PM blocks. As the latter would have minimal impact on PM accuracy, it appears that the non-focal PM instruction induces general threshold changes in addition to PM-specific delay. Horn and Bayen (2015) proposed that PM instructions cause participants to increase general response caution, because they perceive the task to be more complex. In other words, they suggested that the threshold increases underlying PM cost owe to a change in participants' general impression of the task, rather than the threshold increases being implemented for the sake of PM performance on PM trials. Alternatively, non-word thresholds may increase with word trial specific PM because participants do not realize when encoding the PM instruction that only increasing the word decision threshold is most efficient (for further discussion see Strickland et al., 2017). In any case, the magnitude of the effect of non-focal PM on non-word thresholds was much smaller than the effect on word thresholds, and so much of the threshold effect can be attributed to delaying word responses in favor of PM responding.

Unlike previous models, which did not predict PM accuracy (Heathcote, Loft, & Remington, 2015; Horn & Bayen, 2015; Strickland et al., 2017; Ball & Aschenbrenner, 2017), PMDC quantitatively estimates the extent to which proactive control over ongoing task decisions benefits PM accuracy. For both experiments, we found that conditions with higher ongoing task thresholds (non-focal and PM-important), did have higher PM accuracy as a result of those thresholds. However, variations in PM accuracy were not only a result of differences in ongoing task thresholds. In fact, differences between conditions in PM-accumulator parameters were more influential than differences in ongoing task parameters. This is obvious in Experiment 1, because focal PM accuracy was higher than non-focal PM despite lower cost. However, in Experiment 2 PM cost and PM accuracy increased together. Only with modeling could we parse the relative contribution to the advantage of PM-important accuracy of our cost-related mechanism (proactive control over ongoing task decisions) from the contribution of other mechanisms such as differences in PM thresholds and reactive control. This demonstrates a general benefit of quantitative theorizing, in which not only the direction of effects is predicted, but also the magnitude (Pitt, Myung, & Zhang, 2002).

We wish to make one parting point about PMDC's proactive control over thresholds. In contrast to our instantiation of proactive control as increased ongoing task thresholds, another recent paper has proposed that proactive control corresponds to PM target checking (Ball & Brewer, 2017). Ball et al. argued that PM checks caused increases to the μ parameter that they observed when fitting the ex-Gaussian distribution to PM costs. However, although μ costs could be brought about by increased non-decision time (which might indicate a serial PM check), they could also be brought about by increased thresholds (Matzke & Wagenmakers, 2009). Thus, the μ increases that Ball et al. observed actually could have been caused by proactive control over ongoing task decisions as specified by PMDC. We prefer this interpretation, because it is embedded in a clear process theory of PM that has been fit to, and supported by, actual PM accuracy and PM RT data. In contrast, we do not yet know what type of PM checking theory could fit to actual PM accuracy and PM RT data. Ball et al.'s μ costs range between 0.04s and 0.09s, which, given that mean ongoing task RTs were between 0.8s - 0.96s in control conditions, does not seem enough extra time to contain a PM decision that runs in series to the ongoing task decision.

Proactive control was not only evident in ongoing task delay, but also in the PM threshold. In Experiment 2, the PM response threshold was lower in PM-important blocks than in PM-unimportant blocks. This follows straightforwardly from PMDC. Under PMDC, the PM response threshold is malleable, and when PM is important it could be decreased in order to increase the probability that the PM accumulator reaches response selection before the ongoing task accumulators. In Experiment 1, the PM accumulator threshold was higher in non-focal PM blocks than in focal PM blocks. We suggest that this occurred because participants predicted it would be more difficult to correctly distinguish non-focal targets from non-targets, compared to focal targets from non-targets. However, the PMDC framework made no *a priori* prediction about the magnitude of focal vs non-focal PM response thresholds, and, to our knowledge, no current PM theory does. Future work should develop a clear process account of this effect.

That we could quantify PM threshold effects, and the extent to which they were responsible for changes in PM accuracy, exemplifies another benefit of quantitative theorizing about PM processes. Without accounting for the contribution of thresholds to PM accuracy effects, we would not be able to attribute changes in PM accuracy to other PM-related parameters. For example, an extremely low PM threshold could mimic some of the effects of fast PM accumulation on PM trials. Because threshold effects are controlled for in our models, we can safely attribute additional variation in PM accuracy to accumulation rate effects. In our experiments, identifying these accumulation rate effects on PM trials was crucial to revealing PM-induced reactive control.

Reactive control. We found strong evidence of reactive control on PM trials. Unsurprisingly, excitation of the PM accumulator (i.e., accumulation towards the PM response) increased on PM trials, especially when PM targets were focal. More interesting was the finding that the rates of ongoing task response accumulation (i.e., accumulation to the 'word' and 'non-word' responses) also decreased on PM trials, despite stimulus evidence for the ongoing response being the same across PM and non-PM trials. PMDC attributes this effect to feedforward inhibition of ongoing task processing from PM detection processes. Despite very large decreases in ongoing task accumulation on PM trials, we did not observe slowing when we compared raw ongoing task RTs between PM trials and non-PM trials. Our modeling suggests that this results from statistical facilitation from the PM accumulator, such that the slow ongoing task RTs that would be observed on PM trials are absorbed into the PM RT distribution (Raab, 1962). This is consistent with Gilbert et al. (2013), who found the same lack of difference in ongoing task RT in simulation despite specifying lateral inhibition between PM and ongoing task decisions in their model.

