Investigating the effects of ongoing-task bias on prospective memory

Luke Strickland[1], Shayne Loft[2], Andrew Heathcote[3]

[1] The Future of Work Institute,
Curtin University, Australia

[2] The School of Psychological Science,
The University of Western Australia, Australia

[3] The School of Psychology,
The University of Tasmania, Australia

Address for Correspondence
Luke Strickland,
Future of Work Institute,
Curtin University,
78 Murray Street,
6000 Perth, Australia
Email: luke.strickland@curtin.edu.au

Abstract

Event-based prospective memory (PM) refers to the cognitive processes required to perform a planned action upon encountering a future event. Event-based PM studies engage participants in an ongoing task (e.g., lexical decision making) with an instruction to make an alternative PM response to certain items (e.g., items containing 'tor'). The Prospective Memory Decision Control model (PMDC), which provides a quantitative process account of ongoing task and PM decisions, proposes that PM and ongoing-task processes compete in a race to threshold. We use PMDC to test whether, as proposed by the Delay Theory of PM costs, PM can be improved by biasing decision making against a specific ongoing-task choice, so that the PM process is more likely to win the race. We manipulated bias in a lexical decision task with an accompanying PM intention. In one condition, a bias was induced against deciding items were words, and in another a bias was induced against deciding items were non-words. The bias manipulation had little effect on PM accuracy but did affect the types of ongoing-task responses made on missed PM trials. PMDC fit the observed data well and verified that the bias manipulation had the intended effect on ongoing-task processes. Further, although simulations from PMDC could produce an improvement in PM accuracy due to ongoing task bias, this required implausible parameter values. These results illustrate the importance of understanding event-based PM in terms of a comprehensive model of the processes that interact to determine all aspects of task performance.

Prospective Memory (PM) refers to the cognitive processes that allow humans to remember to perform planned actions in the future. PM tasks are prevalent in everyday life, as well as in safety-critical workplace settings such as aviation (Dismukes, 2012; Loft, Dismukes, & Grundgeiger, 2019) and healthcare (Rothschild et al., 2005). Einstein and McDaniel (1990) devised a paradigm to study PM in the laboratory, which is now the basis of a large body of research. The Einstein and McDaniel (1990) paradigm engages participants in an ongoing task such as a lexical decision task (indicate whether letter strings are words or non-words). The PM task is to remember to perform an atypical action at some point during the ongoing task. Many studies examine event-based PM, in which participants must remember to perform the planned action in response to a target event embedded in the ongoing task (e.g., make an alternative response if a presented letter string contains 'tor').

Typically, in PM paradigms participants must either substitute their PM response for an ongoing task response (e.g., press the PM key instead of the ongoing task key if you see a letter string containing 'tor'), and/or their ongoing task response removes the ongoing task stimulus from the screen, ending the opportunity for the PM item to cue the PM response. In either case, ongoing task response selection could interfere with the PM process. Consistent with ongoing and PM processes competing in a race for response selection, Loft and Remington (2013) found that PM performance could be improved by delaying when participants could make their ongoing-task response. Heathcote, Loft, and Remington (2015) elaborated this idea into "Delay Theory", which explained PM costs – slower ongoing-task performance with than without PM demands – as due to participants strategically slowing ongoing-task performance. Delay theory was supported by evidence-accumulation modeling, which found that PM costs were due to an increase in the threshold amount of evidence required to trigger an ongoing-task response.

Heathcote et al., (2015) also found that when PM target stimuli corresponded to only one of two ongoing-task choices (all PM items were words in an ongoing lexical decision task) participants biased responding against that choice by raising its threshold more than the threshold for the other choice (thresholds to make a 'word' decision were elevated more than thresholds to make a 'non-word' decision). They argued that this was consistent with participants using a selective bias strategy in order to improve PM responding. Their idea was that biasing ongoing-task thresholds against a specific ongoing-task choice could improve PM performance to target stimuli associated with that choice (Heathcote et al., 2015). For example, a bias against responding 'word' in a lexical decision task would be expected to improve PM accuracy to word targets. However, Heathcote et al. did not test this proposition empirically and could not show that it worked theoretically because their modeling was limited to the ongoing task (i.e., the PM decision process was not modelled).

Recently, Strickland, Loft, Remington, and Heathcote (2018) developed an evidence-accumulation theory of both ongoing-task and event-based PM decision processes, 'Prospective Memory Decision Control' (PMDC), that provides a comprehensive account for ongoing-task and PM response choices and response times (RTs). Following Loft and Remington (2013), PMDC proposes that parallel PM and ongoing-task decision processes, as modelled by linear ballistic accumulators (LBA; Brown & Heathcote, 2008), race to accumulate evidence towards their respective thresholds. Consistent with Delay Theory, simulations from Strickland et al.'s model indicated that globally increasing ongoing-task decision thresholds (e.g., increasing both word and non-word thresholds) improved PM performance, by reducing the probability that ongoing-task decisions would pre-empt PM processes.

It is difficult to experimentally test delay theory's claim that increased ongoing task

caution improves PM accuracy, because manipulations targeted at ongoing task caution may affect the perceived importance of the PM task, with potentially confounding effects on PM processes. Instead, this study tests the other key claim of delay theory: that biases in ongoing task thresholds can affect PM. If such biases affect PM accuracy, they could provide an efficient strategy to improve PM when PM events only conflict with a specific type of response. For example, suppose one usually turns left at a specific round-a-bout on the way home from work, but one day they must turn right to go shopping. Their success at this PM task may be improved by selectively increasing caution towards turning left at future roundabouts (but not necessarily increasing caution to go straight or in other directions), as this would increase the time available to retrieve their intent to turn right at that specific roundabout in order to go shopping.

In PMDC, selectively raising the threshold to an ongoing-task choice (e.g., raising the threshold to respond 'word') could allow the PM process time to reach threshold on relevant trials (following the example, PM would be less likely to be pre-empted by 'word' decisions on word PM trials). However, although Strickland et al., (2018) found that globally raised ongoing-task thresholds contributed to PM accuracy, PMDC modelling indicated that other processes, including proactive control of PM thresholds and reactive control of evidence accumulation, were more important in determining PM accuracy. Thus, it is not a given that ongoing-task biases can substantially affect PM accuracy. We apply PMDC to our study to test whether it can account for the effects of our bias manipulation. Applying PMDC also allows us to assess how our manipulation affects cognitive processes (e.g., test whether the bias manipulation affects ongoing task thresholds), and how shifts in cognitive processes map to shifts in performance (e.g., determine to what degree shifts in PM accuracy are caused by shifts in ongoing task thresholds). Before reporting the experiment, we introduce PMDC in more detail.

Prospective Memory Decision Control

PMDC assumes that parallel PM and ongoing LBA processes compete in a race to threshold (Figure 1). Each accumulator begins a trial at some start point drawn from a uniform distribution $U [0, A]$, and evidence for each accumulator increases at a speed given by the accumulation rate (drawn from a normal distribution with mean v, standard deviation sv). The first accumulator to reach its threshold, b, decides the overt response. Total RT is determined by total decision time plus some non-decision time included to capture processes that occur outside of the decision stage, such as stimulus encoding and motor responding. Strickland et al. (2018) found that, with these assumptions, PMDC was able to provide a comprehensive account of performance. This included accurate fits, at the level of individual participants, to response choices, mean RTs, as well as variance and skew in RTs on both PM and ongoing-task trials.

PMDC has provided insights into a range of cognitive processes that support PM, including capacity allocation between PM and ongoing tasks, as well as proactive control (Braver, 2012) and reactive control (Braver, 2012) over PM and ongoing-task processes. Below we review PMDC's mechanisms, and existing evidence for each. It is worth noting that PMDC differs from previous verbal theory integrating cognitive control and PM (e.g., Bugg, McDaniel, & Einstein, 2013) in that PMDC provides specific quantitative instantiations of how control processes affect decision making.
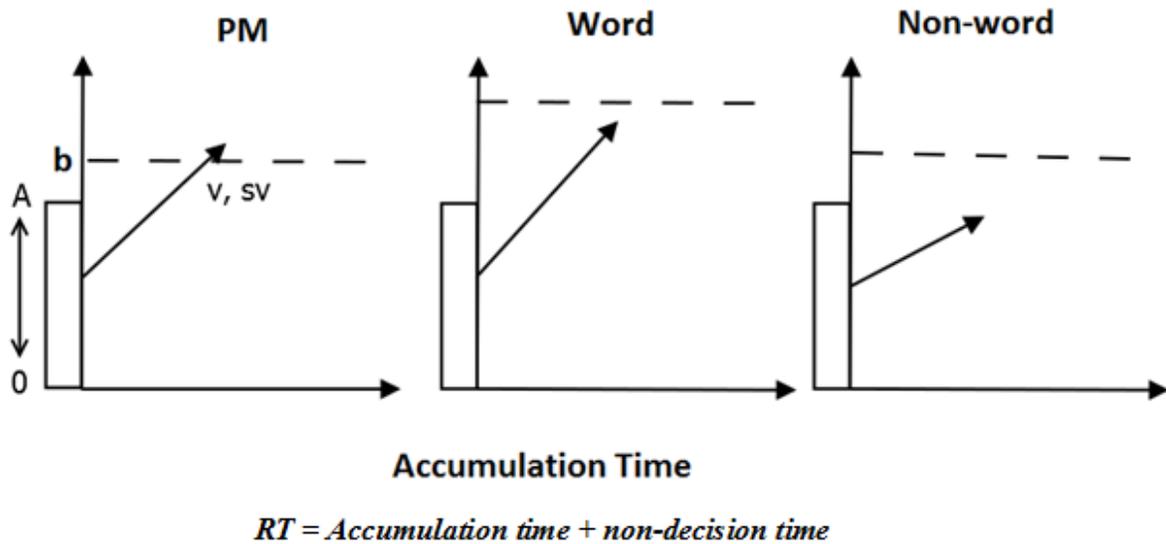
$$RT = Accumulation\ time + non\text{-}decision\ time$$

Figure 1. The PMDC model (Strickland et al., 2018). Evidence for each accumulator begins each trial at some start point drawn from the uniform distribution U[0, A]. Evidence then accumulates towards each accumulators' respective threshold, b, at an accumulation rate drawn from a normal distribution with mean v, standard deviation sv. The first accumulator to reach threshold determines the observed response. Total response time is determined by total time for the first accumulator to reach threshold plus a non-decision time parameter.

Capacity sharing. Many PM studies find that PM demands increase ongoing-task RTs, even on "non-PM trials" where no PM item is presented (Marsh, Hicks, Cook, Hansen, & Pallos, 2003; Smith, 2003). This effect is referred to as *PM cost*. PM theories assumed that cost results from *capacity sharing* between PM and ongoing-task processes (e.g., Einstein et al., 2005; Smith, 2003). The idea is that monitoring for PM items usurps resources from ongoing-task processes, increasing RTs. PMDC measures information-processing capacity with its accumulation rate parameters, making it possible to test the capacity-sharing hypothesis. Ongoing task accumulation rates can be associated with either the 'match' or 'mismatch' accumulator. The *match* accumulation rate measures evidence accumulation in the accumulator corresponding to the correct decision. The *mismatch* accumulation rate refers to evidence in the

accumulator corresponding to the incorrect decision. In Figure 1, for example, the evidence accumulation rate for the 'word' accumulator corresponds to 'match' accumulation on word trials and 'mismatch' accumulation on non-word trials. Similarly, the 'non-word' accumulation rate corresponds to 'match' accumulation on non-word trials and 'mismatch' accumulation on word trials. Evidence accumulation models generally indicate that PM demands do not cost ongoing-task capacity in standard paradigms (e.g., Heathcote et al. 2015; Horn & Bayen, 2015; Strickland, Heathcote, Remington, & Loft, 2017), but that capacity sharing can occur in more demanding paradigms, such as in simulations of air traffic control (Boag, Strickland, Heathcote, Neal, & Loft, 2019; Boag, Strickland, Loft, & Heathcote, 2019) and maritime surveillance (Strickland et al., 2019).

