

Investigating the effects of ongoing task bias on prospective memory

Journal:	<i>Quarterly Journal of Experimental Psychology</i>
Manuscript ID	QJE-STD-19-291.R2
Manuscript Type:	Standard Article
Date Submitted by the Author:	27-Feb-2020
Complete List of Authors:	Strickland, Luke; Curtin University, Future of Work Institute Loft, Shayne; University of Western Australia, Psychological Science Heathcote, Andrew; University of Tasmania, The School of Psychology
Keywords:	prospective memory, linear ballistic accumulator, prospective memory cost, response bias

SCHOLARONE™
Manuscripts

Investigating the effects of ongoing-task bias on prospective memory

Luke Strickland¹, Shayne Loft², Andrew Heathcote³

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

Author Note

This research was supported by the Australian Government through the Australian Research Council (DP160101891). All data and code associated with the manuscript are available at https://github.com/lukestrickland/PM_ongoingtask_bias.

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.

Keywords: *prospective memory, linear ballistic accumulator, prospective memory cost, response bias*

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.

1
2
3
4
5 Heathcote et al., (2015) also found that when PM target stimuli corresponded to only one of two
6
7 ongoing-task choices (all PM items were words in an ongoing lexical decision task) participants
8
9 biased responding against that choice by raising its threshold more than the threshold for the
10
11 other choice (thresholds to make a ‘word’ decision were elevated more than thresholds to make a
12
13 ‘non-word’ decision). They argued that this was consistent with participants using a selective
14
15 bias strategy in order to improve PM responding. Their idea was that biasing ongoing-task
16
17 thresholds against a specific ongoing-task choice could improve PM performance to target
18
19 stimuli associated with that choice (Heathcote et al., 2015). For example, a bias against
20
21 responding ‘word’ in a lexical decision task would be expected to improve PM accuracy to word
22
23 targets. However, Heathcote et al. did not test this proposition empirically and could not show
24
25 that it worked theoretically because their modeling was limited to the ongoing task (i.e., the PM
26
27 decision process was not modelled).
28
29
30
31

32
33 Recently, Strickland, Loft, Remington, and Heathcote (2018) developed an evidence-
34
35 accumulation theory of both ongoing-task and event-based PM decision processes, ‘Prospective
36
37 Memory Decision Control’ (PMDC), that provides a comprehensive account for ongoing-task
38
39 and PM response choices and response times (RTs). Following Loft and Remington (2013),
40
41 PMDC proposes that parallel PM and ongoing-task decision processes, as modelled by linear
42
43 ballistic accumulators (LBA; Brown & Heathcote, 2008), race to accumulate evidence towards
44
45 their respective thresholds. Consistent with Delay Theory, simulations from Strickland et al.’s
46
47 model indicated that globally increasing ongoing-task decision thresholds (e.g., increasing both
48
49 word and non-word thresholds) improved PM performance, by reducing the probability that
50
51 ongoing-task decisions would pre-empt PM processes.
52
53
54

55
56 It is difficult to experimentally test delay theory’s claim that increased ongoing task
57
58
59
60

1
2
3
4
5 caution improves PM accuracy, because manipulations targeted at ongoing task caution may
6
7 affect the perceived importance of the PM task, with potentially confounding effects on PM
8
9 processes. Instead, this study tests the other key claim of delay theory: that biases in ongoing task
10
11 thresholds can affect PM. If such biases affect PM accuracy, they could provide an efficient
12
13 strategy to improve PM when PM events only conflict with a specific type of response. For
14
15 example, suppose one usually turns left at a specific round-a-bout on the way home from work,
16
17 but one day they must turn right to go shopping. Their success at this PM task may be improved
18
19 by selectively increasing caution towards turning left at future roundabouts (but not necessarily
20
21 increasing caution to go straight or in other directions), as this would increase the time available
22
23 to retrieve their intent to turn right at that specific roundabout in order to go shopping.
24
25
26

27
28 In PMDC, selectively raising the threshold to an ongoing-task choice (e.g., raising the
29
30 threshold to respond ‘word’) could allow the PM process time to reach threshold on relevant
31
32 trials (following the example, PM would be less likely to be pre-empted by ‘word’ decisions on
33
34 word PM trials). However, although Strickland et al., (2018) found that globally raised ongoing-
35
36 task thresholds contributed to PM accuracy, PMDC modelling indicated that other processes,
37
38 including proactive control of PM thresholds and reactive control of evidence accumulation,
39
40 were more important in determining PM accuracy. Thus, it is not a given that ongoing-task
41
42 biases can substantially affect PM accuracy. We apply PMDC to our study to test whether it can
43
44 account for the effects of our bias manipulation. Applying PMDC also allows us to assess how
45
46 our manipulation affects cognitive processes (e.g., test whether the bias manipulation affects
47
48 ongoing task thresholds), and how shifts in cognitive processes map to shifts in performance
49
50 (e.g., determine to what degree shifts in PM accuracy are caused by shifts in ongoing task
51
52 thresholds). Before reporting the experiment, we introduce PMDC in more detail.
53
54
55
56
57
58
59
60

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.

--- Insert Figure 1 about here ---

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

1
2
3
4
5 from *capacity sharing* between PM and ongoing-task processes (e.g., Einstein et al., 2005;
6
7 Smith, 2003). The idea is that monitoring for PM items usurps resources from ongoing-task
8
9 processes, increasing RTs. PMDC measures information-processing capacity with its
10
11 accumulation rate parameters, making it possible to test the capacity-sharing hypothesis.
12
13 Ongoing task accumulation rates can be associated with either the ‘match’ or ‘mismatch’
14
15 accumulator. The *match* accumulation rate measures evidence accumulation in the accumulator
16
17 corresponding to the correct decision. The *mismatch* accumulation rate refers to evidence in the
18
19 accumulator corresponding to the incorrect decision. In Figure 1, for example, the evidence
20
21 accumulation rate for the ‘word’ accumulator corresponds to ‘match’ accumulation on word
22
23 trials and ‘mismatch’ accumulation on non-word trials. Similarly, the ‘non-word’ accumulation
24
25 rate corresponds to ‘match’ accumulation on non-word trials and ‘mismatch’ accumulation on
26
27 word trials. Evidence accumulation models generally indicate that PM demands do not cost
28
29 ongoing-task capacity in standard paradigms (e.g., Heathcote et al. 2015; Horn & Bayen, 2015;
30
31 Strickland, Heathcote, Remington, & Loft, 2017), but that capacity sharing can occur in more
32
33 demanding paradigms, such as in simulations of air traffic control (Boag, Strickland, Heathcote,
34
35 Neal, & Loft, 2019; Boag, Strickland, Loft, & Heathcote, 2019) and maritime surveillance
36
37 (Strickland et al., 2019).

38
39
40
41
42
43
44 **Proactive control.** Proactive control refers to cognitive control applied in advance of
45
46 cognitively demanding events, to prepare for their occurrence. Proactive control is applied over
47
48 PMDC’s response thresholds, as thresholds are the locus of *a priori* strategy. Prior to the
49
50 development of PMDC, the delay theory of PM cost (Heathcote et al., 2015) suggested that
51
52 participants raise ongoing-task thresholds so that ongoing-task response selection does not pre-
53
54 empt PM response selection, improving PM. For example, in Figure 1 a high threshold to
55
56
57
58
59
60

1
2
3
4
5 respond word is depicted. Thus, on PM trials that require a ‘word’ ongoing-task response, the
6
7 ‘word’ accumulator will take a relatively longer time to reach threshold than if the threshold was
8
9 lower, and so there would be more time for the PM accumulator to accrue evidence, increasing
10
11 the probability that it reaches its threshold first. Consistent with delay theory, elevated ongoing
12
13 task thresholds have been found to underlie PM costs in many applications of evidence
14
15 accumulation models to PM cost data (e.g., Heathcote et al., 2015; Horn & Bayen, 2015;
16
17 Strickland et al., 2017; Strickland et al., 2018). Thus, PMDC includes proactive control over
18
19 ongoing task thresholds as a possible mechanism for improving PM accuracy.
20
21
22
23

24 Delay theory and PMDC are similar in that both propose a race to response selection
25
26 between PM and ongoing task processes that could potentially be supported by increases in
27
28 ongoing task thresholds. A key difference is that PMDC quantitatively instantiates the PM
29
30 process as an LBA accumulator, whereas delay theory provided only a verbal description of the
31
32 PM process. In fact, simulations reveal that, given PMDC’s assumptions, ongoing task
33
34 thresholds have only a weak effect on PM accuracy (Strickland et al., 2018). Furthermore, delay
35
36 theory proposes only a single mechanism that supports PM, whereas PMDC proposes a range of
37
38 mechanisms other than ongoing task threshold delays that could improve PM accuracy. For
39
40 example, under PMDC proactive control can also apply over the PM threshold. Indeed, when the
41
42 importance of PM is emphasized, it has been found that the PM threshold can be decreased to
43
44 increase the probability of a PM decision (Strickland et al., 2018), and that this form of proactive
45
46 control is critical to supporting PM accuracy.
47
48
49
50
51

52 **Reactive control.** Reactive control refers to cognitive control that occurs “just in time”,
53
54 that is when PM event is processed. PMDC’s reactive control structure is depicted in Figure 2.
55
56
57
58
59
60

1
2
3
4
5 Upon processing a PM item, encoding PM stimulus inputs may cause participants to accrue
6
7 evidence towards the PM decision (reactive excitation), but to inhibit (i.e., slow down)
8
9 accumulation to competing ongoing-task decisions (reactive inhibition). In Strickland et al.,
10
11 (2018), both forms of reactive control were critical to explaining variation in PM accuracy.
12
13