To our knowledge, acute, PM trial-induced inhibition as specified by PMDC (and by Gilbert et al., 2013) is the only mechanism proposed to date that clearly predicts our finding of reduced ongoing task processing rates on PM trials. However, *post-hoc* we can think of at least two alternative explanations for the finding, both of which we have reason to reject. One alternative is that participants engage in acute capacity sharing on PM trials, rather than the ubiquitous capacity sharing specified by previous PM theory that occurs on both non-PM and PM trials (e.g., Smith, 2003). For example, on PM trials processing of PM target features during encoding time may cause participants to shunt subsequent attention at the decision stage towards PM accumulation and away from ongoing accumulation. Although this could equally account for the decreased ongoing task accumulation on PM trials, it is inconsistent with the effects of our focality and importance manipulations. When PM was focal rather than non-focal, there was a larger decrease in ongoing task accumulation on PM trials. As capacity sharing theories argue that focal PM tasks are less capacity demanding than non-focal PM tasks (e.g., McDaniel & Einstein, 2000; Smith, 2008), they would predict that ongoing task accumulation should be faster, not slower, on PM trials in focal blocks compared with non-focal blocks. In contrast, increased reactive control of ongoing accumulation with focality follows directly from our proposed PMDC architecture (more activation of the PM detector increasing input to feedforward inhibition, Figure 2). Further, when PM was important, there was a large decrease in ongoing task accumulation on PM trials, but not substantial effects on PM accumulation. If importance increased the capacity acutely shunted towards the PM task, then the capacity of the PM accumulator (i.e., the PM accumulation rate) should have increased in the PM-important blocks.

The second alternative explanation is that participants might begin to suppress ongoing task accumulation (or devote capacity to PM) when PM had not occurred for a while, if they assume that another PM cue is coming soon. To investigate this possibility, we did follow-up modeling in which we treated the three non-PM trials prior to each PM trial separately from other non-PM trials (see supplementary materials for results). We found not much difference between rate parameters proximal to PM trials, and rate parameters for the rest of the non-PM trials, suggesting that this explanation has little merit.

In summary, we found that the PMDC architecture, that is, a three-accumulator LBA with feedforward excitation and inhibition, provided a quantitatively sufficient and conceptually cohesive account of all features of PM data from two benchmark PM manipulations. Regarding non-PM trials, the key contribution of our analysis was to implicate threshold control in PM cost. This extends and replicates previous findings (Heathcote, Loft, & Remington, 2015; Horn & Bayen, 2015; Strickland et al., 2017; Ball & Aschenbrenner, 2017), and suggests that PM cost reflects delaying of ongoing task decision processes so that the PM process has more time to complete. In addition to this replication, our complete model, that included both ongoing task responses and PM responses, afforded us novel insights into PM. We found that the threshold to make the PM response varied, both as a function of PM focality and PM importance. This indicates that proactive control applies not only to ongoing task decision processes, but also PM decisions. We also found strong evidence of reactive control of ongoing task accumulation on PM trials. This reactive control was stronger when PM was focal, and when PM was important. Focal PM was also associated with increased PM excitation, suggestive of more PM activation from PM stimulus features, whereas important PM was not associated with increased excitation, suggesting a strategic ramping up of reactive inhibition.

No Support for Capacity Sharing

We did not find evidence for capacity sharing between PM processes and the ongoing task in either of our experiments. Although there was some variation by PM block in non-PM trial accumulation rates, in both experiments we found multiple effects that pointed in opposite directions, with no clear net loss of ongoing task capacity. In Experiment 1, neither focal nor non-focal PM caused a clear capacity cost to non-PM trial accumulation, and there was no clear difference in non-PM trial accumulation rates between the two, either. We also did not find evidence of PM cost to non-PM trial accumulation rates in Experiment 2, even when the importance of the PM task was emphasized, which has been argued to increase the magnitude, and likelihood of, PM capacity sharing effects (McDaniel & Einstein, 2000; Smith & Bayen, 2004). Taking our findings together with previous work (Heathcote, Loft, & Remington, 2015; Horn & Bayen, 2015; Strickland et al., 2017; Ball & Aschenbrenner, 2017), eleven modelled data sets have now revealed no cost to non-PM trial accumulation under PM conditions. Thus, a central claim of extant PM theories, that PM cost results from capacity sharing between PM processes and the ongoing task (e.g., McDaniel & Einstein, 2000; Smith, 2003), is not supported by our fine-grained analysis of the latent variables underlying the PM cost effect. In addition, our finding that non-PM trial accumulation rates do not vary over focal and non-focal conditions challenges the assumption of current PM theories that focal PM has lower PM cost and higher PM accuracy due to the PM task requiring less ongoing task capacity (Einstein & McDaniel, 2005; Smith et al. 2007). Our finding that non-PM trial rates do not vary between PM-important and PM-unimportant conditions challenges the assumption that PM importance is associated with higher cost and higher PM accuracy because it causes more ongoing task capacity to be allocated to the PM task (Einstein & McDaniel, 2005; Smith & Bayen, 2004). Furthermore, in Experiment 2 we failed to find any effect of PM importance on non-shared PM capacity, that is PM capacity that is not relevant to ongoing task performance. Had non-shared PM capacity increased when PM importance was emphasized, we would expect a faster PM accumulation rate in PM-important blocks, which we did not find.

We do not, however, argue that PM and ongoing task capacity sharing never occurs under any circumstances. If we venture outside the canonical laboratory PM paradigm devised by Einstein and McDaniel (1990), into paradigms that are somehow more representative of everyday, or safety critical, PM tasks, we may discover evidence of a PM capacity burden on ongoing tasks. For example, in the majority of PM experiments, including those modeled, PM stimulus features are readily apparent and processed with the same perceptual information as the ongoing task. In contrast, safety critical tasks such as air traffic control require that the operator pay attention to many different sources of perceptual information (Loft, 2014). Under these conditions, participants may incur a PM capacity cost to the ongoing task if they are unable to simultaneously attend to PM-relevant and ongoing task-relevant information.