     Proactive control. Proactive control refers to cognitive control applied in advance of cognitively demanding events, to prepare for their occurrence. Proactive control is applied over PMDC's response thresholds, as thresholds are the locus of *a priori* strategy. Prior to the development of PMDC, the delay theory of PM cost (Heathcote et al., 2015) suggested that participants raise ongoing-task thresholds so that ongoing-task response selection does not pre-empt PM response selection, improving PM. For example, in Figure 1 a high threshold to respond word is depicted. Thus, on PM trials that require a 'word' ongoing-task response, the 'word' accumulator will take a relatively longer time to reach threshold than if the threshold was lower, and so there would be more time for the PM accumulator to accrue evidence, increasing the probability that it reaches its threshold first. Consistent with delay theory, elevated ongoing task thresholds have been found to underlie PM costs in many applications of evidence accumulation models to PM cost data (e.g., Heathcote et al., 2015; Horn & Bayen, 2015; Strickland et al., 2017; Strickland et al., 2018). Thus, PMDC includes proactive control over

ongoing task thresholds as a possible mechanism for improving PM accuracy.

Delay theory and PMDC are similar in that both propose a race to response selection between PM and ongoing task processes that could potentially be supported by increases in ongoing task thresholds. A key difference is that PMDC quantitatively instantiates the PM process as an LBA accumulator, whereas delay theory provided only a verbal description of the PM process. In fact, simulations reveal that, given PMDC's assumptions, ongoing task thresholds have only a weak effect on PM accuracy (Strickland et al., 2018). Furthermore, delay theory proposes only a single mechanism that supports PM, whereas PMDC proposes a range of mechanisms other than ongoing task threshold delays that could improve PM accuracy. For example, under PMDC proactive control can also apply over the PM threshold. Indeed, when the importance of PM is emphasized, it has been found that the PM threshold can be decreased to increase the probability of a PM decision (Strickland et al., 2018), and that this form of proactive control is critical to supporting PM accuracy.

Reactive control. Reactive control refers to cognitive control that occurs "just in time", that is when PM event is processed. PMDC's reactive control structure is depicted in Figure 2. Upon processing a PM item, encoding PM stimulus inputs may cause participants to accrue evidence towards the PM decision (reactive excitation), but to inhibit (i.e., slow down) accumulation to competing ongoing-task decisions (reactive inhibition). In Strickland et al., (2018), both forms of reactive control were critical to explaining variation in PM accuracy.
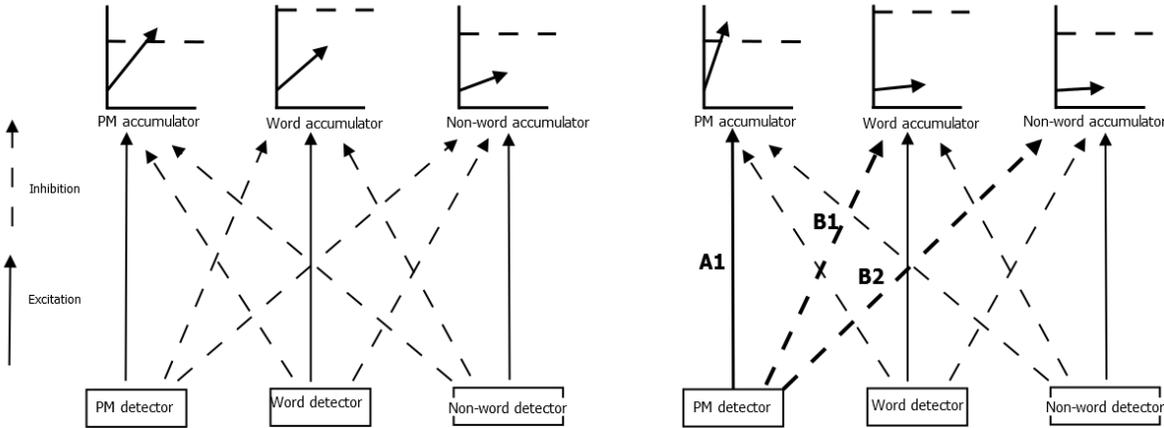
Figure 2. PMDC's reactive control (Strickland et al., 2018). The boxes represent stimulus input detectors. The solid lines represent excitatory connections to the accumulators, and the dashed lines inhibitory connections. Encoding stimulus inputs activates the detectors (e.g., PM-like stimulus inputs activate the PM detector). In turn this excites the relevant accumulator (e.g., the PM accumulator via connection A1), and inhibits competing accumulators (e.g. ongoing-task accumulation via connections B1 and B2).

Testing Delay Theory

Recently, Anderson, Rummel and McDaniel (2018) attempted to isolate, and manipulate, the effect of ongoing-task decision thresholds on PM accuracy. They compared 'standard' event-based PM conditions with a 'delay' condition that instructed participants to be cautious to make ongoing-task decisions and not to monitor for PM targets. They found that the latter did not improve PM performance, and argued that, therefore, proactive control over ongoing-task decisions does not support PM. However, it is not clear that their manipulation selectively affected ongoing-task decision thresholds. Their instruction not to monitor for PM targets may have caused at least some participants to increase their PM threshold, counteracting possible benefits of increased ongoing-task caution. Indeed, PMDC indicates that much of the improvement to PM accuracy under PM importance emphasis is driven by control of the PM threshold (Strickland et al., 2018). In Anderson et al.'s study, the PM threshold was not estimated because too few PM

trials were observed.

Even if Anderson et al.'s (2018) experiment were repeated with no explicit emphasis instruction for the PM task, manipulating ongoing-task caution could still affect PM processes. For example, emphasizing caution on the ongoing task could make the PM task appear less important, leading to a similar confound. Alternatively, if participants try to conserve their overall threshold levels then increasing ongoing task caution could lead to a decrease in the PM threshold. Thus, delay theory cannot be decisively tested by manipulating ongoing task caution. To experimentally test delay theory, it is crucial to manipulate ongoing task thresholds in isolation, without affecting PM processes. In the current study, we attempt such a test. Rather than manipulating overall caution, we manipulate another factor that delay theory claims is important for supporting PM accuracy: bias in ongoing task thresholds (Heathcote et al., 2015).

Bias in ongoing task thresholds refers to threshold levels that advantage one ongoing task decision over another. For example, a bias against word decisions could be implemented by shifting the word threshold up and non-word threshold down. On average, the ongoing-task 'match' accumulator will be faster than the 'mismatch' accumulator, to support better than chance accuracy on the ongoing-task. As a result, delay theory proposes that bias against an ongoing-task choice could potentially improve PM performance to items matching that choice – for example, a bias against 'word' decisions could improve performance to PM items that are words – because it would slow down participants making word responses and thus allow the PM response more time to reach threshold (Heathcote et al., 2015). This claim has been supported by analysis of "stimulus-specific" PM tasks (Lourenço, White, & Maylor, 2013), in which PM is associated with a specific event (e.g., PM items are always words). Under such conditions implementing ongoing-task biases could be an efficient way to improve PM performance without globally slowing down responding.

Importantly, this strategy can be implemented proactively and does not require the participant to make any changes in task strategy on a stimulus-by-stimulus basis, as the strategy differs between accumulators but is the same for all stimuli. Consistent with this, in stimulus-specific tasks participants do implement a bias against the decision that competes with PM (e.g., Heathcote et al., 2015; Strickland et al., 2018).

The findings that stimulus-specific PM tasks induce shifts in ongoing task bias are key to the case for delay theory. They favor delay theory over an alternative theory of ongoing task threshold increases – that PM instructions increase overall perceptions of task complexity (Horn & Bayen, 2015). Although it is plausible that an increase in perceived task complexity would induce a shift in caution, there is no reason to expect it would induce a bias. In contrast, delay theory clearly predicts the shifts in bias, and crucially it makes the claim that such shifts in bias are functional to PM (Heathcote et al., 2015). The claim that bias affects PM performance can be directly tested experimentally because, unlike overall caution, ongoing task bias can be manipulated without confounding from unintended effects on PM processes. Here we present an experiment to test the key claim of delay theory that bias in ongoing-task thresholds affects PM performance. We apply the PMDC model to our experiment, both to contrast it with delay theory and to validate our inferences about latent psychological processes.

The Current Study

Participants performed a lexical decision task with an accompanying PM task to detect items containing a target syllable (e.g., any letter string containing 'tor'). We include a within-subject blocked manipulation of ongoing-task bias. In some blocks we induce a bias against making word decisions, and in the others induce a bias against making non-word decisions. We manipulate bias by discouraging certain types of errors – for example, to induce a bias against

'word' responding, we strongly discourage making 'word' responses on non-word trials. With this manipulation, there is no obvious reason that the relative importance of the PM or ongoing tasks would differ across bias conditions. Thus, the design is suitable to test delay theory without confounding from differences in the perceived relative importance of the ongoing and PM tasks. Word and non-word PM targets are included in both blocks. Thus, we can assess the degree to which bias against word responding benefits PM accuracy to words, and the degree to which bias against non-word responding benefits PM accuracy to non-words. To examine the effects of ongoing-task bias on PM performance, we examine PM performance across our manipulation of PM stimulus type (PM word, PM non-word), and our blocked bias manipulation (bias against word, bias against non-word). In addition to standard analyses, we apply the PMDC model to determine whether it can fit to the effects of this new manipulation. PMDC is also critical to determining whether our bias manipulation is successful in affecting thresholds.

According to the delay theory, we would expect to observe increased PM accuracy to word PM targets when bias is induced against word decisions, and increased PM accuracy to non-word PM targets with bias against non-word decisions. This would occur because bias extends the completion time of the matching ongoing-task accumulator, allowing the PM accumulator more time to reach threshold (Heathcote et al., 2015). Although such a mechanism may be possible under PMDC, previous simulations from the model suggest only a minor role for ongoing task threshold elevation in supporting PM accuracy (Strickland et al., 2018). In addition, alternative mechanisms could reduce the potential benefits to PM accuracy of bias against the matching accumulator. For example, although the matching accumulator is faster on average than the mismatching accumulator in the PMDC model, the mismatching accumulator will not always be at a disadvantage, because rates vary from trial to trial. Thus, bias against the correct ongoing-

task decision may allow the incorrect ongoing-task decision to become competitive with the PM

process, at least on some trials, in which case bias might not improve PM performance but

instead increase the proportion of incorrect ongoing-task responses submitted on PM trials.

Method

Participants

The upper age limit for participation was 35, and English as a first language (the language

spoken in the childhood home) was required. Participants performed two 2 hour sessions, each

on a separate day. The data of three participants was excluded and replaced: two because they

made many very fast (<0.2s) responses in at least one block (one participant had a block with

72% fast responses, the other had a block with 35% fast responses), and one because they made

many slow RTs in a block (9% of responses were over 5 seconds). Remaining were 32

participants (23 females) aging from 17-34 (average = 19.84 years).