14
15 --- Insert Figure 2 about here ---
16
17
18

19 20 **Testing Delay Theory**

21
22 Recently, Anderson, Rummel and McDaniel (2018) attempted to isolate, and manipulate,
23
24 the effect of ongoing-task decision thresholds on PM accuracy. They compared ‘standard’ event-
25
26 based PM conditions with a ‘delay’ condition that instructed participants to be cautious to make
27
28 ongoing-task decisions and not to monitor for PM targets. They found that the latter did not
29
30 improve PM performance, and argued that, therefore, proactive control over ongoing-task
31
32 decisions does not support PM. However, it is not clear that their manipulation selectively
33
34 affected ongoing-task decision thresholds. Their instruction not to monitor for PM targets may
35
36 have caused at least some participants to increase their PM threshold, counteracting possible
37
38 benefits of increased ongoing-task caution. Indeed, PMDC indicates that much of the
39
40 improvement to PM accuracy under PM importance emphasis is driven by control of the PM
41
42 threshold (Strickland et al., 2018). In Anderson et al.’s study, the PM threshold was not
43
44 estimated because too few PM trials were observed.
45
46
47
48

49
50 Even if Anderson et al.’s (2018) experiment were repeated with no explicit emphasis
51
52 instruction for the PM task, manipulating ongoing-task caution could still affect PM processes.
53
54 For example, emphasizing caution on the ongoing task could make the PM task appear less
55
56
57
58
59
60

1
2
3
4
5 important, leading to a similar confound. Alternatively, if participants try to conserve their
6
7 overall threshold levels then increasing ongoing task caution could lead to a decrease in the PM
8
9 threshold. Thus, delay theory cannot be decisively tested by manipulating ongoing task caution.
10
11 To experimentally test delay theory, it is crucial to manipulate ongoing task thresholds in
12
13 isolation, without affecting PM processes. In the current study, we attempt such a test. Rather
14
15 than manipulating overall caution, we manipulate another factor that delay theory claims is
16
17 important for supporting PM accuracy: bias in ongoing task thresholds (Heathcote et al., 2015).
18
19

20
21 Bias in ongoing task thresholds refers to threshold levels that advantage one ongoing task
22
23 decision over another. For example, a bias against word decisions could be implemented by
24
25 shifting the word threshold up and non-word threshold down. On average, the ongoing-task
26
27 ‘match’ accumulator will be faster than the ‘mismatch’ accumulator, to support better than
28
29 chance accuracy on the ongoing-task. As a result, delay theory proposes that bias against an
30
31 ongoing-task choice could potentially improve PM performance to items matching that choice –
32
33 for example, a bias against ‘word’ decisions could improve performance to PM items that are
34
35 words – because it would slow down participants making word responses and thus allow the PM
36
37 response more time to reach threshold (Heathcote et al., 2015). This claim has been supported by
38
39 analysis of “stimulus-specific” PM tasks (Lourenço, White, & Maylor, 2013), in which PM is
40
41 associated with a specific event (e.g., PM items are always words). Under such conditions
42
43 implementing ongoing-task biases could be an efficient way to improve PM performance without
44
45 globally slowing down responding. Importantly, this strategy can be implemented proactively
46
47 and does not require the participant to make any changes in task strategy on a stimulus-by-
48
49 stimulus basis, as the strategy differs between accumulators but is the same for all stimuli.
50
51 Consistent with this, in stimulus-specific tasks participants do implement a bias against the
52
53
54
55
56
57
58
59
60

PROSPECTIVE MEMORY AND ONGOING-TASK BIAS

1
2
3
4
5 decision that competes with PM (e.g., Heathcote et al., 2015; Strickland et al., 2018).

6
7 The findings that stimulus-specific PM tasks induce shifts in ongoing task bias are key to
8 the case for delay theory. They favor delay theory over an alternative theory of ongoing task
9 threshold increases – that PM instructions increase overall perceptions of task complexity (Horn
10 & Bayen, 2015). Although it is plausible that an increase in perceived task complexity would
11 induce a shift in caution, there is no reason to expect it would induce a bias. In contrast, delay
12 theory clearly predicts the shifts in bias, and crucially it makes the claim that such shifts in bias
13 are functional to PM (Heathcote et al., 2015). The claim that bias affects PM performance can be
14 directly tested experimentally because, unlike overall caution, ongoing task bias can be
15 manipulated without confounding from unintended effects on PM processes. Here we present an
16 experiment to test the key claim of delay theory that bias in ongoing-task thresholds affects PM
17 performance. We apply the PMDC model to our experiment, both to contrast it with delay theory
18 and to validate our inferences about latent psychological processes.

19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35 **The Current Study**
36

37 Participants performed a lexical decision task with an accompanying PM task to detect
38 items containing a target syllable (e.g., any letter string containing ‘tor’). We include a within-
39 subject blocked manipulation of ongoing-task bias. In some blocks we induce a bias against
40 making word decisions, and in the others induce a bias against making non-word decisions. We
41 manipulate bias by discouraging certain types of errors – for example, to induce a bias against
42 ‘word’ responding, we strongly discourage making ‘word’ responses on non-word trials. With
43 this manipulation, there is no obvious reason that the relative importance of the PM or ongoing
44 tasks would differ across bias conditions. Thus, the design is suitable to test delay theory without
45 confounding from differences in the perceived relative importance of the ongoing and PM tasks.

46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

PROSPECTIVE MEMORY AND ONGOING-TASK BIAS

12

1
2
3
4
5 Word and non-word PM targets are included in both blocks. Thus, we can assess the degree to
6
7 which bias against word responding benefits PM accuracy to words, and the degree to which bias
8
9 against non-word responding benefits PM accuracy to non-words. To examine the effects of
10
11 ongoing-task bias on PM performance, we examine PM performance across our manipulation of
12
13 PM stimulus type (PM word, PM non-word), and our blocked bias manipulation (bias against
14
15 word, bias against non-word). In addition to standard analyses, we apply the PMDC model to
16
17 determine whether it can fit to the effects of this new manipulation. PMDC is also critical to
18
19 determining whether our bias manipulation is successful in affecting thresholds.
20
21
22

23 According to the delay theory, we would expect to observe increased PM accuracy to word
24
25 PM targets when bias is induced against word decisions, and increased PM accuracy to non-word
26
27 PM targets with bias against non-word decisions. This would occur because bias extends the
28
29 completion time of the matching ongoing-task accumulator, allowing the PM accumulator more
30
31 time to reach threshold (Heathcote et al., 2015). Although such a mechanism may be possible
32
33 under PMDC, previous simulations from the model suggest only a minor role for ongoing task
34
35 threshold elevation in supporting PM accuracy (Strickland et al., 2018). In addition, alternative
36
37 mechanisms could reduce the potential benefits to PM accuracy of bias against the matching
38
39 accumulator. For example, although the matching accumulator is faster on average than the
40
41 mismatching accumulator in the PMDC model, the mismatching accumulator will not always be
42
43 at a disadvantage, because rates vary from trial to trial. Thus, bias against the correct ongoing-
44
45 task decision may allow the incorrect ongoing-task decision to become competitive with the PM
46
47 process, at least on some trials, in which case bias might not improve PM performance but
48
49 instead increase the proportion of incorrect ongoing-task responses submitted on PM trials.
50
51
52
53
54
55
56
57
58
59
60

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

PROSPECTIVE MEMORY AND ONGOING-TASK BIAS

14

1
2
3
4
5 one block of each two bias conditions on each day. As explained in detail in the procedure
6
7 section, we manipulated bias by instructing participants either to be cautious to make word
8
9 responses (word caution condition, *Wc*), or cautious to make non-word responses (non-word
10
11 caution condition, *Nc*). The bias condition order used in day 1 was reversed for day 2, e.g. if the
12
13 *Wc* condition was block one of day one then it would be block 2 of day 2. For each day of the
14
15 experiment for each participant, one substring (*tor*, *ver*) was the PM target for the *Wc* block and
16
17 the other substring was the PM target for the *Nc* block. The assignment of PM target substring to
18
19 condition was reversed for each participant between day 1 and day 2. As condition block order
20
21 was also reversed for each participant between day 1 and day 2, substring block order was the
22
23 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)
24
25 then *tor* had to also be the target in the first block of day two (following the example, it would be
26
27 a *Nc* block). The four different ways in which the block order and PM target substrings could be
28
29 matched (while satisfying the above conditions) were counterbalanced across the 32 participants.
30
31
32
33
34

35 In each block participants were presented with 309 non-target non-words and 309 non-
36
37 target words, as well as 21 PM target non-words and 21 PM target words. For each participant,
38
39 the 21 PM target words and 21 non-words used for the *Wc* condition for a given substring were
40
41 drawn randomly, without replacement, from the total 42 words and 42 non-words which
42
43 contained that substring. The other 21-word PM targets and 21 non-word PM targets were used
44
45 for the *Nc* condition. The order in which participants were presented their non-target stimuli was
46
47 random within each block. In order to reduce fatigue effects participants were given five 1-
48
49 minute breaks within each 660 trial block. The breaks occurred after participants completed each
50
51 110 trial segment, so after trials 110, 220, 330, 440 and 550. Thus, blocks were divided into
52
53 sixths. PM targets (in both *Wc* and *Nc* blocks) were presented 42 times per 660 trial block;
54
55
56
57
58
59
60

1
2
3
4
5 randomly between trials 6-20, 21-35, 36-50, 51-65, 66-80, 81-95, 96-110 of each sixth of a
6
7 block. Therefore, the ratio of PM trials to nontarget trials was 1:14. Target trials were separated
8
9 by at least 4 lexical decision trials. The order in which the PM targets filled the chosen positions
10
11 was random.