Limitations and Future Directions

Because specifying a comprehensive cognitive model requires taking a stance on difficult meta-theoretical issues, our approach invites criticisms that verbal PM theory has avoided. One such criticism of PMDC is that PM and ongoing task processes may occur in series, (e.g., the PM process occurs followed by the ongoing task process, or vise versa), rather than in parallel as we specified. Other than pointing out that our parallel architecture fit the data well, we would also point out that previous PM theories, although not forthcoming with their underpinning architectural assumptions, seem to be consistent with parallel PM task and ongoing task processing. The preparatory attentional and memory processes theory proposes that preparatory PM processes occur "on the periphery of our attentional focus" while the ongoing task is being performed (Smith, 2010). The Multiprocess view claims that spontaneous processes occur when "cue-driven thoughts unrelated to the ongoing task" enter consciousness during ongoing task processing (Einstein & McDaniel, 2010). The Associative Activation Theory of PM (Nowinski & Dismukes, 2005) explicitly specifies a parallel architecture, in which activation of PM and ongoing task intentions accumulate to threshold, at a rate driven by the contents of "focal attention". More generally, associating PM cost with capacity sharing requires assuming parallel processing, as serial processes would extend response latency whether they are capacity consuming or not (Navon, 1984).

A reviewer pointed out that there is a serial component in the PMDC model, namely non-decision time, and that PM processes might increase this time. Indeed, Horn and Bayen (2015) argued that DDM estimates of non-decision time increase with PM cost because participants perform a sequential PM check. We initially did not model differences in non-decision time across PM conditions, as we were concerned about allowing too much flexibility in our model, and as our previous work indicated that this parameter plays no role in the LBA account of PM cost (Heathcote, Loft, & Remington, 2015; Strickland et al., 2017). However, to address the possibility of a serial PM check within the PMDC architecture, we re-fitted the model with non-decision time permitted to vary over PM condition. Supplementary materials include the resulting parameter estimates. Overall, we did not find compelling evidence for PM checking during non-decision time. Non-decision time actually *decreased* under PM conditions, and decreased most when PM was important (32ms). We did find one increase in non-decision time of 12ms for non-focal conditions in Experiment 1. However, as most other effects went in the opposite direction, and as 12ms is a very small amount of time relative to ongoing task decision time, we think it is very unlikely that this sole effect reflected as sequential PM check. Given the largely unexpected direction of these non-decision time effects, we suspect that, indeed, flexibility in non-decision time led to 'over-fitting', and thus in the main body of the text we maintained our PMDC account of the data absent that flexibility.

Evidence for our reported PMDC mechanisms largely remained even with non-decision time effects included in the model (see supplementary materials). However, there were a couple of differences for Experiment 1. One, focal cost to word thresholds increased up to meet non-focal levels of cost, likely due to trading off with decreases in non-decision times in focal conditions (below control). Given this results from non-decision time decreasing with focal PM, an implausible effect, we suspect over-fitting. The second, perhaps more interesting, difference is that non-word threshold elevations under non-focal conditions were absent. This threshold effect was the only one that does not fall directly out of PMDC theory - as all PM targets were words in this experiment, the non-word decision was a weak competitor to the PM decision, and thus raising its threshold should have little benefit to PM. Thus, we recommend it be viewed with more skepticism than the other reported effects.

We also recommend further investigation into the issue of serial vs parallel PM and ongoing task processes. Adjudicating empirically between serial and parallel architectures with RT and accuracy data alone is difficult, particularly when cognitive capacity allocation may vary. In some highly controlled paradigms, it is possible with an array of factorial comparisons of observed RT and accuracy, given some modest assumptions(i.e., Systems Factorial Technology: Fific, Nosofsky, & Townsend, 2008). It could be very interesting to develop PM experiments along these lines, although the PM paradigm would need to be modified. The other solution, in line with our approach, is to examine specific computational theories that commit to a serial or parallel architecture. As differing computational theories emerge, they can be compared in terms of the theoretical leverage they offer, or on quantitative grounds (i.e., model selection). For example, it might be possible to instantiate Horn and Bayen (2015)'s target checking hypothesis in a computational model. We could then quantitatively compare the resulting model with PMDC.