Materials

The lexical decision task was programmed in E-prime. 1236 words and 1236 non-words

were randomly selected from Strickland et al. (2018)'s second experiment. Word stimuli

occurred 1-7 times per million in the TMSH database (Dennis, 1995) (low frequency). Non-word

stimuli were created using the Wuggy algorithm (Keuleers & Brysbaert, 2010). Wuggy was set

to replace two out of three subsyllabic segments of the words, while matching segment lengths

and transition frequencies. The PM task was to detect a target substring (either *tor* or *ver*). 28

word PM targets and 28 non-word PM targets containing *tor* were taken from Strickland et al.

(2018), and the same for PM targets containing *ver*. An additional 14 PM words and 14 PM non-

words were obtained for each of the two substrings (using the same stimulus selection methods).

Thus, the total list of the study's PM targets included 42 word PM targets and 42 non-word PM

targets containing *tor* and another 42 word PM targets and 42 non-word PM targets containing

ver. Each stimulus was presented once each to all participants.

Participants performed 4 blocks of 660 trials over 2 days (day one, day two). This included

one block of each two bias conditions on each day. As explained in detail in the procedure

section, we manipulated bias by instructing participants either to be cautious to make word

responses (word caution condition, *Wc*), or cautious to make non-word responses (non-word

caution condition, *Nc*). The bias condition order used in day 1 was reversed for day 2, e.g. if the

Wc condition was block one of day one then it would be block 2 of day 2. For each day of the

experiment for each participant, one substring (*tor, ver*) was the PM target for the Wc block and

the other substring was the PM target for the Nc block. The assignment of PM target substring to

condition was reversed for each participant between day 1 and day 2. As condition block order

was also reversed for each participant between day 1 and day 2, substring block order was the

same for days 1 and 2, i.e. if *tor* was the target in the first block on day one (e.g., a Wc block)

then *tor* had to also be the target in the first block of day two (following the example, it would be

a Nc block). The four different ways in which the block order and PM target substrings could be

matched (while satisfying the above conditions) were counterbalanced across the 32 participants.

In each block participants were presented with 309 non-target non-words and 309 non-

target words, as well as 21 PM target non-words and 21 PM target words. For each participant,

the 21 PM target words and 21 non-words used for the Wc condition for a given substring were

drawn randomly, without replacement, from the total 42 words and 42 non-words which

contained that substring. The other 21-word PM targets and 21 non-word PM targets were used

for the Nc condition. The order in which participants were presented their non-target stimuli was

random within each block. In order to reduce fatigue effects participants were given five 1-

minute breaks within each 660 trial block. The breaks occurred after participants completed each 110 trial segment, so after trials 110, 220, 330, 440 and 550. Thus, blocks were divided into sixths. PM targets (in both Wc and Nc blocks) were presented 42 times per 660 trial block; randomly between trials 6-20, 21-35, 36-50, 51-65, 66-80, 81-95, 96-110 of each sixth of a block. Therefore, the ratio of PM trials to nontarget trials was 1:14. Target trials were separated by at least 4 lexical decision trials. The order in which the PM targets filled the chosen positions was random.

Procedure

Participants first performed practice lexical decision trials. They 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'). They were asked to make their responses as quickly and accurately as possible. For the experimental blocks (the PM blocks, both Wc and Nc), 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, e.g. "*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'.* 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 four response key orders were also counterbalanced with bias condition block order and PM target substring block order, so that each combination was used for two of the 32 participants. Participants were instructed before the commencement of each sixth of a block to rest their fingers on their assigned response key combination: one hand resting the index and middle finger

on the lexical decision keys (e.g. left hand index on *d*, left hand middle on *s*), and the other hand resting the index finger on the PM key (following the example, right hand index on *j*).

Each day, participants first completed their 24 practice lexical decision trials and received percentage feedback on their accuracy (e.g., "87.50% correct"). They then proceeded to the experimental blocks and were presented with either Wc or Nc instructions. For the Wc blocks, participants were instructed to be careful about making word responses, e.g. "*In the next block of trials try to respond quickly and accurately to all items, but note that it is extra important to avoid errors where you incorrectly respond WORD ('s') to non-word items. That is, only press the 's' key when you are absolutely sure an item is a word. If you do incorrectly classify an item as a word, you will be presented a special 'incorrect' message which delays the task more than if you incorrectly classify an item as a non-word.*" For the Nc blocks, the opposite instruction was presented (in the example substitute non-word for word and word for non-word). Each time participants received a bias instruction, they repeated the instruction to the experimenter. After receiving the bias instructions, participants were given their PM instruction to make an alternative response to target substrings (see example from the paragraph above) and repeated the instruction to the experimenter. Participants next completed a three minute distractor puzzle, after which they began the first block of experimental trials. After completion of each sixth of a block, participants were presented feedback on the accuracy of their responses (%) to the lexical decision task. After completing their first bias condition for the day (e.g., the Wc condition), participants were instructed that their bias and PM instructions no longer applied, e.g. "*Please note that the instruction you received to prioritize avoiding incorrectly making word responses to non-words no longer applies. These errors will not trigger a longer 'incorrect' message anymore. You also do not need to make a special response to items containing 'ver' in the next*

block of trials. In fact, no item containing 'ver' will be presented." When participants received

their second block's PM instructions (after they had already received their new caution

instructions), they were again reminded that their old PM target was no longer relevant, e.g.

"*Please be reminded that you no longer need to press 'j' if you see an item containing 'ver'. In*

fact, no item containing 'ver' will be presented in the next block of trials." In addition to the

breaks within blocks, participants were instructed to rest for 2 minutes between blocks.

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 (size 18 courier new font) which remained on the black

screen until the participant made a response. If the participant made a correct word/non-word

response (including on PM trials), or a correct PM response, the subsequent trial immediately

began (next fixation cross). If the participant made an incorrect response, the feedback they

received varied depending on bias condition, with longer delays for the discouraged ongoing-

task errors (e.g., for word responses to non-word trials in the Wc condition). This was included

to increase the strength of the bias manipulation[1]. In the Wc condition, word responses to non-

word trials triggered a screen which displayed 'INCORRECT!!! ' in size 44 courier new font for

15 seconds. Non-word responses to word trials in the Wc condition triggered a screen which

displayed 'INCORRECT' in size 18 courier new font for 1 second. In Nc blocks the reverse was

true, non-word responses to word trials would trigger the 15 second, size 44 'INCORRECT!!!'

whereas word responses to non-word trials would trigger the 1 second, size 18 'INCORRECT'

---

[1] We ran an initial pilot, $N = 9$, with 5 second timeouts, and found that this failed to affect ongoing-task RTs and

accuracies. Thus, we extended the timeout to 15 seconds.

message. In both conditions, any other incorrect responses (random key presses or PM false

alarms) triggered the 1 second, size 18 'INCORRECT' screen. Correct lexical decision responses

to PM trials (PM misses) did not trigger a feedback screen. The subsequent trial would begin

immediately (next fixation cross) after either feedback screen was displayed.

Results

An alpha level of 0.05 was used in all analyses. The first two trials after each rest period

(1.8% of trials) were excluded from analyses, as were trials where participants responded with a

key not corresponding to their PM or lexical decision task (0.02% of responses). The two trials

after each PM target and PM false alarm (12.9% of trials) were excluded, which is common

practice in PM studies, in order to avoid contamination from post-PM slowing (e.g., Meier &

Rey-Mermet, 2012). If participants submitted a discouraged lexical decision error (e.g. a 'word'

response to a non-word in the Wc condition), they were presented a 15s feedback screen, during

which they might become distracted. Thus, we excluded any trials which immediately followed

this long timeout (2.7% of trials). Following these exclusions, we cut out any remaining trials

with outlying RTs (< 0.2s or > mean RT plus 3 times the interquartile range / 1.349, 4.82% of

remaining trials). From the original 2640 trials, this left on average 2086 trials remaining (range

2005-2152) for data analysis and PMDC modelling, corresponding to an average of 39 out of the

42 PM trials remaining for each PM stimulus type for each bias condition.

We conducted mixed effects model analyses using the R programming language (R Core

Team, 2019) and the 'lme4' package (Bates, Mächler, Bolker, & Walker, 2015). These models

included a random participant intercept term, but not random participant slopes. No other

random effects were included. To analyse accuracy, we fit generalized linear models to each

observed response with a probit link function. To analyse RT, we fit linear mixed effects models

to participant mean RTs. In addition to stimulus type (non-word, word) and bias condition (Nc, Wc), the reported models included a day order factor (day one, day two) to capture effects of task repetition. We fit models including all factors and interactions, testing for significance with Wald chi-squared tests. The null model for the significance tests of each term included all other terms, except for higher order interactions including the term (e.g., the interaction between stimulus type and bias condition would be ignored when testing the main effect of stimulus type). The outcomes of these tests are tabulated in the supplementary materials. In text, we report descriptive statistics broken down by the factors that we found were significant, as well as follow-up pair bed samples t-tests. The t-tests were calculated using participant mean RTs and accuracies, averaged for each participant over any factors not relevant to the test. Effect sizes are reported in terms of Cohen's d. We report within-subject standard errors calculated with the Morey (2008) bias-corrected method.

Lexical Decision Task

We first assess whether our bias manipulation had the intended effect on ongoing task accuracy. A bias against responding non-word (Nc condition) would be expected to improve accuracy to word stimuli, and a bias against responding word (Wc condition) would be expected to improve accuracy towards non-word stimuli. As presented in the first two rows of Table 1, we found both these effects. Accuracy to non-word stimuli was higher in the Wc condition than the Nc condition, $t(31) = 4.84$, p $<.001$, $d = 0.86$, and accuracy to word stimuli was higher in the Nc condition than the Wc condition, $t(31) = 6.06$, $p <.001$, $d = 1.07$. This indicates that we successfully manipulated ongoing-task bias. We also found that accuracy was marginally higher on day one ($M = 91.8\%$, $SE = 1.3\%$) than day two ($M = 91.1\%$, $SE = 1.5\%$), $t(31) = 1.80$, $p =.08$, $d = 0.32$.

Table 1. *Accuracy and correct RT for the ongoing and prospective memory tasks. The brackets contain within-subject standard errors, calculated with the Morey (2008) bias-corrected method.*

	Wc		Nc	
Stimulus Type	Accuracy (%)	Correct RT (seconds)	Accuracy	Correct RT (seconds)
Word	88.2 (0.6)	0.846 (0.017)	93.3 (0.8)	0.808 (0.015)
Non-word	96.1 (0.8)	0.867 (0.019)	88.3 (1.3)	0.948 (0.017)
PM Word	61.4 (3.8)	0.966 (0.024)	61.4 (3.5)	0.945 (0.021)
PM Non-word	56.7 (4.2)	0.922 (0.022)	59.8 (3.4)	0.959 (0.025)

Bias would also be expected to affect RTs: a bias against responding non-word (Nc) should result in slower non-word RTs, and a bias against responding word (Wc) should result in slower word RTs. We found that correct RTs were slower to non-word stimuli in the Nc condition than the Wc condition, $t(31) = 6.46$, $p < .001$, $d = 1.14$, and slower to word stimuli in the Wc condition than the Nc condition, $t(31) = 4.37$, $p < .001$, $d = 0.77$, indicating that our bias manipulation was successful. Correct RTs were also slower on day one ($M = 0.90$s, $SE = 0.02$s) than on day two ($M = 0.83$s, $SE = 0.02$s), $t(31) = 5.13$, $p < .001$, $d = 0.91$. Due to our high trial numbers, we also analyzed ongoing task error RTs. Error RTs were slower in the Nc ($M = 0.975$s, $SE = 0.028$s) condition than the Wc condition ($M = 0.934$s, $SE = 0.03$s), $t(31) = 2.26$, $p = .03$, $d = 0.40$, and slower on day one ($M = 0.998$s, $SE = 0.03$s) than on day two ($M = 0.910$s, $SE = 0.024$s), $t(31) = 6.29$, $p < .001$, $d = 1.11$. However, there was no interaction between stimulus type and bias condition, and thus our bias manipulation appeared not to have a strong effect on error RTs. This might owe to relatively poor measurement of error RTs – ongoing task accuracy was quite high and so few error RTs were observed.