14 Procedure

16 Participants first performed practice lexical decision trials. They were instructed that they
17
18 would be presented with letter strings and that they should press a key to indicate whether strings
19
20 were words or non-words (e.g. press 's' for word, 'd' for non-word'). They were asked to make
21
22 their responses as quickly and accurately as possible. For the experimental blocks (the PM
23
24 blocks, both Wc and Nc), participants were additionally instructed to press an alternative key
25
26 instead of their word or non-word response when they encountered items containing a target
27
28 substring, e.g. *"In the next block of lexical decision trials, if you see ANY item that contains 'tor'*
29
30 *then press 'j' INSTEAD of 's' or 'd'. For example, if you see 'indicator' then press 'j' instead of 's'*
31
32 *or if you see 'botoraty' then press 'j' instead of 'd'.* Four response key assignments were
33
34 counterbalanced across participants; 1) $s = \text{word}, d = \text{non-word}, j = \text{PM}$, 2) $d = \text{word}, s = \text{non-}$
35
36 $\text{word}, j = \text{PM}$, 3) $k = \text{word}, j = \text{non-word}, d = \text{PM}$, and 4) $j = \text{word}, k = \text{non-word}, d = \text{PM}$. The
37
38 four response key orders were also counterbalanced with bias condition block order and PM
39
40 target substring block order, so that each combination was used for two of the 32 participants.
41
42 Participants were instructed before the commencement of each sixth of a block to rest their
43
44 fingers on their assigned response key combination: one hand resting the index and middle finger
45
46 on the lexical decision keys (e.g. left hand index on d , left hand middle on s), and the other hand
47
48 resting the index finger on the PM key (following the example, right hand index on j).
49
50
51
52
53
54

55 Each day, participants first completed their 24 practice lexical decision trials and received
56
57
58
59
60

PROSPECTIVE MEMORY AND ONGOING-TASK BIAS

16

1
2
3
4
5 percentage feedback on their accuracy (e.g., “87.50% correct”). They then proceeded to the
6
7 experimental blocks and were presented with either Wc or Nc instructions. For the Wc blocks,
8
9 participants were instructed to be careful about making word responses, e.g. *“In the next block of*
10
11 *trials try to respond quickly and accurately to all items, but note that it is extra important to*
12
13 *avoid errors where you incorrectly respond WORD ('s') to non-word items. That is, only press*
14
15 *the 's' key when you are absolutely sure an item is a word. If you do incorrectly classify an item*
16
17 *as a word, you will be presented a special 'incorrect' message which delays the task more than if*
18
19 *you incorrectly classify an item as a non-word.”* For the Nc blocks, the opposite instruction was
20
21 presented (in the example substitute non-word for word and word for non-word). Each time
22
23 participants received a bias instruction, they repeated the instruction to the experimenter. After
24
25 receiving the bias instructions, participants were given their PM instruction to make an
26
27 alternative response to target substrings (see example from the paragraph above) and repeated
28
29 the instruction to the experimenter. Participants next completed a three minute distractor puzzle,
30
31 after which they began the first block of experimental trials. After completion of each sixth of a
32
33 block, participants were presented feedback on the accuracy of their responses (%) to the lexical
34
35 decision task. After completing their first bias condition for the day (e.g., the Wc condition),
36
37 participants were instructed that their bias and PM instructions no longer applied, e.g. *“Please*
38
39 *note that the instruction you received to prioritize avoiding incorrectly making word responses*
40
41 *to non-words no longer applies. These errors will not trigger a longer 'incorrect' message*
42
43 *anymore. You also do not need to make a special response to items containing 'ver' in the next*
44
45 *block of trials. In fact, no item containing 'ver' will be presented.”* When participants received
46
47 their second block’s PM instructions (after they had already received their new caution
48
49 instructions), they were again reminded that their old PM target was no longer relevant, e.g.
50
51
52
53
54
55
56
57
58
59
60

PROSPECTIVE MEMORY AND ONGOING-TASK BIAS

17

1
2
3
4
5 “Please be reminded that you no longer need to press ‘j’ if you see an item containing ‘ver’. In
6
7 fact, no item containing ‘ver’ will be presented in the next block of trials.” In addition to the
8
9 breaks within blocks, participants were instructed to rest for 2 minutes between blocks.

11 Each trial began with a fixation cross ‘+’, displayed in white on a black background for
12
13 0.5s. The fixation cross was then replaced by a blank screen for 0.25s, which was followed by
14
15 the presentation of a white letter string (size 18 courier new font) which remained on the black
16
17 screen until the participant made a response. If the participant made a correct word/non-word
18
19 response (including on PM trials), or a correct PM response, the subsequent trial immediately
20
21 began (next fixation cross). If the participant made an incorrect response, the feedback they
22
23 received varied depending on bias condition, with longer delays for the discouraged ongoing-
24
25 task errors (e.g., for word responses to non-word trials in the Wc condition). This was included
26
27 to increase the strength of the bias manipulation¹. In the Wc condition, word responses to non-
28
29 word trials triggered a screen which displayed ‘INCORRECT!!!’ in size 44 courier new font for
30
31 15 seconds. Non-word responses to word trials in the Wc condition triggered a screen which
32
33 displayed ‘INCORRECT’ in size 18 courier new font for 1 second. In Nc blocks the reverse was
34
35 true, non-word responses to word trials would trigger the 15 second, size 44 ‘INCORRECT!!!’
36
37 whereas word responses to non-word trials would trigger the 1 second, size 18 ‘INCORRECT’
38
39 message. In both conditions, any other incorrect responses (random key presses or PM false
40
41 alarms) triggered the 1 second, size 18 ‘INCORRECT’ screen. Correct lexical decision responses
42
43 to PM trials (PM misses) did not trigger a feedback screen. The subsequent trial would begin
44
45
46
47
48
49
50
51

52
53 ¹ We ran an initial pilot, $N = 9$, with 5 second timeouts, and found that this failed to affect ongoing-task RTs and
54
55 accuracies. Thus, we extended the timeout to 15 seconds.
56
57
58
59
60

1
2
3
4
5 immediately (next fixation cross) after either feedback screen was displayed.
6

7 **Results**

8
9 An alpha level of 0.05 was used in all analyses. The first two trials after each rest period
10 (1.8% of trials) were excluded from analyses, as were trials where participants responded with a
11 key not corresponding to their PM or lexical decision task (0.02% of responses). The two trials
12 after each PM target and PM false alarm (12.9% of trials) were excluded, which is common
13 practice in PM studies, in order to avoid contamination from post-PM slowing (e.g., Meier &
14 Rey-Mermet, 2012). If participants submitted a discouraged lexical decision error (e.g. a ‘word’
15 response to a non-word in the Wc condition), they were presented a 15s feedback screen, during
16 which they might become distracted. Thus, we excluded any trials which immediately followed
17 this long timeout (2.7% of trials). Following these exclusions, we cut out any remaining trials
18 with outlying RTs ($< 0.2s$ or $> \text{mean RT} + 3 \times \text{interquartile range} / 1.349$, 4.82% of
19 remaining trials). From the original 2640 trials, this left on average 2086 trials remaining (range
20 2005-2152) for data analysis and PMDC modelling, corresponding to an average of 39 out of the
21 42 PM trials remaining for each PM stimulus type for each bias condition.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38

39 We conducted mixed effects model analyses using the R programming language (R Core
40 Team, 2019) and the ‘lme4’ package (Bates, Mächler, Bolker, & Walker, 2015). These models
41 included a random participant intercept term, but not random participant slopes. No other
42 random effects were included. To analyse accuracy, we fit generalized linear models to each
43 observed response with a probit link function. To analyse RT, we fit linear mixed effects models
44 to participant mean RTs. In addition to stimulus type (non-word, word) and bias condition (Nc,
45 Wc), the reported models included a day order factor (day one, day two) to capture effects of
46 task repetition. We fit models including all factors and interactions, testing for significance with
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 Wald chi-squared tests. The null model for the significance tests of each term included all other
6
7 terms, except for higher order interactions including the term (e.g., the interaction between
8
9 stimulus type and bias condition would be ignored when testing the main effect of stimulus
10
11 type). The outcomes of these tests are tabulated in the supplementary materials. In text, we report
12
13 descriptive statistics broken down by the factors that we found were significant, as well as
14
15 follow-up pair bed samples t-tests. The t-tests were calculated using participant mean RTs and
16
17 accuracies, averaged for each participant over any factors not relevant to the test. Effect sizes are
18
19 reported in terms of Cohen's d . We report within-subject standard errors calculated with the
20
21 Morey (2008) bias-corrected method.
22
23
24

25 **Lexical Decision Task**

26
27
28 We first assess whether our bias manipulation had the intended effect on ongoing task
29
30 accuracy. A bias against responding non-word (Nc condition) would be expected to improve
31
32 accuracy to word stimuli, and a bias against responding word (Wc condition) would be expected
33
34 to improve accuracy towards non-word stimuli. As presented in the first two rows of Table 1,
35
36 we found both these effects. Accuracy to non-word stimuli was higher in the Wc condition than
37
38 the Nc condition, $t(31) = 4.84$, $p < .001$, $d = 0.86$, and accuracy to word stimuli was higher in the
39
40 Nc condition than the Wc condition, $t(31) = 6.06$, $p < .001$, $d = 1.07$. This indicates that we
41
42 successfully manipulated ongoing-task bias. We also found that accuracy was marginally higher
43
44 on day one ($M = 91.8\%$, $SE = 1.3\%$) than day two ($M = 91.1\%$, $SE = 1.5\%$), $t(31) = 1.80$, $p = .08$,
45
46
47
48
49 $d = 0.32$.