One limitation of the current work is that we only modeled one type of PM response mode. Our PM instruction was identical to that of Strickland et al. (2017), which told participants to make the PM response instead of the ongoing task response. Horn and Bayen (2015) used a similar instruction to "press the target key immediately when presented with a target". Heathcote et al. and others have used a slightly different instruction to simply make the PM response when the target was presented (e.g., Einstein et al., 2005; Loft & Humphreys, 2012; Loft & Remington, 2013; Scullin, McDaniel, & Einstein, 2010; Scullin, McDaniel, Shelton, & Lee, 2010; Smith, 2003; Smith & Bayen, 2004; Taylor, Marsh, Hicks, & Hancock, 2004). Although this latter instruction very typically results in participants choosing to respond by pressing the PM response key instead of the ongoing task response key, it is conceivable that PM actions may be retrieved during the "down time" before the next trial, or during the next trial itself. However, late PM responses have been rare when participants are specifically instructed that late PM responses can be made. This suggests it is critical that PM response selection is made before ongoing task response selection. For this reason we are confident that conclusions from the current modeling architecture would generalize to these similar forms of PM response mode instructions. A further type of instruction is to explicitly tell the participant to make the ongoing task response before the PM response (Loft, Kearney, & Remington, 2008; Loft & Yeo, 2007; Marsh et al., 2005; Marsh et al., 2003; Meeks, Hicks, & Marsh, 2007; Meeks & Marsh, 2010), although it might be argued to be less relevant to PM tasks outside the laboratory, which usually interrupt other activities (Einstein & McDaniel, 1996; Loft et al., 2013). As outlined in more detail by Heathcote et al. (2015), terminating stimulus processing (response selection) to make an ongoing task response would still interrupt 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 and can no longer be assessed for PM response selection), and thus we would expect the current modeling architecture to generalize to this PM response mode. In fact, costs have been accounted for by threshold, and not accumulation rates, across all three types of PM instruction manipulations. That is, the Heathcote et al. (2015) experiments used the instruction for participants to make the PM response when the target was presented, Strickland et al. used the 'respond instead' instruction, and Heathcote et al. found that response threshold accounted for PM costs in the Lourenço et al. experiment, in which participants were specifically instructed to perform their PM response after their ongoing task response. However, as PM trials were not modeled in these studies, it would be worthwhile for future research to investigate how PM thresholds and reactive control will operate using the alternate PM response mode. That said, because PM RTs in the 'PM after' paradigm are confounded by ongoing task response production time, it would be difficult to investigate this with decision process models of RT like the LBA.

Another possible criticism of our model is that we have not invoked the assumption that PM errors can occur because of a complete failure of the PM process to initiate. In the modeling we present, the PM accumulator runs on every trial, and PM errors only occur because the PM accumulator loses the race to response selection. It is more common to assume that at least some PM errors are caused by the PM process completely failing to occur (e.g., McDaniel & Einstein, 2000; Smith & Bayen, 2004). It appears this assumption was not necessary to predict PM accuracies in our experiments. However, this may owe in part to our PM frequency (1:14), which, although within the bounds of previous research (where PM is as high as 20%, most typically the limit is 10%), is on the higher end. Recently, authors have suggested the dynamic multiprocess framework (Scullin et al., 2013), in which PM monitoring processes only truly dissipate when PM tasks are extremely infrequent, for example when PM ratios are around 1:100. This suggests a distinction between the PM errors typically observed in laboratory paradigms, where PM frequency tends to be much higher than 1:100, and PM errors in everyday work or personal life, where a PM task may only be need to be performed rarely. In the dynamic multiprocess view, these failures of monitoring have referred to failures to devote ongoing capacity to the PM task, but, giving the lack of support for capacity sharing between PM and ongoing tasks in our analyses, it might be better to conceptualize the absence of the PM accumulator another way, for example, as a failure to maintain the intent to treat ongoing task items as potential retrieval cues (Tulving, 1985)

Identifying whether PM errors occurred because the PM decision process lost a race to the ongoing task decision process, as compared with errors that occurred because the PM decision process failed to even enter the race, requires a process theory of PM such as our PMDC model. The LBA has already been successfully used to identify the proportion of trials on which such failures occur in the context of the stop-signal task, where they are referred to as "trigger failures" (Matzke, Hughes, Badcock, Michie, & Heathcote, 2017; Matzke, Love, & Heathcote, 2017). In the context of PM such trigger failures would correspond to trials in which either the intent to detect PM targets is not maintained (goal neglect; Duncan, Emslie, Williams, Johnson, & Freer, 1996) and thus there is a subsequent complete lack of the recognition of any of the PM features contained in the stimulus, and as a consequence the PM decision process never enters the race. Ideally, to identify such failures of the PM accumulator to run, an experiment would include a manipulation that targets PM trigger failures (e.g., introducing an unfamiliar ongoing task context not associated with the PM cue and associated PM response; Smith, 2017). We recommend future PM work pursues paradigms along these lines that, in apparent contrast to the paradigms considered here, require the PMDC model to include failures of the PM accumulator to run.

Another future challenge for our framework is to distinguish the retrospective component of PM - remembering *what* to do and *when* - from the prospective component of PM - remembering that some PM action must be performed (Einstein & McDaniel, 1990). Smith and Bayen (2004) demonstrated that their multinomial processing tree model of PM could disentangle the contribution of prospective and retrospective memory processes to categorical response data (e.g., PM accuracy, ongoing task accuracy). It may be possible to do the same, while also incorporating RT data, by building on the PMDC model. Regarding memory for *what* to do, our paradigms in the current work required a very simple PM response (as is typical in the PM literature) - e.g., press the 'j' key. This intention probably does not burden retrospective memory, and so for our studies it made sense to model the entire PM decision as an evidence accumulation process, with no subsequent failures to forget how to perform the PM action. However, in the real world, PM tasks can be more complex, and it conceivable that in some instances people might remember that something needs to be done (i.e., the PM accumulator wins the race to threshold), but subsequently fail to remember what to do. Incorporating this type of retrospective memory failure, in which the PM intention breaches threshold but the PM response cannot be retrieved, may require adding a parameter for the probability of a post-decision response-production-failure to the PMDC model.