Prospective Memory Task (Hits)

PM response false alarms were rare, ranging from 0 to 1.3% of trials, and thus are not analyzed further. PM responses were scored as correct (as PM 'hits') if the participant pressed the PM key instead of a lexical decision key on the target trial. Delay theory predicts that a bias against the 'correct' ongoing task response to a PM stimulus should improve PM accuracy. Thus, PM accuracy would be expected to be higher for non-words in the Nc condition, and higher for words in the Wc condition. However, as displayed in Table 1, we did not find that bias substantially affected PM accuracy. No effects or interactions regarding the bias manipulation reached significance in our model of PM accuracy. However, there was an effect of "day". PM accuracy was lower on day one ($M = 53.2\%$, $SE = 4.7\%$) than on day two ($M = 66.5\%$, $SE = 3.7\%$), $t(31) = 3.99$, $p < .001$, $d = 0.71$. Our mixed effects model also revealed a small effect of PM stimulus type. Accuracy to PM words was marginally larger than accuracy to PM non-words, $t(31) = 1.60$, $p = 0.12$, $d = 0.28$.

If ongoing task bias could sufficiently delay the ongoing task to support PM, then slower PM processes would be able to complete before the ongoing task does, increasing PM RT due to decreased 'statistical facilitation' (Raab, 1962) from ongoing task processes. However, as PM accuracy was not improved by the bias manipulation, this was not expected. Indeed, we did not find any effects of stimulus type or bias condition on PM RT. However, we did find that PM responses were slower on day one ($M = 1.010$s, $SE = 0.028$s) than on day two ($M = 0.893$s, $SE = 0.023$s), $t(31) = 4.33$, $p < .001$, $d = 0.77$.

Prospective Memory Task (Misses)

To further investigate why PM performance was not affected by ongoing-task bias, we examined whether bias affected the type of ongoing-task responses submitted on PM trials. We

created a PM "miss type" factor that denoted whether PM misses were word or non-word responses. We fit a linear mixed effects model to mean response proportions on PM error trials that included miss type and all other potentially relevant experimental factors (stimulus type, bias condition, day). Note that here we examined response proportion rather than predicting every individual response with a generalized linear model because responses are confounded by the PM miss type factor (e.g. when the miss type factor is word then the DV will always be equal to word). We also examined mean PM error RTs with a linear mixed effects model including all potentially relevant factors (miss type, stimulus type, bias condition, day). The supplementary materials contain summaries of our mixed model analyses of PM error type and error RTs, and the major results are discussed in text along with descriptive statistics and follow-up tests.

The previous analyses suggest that our bias manipulation successfully affected ongoing task bias yet failed to affect PM accuracy. If this is the case, we would expect the type of PM error made to change, with fewer 'word' responses submitted on PM trials in the Wc condition, and fewer 'non-word' responses submitted on PM trials in the Nc condition. As displayed in Table 2, we did find that there were fewer word responses on PM trials than non-word responses in the Wc condition, $t (31) = 2.92, p < .01, d = 0.52$. In addition, there was a trend in the reverse direction in the Nc condition; word responses were more common on PM trials than non-word PM responses, $t (31) = 1.82, p = .08, d = 0.32$.

A successful manipulation of ongoing task bias would also be expected to influence RTs of PM errors, with slower word RTs predicted in the Wc condition and slower non-word RTs in the Nc condition. As displayed in Table 2, we found some evidence of such an effect. The effects of the bias manipulation and PM stimulus type interacted. Non-word responses were slower than word responses on PM trials for both Wc and Nc conditions. Non-word PM error responses were

slower in the Nc condition than the Wc condition, t (31) = 2.53, p =.02, d = 0.45. Word

responses to PM trials were numerically slower in the Wc condition than the Nc condition, but

this difference was not near statistical significance t (31) = 1.19, p = 0.24, d = 0.21.

Table 2. *PM Miss Type and Miss RT for Experiment Two. The brackets contain standard errors, calculated with the Morey (2008) bias-corrected method.*

Bias Condition	PM Miss Type	Response Proportion	RT
Nc	Word	21.5 (4.70)	0.80 (0.04)
	Non-word	17.9 (4.56)	0.974 (0.03)
Wc	Word	17.8 (4.71)	0.82 (0.04)
	Non-word	23.1 (5.32)	0.881 (0.04)

We now summarize our results thus far. Our analyses of ongoing task performance

indicate that the bias manipulation was effective in making the discouraged ongoing-task

response both slower and less common. Under delay theory, the effects of PM stimulus type and

the bias manipulation on PM accuracy were expected to interact, such that the Wc condition

would display higher PM accuracy to word PM trials and the Nc condition to non-word PM

trials. However, no evidence was found for such an effect. Rather than PM accuracy to a

stimulus (e.g., PM word stimulus) benefitting from bias against the matching ongoing-task

response (e.g., bias against 'word'), the type of ongoing-task error observed on PM trials

changed (e.g., bias against responding word induced more non-word responses on word PM

trials). Taken together, these results amount to a failed prediction from the delay theory.

In the next section, we present a PMDC analysis of behavior in our experiment. This

analysis has several critical goals. One is to determine whether the PMDC model can provide an

appropriate fit to the effects of our bias manipulation, despite the prediction of delay theory

failing. Another is to examine the latent cognitive processes that were affected by our bias

manipulation. This includes verifying that the bias manipulation did successfully affect

thresholds and verifying that it did not substantially affect PM processes. A final goal was to

determine the scope of PMDC's predictions, and whether the model would have been readily

compatible with a finding of a strong bias effect, if we had obtained one.

Model Analysis

The basic architecture of our model is depicted in Figure 1. We estimated thresholds in terms

of B, which is $b - A$. Model parameters could vary over stimulus type (word, non-word, PM

word, PM non-word), bias condition (Nc, Wc), day (one, two), and accumulator (word, non-

word, PM). To simplify the model, the start-point noise (A) and non-decision time ($t0$)

parameters were fixed across all factors[2]. We estimated one sv for the accumulator matching the

correct response on all trials (e.g., word accumulator on word trial, PM accumulator on PM trial).

The sv for the mismatching accumulators (e.g., non-word accumulator on word trial or PM trial)

was fixed at 1 as a scaling parameter. Thresholds could vary over day, condition and latent

accumulator, but were fixed across stimulus type, as is conventional. Although we have not

allowed mean evidence accumulation rates (v) to vary by day in our previous modelling (e.g.,

Strickland et al., 2018), we did in the current case to account for the observed practice effect on

PM performance. Due to a low number of PM false alarms, we only estimated one PM false

---

[2] A reviewer suggested estimating a separate non-decision time for the second day of the experiment to account for

practice effects. We tried this model and did not find evidence for such an effect, and so we report the simpler model

in text. Summaries of the parameter estimates of the model with varying non-decision time over day can be found in

the supplementary materials.

alarm accumulation rate across all design cells. With these constraints, the most flexible model

we fit had 56 free parameters: one A, one $t0$, one sv, twelve Bs, and 41 vs.

Sampling

We applied Bayesian methods to estimate the posterior probability distribution of our model

parameters. Because we obtained over 2000 trials per participant, we were able to separately

estimate each individual participants' parameters. We could have also fit a hierarchical model,

including a population level distribution, but with our large data sets this posed computational

difficulties (i.e., model-fitting times of many weeks). Bayesian estimation requires specifying

prior beliefs about parameter values, in the form of prior distributions. Our priors are displayed

in Table 3. These relatively uninformative priors are the same as those in Strickland et al. (2018).

No parameters in the prior vary between the Wc and Nc conditions. Differences between

matching and mismatching accumulation rates are included in the prior, to accord with our

expectation that accuracy would be far higher than chance. The prior for the PM false alarm

mean accumulation rate is set low, to encode our expectation that PM false alarms would be very

rare in our study, as they generally are. One notable prior setting is the non-decision time lower

bound of 0.1 seconds and upper bound of 1 second. The lower bound was chosen to avoid non-

decision time estimates that are an implausibly low amount of time to include both stimulus

encoding and response execution, and the upper bound for implausibly high values. Generally,

our choice of priors had little influence on the resulting posterior parameter estimates.

Table 3. *Prior distributions*

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

Our posterior sampling was performed using the Dynamic Models of Choice R suite (Heathcote et al., 2019). We applied the differential evolution Markov Chain Monte Carlo algorithm (Turner, Sederberg, Brown, & Steyvers, 2013), an effective technique for sampling evidence accumulation model parameters. For each sampled model, we ran three times as many chains as there were parameters (e.g., for the most flexible model with 56 parameters we ran 168 chains). To reduce memory requirements, posterior samples were 'thinned' such that we only retained every 20th sample. We obtained 180 total posterior samples (corresponding to 3600 iterations). We ran posterior sampling until the samples were adequately stable, and the chains converged and mixed. This was confirmed with visual inspection and Gelman's multivariate potential scale reduction factor (<1.1, Gelman et al., 2013)

Model Assessment

To determine whether constraining our model further was justified, we compared the most flexible model (the 'top' model) to simpler models using the deviance information criterion

(DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). DIC measures model fit while

punishing model complexity, with lower DIC values indicating a better model. DIC values were

summed across participants. Various guidelines exist to determine how large of a DIC difference

is substantial. One interpretation of DIC difference is in terms of model weights, corresponding

to the probability that the selected model is the best model, analogous to Akaike Information

Criterion weights (e.g., see Wagenmakers & Farrell, 2004). For a set of two models, a DIC

difference of greater than 10 corresponds to over a 99% probability that the selected model is the

best. All DIC differences discussed below all far exceed this, suggesting strong support for the

selected model in each comparison. The top model, with 56 parameters, had a DIC value of

13125. We compared the top model with one that fixed thresholds over bias conditions (50

parameters). We found that the fixed threshold model had a substantially larger DIC than the top

model (13634), suggesting that shifts in thresholds were necessary to account for the

manipulation of bias. A model that fixed accumulation rates over bias conditions (36 parameters)

also had a much larger DIC value (13666) than the top model, suggesting that shifts in

accumulation rates were also necessary to account for the effects of the bias manipulation. We

also attempted fixing accumulation rates over the 'day' factor, to test whether our choice to allow

accumulation rates to vary over day was justified. The model with fixed accumulation rates had a

substantially larger DIC (13704) than the top model, suggesting that varying rates over day was

necessary.

We now examine the fit of the selected 'top' model. To obtain posterior predictive model

fit, we simulated data for each participant from each of the observed posterior samples. As

displayed in Figure 3, PMDC fit the observed non-PM trial performance well, including the

effects of our bias manipulation. Furthermore, as demonstrated in Figures 4 and 5, the model was

able to accurately fit the effects of our bias manipulation on PM trial performance, including the

shift in PM error type. Figure 4 depicts some minor miss-fit to the rate of ongoing task errors on

PM trials, with the model slightly over-estimating the frequency of such errors. Such miss-fit

may indicate minor discrepancies between the relationship between RT and choice specified by

our model, and that observed in the data. However, the miss-fit is relatively small, and the model

captures the effect of the bias manipulation on these error rates well. Thus, overall, despite the

observed effects of ongoing task bias being unanticipated by delay theory, they fit well with

PMDC. As the model provided a good account of the observed trends in our data, we proceed to

explore the model mechanisms responsible for the fit.