50
51 --- Insert Table 1 about here ---
52
53
54
55

56 Bias would also be expected to affect RTs: a bias against responding non-word (Nc)
57
58
59
60

1
2
3
4
5 should result in slower non-word RTs, and a bias against responding word (Wc) should result in
6
7 slower word RTs. We found that correct RTs were slower to non-word stimuli in the Nc
8
9 condition than the Wc condition, $t(31) = 6.46, p < .001, d = 1.14$, and slower to word stimuli in
10
11 the Wc condition than the Nc condition, $t(31) = 4.37, p < .001, d = 0.77$, indicating that our bias
12
13 manipulation was successful. Correct RTs were also slower on day one ($M = 0.90s, SE = 0.02s$)
14
15 than on day two ($M = 0.83s, SE = 0.02s$), $t(31) = 5.13, p < .001, d = 0.91$. Due to our high trial
16
17 numbers, we also analyzed ongoing task error RTs. Error RTs were slower in the Nc ($M =$
18
19 $0.975s, SE = 0.028s$) condition than the Wc condition ($M = 0.934s, SE = 0.03s$), $t(31) = 2.26, p$
20
21 $= .03, d = 0.40$, and slower on day one ($M = 0.998s, SE = 0.03s$) than on day two ($M = 0.910s,$
22
23 $SE = 0.024s$), $t(31) = 6.29, p < .001, d = 1.11$. However, there was no interaction between
24
25 stimulus type and bias condition, and thus our bias manipulation appeared not to have a strong
26
27 effect on error RTs. This might owe to relatively poor measurement of error RTs – ongoing task
28
29 accuracy was quite high and so few error RTs were observed.

34 35 **Prospective Memory Task (Hits)**

36
37 PM response false alarms were rare, ranging from 0 to 1.3% of trials, and thus are not
38
39 analyzed further. PM responses were scored as correct (as PM ‘hits’) if the participant pressed
40
41 the PM key instead of a lexical decision key on the target trial. Delay theory predicts that a bias
42
43 against the ‘correct’ ongoing task response to a PM stimulus should improve PM accuracy. Thus,
44
45 PM accuracy would be expected to be higher for non-words in the Nc condition, and higher for
46
47 words in the Wc condition. However, as displayed in Table 1, we did not find that bias
48
49 substantially affected PM accuracy. No effects or interactions regarding the bias manipulation
50
51 reached significance in our model of PM accuracy. However, there was an effect of “day”. PM
52
53 accuracy was lower on day one ($M = 53.2%, SE = 4.7%$) than on day two ($M = 66.5%, SE =$
54
55
56
57
58
59
60

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 ($M = 1.010s$, $SE = 0.028s$) than on day two ($M = 0.893s$, $SE = 0.023s$), $t(31) = 4.33$, $p < .001$, $d = 0.77$.

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.

1
2
3
4
5 The previous analyses suggest that our bias manipulation successfully affected ongoing
6 task bias yet failed to affect PM accuracy. If this is the case, we would expect the type of PM
7 error made to change, with fewer ‘word’ responses submitted on PM trials in the Wc condition,
8 and fewer ‘non-word’ responses submitted on PM trials in the Nc condition. As displayed in
9 Table 2, we did find that there were fewer word responses on PM trials than non-word responses
10 in the Wc condition, $t(31) = 2.92, p < .01, d = 0.52$. In addition, there was a trend in the reverse
11 direction in the Nc condition; word responses were more common on PM trials than non-word
12 PM responses, $t(31) = 1.82, p = .08, d = 0.32$.

13
14
15
16
17
18
19
20
21
22
23 A successful manipulation of ongoing task bias would also be expected to influence RTs of
24 PM errors, with slower word RTs predicted in the Wc condition and slower non-word RTs in the
25 Nc condition. As displayed in Table 2, we found some evidence of such an effect. The effects of
26 the bias manipulation and PM stimulus type interacted. Non-word responses were slower than
27 word responses on PM trials for both Wc and Nc conditions. Non-word PM error responses were
28 slower in the Nc condition than the Wc condition, $t(31) = 2.53, p = .02, d = 0.45$. Word
29 responses to PM trials were numerically slower in the Wc condition than the Nc condition, but
30 this difference was not near statistical significance $t(31) = 1.19, p = 0.24, d = 0.21$.

31
32
33
34
35
36
37
38
39
40
41
42 --- *Insert Table 2 about here* ---
43
44
45

46 We now summarize our results thus far. Our analyses of ongoing task performance
47 indicate that the bias manipulation was effective in making the discouraged ongoing-task
48 response both slower and less common. Under delay theory, the effects of PM stimulus type and
49 the bias manipulation on PM accuracy were expected to interact, such that the Wc condition
50 would display higher PM accuracy to word PM trials and the Nc condition to non-word PM
51
52
53
54
55
56
57
58
59
60

1
2
3
4
5 trials. However, no evidence was found for such an effect. Rather than PM accuracy to a
6
7 stimulus (e.g., PM word stimulus) benefitting from bias against the matching ongoing-task
8
9 response (e.g., bias against ‘word’), the type of ongoing-task error observed on PM trials
10
11 changed (e.g., bias against responding word induced more non-word responses on word PM
12
13 trials). Taken together, these results amount to a failed prediction from the delay theory.
14
15

16
17 In the next section, we present a PMDC analysis of behavior in our experiment. This
18
19 analysis has several critical goals. One is to determine whether the PMDC model can provide an
20
21 appropriate fit to the effects of our bias manipulation, despite the prediction of delay theory
22
23 failing. Another is to examine the latent cognitive processes that were affected by our bias
24
25 manipulation. This includes verifying that the bias manipulation did successfully affect
26
27 thresholds and verifying that it did not substantially affect PM processes. A final goal was to
28
29 determine the scope of PMDC’s predictions, and whether the model would have been readily
30
31 compatible with a finding of a strong bias effect, if we had obtained one.
32
33

34 35 **Model Analysis**

36
37 The basic architecture of our model is depicted in Figure 1. We estimated thresholds in terms
38
39 of B , which is $b - A$. Model parameters could vary over stimulus type (word, non-word, PM
40
41 word, PM non-word), bias condition (N_c , W_c), day (one, two), and accumulator (word, non-
42
43 word, PM). To simplify the model, the start-point noise (A) and non-decision time (t_0)
44
45 parameters were fixed across all factors². We estimated one sv for the accumulator matching the
46
47
48
49

50
51 ² A reviewer suggested estimating a separate non-decision time for the second day of the experiment to account for
52
53 practice effects. We tried this model and did not find evidence for such an effect, and so we report the simpler model
54
55 in text. Summaries of the parameter estimates of the model with varying non-decision time over day can be found in
56
57
58
59
60

1
2
3
4
5 correct response on all trials (e.g., word accumulator on word trial, PM accumulator on PM trial).
6
7 The sv for the mismatching accumulators (e.g., non-word accumulator on word trial or PM trial)
8
9 was fixed at 1 as a scaling parameter. Thresholds could vary over day, condition and latent
10
11 accumulator, but were fixed across stimulus type, as is conventional. Although we have not
12
13 allowed mean evidence accumulation rates (v) to vary by day in our previous modelling (e.g.,
14
15 Strickland et al., 2018), we did in the current case to account for the observed practice effect on
16
17 PM performance. Due to a low number of PM false alarms, we only estimated one PM false
18
19 alarm accumulation rate across all design cells. With these constraints, the most flexible model
20
21 we fit had 56 free parameters: one A , one $t\theta$, one sv , twelve B s, and 41 v s.
22
23
24

25 **Sampling**

26
27 We applied Bayesian methods to estimate the posterior probability distribution of our model
28
29 parameters. Because we obtained over 2000 trials per participant, we were able to separately
30
31 estimate each individual participants' parameters. We could have also fit a hierarchical model,
32
33 including a population level distribution, but with our large data sets this posed computational
34
35 difficulties (i.e., model-fitting times of many weeks). Bayesian estimation requires specifying
36
37 prior beliefs about parameter values, in the form of prior distributions. Our priors are displayed
38
39 in Table 3. These relatively uninformative priors are the same as those in Strickland et al. (2018).
40
41
42 No parameters in the prior vary between the Wc and Nc conditions. Differences between
43
44 matching and mismatching accumulation rates are included in the prior, to accord with our
45
46 expectation that accuracy would be far higher than chance. The prior for the PM false alarm
47
48 mean accumulation rate is set low, to encode our expectation that PM false alarms would be very
49
50
51
52

53
54
55 the supplementary materials.
56
57
58
59
60

1
2
3
4
5 rare in our study, as they generally are. One notable prior setting is the non-decision time lower
6
7 bound of 0.1 seconds and upper bound of 1 second. The lower bound was chosen to avoid non-
8
9 decision time estimates that are an implausibly low amount of time to include both stimulus
10
11 encoding and response execution, and the upper bound for implausibly high values. Generally,
12
13 our choice of priors had little influence on the resulting posterior parameter estimates.
14
15