Regarding memory for *when* to perform the PM response, that is memory for which items are PM targets, our studies were also relatively simple: we required participants respond to either a single target word, single target syllable, or single target category. In contrast, some other PM paradigms require PM responses to a multi-item PM target list (e.g., Cohen, Jaudas, & Gollwitzer, 2008; Hicks et al., 2005; Loft, Humphreys, & Whitney, 2008). Whereas it should be relatively easy to maintain a single PM target in memory, maintaining a list of targets might require either rehearsal of the list, or perhaps maintaining a pointer to the stored multi-target list (Humphreys, Murray, & Maguire, 2009). We foresee two ways PMDC might identify the effects of these processes. First, high retrospective memory demands might lead to weaker PM activation of a target from a multi-target list when that target is encountered in the ongoing task, without memory for the target failing entirely. This could manifest in the capacity of the PM accumulator, with a long target list leading to less activation for any given single processed target due to noise from other targets in the list, slowing the PM accumulation rate. Second, high retrospective memory demands might result in retrospective memory lapses or failures (i.e., forgetting that an target is in the PM set). In this case, virtually no PM evidence at all would accrue to PM trials that present forgotten targets. This might manifest in the model as 'trigger failure' (i.e., the PM accumulator failing to enter the race), and so require the addition of a trigger-failure-probability parameter.

Two further new directions may be important for future work. One is to model the effects of PM "lures" (Knight et al., 2011; Scullin et al., 2009; Scullin, McDaniel, & Einstein, 2010). To recap, lures can be trials that match some, but not all, of the PM detection rule in PM blocks of trials. They can also be trials in control blocks that present PM targets corresponding to a suspended PM intention (i.e., a PM intention that should be performed later, but not in that block of trials). Responses are slower to both types of lure trials than to other non-PM trials, which we conjecture is due to reactive control of ongoing task accumulation. It would be useful to test this quantitatively by fitting PMDC to a design with lure trials. If lure effects are due to reactive inhibition, lure trials should result in lower ongoing task accumulation rates than typical non-PM trials.

Another important direction for future research is to connect PMDC to neural mechanisms. The PM literature has already accrued an interesting set of neurological findings (Cona, Scarpazza, Sartori, Moscovitch, & Bisiacchi, 2015). For example, the neuroimaging has revealed different patterns of activation in the anterior prefrontal cortex and the broader frontoparietal network in PM blocks of trials when compared with control blocks of trials (e.g., McDaniel et al., 2013). In non-PM related neuroimaging studies that have fitted the LBA, threshold adjustments have been associated with the striatum and the pre-supplementary motor area (e.g., Forstmann et al., 2010). It would be interesting to see whether these regions interact when participants shift their thresholds in PM blocks, for example if PM-related regions (anterior prefrontal cortex and the broader frontoparietal network) send signals to the striatum and the pre-supplementary motor area to control thresholds. PM studies have also found a diverse set of acute, transient patterns of neurological activation on PM trials (e.g., McDaniel et al., 2013), which have been attributed to spontaneous PM processes, and/or reactive control. It would be worth investigating which of these patterns of activation are associated with the reactive inhibition of ongoing task processing in the PMDC model. To explore these issues further, a 'model-based cognitive neuroscience' (Forstmann & Wagenmakers, 2015) approach, in which neural measurements are mapped to parameter estimates from models, would be useful. Practically speaking, mathematical and neural approaches to PM go together well, because both require collecting many PM trials.

References

Appelbaum, L. G., Boehler, C. N., Davis, L. A., Won, R. J., & Woldorff, M. G. (2014). The dynamics of proactive and reactive cognitive control processes in the human brain. *Journal of Cognitive Neuroscience*, *26*, 1021–1038. doi:10.1162/jocn_a_00542

Arnal, J. (2008). *A stochastic model of prospective memory* (Doctoral dissertation, University of Arkansas).

Ball, B. H. (2015). *Cognitive control processes underlying continuous and transient monitoring processes in event-based prospective memory* (Doctoral dissertation, Arizona State University). Retrieved from https://repository.asu.edu/attachments/150526/content/Ball_asu_0010E_14793.pdf

Ball, B. H., & Aschenbrenner, A. (2017). The importance of age-related differences in prospective memory: Evidence from diffusion model analyses. *Psychonomic Bulletin & Review*. doi:10.3758/s13423-017-1318-4

Ball, B. H., & Brewer, G. A. (2017). Proactive control processes in event-based prospective memory: Evidence from intraindividual variability and ex-Gaussian analyses. *Journal of Experimental Psychology. Learning, Memory, and Cognition*. doi:10.1037/xlm0000489

Boehler, C. N., Schevernels, H., Hopf, J.-M., Stoppel, C. M., & Krebs, R. M. (2014). Reward prospect rapidly speeds up response inhibition via reactive control. *Cognitive, Affective, & Behavioral Neuroscience*, *14*, 593–609. doi:10.3758/s13415-014-0251-5

Bowden, V. K., Visser, T. A., & Loft, S. (2017). Forgetting induced speeding: Can prospective memory failure account for drivers exceeding the speed limit? *Journal of Experimental Psychology: Applied*, *23*, 180–190.

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

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*, 106–113. doi:10.1016/j.tics.2011.12.010

Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: Effects of frequency, inhibition and errors. *Cerebral Cortex*, *11*, 825–836. doi:10.1093/cercor/11.9.825

Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., ... Mumenthaler, M. S. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General*, *130*, 746. doi:10.1037/0096-3445.130.4.746

Braver, T. S., Gray, J. R., & Burgess, G. C. (2007). Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 76–106). New York: Oxford University Press.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, *7*, 434–455. doi:10.1080/10618600.1998.10474787

Brown, S., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178. doi:10.1016/j.cogpsych.2007.12.002

Bugg, J. M., McDaniel, M. A., & Einstein, G. O. (2013). Event-based prospective remembering: An integration of prospective memory and cognitive control theories. In D. Reisberg (Ed.), *The oxford handbook of cognitive psychology* (pp. 267–283). Oxford, England: Oxford University Press. doi:10.1093/oxfordhb/9780195376746.013.0018

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523–547. doi:10.1037/0033-295x.97.4.523