Model Mechanisms

In this section, we review how our selected model accounted for observed performance,

with a focus on the manipulation of ongoing-task bias. To summarize model parameters across

the group, we averaged each posterior sample across participants. The resulting participants-

averaged distribution was used for data summaries and posterior inference. Throughout this

section, we report the posterior means (*M*) and standard deviations (*SD*) of this group-averaged

distribution. The posterior mean of the participants-averaged non-decision times (*t0*) was 0.14s

(*SD* = .003s). The posterior mean of the start-point noise parameter (*A*) was 0.44 (*SD* = 0.02),

and the mean of the standard deviation of match accumulation rates (*sv*) was 0.59 (*SD* = .006). In

the sections below, we discuss the estimates of threshold and mean accumulation rate

parameters. To test for differences across conditions, we constructed difference distributions by

calculating the difference between the parameters for every posterior sample. To summarize

these distributions, we report a *Z* score (mean/SD of the difference distribution) and one tailed

posterior *p* value, with lower *p* values indicating a more substantial probability of a difference

between parameters. For the latter we report min (p, 1-p), corresponding to the lowest probability

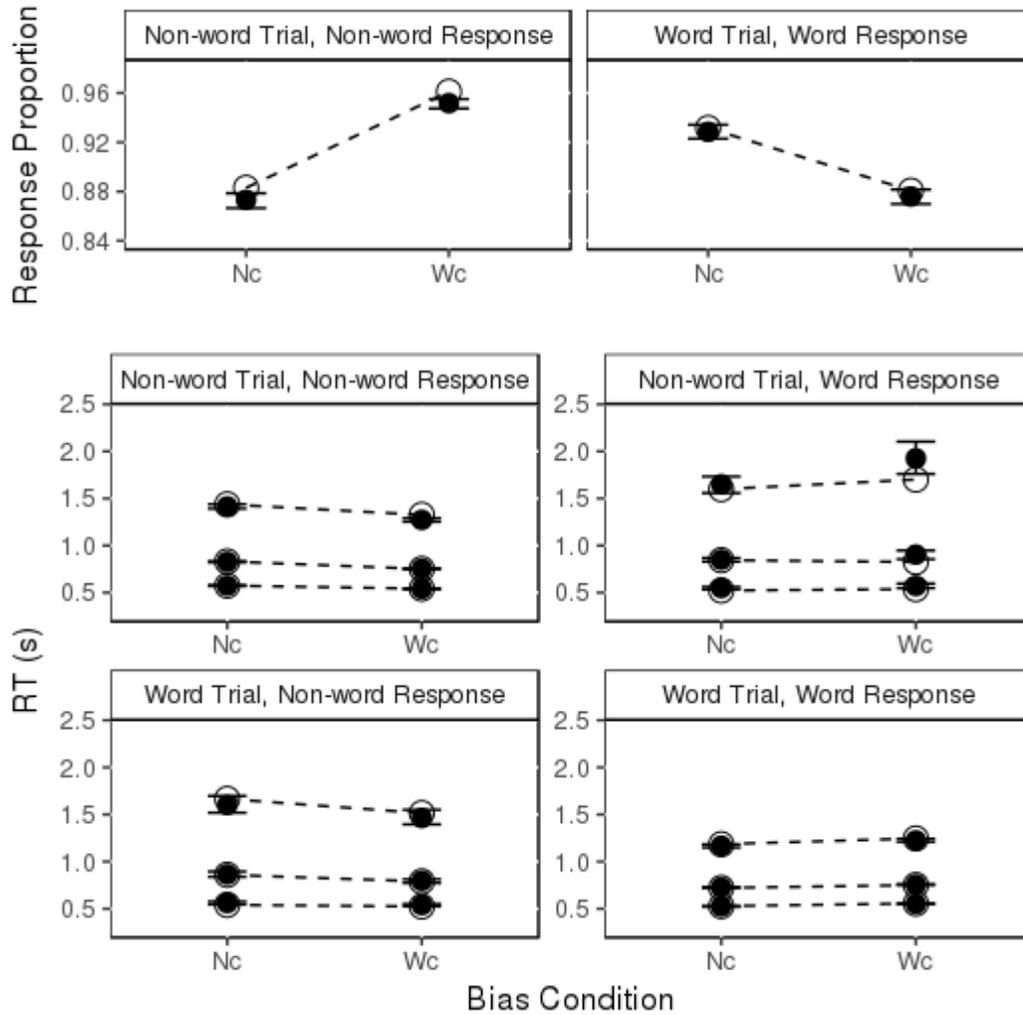of a difference in either direction.

Figure 3. Model fits to the non-PM trial response data. The white dots indicate the observed data. The black

dots indicate the mean prediction of the model. The black bars indicate the 95% credible intervals of the

posterior predictions. For the purposes of these graphs, we concatenated all participants' data into a single

data frame, and posterior predictions for all participants into a single data frame, and then calculated

summary statistics on these data frames. To demonstrate fit to RTs, the RT graphs illustrate fits to three

quantiles: the fastest ten percent of RTs, the median RTs and the ten percent of slowest RTs. Nc refers to

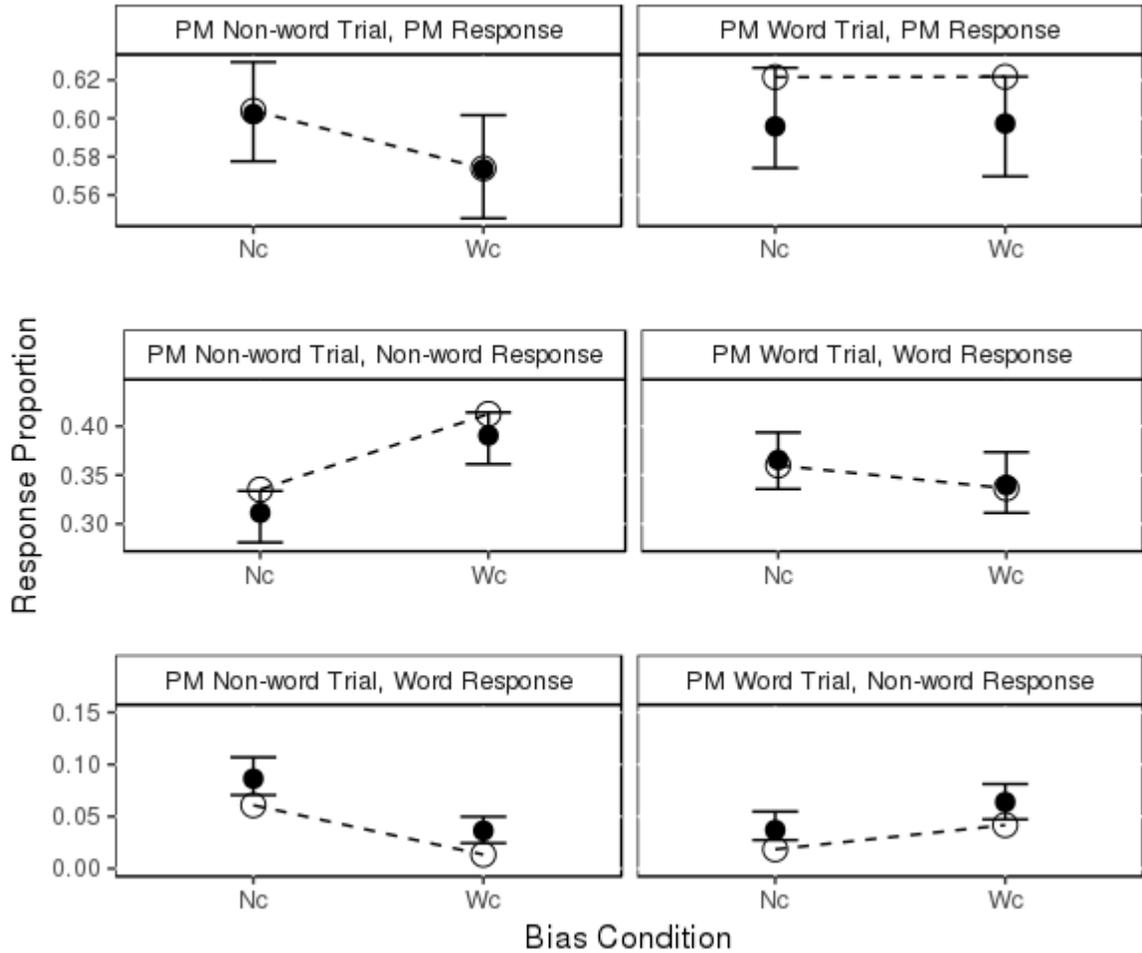the condition with a bias against non-word decisions, and Wc to bias against word decisions.

Figure 4. Model fits to PM trial response proportions. The white dots indicate the observed data. The black dots indicate the posterior mean prediction of the model. The black bars indicate the 95% credible intervals of the posterior predictions. For the purposes of these graphs, we concatenated all participants' data into a single data frame, and posterior predictions for all participants into a single data frame, and then calculated response proportions with these data frames. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against word decisions.
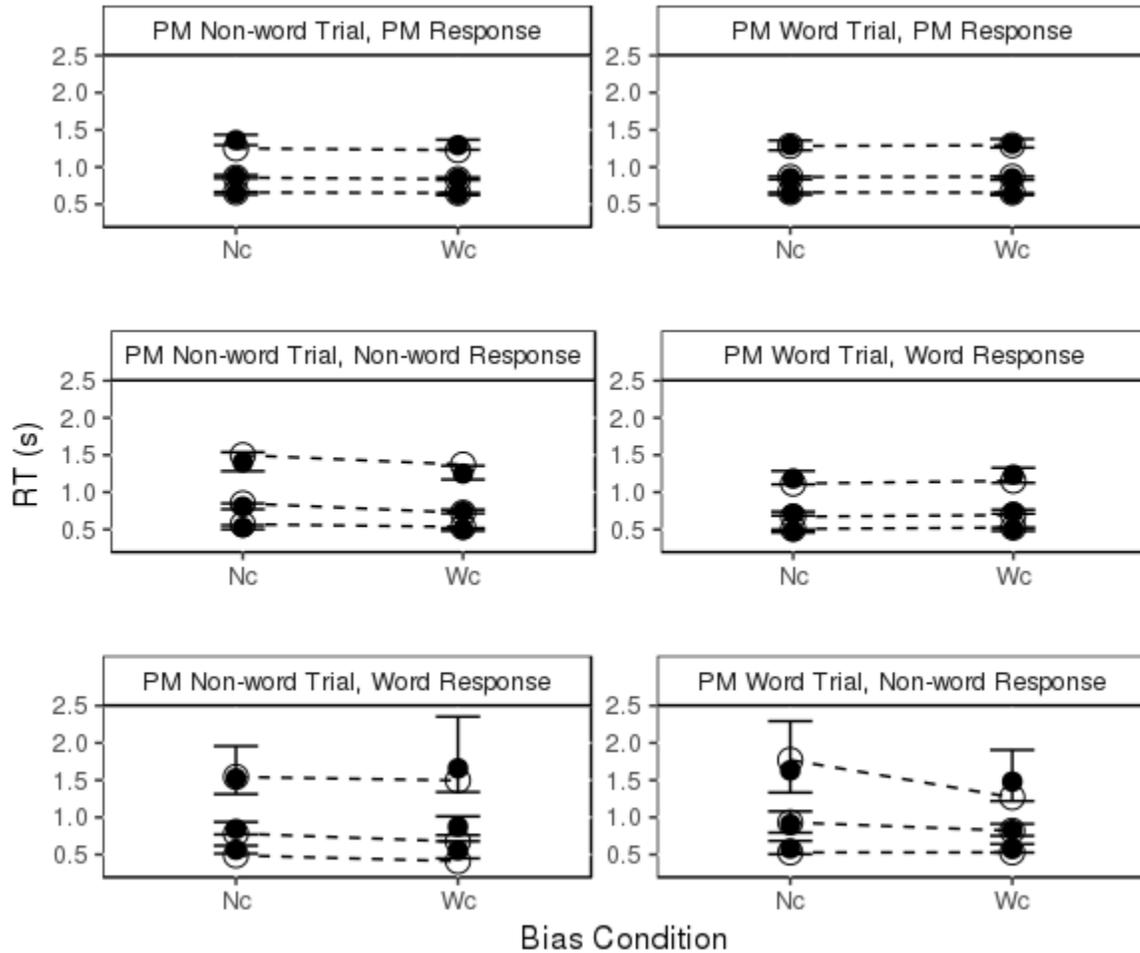
Figure 5. Model fits to PM trial response times. The white dots indicate the observed data. The black dots indicate the mean prediction of the model. The black bars indicate the 95% credible intervals of the posterior predictions. For the purposes of these graphs, we concatenated all participants' data into a single data frame, and posterior predictions for all participants into a single data frame, and then calculated summary statistics on these data frames. To demonstrate fit to RTs, the graphs illustrate fits to three quantiles: the fastest ten percent of RTs, the median RTs and the ten percent of slowest RTs. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against word decisions.