16
17 --- Insert Table 3 about here ---
18
19

20
21 Our posterior sampling was performed using the Dynamic Models of Choice R suite
22
23 (Heathcote et al., 2019). We applied the differential evolution Markov Chain Monte Carlo
24
25 algorithm (Turner, Sederberg, Brown, & Steyvers, 2013), an effective technique for sampling
26
27 evidence accumulation model parameters. For each sampled model, we ran three times as many
28
29 chains as there were parameters (e.g., for the most flexible model with 56 parameters we ran 168
30
31 chains). To reduce memory requirements, posterior samples were ‘thinned’ such that we only
32
33 retained every 20th sample. We obtained 180 total posterior samples (corresponding to 3600
34
35 iterations). We ran posterior sampling until the samples were adequately stable, and the chains
36
37 converged and mixed. This was confirmed with visual inspection and Gelman’s multivariate
38
39 potential scale reduction factor (<1.1, Gelman et al., 2013)
40
41
42

43 **Model Assessment**

44
45
46 To determine whether constraining our model further was justified, we compared the most
47
48 flexible model (the ‘top’ model) to simpler models using the deviance information criterion
49
50 (DIC; Spiegelhalter, Best, Carlin, & Van Der Linde, 2002). DIC measures model fit while
51
52 punishing model complexity, with lower DIC values indicating a better model. DIC values were
53
54 summed across participants. Various guidelines exist to determine how large of a DIC difference
55
56
57
58
59
60

1
2
3
4
5 is substantial. One interpretation of DIC difference is in terms of model weights, corresponding
6
7 to the probability that the selected model is the best model, analogous to Akaike Information
8
9 Criterion weights (e.g., see Wagenmakers & Farrell, 2004). For a set of two models, a DIC
10
11 difference of greater than 10 corresponds to over a 99% probability that the selected model is the
12
13 best. All DIC differences discussed below all far exceed this, suggesting strong support for the
14
15 selected model in each comparison. The top model, with 56 parameters, had a DIC value of
16
17 13125. We compared the top model with one that fixed thresholds over bias conditions (50
18
19 parameters). We found that the fixed threshold model had a substantially larger DIC than the top
20
21 model (13634), suggesting that shifts in thresholds were necessary to account for the
22
23 manipulation of bias. A model that fixed accumulation rates over bias conditions (36 parameters)
24
25 also had a much larger DIC value (13666) than the top model, suggesting that shifts in
26
27 accumulation rates were also necessary to account for the effects of the bias manipulation. We
28
29 also attempted fixing accumulation rates over the ‘day’ factor, to test whether our choice to allow
30
31 accumulation rates to vary over day was justified. The model with fixed accumulation rates had a
32
33 substantially larger DIC (13704) than the top model, suggesting that varying rates over day was
34
35 necessary.
36
37
38
39
40

41 We now examine the fit of the selected ‘top’ model. To obtain posterior predictive model
42
43 fit, we simulated data for each participant from each of the observed posterior samples. As
44
45 displayed in Figure 3, PMDC fit the observed non-PM trial performance well, including the
46
47 effects of our bias manipulation. Furthermore, as demonstrated in Figures 4 and 5, the model was
48
49 able to accurately fit the effects of our bias manipulation on PM trial performance, including the
50
51 shift in PM error type. Figure 4 depicts some minor miss-fit to the rate of ongoing task errors on
52
53 PM trials, with the model slightly over-estimating the frequency of such errors. Such miss-fit
54
55
56
57
58
59
60

1
2
3
4
5 may indicate minor discrepancies between the relationship between RT and choice specified by
6
7 our model, and that observed in the data. However, the miss-fit is relatively small, and the model
8
9 captures the effect of the bias manipulation on these error rates well. Thus, overall, despite the
10
11 observed effects of ongoing task bias being unanticipated by delay theory, they fit well with
12
13 PMDC. As the model provided a good account of the observed trends in our data, we proceed to
14
15 explore the model mechanisms responsible for the fit.
16
17

18 **Model Mechanisms**

19
20
21 In this section, we review how our selected model accounted for observed performance,
22
23 with a focus on the manipulation of ongoing-task bias. To summarize model parameters across
24
25 the group, we averaged each posterior sample across participants. The resulting participants-
26
27 averaged distribution was used for data summaries and posterior inference. Throughout this
28
29 section, we report the posterior means (M) and standard deviations (SD) of this group-averaged
30
31 distribution. The posterior mean of the participants-averaged non-decision times (t_0) was 0.14s
32
33 ($SD = .003s$). The posterior mean of the start-point noise parameter (A) was 0.44 ($SD = 0.02$),
34
35 and the mean of the standard deviation of match accumulation rates (sv) was 0.59 ($SD = .006$). In
36
37 the sections below, we discuss the estimates of threshold and mean accumulation rate
38
39 parameters. To test for differences across conditions, we constructed difference distributions by
40
41 calculating the difference between the parameters for every posterior sample. To summarize
42
43 these distributions, we report a Z score (mean/ SD of the difference distribution) and one tailed
44
45 posterior p value, with lower p values indicating a more substantial probability of a difference
46
47 between parameters. For the latter we report $\min(p, 1-p)$, corresponding to the lowest probability
48
49 of a difference in either direction.
50
51
52
53
54

55 --- Insert Figure 3 about here ---
56
57
58
59
60

1
2
3
4
5 --- Insert Figure 4 about here ---
6

7 --- Insert Figure 5 about here ---
8

9 **Thresholds.** Thresholds are plotted in Figure 6. Overall, there was a large bias against non-
10 word responding, with thresholds much higher to respond non-word than to respond word, $Z =$
11 $17.88, p < .001$. Determining that thresholds were affected by our bias manipulation is a critical
12 manipulation check. Consistent with our manipulation affecting threshold bias, non-word
13 thresholds were higher in the Nc condition (day one $M = 1.34, SD = 0.02$; day two $M = 1.22, SD$
14 $= 0.02$) than the Wc condition (day one $M = 1.27, SD = 0.02$; day two $M = 1.13, SD = 0.02$), $Z =$
15 $8.05, p < .001$. Similarly, word thresholds were higher in the Wc condition (day one $M = 1.19, SD$
16 $= 0.02$; day two $M = 1.13, SD = 0.02$) than the Nc condition (day one $M = 1.10, SD = 0.02$; day
17 two $M = 1.04, SD = 0.02$), $Z = 10.3, p < .001$. Thus, the model indicates that our manipulation
18 was successful in inducing a threshold bias in the expected directions (higher word thresholds in
19 Wc blocks, higher non-word thresholds in Nc blocks). Consistent with our manipulation
20 selectively affecting ongoing-task processes, PM thresholds in the Nc condition were not
21 substantially different from the Wc condition, $Z = 0.67, p = .252$. On day one, PM thresholds
22 were numerically, but not substantially, higher in the Nc condition ($M = 1.76, SD = 0.05$) than
23 the Wc condition ($M = 1.72, SD = 0.05$), $Z = 0.58, p = 0.28$, whereas on day two PM thresholds
24 were higher in the Wc condition ($M = 1.64, SD = 0.04$) than the Nc condition ($M = 1.55, SD =$
25 0.04), $Z = 1.7, p = 0.045$.
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47

48 --- Insert Figure 6 about here ---
49
50
51
52

53 **Accumulation rates (non-PM trials).** Although thresholds are the parameter traditionally
54 associated with bias, it is possible that bias also affects the speed of evidence accumulation.
55
56
57
58
59
60

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 particularly 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

1
2
3
4
5 towards the PM response on PM non-word trials was not substantially different in the Nc
6
7 condition (day one $M = 1.87$, $SD = 0.07$; day two $M = 2.06$, $SD = 0.07$) and Wc conditions (day
8
9 one $M = 1.84$, $SD = 0.09$; day two $M = 2.20$, $SD = 0.08$), $Z = -0.76$, $p = .22$. Similarly, PM
10
11 accumulation towards PM words was not substantially different in the Nc condition (day one M
12
13 $= 1.97$, $SD = 0.08$; day two $M = 2.20$, $SD = 0.07$) and the Wc condition (day one $M = 1.88$, $SD =$
14
15 0.08 ; day two $M = 2.36$, $SD = 0.06$), $Z = -0.54$, $p = .30$. However, PM accumulation rates did
16
17 increase from day one to day two for both PM word trials, $Z = 4.90$, $p < .001$, and PM non-word
18
19 trials, $Z = 3.64$, $p < .001$. This increase in PM accumulation is consistent with practice improving
20
21 performance on the PM task. Consistent with reactive inhibition, all ongoing-task accumulation
22
23 rates were far lower on PM trials than on non-PM trials (see contrasts in Table 4). These reactive
24
25 inhibitory control effects were very large, replicating previous findings (Strickland et al. 2018).
26
27
28
29

30 --- Insert Figure 7 about here ---

31 --- Insert Figure 8 about here ---

32 --- Insert Table 4 about here ---

33
34
35
36
37
38 ***Accumulation rates (PM trials). Bias manipulation.*** The bias manipulation was
39
40 expected to affect ongoing task accumulation rates on PM trials similarly to how it affected non-
41
42 PM trials accumulation rates, and it did. Non-word accumulation was substantially slower to PM
43
44 non-words in the Nc condition (day one $M = 1.00$, $SD = 0.06$; day two $M = 0.58$, $SD = 0.08$)
45
46 than in the Wc condition (day one $M = 1.24$, $SD = 0.06$; day two $M = 0.80$, $SD = 0.07$), $Z = 3.49$,
47
48 $p < .001$. For PM word trials, there was a trend towards non-word accumulation being slower for
49
50 the Nc condition (day one $M = -1.21$, $SD = 0.16$; day two $M = -1.29$, $SD = 0.17$) than the Wc
51
52 condition (day one $M = -0.96$, $SD = 0.14$; day two $M = -1.08$, $SD = 0.15$), $Z = 1.5$, $p = 0.07$.
53
54
55
56
57
58
59
60

1
2
3
4
5 Word accumulation was not substantially slower to PM words in the Wc condition (day one $M =$
6 0.90 , $SD = 0.07$; day two $M = 0.82$, $SD = 0.07$) than the Nc condition (day one $M = 0.98$, $SD =$
7 0.06 ; day two $M = 0.61$, $SD = 0.08$), $Z = -0.97$, $p = .17$. However, it was slower towards PM
8
9 non-words in the Wc condition (day one $M = -1.43$, $SD = 0.16$; day two $M = -1.54$, $SD = 0.17$)
10
11 than the Nc condition (day one $M = -1.25$, $SD = 0.14$; day two $M = -1.09$, $SD = 0.14$), $Z = -2.07$,
12
13
14
15
16 $p = .02$.