Burgess, G. C., & Braver, T. S. (2010). Neural mechanisms of interference control in working memory: Effects of interference expectancy and fluid intelligence. *Plos One*, *5*, e12861. doi:10.1371/journal.pone.0012861

Chevalier, N., Martis, S., Curran, T., & Munakata, Y. (2015). Metacognitive processes in executive control development: The case of reactive and proactive control. *Journal of Cognitive Neuroscience*, *27*, 1125–1136. doi:10.1162/jocn_a_00782

Cohen, A. L., Jaudas, A., & Gollwitzer, P. M. (2008). Number of cues influences the cost of remembering to remember. *Memory & Cognition*, *36*, 149–156. doi:0.3758/MC.36.1.149

Cona, G., Scarpazza, C., Sartori, G., Moscovitch, M., & Bisiacchi, P. S. (2015). Neural bases of prospective memory: A meta-analysis and the “attention to delayed intention”(AtoDI) model. *Neuroscience & Biobehavioral Reviews*, *52*, 21–37.

Dennis, S. (1995). The Sydney Morning Herald word database. In *Noetica: Open forum* (Vol. 1). Retrieved from http://psy.uq.edu.au/CogPsych/Noetica

Dismukes, R. (2012). Prospective memory in workplace and everyday situations. *Current Directions in Psychological Science*, *21*, 215–220. doi:10.1177/0963721412447621

Dismukes, R., & Nowinski, J. (2006). Prospective memory, concurrent task management, and pilot error. In A. Kramer, D. Wiegmann, & A. Kirlik (Eds.), *Attention: From theory to practice* (pp. 225–236). New York: Oxford University Press. doi:10.1093/acprof:oso/9780195305722.003.0016

Donkin, C., Brown, S., & Heathcote, A. (2011). Drawing conclusions from choice response time models: A tutorial using the linear ballistic accumulator. *Journal of Mathematical Psychology*, *55*, 140–151. doi:10.1016/j.jmp.2010.10.001

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology*, *30*, 257–303. doi:10.1006/cogp.1996.0008

Eidels, A., Donkin, C., Brown, S. D., & Heathcote, A. (2010). Converging measures of workload capacity. *Psychonomic Bulletin & Review*, *17*, 763–771. doi: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*, 717–726. doi: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. doi:10.1037/a0019184

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

Ellis, J., & Milne, A. (1996). Retrieval cue specificity and the realization of delayed intentions. *The Quarterly Journal of Experimental Psychology*, *49*(A), 862–887. doi:10.1080/713755662

Fific, M., Nosofsky, R. M., & Townsend, J. T. (2008). Information-processing architectures in multidimensional classification: A validation test of the systems factorial technology. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 356. doi:10.1037/0096-1523.34.2.356

Forstmann, B. U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E.-J., … Turner, R. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences*, *107*, 15916–15920.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., Von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*, *105*, 17538–17542. doi:10.1073/pnas.0805903105

Forstmann, B. U., & Wagenmakers, E.-J. (2015). *An introduction to model-based cognitive neuroscience*. Springer.

Gawande, A. A., Studdert, D. M., Orav, E. J., Brennan, T. A., & Zinner, M. J. (2003). Risk factors for retained instruments and sponges after surgery. *New England Journal of Medicine*, *348*, 229–235. doi:10 . 1056 / NEJMsa021721

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian data analysis, third edition*. Chapman & Hall/CRC Texts in Statistical Science. London: Taylor & Francis. Retrieved from https://books.google.com.au/books?id=ZXL6AQAAQBAJ

Gilbert, S. J., Hadjipavlou, N., & Raoelison, M. (2013). Automaticity and control in prospective memory: A computational model. *Plos One*, *8*, e59852. doi:10.1371/journal.pone.0059852

Gobell, J. L., Tseng, C.-h., & Sperling, G. (2004). The spatial distribution of visual attention. *Vision Research*, *44*, 1273–1296. doi:10.1016/j.visres.2004.01.012

Guynn, M. (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. doi:10 . 1080 / 00207590344000178

Harrison, T., & Einstein, G. O. (2010). Prospective memory: Are preparatory attentional processes necessary for a single focal cue? *Memory & Cognition*, *38*, 860–867. doi:10.3758/MC.38.7.860

Heathcote, A., Brown, S. D., & Wagenmakers, E.-J. (2015). An introduction to good practices in cognitive modeling. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 25–48). New York: Springer.

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. doi:10.1037/a0038952

Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology*, *3*, 292. doi:10.3389/fpsyg.2012.00292

Hicks, J. L., Marsh, R., & 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. doi:https://doi.org/10.1016/j.jml.2005.04.001

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. doi:10.1037/a0022808

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. doi:10 . 1037 / a0037676

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.

Humphreys, M. S., Murray, K. L., & Maguire, A. M. (2009). Contexts and control operations used in accessing list-specific, generalized, and semantic memories. *Cognitive Psychology*, *58*, 311–337.

Irlbacher, K., Kraft, A., Kehrer, S., & Brandt, S. A. (2014). Mechanisms and neuronal networks involved in reactive and proactive cognitive control of interference in working memory. *Neuroscience & Biobehavioral Reviews*, *46*, 58–70. doi:10.1016/j.neubiorev.2014.06.014

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

Kalanthroff, E., Avnit, A., Henik, A., Davelaar, E. J., & Usher, M. (2015). Stroop proactive control and task conflict are modulated by concurrent working memory load. *Psychonomic Bulletin & Review*, *22*, 869–875. doi:10.3758/s13423-014-0735-x

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*, 627–633. doi:10.3758/BRM.42.3.627

Kidder, D. P. (1999). *Effects of switching attention between tasks on age differences in prospective memory* (Doctoral dissertation, Georgia Institute of Technology).