Thresholds. Thresholds are plotted in Figure 6. Overall, there was a large bias against non-word responding, with thresholds much higher to respond non-word than to respond word, $Z = 17.88$, $p < .001$. Determining that thresholds were affected by our bias manipulation is a critical manipulation check. Consistent with our manipulation affecting threshold bias, non-word

thresholds were higher in the Nc condition (day one $M = 1.34$, $SD = 0.02$; day two $M = 1.22$, $SD = 0.02$) than the Wc condition (day one $M = 1.27$, $SD = 0.02$; day two $M = 1.13$, $SD = 0.02$), $Z = 8.05$, $p <.001$. Similarly, word thresholds were higher in the Wc condition (day one $M = 1.19$, $SD = 0.02$; day two $M = 1.13$, $SD = 0.02$) than the Nc condition (day one $M = 1.10$, $SD = 0.02$; day two $M = 1.04$, $SD = 0.02$), $Z = 10.3$, $p <.001$. Thus, the model indicates that our manipulation was successful in inducing a threshold bias in the expected directions (higher word thresholds in Wc blocks, higher non-word thresholds in Nc blocks). Consistent with our manipulation selectively affecting ongoing-task processes, PM thresholds in the Nc condition were not substantially different from the Wc condition, $Z = 0.67$, $p = .252$. On day one, PM thresholds were numerically, but not substantially, higher in the Nc condition ($M = 1.76$, $SD = 0.05$) than the Wc condition ($M = 1.72$, $SD = 0.05$), $Z = 0.58$, $p = 0.28$, whereas on day two PM thresholds were higher in the Wc condition ($M = 1.64$, $SD = 0.04$) than the Nc condition ($M = 1.55$, $SD = 0.04$), $Z = 1.7$, $p = 0.045$.
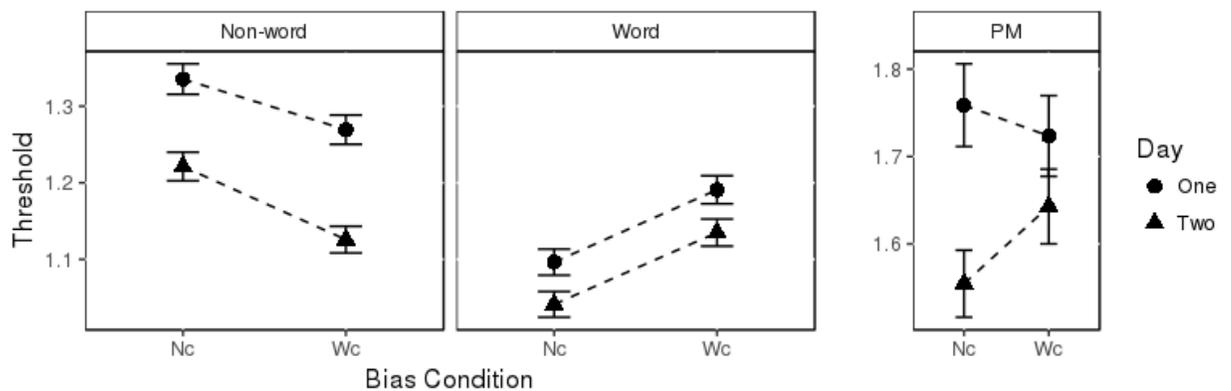

Figure 6. Participant-averaged thresholds. The symbols correspond to the posterior mean, and the error bars correspond to plus and minus one standard deviation from the mean. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against word decisions.

Accumulation rates (non-PM trials). Although thresholds are the parameter traditionally associated with bias, it is possible that bias also affects the speed of evidence accumulation. Mean accumulation rates for non-PM trials are plotted in Figure 7. There was some evidence of the bias manipulation inducing a bias in accumulation rates. Non-word accumulation to non-word items was substantially slower in the Nc condition (day one $M = 1.97$, $SD = 0.02$; day two $M = 2.00$, $SD = 0.02$) than the Wc condition (day one $M = 2.11$, $SD = 0.02$; day two $M = 2.14$, $SD = 0.02$), $Z = 8.95$, p <.001. Similarly, non-word accumulation was slower to word items in the Nc condition (day one $M = 0.01$, $SD = 0.05$; day two $M = 0.15$, $SD = 0.04$) than the Wc condition (day one $M = 0.21$, $SD = 0.04$; day two $M = 0.33$, $SD = 0.04$), $Z = 5.58$, p <.001. Word accumulation to word items was not substantially slower in the Wc (day one $M = 2.02$, $SD = 0.02$; day two $M = 2.10$, $SD = 0.02$) condition than the Nc condition (day one $M = 2.02$, $SD = 0.02$; day two $M = 2.09$, $SD = 0.02$), $Z = -0.27$, $p = .40$. However, word accumulation to non-word items was substantially slower in the Wc condition (day one $M = -0.86$, $SD = 0.06$; day two $M = -0.67$, $SD = 0.06$) than the Nc condition (day one $M = -0.41$, $SD = 0.04$; day two $M = -0.33$, $SD = 0.05$), $Z = 8.61$, p <.001. In summary, we found three out of four accumulation rate effects in line with a bias in accumulation rates consistent with our manipulation, and one indicating little difference. We did not anticipate these accumulation rate effects, but particiularly for non-word, they are consistent with control over stimulus bias (White & Poldrack, 2014), with a higher criterion set for what counts as evidence towards discouraged decisions.

Accumulation rates (PM trials). *Reactive control.* Our experiment provides another opportunity to test for the presence of PMDC's reactive control mechanisms, by examining the differences between PM trial accumulation and non-PM trial accumulation. PM trial accumulation rates are plotted in Figure 8. Trivially, PM accumulation rates on PM trials were

much higher than the PM accumulation rate on non-PM trials, consistent with reactive excitation.

Consistent with our bias manipulation selectively affecting ongoing-task processes, accumulation

towards the PM response on PM non-word trials was not substantially different in the Nc

condition (day one M = 1.87, SD = 0.07; day two M = 2.06, SD = 0.07) and Wc conditions (day

one M = 1.84, SD = 0.09; day two M = 2.20, SD = 0.08), Z = -0.76, p = .22. Similarly, PM

accumulation towards PM words was not substantially different in the Nc condition (day one M

= 1.97, SD = 0.08; day two M = 2.20, SD = 0.07) and the Wc condition (day one M = 1.88, SD =

0.08; day two M = 2.36, SD = 0.06), Z = -0.54, p = .30. However, PM accumulation rates did

increase from day one to day two for both PM word trials, Z = 4.90, p <.001, and PM non-word

trials, Z = 3.64, p <.001. This increase in PM accumulation is consistent with practice improving

performance on the PM task. Consistent with reactive inhibition, all ongoing-task accumulation

rates were far lower on PM trials than on non-PM trials (see contrasts in Table 4). These reactive