17 18 19 **Posterior Exploration**

20
21 Simulations from PMDC can be useful to understand why the model fitted the observed
22
23 data, and to explore other types of data that the model could potentially fit. In the supplementary
24
25 materials, we report simulations that break down in detail exactly how the bias manipulation
26
27 affected various aspects of ongoing task performance. Here, we report a simulation that answers
28
29 one particularly pertinent question: whether there are conditions under which PMDC *would*
30
31 predict that ongoing-task bias substantially affects PM accuracy. To do so, we examine predicted
32
33 effects of bias on PM accuracy caused by manipulating some parameters in our model while
34
35 maintaining others at their values estimated from the current study.
36
37
38

39
40 To calculate an overall measure of the effect of bias condition on PM accuracy, we
41
42 summed the increase in PM accuracy to PM non-words in the Nc condition with the increase in
43
44 PM accuracy to PM words in the Wc condition. We plot detailed posterior predictions of this
45
46 measure in the supplementary materials and summarize here by discussing the posterior mean
47
48 predictions. Bias condition did not induce a substantial shift in PM accuracy across PM word and
49
50 non-word targets in the data (total summed PM accuracy shift = 0.03), and this lack of effect was
51
52 fit closely by the earlier presented full model (0.025). However, we identified two ways in which
53
54 our model *could* simulate some degree of bias-induced shift in PM accuracy. For one,
55
56
57
58
59
60

1
2
3
4
5 substantially amplifying the effect that the bias manipulation had on thresholds could produce a
6
7 PM accuracy shift. However, this method was very inefficient in improving PM accuracy. For
8
9 example, when we more than doubled the observed mean bias effects, by adding 0.1 to the word
10
11 threshold in the Wc condition and 0.1 to the non-word threshold in the Nc condition, the
12
13 predicted bias effect on PM accuracy (i.e., the sum of PM word accuracy advantage in Wc over
14
15 Nc and PM non-word advantage in Nc over Wc) was only 0.07 (as compared with 0.03 in the
16
17 data). Such large shifts in bias could slow performance and potentially impose ongoing-task
18
19 accuracy costs, making it unlikely that participants would be prepared to implement them.
20
21
22

23 Another way the model could predict a bias benefit to PM accuracy was by reducing the
24
25 variability in rates effectively to zero. This makes it virtually impossible for the mismatching
26
27 ongoing task accumulator (e.g., the non-word accumulator on a PM word trial) to draw an
28
29 accumulation rate that allows it to compete with the PM accumulator. Reducing rate variability
30
31 this way led to our model predicting a summed bias benefit to PM accuracy of around 0.06.
32
33 However, as rate variability was set to implausible levels, it is unlikely such an effect could be
34
35 induced experimentally. Still, this result is informative about mechanisms in the full model, as it
36
37 indicates that rate variability reduces the effects of bias on PM accuracy. It appears that rate
38
39 variability allows the ongoing-task mismatch accumulator to become competitive with the PM
40
41 accumulator on some trials, and so the mismatch accumulator can pre-empt PM when favored by
42
43 bias.
44
45
46
47

48 In summary, simulations from PMDC predicted that inducing a bias benefit to PM
49
50 accuracy would require inflating the bias effects on the ongoing task to more than double that
51
52 observed in the data, or removing rate variability from the model, both of which seem unlikely to
53
54 occur in practice.
55
56
57
58
59
60

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

1
2
3
4
5 thresholds, it also successfully accounted for the fact that the bias manipulation did not affect PM
6
7 accuracy. This is in part because the bias manipulation allowed the ‘mismatching’ lexical-
8
9 decision accumulator to compete with the PM accumulator on some PM trials (e.g., the non-
10
11 word accumulator became more likely to reach threshold on PM word trials), reducing any
12
13 potential benefits of delaying the matching accumulator. Simulations suggested that threshold
14
15 bias could conceivably affect PM if the effect on threshold was much larger than observed.
16
17 However, to get even a small PM benefit to bias, we simulated bias effects twice as large as
18
19 those observed, which could interfere with performance more generally (e.g., by slowing
20
21 ongoing-task RTs and reducing accuracy). Thus, it seems that participants would be unlikely to
22
23 implement ongoing-task bias increases that could effectively support PM. Our finding that
24
25 ongoing-task bias was not effective in improving PM may shed some light on previous findings
26
27 regarding “stimulus-specific” PM instructions. Under such instructions, in which participants are
28
29 informed that PM targets only appear in one type of ongoing-task item, Heathcote et al. (2015)
30
31 argued in their delay theory that PM could be supported by selectively raising the ongoing-task
32
33 threshold corresponding to a matching response to that item. However, the current findings
34
35 suggest that such selective biases may not be as effective as they anticipated. This could explain
36
37 the findings of Horn and Bayen (2015), and Strickland et al., (2017, 2018) that participants
38
39 increased both ongoing-task thresholds with PM, even when explicitly informed that their PM
40
41 task was stimulus-specific.
42
43
44
45
46
47

48
49 Our finding that ongoing-task bias does not affect PM accuracy is inconsistent with
50
51 Heathcote et al.’s (2015) prediction based on Delay Theory, underscoring the difficulty with
52
53 anticipating exactly how control processes will influence the output of complex cognitive
54
55 models. They reasoned that bias might improve PM performance because delaying the ongoing-
56
57
58
59
60

1
2
3
4
5 task accumulator that is most competitive with PM should allow the PM accumulator more time
6
7 to accumulate. However, our simulations from PMDC predict that effects of ongoing-task bias
8
9 on PM are not substantial, in part because the beneficial effects of delays in the matching
10
11 accumulator are offset by the mismatching accumulator competing with the PM accumulator.
12
13 This was not anticipated in Heathcote et al.'s work that focused solely on PM cost because,
14
15 without a full model of both ongoing-task and PM task decision processes, it was not possible to
16
17 directly examine how cost-related mechanisms impacted PM performance. This illustrates how
18
19 researcher's intuitions cannot be assumed to accord with the function of a cognitive model
20
21 (Farrell & Lewandowsky, 2010), highlighting the importance of directly simulating quantitative
22
23 model predictions where possible. In a similar vein, Anderson et al. (2018)'s finding that
24
25 increased ongoing-task thresholds did not improve PM accuracy was potentially undermined by
26
27 the fact that they did not measure, or account for, shifts in PM accumulator control processes
28
29 such as PM thresholds.
30
31
32
33

34
35 The current results, and those of Anderson et al.'s (2018) , indicate that ongoing-task
36
37 thresholds play a relatively minor role in supporting PM accuracy, with the potential to be
38
39 overpowered or nullified by other PM decision processes. Taken together, these findings
40
41 illustrate that delay theory, at least when interpreted in isolation, fails to accurately predict PM
42
43 accuracy. However, ongoing-task thresholds are by far the largest underlying component of PM
44
45 cost in a range of studies. Thus, delay theory provides a compelling account of PM costs, but a
46
47 poor account of PM accuracy, illustrating an important distinction between these measures.
48
49

50
51 Given that increased ongoing-task thresholds underlie a large part of PM cost, and are
52
53 largely comprised of threshold elevations that appear to have little effect on PM accuracy, we
54
55 believe the PM cost measure deserves less focus than it has had in the literature. In the past,
56
57
58
59
60

1
2
3
4
5 influential PM theories such as the Preparatory Attentional and Memory processes theory
6
7 (Smith, 2003), and the Multi-process view (Einstein et al. 2005), have used costs to infer the PM
8
9 processes responsible for variation in PM accuracy. For example, they assume that non-focal PM
10
11 accuracy is poorer than focal PM accuracy because it is reliant on PM monitoring that shares
12
13 capacity with the ongoing task, whereas focal PM is less reliant (and in the case of the Multi-
14
15 process view, is fully reliant on spontaneous retrieval), and that PM importance emphasis causes
16
17 more capacity to be allocated to the PM task, resulting in increased cost and increased PM
18
19 accuracy (Einstein et al., 2005; Smith & Bayen, 2004). Whether one adopts our theoretical
20
21 position that costs reflect ongoing-task thresholds, or other theoretical positions that costs reflect
22
23 monitoring, our modeling results indicate neither cognitive process is the primary determinant of
24
25 PM performance. We acknowledge that these conclusions have far reaching implications for the
26
27 way PM research is done, and further research is required to verify them. It will be important for
28
29 future research to test whether our findings hold for a broader range of bias manipulations and
30
31 task implementations. For example, in the current study, our instructions' emphasis on the
32
33 ongoing task, and associated delays punishing ongoing task errors, may have led to the
34
35 perception that the PM task is of secondary importance. This could have reduced the attention or
36
37 effort that some participants paid to the PM task. However, as both our experimental conditions
38
39 include a bias instruction, any effects of bias on PM importance would be expected to be equal
40
41 across our conditions and thus there is no reason to expect this would confound any of our
42
43 comparisons. Still, it would be interesting to examine how the findings apply at different levels
44
45 of PM importance.
46
47
48
49
50
51