Kidder, D. P., Park, D. C., Hertzog, C., & Morrell, R. W. (1997). Prospective memory and aging: The effects of working memory and prospective memory task load. *Aging, Neuropsychology, and Cognition*, *4*, 93–112. doi:10.1080/13825589708256639

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. doi:10.1080/09658210344000099

Kliegel, M., McDaniel, M. A., & Einstein, G. O. (2008). *Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives*. Mahwah, NJ: Lawrence Erlbaum.

Knight, J. B., Meeks, J. T., Marsh, R. L., Cook, G. I., Brewer, G. A., & Hicks, J. L. (2011). An observation on the spontaneous noticing of prospective memory event-based cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 298–307. doi:10.1037/a0021969

Locke, H. S., & Braver, T. S. (2008). Motivational influences on cognitive control: Behavior, brain activation, and individual differences. *Cognitive, Affective, & Behavioral Neuroscience*, *8*, 99–112. doi:10.3758/CABN.8.1.99

Loft, S. (2014). Applying Psychological Science to Examine Prospective Memory in Simulated Air Traffic Control. *Current Directions in Psychological Science*, *23*, 326–331. doi: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*, 1142–1147. doi: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*, 465–479. doi:10.1016/j.jml.2007.04.002

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

Loft, S., & Remington, R. (2010). Prospective memory and task interference in a continuous monitoring dynamic display task. *Journal of Experimental Psychology: Applied*, *16*, 145–157.

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*, 1432–1447. doi:10.1080/17470218.2012.750677

Loft, S., & Yeo, G. (2007). An investigation into the resource requirements of event-based prospective memory. *Memory & Cognition*, *35*, 263–274. doi:10.3758/BF03193447

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. doi:10.1037/a0033702

Lucenet, J., Blaye, A., Chevalier, N., & Kray, J. (2014). Cognitive control and language across the life span: Does labeling improve reactive control? *Developmental Psychology*, *50*, 1620–1627. doi:10.1037/a0035867

Ly, A., Boehm, U., Heathcote, A., Turner, B. M., Forstmann, B., Marsman, M., & Matzke, D. (2017). A flexible and efficient hierarchical bayesian approach to the exploration of individual differences in cognitive-model-based neuroscience. In A. A. Moustafa (Ed.), *Computational models of brain and behavior* (pp. 467–480). Wiley Blackwell.

Ly, A., Marsman, M., & Wagenmakers, E.-J. (in press). Analytic posteriors for pearson's correlation coefficient. *Statistica Neerlandica*, *72*, 4–13. doi:10.1111/stan.12111

Marklund, P., & Persson, J. (2012). Context-dependent switching between proactive and reactive working memory control mechanisms in the right inferior frontal gyrus. *Neuroimage*, *63*, 1552–1560. doi:10.1016/j.neuroimage.2012.08.016

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

Marsh, R., Hicks, J., & Cook, G. (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. doi:10.1037/0278-7393.31.1.68

Marsh, R., Hicks, J., Cook, G., Hansen, J., & Pallos, A. (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. doi:10.1037/0278-7393.29.5.861

Marsh, R., Hicks, J., & Watson, V. (2002). The dynamics of intention retrieval and coordination of action in event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 652–659. doi:10.1037//0278-7393.28.4.652

Matzke, D., Hughes, M., Badcock, J. C., Michie, P., & Heathcote, A. (2017). Failures of cognitive control or attention? The case of stop-signal deficits in schizophrenia. *Attention, Perception, & Psychophysics*, *79*, 1078–1086. doi:10.3758/s13414-017-1287-8

Matzke, D., Love, J., & Heathcote, A. (2017). A Bayesian approach for estimating the probability of trigger failures in the stop-signal paradigm. *Behavior Research Methods*, 267–281. doi:10.3758/s13428-015-0695-8

Matzke, D., & Wagenmakers, E.-J. (2009). Psychological interpretation of the ex-Gaussian and shifted Wald parameters: A diffusion model analysis. *Psychonomic Bulletin & Review*, *16*, 798–817.

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

McDaniel, M. A., LaMontagne, P., Beck, S. M., Scullin, M. K., & Braver, T. S. (2013). Dissociable neural routes to successful prospective memory. *Psychological Science*, *24*, 1791–1800. doi:10 . 1177 / 0956797613481233

Meeks, J., Hicks, J. L., & Marsh, R. (2007). Metacognitive awareness of event-based prospective memory. *Consciousness and Cognition*, *16*, 997–1004. doi:10.1016/j.concog.2006.09.005

Meeks, J., & Marsh, R. (2010). Implementation intentions about nonfocal event-based prospective memory tasks. *Psychological Research*, *74*, 82–89. doi:10 . 1007 / s00426-008-0223-x

Meng, X.-L. (1994). Posterior predictive p-values. *The Annals of Statistics*, 1142–1160. doi:10 . 1214 / aos / 1176325622

Millar, R. B. (2017). Conditional vs marginal estimation of the predictive loss of hierarchical models using waic and cross-validation. *Statistics and Computing*. doi:10.1007/s11222-017-9736-8

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202. doi:10.1146/annurev.neuro.24.1.167

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. doi:10.1006/cogp.1999.0734

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods in Psychology*, *4*, 61–64. doi:10.20982/tqmp.04.2.p061

Navon, D. (1984). Resources—A theoretical soup stone? *Psychological Review*, *91*, 216–234. doi:10 . 1037 / 0033-295X.91.2.216