inhibitory control effects were very large, replicating previous findings (Strickland et al. 2018).
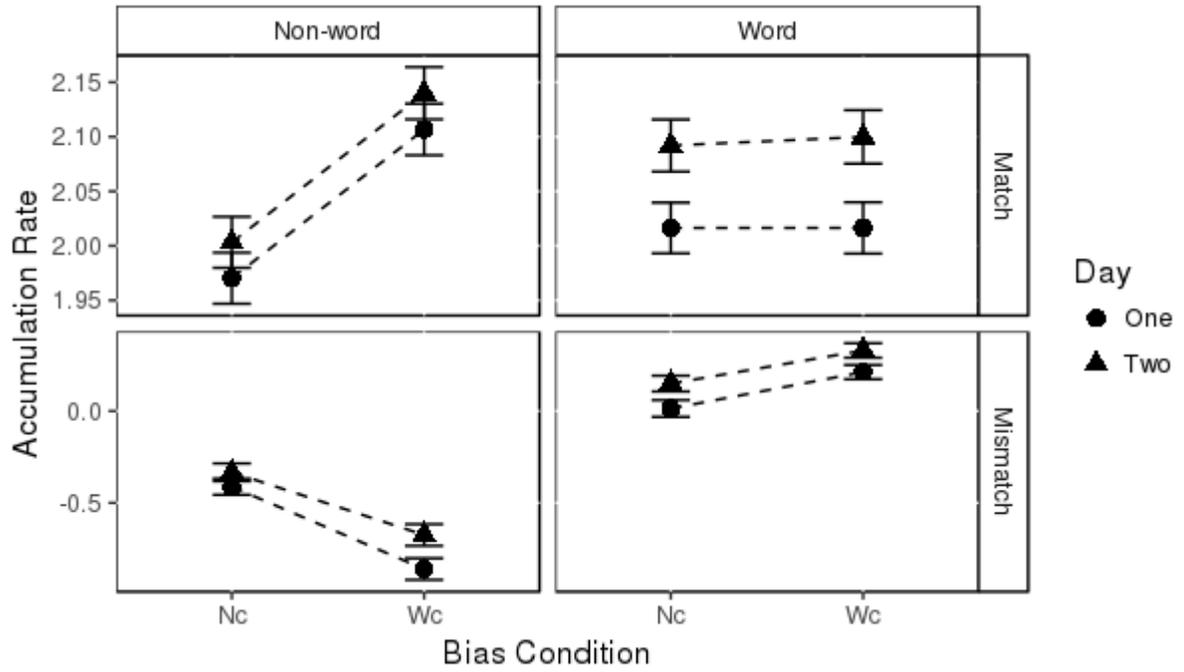
Figure 7. Participant-averaged non-PM trial accumulation rates for word and non-word stimuli (columns). The symbols correspond to the posterior mean, and the error bars correspond to plus and minus one standard deviation from the mean. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against word decisions. Accumulation to PM decisions on non-PM trials ($M = -2.79$, $SD = 0.11$), that is PM false alarm accumulation, is not plotted.
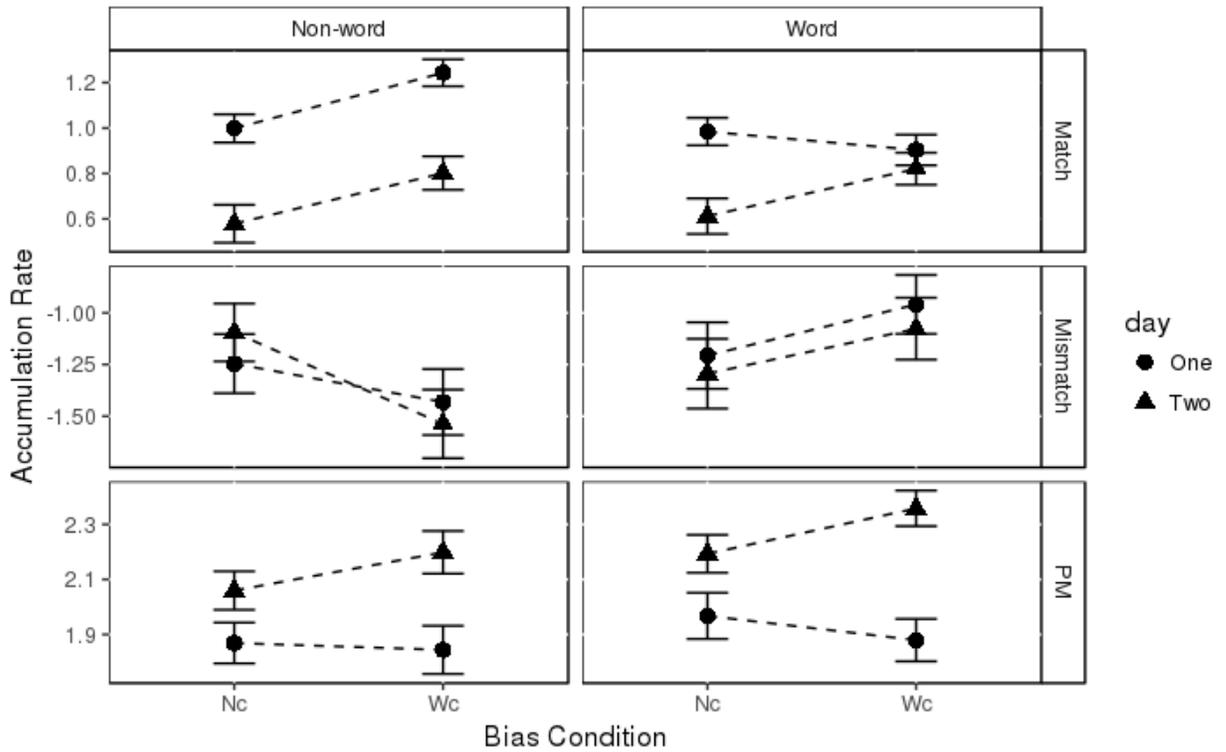
Figure 8. Participant-averaged PM trial accumulation rates. The symbols correspond to the posterior mean, and the error bars correspond to plus and minus one standard deviation from the mean. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against word decisions.

Table 4. *Contrasts relevant to reactive inhibitory control. Z values of the posterior difference distributions (corresponding one-tailed p values in brackets)*

Contrast	Word Accumulator	Non-word Accumulator
Word trials		
Nc: non-PM - PM	27.12 (<.001)	11.35 (<.001)
Wc: non-PM - PM	26.26 (<.001)	12.38 (<.001)
Non-word trials		
Nc: non-PM - PM	7.87 (<.001)	24.58 (<.001)
Wc: non-PM - PM	5.94 (<.001)	24.91 (<.001)

Accumulation rates (PM trials). Bias manipulation. The bias manipulation was expected to affect ongoing task accumulation rates on PM trials similarly to how it affected non-PM trials accumulation rates, and it did. Non-word accumulation was substantially slower to PM non-words in the Nc condition (day one $M = 1.00$, $SD = 0.06$; day two $M = 0.58$, $SD = 0.08$) than in the Wc condition (day one $M = 1.24$, $SD = 0.06$; day two $M = 0.80$, $SD = 0.07$), $Z = 3.49$, $p < .001$. For PM word trials, there was a trend towards non-word accumulation being slower for the Nc condition (day one $M = -1.21$, $SD = 0.16$; day two $M = -1.29$, $SD = 0.17$) than the Wc condition (day one $M = -0.96$, $SD = 0.14$; day two $M = -1.08$, $SD = 0.15$), $Z = 1.5$, $p = 0.07$. Word accumulation was not substantially slower to PM words in the Wc condition (day one $M = 0.90$, $SD = 0.07$; day two $M = 0.82$, $SD = 0.07$) than the Nc condition (day one $M = 0.98$, $SD = 0.06$; day two $M = 0.61$, $SD = 0.08$), $Z = -0.97$, $p = .17$. However, it was slower towards PM non-words in the Wc condition (day one $M = -1.43$, $SD = 0.16$; day two $M = -1.54$, $SD = 0.17$) than the Nc condition (day one $M = -1.25$, $SD = 0.14$; day two $M = -1.09$, $SD = 0.14$), $Z = -2.07$, $p = .02$.

Posterior Exploration

Simulations from PMDC can be useful to understand why the model fitted the observed data, and to explore other types of data that the model could potentially fit. In the supplementary materials, we report simulations that break down in detail exactly how the bias manipulation affected various aspects of ongoing task performance. Here, we report a simulation that answers one particularly pertinent question: whether there are conditions under which PMDC *would* predict that ongoing-task bias substantially affects PM accuracy. To do so, we examine predicted effects of bias on PM accuracy caused by manipulating some parameters in our model while maintaining others at their values estimated from the current study.

To calculate an overall measure of the effect of bias condition on PM accuracy, we summed the increase in PM accuracy to PM non-words in the Nc condition with the increase in PM accuracy to PM words in the Wc condition. We plot detailed posterior predictions of this measure in the supplementary materials and summarize here by discussing the posterior mean predictions. Bias condition did not induce a substantial shift in PM accuracy across PM word and non-word targets in the data (total summed PM accuracy shift = 0.03), and this lack of effect was fit closely by the earlier presented full model (0.025). However, we identified two ways in which our model *could* simulate some degree of bias-induced shift in PM accuracy. For one, substantially amplifying the effect that the bias manipulation had on thresholds could produce a PM accuracy shift. However, this method was very inefficient in improving PM accuracy. For example, when we more than doubled the observed mean bias effects, by adding 0.1 to the word threshold in the Wc condition and 0.1 to the non-word threshold in the Nc condition, the predicted bias effect on PM accuracy (i.e., the sum of PM word accuracy advantage in Wc over Nc and PM non-word advantage in Nc over Wc) was only 0.07 (as compared with 0.03 in the data). Such large shifts in bias could slow performance and potentially impose ongoing-task accuracy costs, making it unlikely that participants would be prepared to implement them.

Another way the model could predict a bias benefit to PM accuracy was by reducing the variability in rates effectively to zero. This makes it virtually impossible for the mismatching ongoing task accumulator (e.g., the non-word accumulator on a PM word trial) to draw an accumulation rate that allows it to compete with the PM accumulator. Reducing rate variability this way led to our model predicting a summed bias benefit to PM accuracy of around 0.06. However, as rate variability was set to implausible levels, it is unlikely such an effect could be induced experimentally. Still, this result is informative about mechanisms in the full model, as it

indicates that rate variability reduces the effects of bias on PM accuracy. It appears that rate variability allows the ongoing-task mismatch accumulator to become competitive with the PM accumulator on some trials, and so the mismatch accumulator can pre-empt PM when favored by bias.

In summary, simulations from PMDC predicted that inducing a bias benefit to PM accuracy would require inflating the bias effects on the ongoing task to more than double that observed in the data, or removing rate variability from the model, both of which seem unlikely to occur in practice.

Discussion

We manipulated bias towards word and non-word responding in a lexical decision task and examined resulting effects on PM performance to word and non-word PM trials. Our manipulation was successful in affecting ongoing-task bias, without substantially influencing confounding processes such as PM thresholds or PM accumulation rates across conditions. We observed a lower proportion of word responses in the Wc condition, and a lower proportion of non-word responses in the Nc condition. Furthermore, RTs increased to word responses in the Wc condition and to non-word responses in the Nc condition. However, we did not find that ongoing-task bias affected PM accuracy or PM hit RTs. Instead, it affected the type of PM errors submitted, with a lower proportion of non-word errors in the Nc condition and a trend towards less word PM errors submitted in the Wc condition. Ongoing-task bias also affected the RTs of PM errors, with slower PM miss non-word responses in the Nc condition.

We found that PMDC provided an accurate and informative account of our data. The model indicated a threshold bias against word responding in the Wc condition and against non-word responding in the Nc condition, consistent with response bias. The bias manipulation also

affected ongoing-task accumulation rates, such that non-word accumulation rates were reduced in the Nc condition and word accumulation rates reduced in the Wc condition. These shifts in ongoing-task accumulation, although unanticipated, are consistent with stimulus bias effects reported by White and Poldrack (2014) (see also Starns & Ratcliff, 2010), whereby participants are more stringent in accepting that a stimulus provides evidence for a choice when biased against that choice. In any case, these effects on ongoing-task accumulation were not the focus on the study, and do not interfere with our PM-related conclusions.

Although PMDC indicated clearly that the bias manipulation affected ongoing-task thresholds, it also successfully accounted for the fact that the bias manipulation did not affect PM accuracy. This is in part because the bias manipulation allowed the 'mismatching' lexical-decision accumulator to compete with the PM accumulator on some PM trials (e.g., the non-word accumulator became more likely to reach threshold on PM word trials), reducing any potential benefits of delaying the matching accumulator. Simulations suggested that threshold bias could conceivably affect PM if the effect on threshold was much larger than observed. However, to get even a small PM benefit to bias, we simulated bias effects twice as large as those observed, which could interfere with performance more generally (e.g., by slowing ongoing-task RTs and reducing accuracy). Thus, it seems that participants would be unlikely to implement ongoing-task bias increases that could effectively support PM. Our finding that ongoing-task bias was not effective in improving PM may shed some light on previous findings regarding "stimulus-specific" PM instructions. Under such instructions, in which participants are informed that PM targets only appear in one type of ongoing-task item, Heathcote et al. (2015) argued in their delay theory that PM could be supported by selectively raising the ongoing-task threshold corresponding to a matching response to that item. However, the current findings

suggest that such selective biases may not be as effective as they anticipated. This could explain the findings of Horn and Bayen (2015), and Strickland et al., (2017, 2018) that participants increased both ongoing-task thresholds with PM, even when explicitly informed that their PM task was stimulus-specific.

Our finding that ongoing-task bias does not affect PM accuracy is inconsistent with Heathcote et al.'s (2015) prediction based on Delay Theory, underscoring the difficulty with anticipating exactly how control processes will influence the output of complex cognitive models. They reasoned that bias might improve PM performance because delaying the ongoing-task accumulator that is most competitive with PM should allow the PM accumulator more time to accumulate. However, our simulations from PMDC predict that effects of ongoing-task bias on PM are not substantial, in part because the beneficial effects of delays in the matching accumulator are offset by the mismatching accumulator competing with the PM accumulator. This was not anticipated in Heathcote et al.'s work that focused solely on PM cost because, without a full model of both ongoing-task and PM task decision processes, it was not possible to directly examine how cost-related mechanisms impacted PM performance. This illustrates how researcher's intuitions cannot be assumed to accord with the function of a cognitive model (Farrell & Lewandowsky, 2010), highlighting the importance of directly simulating quantitative model predictions where possible. In a similar vein, Anderson et al. (2018)'s finding that increased ongoing-task thresholds did not improve PM accuracy was potentially undermined by the fact that they did not measure, or account for, shifts in PM accumulator control processes such as PM thresholds.