52
53 It is worth noting that although we analyze the results of only one experiment here, the
54
55 quality of measurement was high, with over 2000 trials modelled for each participant. Hence, our
56
57
58
59
60

1
2
3
4
5 single experiment with 32 participants yields more data than 160 participants would in a typical
6
7 PM experiment (assuming 400 trials per participant), and substantially more data than previous
8
9 large-scale studies on PM cost (e.g., Anderson et al., 2018). Although PM trials comprise only a
10
11 fraction of this data, PM trials were a larger proportion of total trials in this study than in most
12
13 previous studies, and so the number of PM trials we observed is also substantially larger than is
14
15 typical. This focus on trial numbers is necessary for process modelling, where reliable inference
16
17 depends on the number of trials per participant, rather than the number of participants (Kolossa
18
19 & Kopp, 2018), and has been argued to underpin the most reproducible findings in psychology
20
21 (Smith & Little, 2018). Given this, we believe it worth considering that rather than relying on
22
23 PM cost for inferences about PM processes, it is much more effective to use a cognitive model
24
25 like PMDC that directly measures the psychological processes underlying PM performance
26
27 (Strickland et al. 2018). In this vein, we have recently used the PMDC framework to develop and
28
29 test a detailed theory of how PM and ongoing processes can share capacity in cognitively
30
31 demanding, complex tasks such as air traffic control (Boag et al., 2019) and maritime
32
33 surveillance (Strickland et al., 2019). One again, these experiments have an order of magnitude
34
35 more data per participant than typical PM experiments, and in the case of Boag et al.'s (2019)
36
37 experiment, 246 participants.
38
39
40
41
42

43
44 Although increases in ongoing-task thresholds do not effectively support PM in the
45
46 standard laboratory paradigm we used here, they may do so in different paradigms. For example,
47
48 we have found that proactive control over ongoing-task thresholds more substantially supports
49
50 PM accuracy in simulated air traffic control (e.g., Boag et al., 2019), perhaps owing to the longer
51
52 time scales of the simulated air traffic control decisions. In addition, even in standard laboratory
53
54 paradigms, some methods of delaying the ongoing-task can improve PM performance. For
55
56
57
58
59
60

1
2
3
4
5 example, Loft and Remington (2013) found that preventing participants from submitting
6
7 responses for around 1 second could bring PM accuracy to almost ceiling. However, this
8
9 manipulation was much stronger than any delay likely to be imposed by threshold control - in the
10
11 current paradigm, a 1 second delay would more than double RTs. It is also possible that the delay
12
13 theory mechanism could potentially play a secondary, minor role in supporting PM accuracy
14
15 when complemented by decreased PM thresholds and increases in reactive control, as was shown
16
17 in Strickland et al. (2018). With PM importance emphasis, increases in ongoing-task thresholds,
18
19 decreases in PM thresholds and increases in PM-induced reactive inhibitory control acted
20
21 together to enhance PM performance. The major difference between Strickland et al. and the
22
23 current study is that the former manipulated participants' motivation towards PM and the
24
25 ongoing-task directly, allowing them to adjust their cognitive control processes as they saw fit to
26
27 achieve desired outcomes. By contrast, in the current study we attempted a selective
28
29 manipulation of ongoing-task thresholds. Although our manipulation also affected ongoing-task
30
31 rates it did not affect parameters associated with the PM process, PM thresholds and
32
33 accumulation rates associated with reactive control. Thus, our results combined with those of
34
35 Strickland et al. suggest that it is the parameters associated with the PM process that should be
36
37 the target of manipulations that attempt to improve PM performance.
38
39
40
41
42
43

44 One interesting, but unanticipated, finding was that PM accuracy was substantially higher
45
46 on day two of the study than day one. This contrasts with Strickland et al. (2018), where PM
47
48 accuracy decreased on later days of the study. A difference between the current study and
49
50 previous studies was that here we re-used the same PM target letter strings (i.e., the letter strings
51
52 'tor' and 'ver') on days one and day two. Over time, practice at the PM task may have led
53
54 participants to develop familiarity with the PM letter strings, improving the PM-related evidence
55
56
57
58
59
60

1
2
3
4
5 extracted from the stimulus. Consistent with this explanation, our model indicated PM
6
7 accumulation rates were higher on day two than day one. The role of stimulus familiarity in PM
8
9 was also highlighted by a recent study that found differences between focal and non-focal PM
10
11 accuracy could be eliminated with repeated exposure to non-focal PM targets (Hicks, Franks, &
12
13 Spitler, 2017). The role of learning and familiarity in PM processes awaits further investigation.
14
15

16
17 In summary, the current study indicated that manipulating ongoing-task bias has little
18
19 effect on PM accuracy, because bias against one ongoing-task decision allows the other ongoing-
20
21 task decision to effectively pre-empt PM. It is theoretically possible that extreme ongoing-task
22
23 bias may affect PM accuracy, but the deleterious effect of such bias on ongoing-task
24
25 performance makes it impractical to implement. Further, the current results suggest that because
26
27 it is based on a comprehensive characterization, the PMDC model (Strickland et al. 2018), rather
28
29 than Delay Theory (Heathcote et al. 2015), should be used as a basis for understanding, and
30
31 making predictions about event-based PM. Similarly, but in terms of empirical measures, our
32
33 results indicate that ongoing-task RTs provide an incomplete and potentially misleading guide to
34
35 PM-related processing in general, and PM accuracy in particular.
36
37
38
39
40
41

42 **Supplementary Material**

43
44 The Supplementary Material is available at: qjep.sagepub.com
45
46
47

48 **References**

49
50
51 Anderson, F. T., Rummel, J., & McDaniel, M. A. (2018). Proceeding with care for successful
52
53 prospective memory: Do we delay ongoing responding or actively monitor for cues?
54
55 *Journal of Experimental Psychology: Learning, Memory, and Cognition.*
56
57
58
59
60

- 1
2
3
4
5 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models
6
7 Using lme4. *Journal of Statistical Software; Vol 1, Issue 1 (2015)*. Retrieved from
8
9 <https://www.jstatsoft.org/v067/i01>
10
11
12 Boag, R. J., Strickland, L., Heathcote, A., Neal, A., & Loft, S. (2019). Cognitive control and
13
14 capacity for prospective memory in complex dynamic environments. *Journal of*
15
16 *Experimental Psychology. General*. <https://doi.org/10.1037/xge0000599>
17
18
19 Boag, R. J., Strickland, L., Loft, S., & Heathcote, A. (in press). Strategic attention and decision
20
21 control support prospective memory in a complex dual-task environment. *Cognition*.
22
23
24 Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework.
25
26 *Trends in Cognitive Sciences, 16(2)*, 106–113.
27
28
29 Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time:
30
31 Linear ballistic accumulation. *Cognitive Psychology, 57(3)*, 153–178.
32
33
34 Bugg, J. M., McDaniel, M. A., & Einstein, G. O. (2013). Event-based prospective remembering:
35
36 An integration of prospective memory and cognitive control theories. In D. Reisberg
37
38 (Ed.), *The Oxford handbook of cognitive psychology* (pp. 267–283). Oxford University
39
40 Press.
41
42
43 Dennis, S. (1995). *The Sydney Morning Herald word database*.
44
45
46 Dismukes, R. K. (2012). Prospective memory in workplace and everyday situations. *Current*
47
48 *Directions in Psychological Science, 21(4)*, 215–220.
49
50 <https://doi.org/10.1177/0963721412447621>
51
52
53 Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of*
54
55 *Experimental Psychology: Learning, Memory, and Cognition, 16(4)*, 717–726.
56
57
58 Einstein, G. O., McDaniel, M. A., Thomas, R., Mayfield, S., Shank, H., Morrisette, N., &
59
60

- 1
2
3
4
5 Breneiser, J. (2005). Multiple processes in prospective memory retrieval: Factors
6
7 determining monitoring versus spontaneous retrieval. *Journal of Experimental*
8
9 *Psychology: General*, 134(3), 327–342.
- 10
11 Gelman, A., Carlin, J., Stern, H., Dunson, D., Vehtari, A., & Rubin, D. (2013). *Bayesian Data*
12
13 *Analysis* (3rd ed.). London, England: Taylor & Francis.
- 14
15
16 Heathcote, A., Lin, Y., Reynolds, A., Strickland, L., Gretton, M., & Matzke, D. (2019). Dynamic
17
18 models of choice. *Behavior Research Methods*, 51(2), 961–985.
- 19
20
21 Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A
22
23 delay theory of prospective memory costs. *Psychological Review*, 122(2), 376–410.
- 24
25
26 Hicks, J. L., Franks, B. A., & Spitler, S. N. (2017). Prior task experience and comparable
27
28 stimulus exposure nullify focal and nonfocal prospective memory retrieval differences.
29
30 *The Quarterly Journal of Experimental Psychology*, 70(10), 1997–2006.
- 31
32
33 Horn, S. S., & Bayen, U. J. (2015). Modeling criterion shifts and target checking in prospective
34
35 memory monitoring. *Journal of Experimental Psychology: Learning, Memory, and*
36
37 *Cognition*, 41(1), 95–117.
- 38
39
40 Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior*
41
42 *Research Methods*, 42(3), 627–633.
- 43
44
45 Kolossa, A., & Kopp, B. (2018). Data quality over data quantity in computational cognitive
46
47 neuroscience. *NeuroImage*, 172, 775–785.
- 48
49
50 Loft, S., Dismukes, R. K., & Grundgeiger, T. (2019). Prospective memory in safety-critical work
51
52 contexts. In J. Rummel & M. A. McDaniel (Eds.), *Prospective Memory* (pp. 170–185).
53
54 Routledge.
- 55
56
57 Loft, S., & Remington, R. W. (2013). Wait a second: Brief delays in responding reduce focality
58
59
60