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, *86*, 214–255. doi:10.1037/0033-295x.86.3.214

Norman, D. (1981). Categorization of action slips. *Psychological Review*, *88*, 1–15. doi:10.1037/0033-295X.88.1.1

Nowinski, J. L., & Dismukes, R. (2005). Effects of ongoing task context and target typicality on prospective memory performance: The importance of associative cueing. *Memory*, *13*, 649–657. doi:10 . 1080 / 09658210444000313

Park, D. C., Hertzog, C., Kidder, D. P., Morrell, R. W., & Mayhorn, C. B. (1997). Effect of age on event-based and time-based prospective memory. *Psychology and Aging*, *12*, 314. doi:10.1037/0882-7974.12.2.314

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 358–377. doi:10.1037/0096-1523.10.3.358

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*, 472–490. doi:10.1037/0033-295X.109.3.472

Poulton, E. (1982). Influential companions: Effects of one strategy on another in the within-subjects designs of cognitive psychology. *Psychological Bulletin*, *91*, 673–690. doi:10.1037/0033-2909.91.3.673

Provost, A., & Heathcote, A. (2015). Titrating decision processes in the mental rotation task. *Psychological Review*, *122*, 735–754. doi:10.1037/a0039706

Raab, D. H. (1962). Statistical facilitation of simple reaction time. *Transactions of the New York Academy of Sciences*, *24*, 574–590.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review*, *111*, 159–182. doi:10.1037/0033-295X.111.1.159

Reason, J. (1990). *Human error*. Cambridge, England: Cambridge University Press.

Ruthruff, E., Remington, R. W., & Johnston, J. C. (2001). Switching between simple cognitive tasks: The interaction of top-down and bottom-up factors. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 1404–1419. doi:10.1037/0096-1523.27.6.1404

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Incorporated.

Scullin, M. K., Einstein, G. O., & McDaniel, M. A. (2009). Evidence for spontaneous retrieval of suspended but not finished prospective memories. *Memory & Cognition*, *37*, 425–433. doi:10.3758/MC.37.4.425

Scullin, M. K., McDaniel, M. A., & Einstein, G. O. (2010). Control of cost in prospective memory: Evidence for spontaneous retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 190–203. doi:10.1037/a0017732

Scullin, M. K., McDaniel, M. A., Shelton, J., & Lee, J. (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. doi:10.1037/a0018971

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

Shorrock, S. T. (2005). Errors of memory in air traffic control. *Safety Science*, *43*, 571–588. doi:10.1016/j.ssci.2005.04.001

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. doi:10.1037/0278-7393.29.3.347

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

Smith, R. E. (2017). Prospective memory in context. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 66, pp. 211–249). New York: Academic Press.

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. doi:10.1037/0278-7393.30.4.756

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. doi:10.1037/0278-7393.33.4.734

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. doi:10.1037/xlm0000400

Stuphorn, V., & Emeric, E. (2012). Proactive and reactive control by the medial frontal cortex. *Frontiers in Neuroengineering*, *5*, 9. doi:10.3389/fneng.2012.00009

Taylor, R., Marsh, R., Hicks, J., & Hancock, T. (2004). The influence of partial-match cues on event-based prospective memory. *Memory*, *12*, 203–213. doi:10.1080/09658210244000559

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological Review*, *121*, 179–205. doi:10.1037/a0036137

Tulving, E. (1985). *Elements of episodic memory*. London: Oxford University Press.

Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological Methods*, *18*, 368–384. doi:10.1037/a0032222

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592. doi:10.1037/0033-295X.108.3.550

van Ravenzwaaij, D., Brown, S. D., Marley, A. A., & Heathcote, A. (n.d.). Accumulating advantages: A new approach to multialternative forced choice tasks.

van Wouwe, N. C., Band, G. P., & Ridderinkhof, K. R. (2011). Positive affect modulates flexibility and evaluative control. *Journal of Cognitive Neuroscience*, *23*, 524–539. doi:10.1162/jocn.2009.21380

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing*, *27*, 1413–1432. doi:10.1007/s11222-016-9696-4

Verbruggen, F., & Logan, G. D. (2009). Proactive adjustments of response strategies in the stop-signal paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 835–854. doi:10.1037/a0012726

Voss, A., Voss, J., & Klauer, K. C. (2010). Separating response-execution bias from decision bias: Arguments for an additional parameter in Ratcliff's diffusion model. *British Journal of Mathematical and Statistical Psychology*, *63*, 539–555. doi:10.1348/000711009X477581

Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research*, *14*, 867–897.

Welford, A. T. (1952). The ‘psychological refractory period’and the timing of high-speed performance—a review and a theory. *British Journal of Psychology*, *43*, 2–19. doi:10.1111/j.2044-8295.1952.tb00322.x

West, R., Choi, P., & Travers, S. (2010). The influence of negative affect on the neural correlates of cognitive control. *International Journal of Psychophysiology*, *76*, 107–117. doi:10.1016/j.ijpsycho.2010.03.002

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. doi:10.1037/0882-7974.14.2.264

West, R., Krompinger, J., & Bowry, R. (2005). Disruptions of preparatory attention contribute to failures of prospective memory. *Psychonomic Bulletin & Review*, *12*, 502–507. doi:10.3758/BF03193795

Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–257). Hillsdale, NJ: Erlbaum.

Woods, S. P., Weinborn, M., Velnoweth, A., Rooney, A., & Bucks, R. S. (2012). Memory for intentions is uniquely associated with instrumental activities of daily living in healthy older adults. *Journal of the International Neuropsychological Society*, *18*, 134–138. doi:10.1017/S1355617711001263