The current results, and those of Anderson et al.'s (2018) , indicate that ongoing-task thresholds play a relatively minor role in supporting PM accuracy, with the potential to be

overpowered or nullified by other PM decision processes. Taken together, these findings illustrate that delay theory, at least when interpreted in isolation, fails to accurately predict PM accuracy. However, ongoing-task thresholds are by far the largest underlying component of PM cost in a range of studies. Thus, delay theory provides a compelling account of PM costs, but a poor account of PM accuracy, illustrating an important distinction between these measures.

Given that increased ongoing-task thresholds underlie a large part of PM cost, and are largely comprised of threshold elevations that appear to have little effect on PM accuracy, we believe the PM cost measure deserves less focus than it has had in the literature. In the past, influential PM theories such as the Preparatory Attentional and Memory processes theory (Smith, 2003), and the Multi-process view (Einstein et al. 2005), have used costs to infer the PM processes responsible for variation in PM accuracy. For example, they assume that non-focal PM accuracy is poorer than focal PM accuracy because it is reliant on PM monitoring that shares capacity with the ongoing task, whereas focal PM is less reliant (and in the case of the Multi-process view, is fully reliant on spontaneous retrieval), and that PM importance emphasis causes more capacity to be allocated to the PM task, resulting in increased cost and increased PM accuracy (Einstein et al., 2005; Smith & Bayen, 2004). Whether one adopts our theoretical position that costs reflect ongoing-task thresholds, or other theoretical positions that costs reflect monitoring, our modeling results indicate neither cognitive process is the primary determinant of PM performance. We acknowledge that these conclusions have far reaching implications for the way PM research is done, and further research is required to verify them. It will be important for future research to test whether our findings hold for a broader range of bias manipulations and task implementations. For example, in the current study, our instructions' emphasis on the ongoing task, and associated delays punishing ongoing task errors, may have led to the

perception that the PM task is of secondary importance. This could have reduced the attention or

effort that some participants paid to the PM task. However, as both our experimental conditions

include a bias instruction, any effects of bias on PM importance would be expected to be equal

across our conditions and thus there is no reason to expect this would confound any of our

comparisons. Still, it would be interesting to examine how the findings apply at different levels

of PM importance.

It is worth noting that although we analyze the results of only one experiment here, the

quality of measurement was high, with over 2000 trials modelled for each participant. Hence, our

single experiment with 32 participants yields more data than 160 participants would in a typical

PM experiment (assuming 400 trials per participant), and substantially more data than previous

large-scale studies on PM cost (e.g., Anderson et al., 2018). Although PM trials comprise only a

fraction of this data, PM trials were a larger proportion of total trials in this study than in most

previous studies, and so the number of PM trials we observed is also substantially larger than is

typical. This focus on trial numbers is necessary for process modelling, where reliable inference

depends on the number of trials per participant, rather than the number of participants (Kolossa

& Kopp, 2018), and has been argued to underpin the most reproducible findings in psychology

(Smith & Little, 2018). Given this, we believe it worth considering that rather than relying on

PM cost for inferences about PM processes, it is much more effective to use a cognitive model

like PMDC that directly measures the psychological processes underlying PM performance

(Strickland et al. 2018). In this vein, we have recently used the PMDC framework to develop and

test a detailed theory of how PM and ongoing processes can share capacity in cognitively

demanding, complex tasks such as air traffic control (Boag et al., 2019) and maritime

surveillance (Strickland et al., 2019). One again, these experiments have an order of magnitude

more data per participant than typical PM experiments, and in the case of Boag et al.'s (2019) experiment, 246 participants.

Although increases in ongoing-task thresholds do not effectively support PM in the standard laboratory paradigm we used here, they may do so in different paradigms. For example, we have found that proactive control over ongoing-task thresholds more substantially supports PM accuracy in simulated air traffic control (e.g., Boag et al., 2019), perhaps owing to the longer time scales of the simulated air traffic control decisions. In addition, even in standard laboratory paradigms, some methods of delaying the ongoing-task can improve PM performance. For example, Loft and Remington (2013) found that preventing participants from submitting responses for around 1 second could bring PM accuracy to almost ceiling. However, this manipulation was much stronger than any delay likely to be imposed by threshold control - in the current paradigm, a 1 second delay would more than double RTs. It is also possible that the delay theory mechanism could potentially play a secondary, minor role in supporting PM accuracy when complemented by decreased PM thresholds and increases in reactive control, as was shown in Strickland et al. (2018). With PM importance emphasis, increases in ongoing-task thresholds, decreases in PM thresholds and increases in PM-induced reactive inhibitory control acted together to enhance PM performance. The major difference between Strickland et al. and the current study is that the former manipulated participants' motivation towards PM and the ongoing-task directly, allowing them to adjust their cognitive control processes as they saw fit to achieve desired outcomes. By contrast, in the current study we attempted a selective manipulation of ongoing-task thresholds. Although our manipulation also affected ongoing-task rates it did not affect parameters associated with the PM process, PM thresholds and accumulation rates associated with reactive control. Thus, our results combined with those of

Strickland et al. suggest that it is the parameters associated with the PM process that should be the target of manipulations that attempt to improve PM performance.

One interesting, but unanticipated, finding was that PM accuracy was substantially higher on day two of the study than day one. This contrasts with Strickland et al. (2018), where PM accuracy decreased on later days of the study. A difference between the current study and previous studies was that here we re-used the same PM target letter strings (i.e., the letter strings 'tor' and 'ver') on days one and day two. Over time, practice at the PM task may have led participants to develop familiarity with the PM letter strings, improving the PM-related evidence extracted from the stimulus. Consistent with this explanation, our model indicated PM accumulation rates were higher on day two than day one. The role of stimulus familiarity in PM was also highlighted by a recent study that found differences between focal and non-focal PM accuracy could be eliminated with repeated exposure to non-focal PM targets (Hicks, Franks, & Spitler, 2017). The role of learning and familiarity in PM processes awaits further investigation.

In summary, the current study indicated that manipulating ongoing-task bias has little effect on PM accuracy, because bias against one ongoing-task decision allows the other ongoing-task decision to effectively pre-empt PM. It is theoretically possible that extreme ongoing-task bias may affect PM accuracy, but the deleterious effect of such bias on ongoing-task performance makes it impractical to implement. Further, the current results suggest that because it is based on a comprehensive characterization, the PMDC model (Strickland et al. 2018), rather than Delay Theory (Heathcote et al. 2015), should be used as a basis for understanding, and making predictions about event-based PM. Similarly, but in terms of empirical measures, our results indicate that ongoing-task RTs provide an incomplete and potentially misleading guide to PM-related processing in general, and PM accuracy in particular.

References

Anderson, F. T., Rummel, J., & McDaniel, M. A. (2018). Proceeding with care for successful prospective memory: Do we delay ongoing responding or actively monitor for cues? *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software; Vol 1, Issue 1 (2015)*. Retrieved from https://www.jstatsoft.org/v067/i01

Boag, R. J., Strickland, L., Heathcote, A., Neal, A., & Loft, S. (2019). Cognitive control and capacity for prospective memory in complex dynamic environments. *Journal of Experimental Psychology. General*. https://doi.org/10.1037/xge0000599

Boag, R. J., Strickland, L., Loft, S., & Heathcote, A. (in press). Strategic attention and decision control support prospective memory in a complex dual-task environment. *Cognition*.

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*(2), 106–113.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178.

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 University Press.

Dennis, S. (1995). *The Sydney Morning Herald word database*.

Dismukes, R. K. (2012). Prospective memory in workplace and everyday situations. *Current Directions in Psychological Science*, *21*(4), 215–220.

https://doi.org/10.1177/0963721412447621

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 717–726.

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

Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian Data Analysis* (3rd ed.). London, England: Taylor & Francis.

Heathcote, A., Lin, Y., Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2019). Dynamic models of choice. *Behavior Research Methods*, *51*(2), 961–985.

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

Hicks, J. L., Franks, B. A., & Spitler, S. N. (2017). Prior task experience and comparable stimulus exposure nullify focal and nonfocal prospective memory retrieval differences. *The Quarterly Journal of Experimental Psychology*, *70*(10), 1997–2006.

Horn, S. S., & Bayen, U. J. (2015). Modeling criterion shifts and target checking in prospective memory monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(1), 95–117.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633.

Kolossa, A., & Kopp, B. (2018). Data quality over data quantity in computational cognitive neuroscience. *NeuroImage*, *172*, 775–785.

Loft, S., Dismukes, R. K., & Grundgeiger, T. (2019). Prospective memory in safety-critical work

    contexts. In J. Rummel & M. A. McDaniel (Eds.), *Prospective Memory* (pp. 170–185).

    Routledge.

Loft, S., & Remington, R. W. (2013). Wait a second: Brief delays in responding reduce focality

    effects in event-based prospective memory. *The Quarterly Journal of Experimental*

    Psychology, *66*(7), 1432–1447.

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*(6), 1757–1764.

Marsh, R. L., Hicks, J. L., Cook, G. I., Hansen, J. S., & Pallos, A. L. (2003). Interference to

    ongoing activities covaries with the characteristics of an event-based intention. *Journal of*

    Experimental Psychology: Learning, Memory, and Cognition, *29*(5), 861–870.

Meier, B., & Rey-Mermet, A. (2012). Beyond monitoring: After-effects of responding to

    prospective memory targets. *Consciousness and Cognition*, *21*(4), 1644–1653.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau

    (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64.

R Core Team. (2019). *R: A language and environment for statistical computing.* Retrieved from

    https://www.R-project.org/

Raab, D. H. (1962). Statistical facilitation of simple reaction time. *Transctions of the New York*

    Academy of Sciences, *24*, 574–590.

Rothschild, J. M., Landrigan, C. P., Cronin, J. W., Kaushal, R., Lockley, S. W., Burdick, E., …

    Czeisler, C. A. (2005). The Critical Care Safety Study: The incidence and nature of

    adverse events and serious medical errors in intensive care*. *Critical Care Medicine*,

33(8), 1694–1700.

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review*, *25*(6), 2083–2101.

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 347–361.

Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 756–777.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series b (Statistical Methodology)*, *64*(4), 583–639.

Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed–accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging*, *25*(2), 377–390.

Strickland, L., Elliott, D., Wilson, M. D., Loft, S., Neal, A., & Heathcote, A. (2019). Prospective memory in the red zone: Cognitive control and capacity sharing in a complex, multi-stimulus task. *Journal of Experimental Psychology. Applied*. https://doi.org/10.1037/xap0000224

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*(10), 1616–1629.

Strickland, L., Loft, S., Remington, R. W., & Heathcote, A. (2018). Racing to remember: A theory of decision control in event-based prospective memory. *Psychological Review*, *125*(6), 851–887.

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*(3),

    368–384.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights.

    Psychonomic Bulletin & Review, *11*(1), 192–196.

White, C. N., & Poldrack, R. A. (2014). Decomposing bias in different types of simple decisions.

    Journal of Experimental Psychology: Learning, Memory, and Cognition, *40*(2), 385–398.