- 1
2
3
4
5 effects in event-based prospective memory. *The Quarterly Journal of Experimental*
6
7 *Psychology*, 66(7), 1432–1447.
- 8
9 Lourenço, J. S., White, K., & Maylor, E. A. (2013). Target context specification can reduce costs
10
11 in nonfocal prospective memory. *Journal of Experimental Psychology: Learning,*
12
13 *Memory, and Cognition*, 39(6), 1757–1764.
- 14
15
16 Marsh, R. L., Hicks, J. L., Cook, G. I., Hansen, J. S., & Pallos, A. L. (2003). Interference to
17
18 ongoing activities covaries with the characteristics of an event-based intention. *Journal of*
19
20 *Experimental Psychology: Learning, Memory, and Cognition*, 29(5), 861–870.
- 21
22
23 Meier, B., & Rey-Mermet, A. (2012). Beyond monitoring: After-effects of responding to
24
25 prospective memory targets. *Consciousness and Cognition*, 21(4), 1644–1653.
- 26
27
28 Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau
29
30 (2005). *Tutorials in Quantitative Methods for Psychology*, 4(2), 61–64.
- 31
32
33 R Core Team. (2019). *R: A language and environment for statistical computing*. Retrieved from
34
35 <https://www.R-project.org/>
- 36
37 Raab, D. H. (1962). Statistical facilitation of simple reaction time. *Transactions of the New York*
38
39 *Academy of Sciences*, 24, 574–590.
- 40
41
42 Rothschild, J. M., Landrigan, C. P., Cronin, J. W., Kaushal, R., Lockley, S. W., Burdick, E., ...
43
44 Czeisler, C. A. (2005). The Critical Care Safety Study: The incidence and nature of
45
46 adverse events and serious medical errors in intensive care*. *Critical Care Medicine*,
47
48 33(8), 1694–1700.
- 49
50
51 Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design.
52
53 *Psychonomic Bulletin & Review*, 25(6), 2083–2101.
- 54
55
56 Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory:
57
58
59
60

- 1
2
3
4
5 Investigating the capacity demands of delayed intention performance. *Journal of*
6
7 *Experimental Psychology: Learning, Memory, and Cognition*, 29(3), 347–361.
8
9 Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory.
10
11 *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 756–777.
12
13 Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures
14
15 of model complexity and fit. *Journal of the Royal Statistical Society: Series b (Statistical*
16
17 *Methodology)*, 64(4), 583–639.
18
19 Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed–accuracy compromise:
20
21 Boundary optimality in the diffusion model. *Psychology and Aging*, 25(2), 377–390.
22
23 Strickland, L., Elliott, D., Wilson, M. D., Loft, S., Neal, A., & Heathcote, A. (2019). Prospective
24
25 memory in the red zone: Cognitive control and capacity sharing in a complex, multi-
26
27 stimulus task. *Journal of Experimental Psychology. Applied*.
28
29 <https://doi.org/10.1037/xap0000224>
30
31
32
33
34 Strickland, L., Heathcote, A., Remington, R. W., & Loft, S. (2017). Accumulating evidence
35
36 about what prospective memory costs actually reveal. *Journal of Experimental*
37
38 *Psychology: Learning, Memory, and Cognition*, 43(10), 1616–1629.
39
40 Strickland, L., Loft, S., Remington, R. W., & Heathcote, A. (2018). Racing to remember: A
41
42 theory of decision control in event-based prospective memory. *Psychological Review*,
43
44 125(6), 851–887.
45
46
47 Turner, B. M., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently
48
49 sampling from distributions with correlated dimensions. *Psychological Methods*, 18(3),
50
51 368–384.
52
53
54 Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights.
55
56
57
58
59
60

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.

Figure Captions

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.

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).

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

PROSPECTIVE MEMORY AND ONGOING-TASK BIAS

45

1
2
3
4
5 of RTs, the median RTs and the ten percent of slowest RTs. Nc refers to the condition with a bias
6
7 against non-word decisions, and Wc to bias against word decisions.

9
10 Figure 4. Model fits to PM trial response proportions. The white dots indicate the observed
11 data. The black dots indicate the posterior mean prediction of the model. The black bars indicate
12 the 95% credible intervals of the posterior predictions. For the purposes of these graphs, we
13 concatenated all participants' data into a single data frame, and posterior predictions for all
14 participants into a single data frame, and then calculated response proportions with these data
15 frames. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against
16 word decisions.

17
18
19
20
21
22
23
24
25 Figure 5. Model fits to PM trial response times. The white dots indicate the observed data.
26 The black dots indicate the mean prediction of the model. The black bars indicate the 95%
27 credible intervals of the posterior predictions. For the purposes of these graphs, we concatenated
28 all participants' data into a single data frame, and posterior predictions for all participants into a
29 single data frame, and then calculated summary statistics on these data frames. To demonstrate
30 fit to RTs, the graphs illustrate fits to three quantiles: the fastest ten percent of RTs, the median
31 RTs and the ten percent of slowest RTs. Nc refers to the condition with a bias against non-word
32 decisions, and Wc to bias against word decisions.

33
34
35
36
37
38
39
40
41
42
43
44 Figure 6. Participant-averaged thresholds. The symbols correspond to the posterior mean,
45 and the error bars correspond to plus and minus one standard deviation from the mean. Nc refers
46 to the condition with a bias against non-word decisions, and Wc to bias against word decisions.

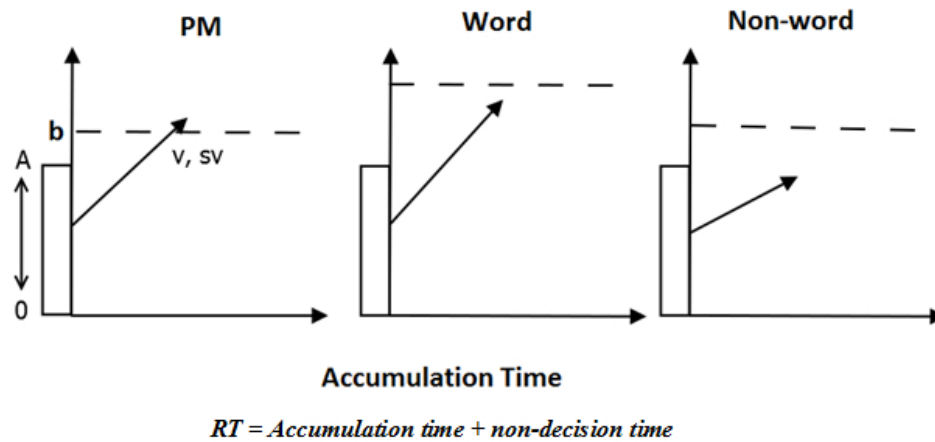
47
48
49
50
51 Figure 7. Participant-averaged non-PM trial accumulation rates for word and non-word
52 stimuli (columns). The symbols correspond to the posterior mean, and the error bars correspond
53 to plus and minus one standard deviation from the mean. Nc refers to the condition with a bias
54
55
56
57
58
59
60

PROSPECTIVE MEMORY AND ONGOING-TASK BIAS

46

1
2
3
4
5 against non-word decisions, and Wc to bias against word decisions. Accumulation to PM
6
7 decisions on non-PM trials ($M = -2.79$, $SD = 0.11$), that is PM false alarm accumulation, is not
8
9 plotted.

10
11 Figure 8. Participant-averaged PM trial accumulation rates. The symbols correspond to the
12
13 posterior mean, and the error bars correspond to plus and minus one standard deviation from the
14
15 mean. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against
16
17 word decisions.
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



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

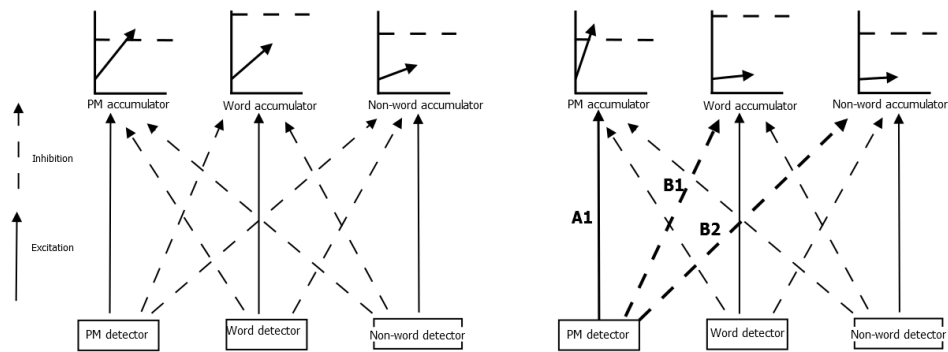


Figure 2. PMDC's reactive control (Strickland et al., 2018). The boxes represent stimulus input detectors.

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).

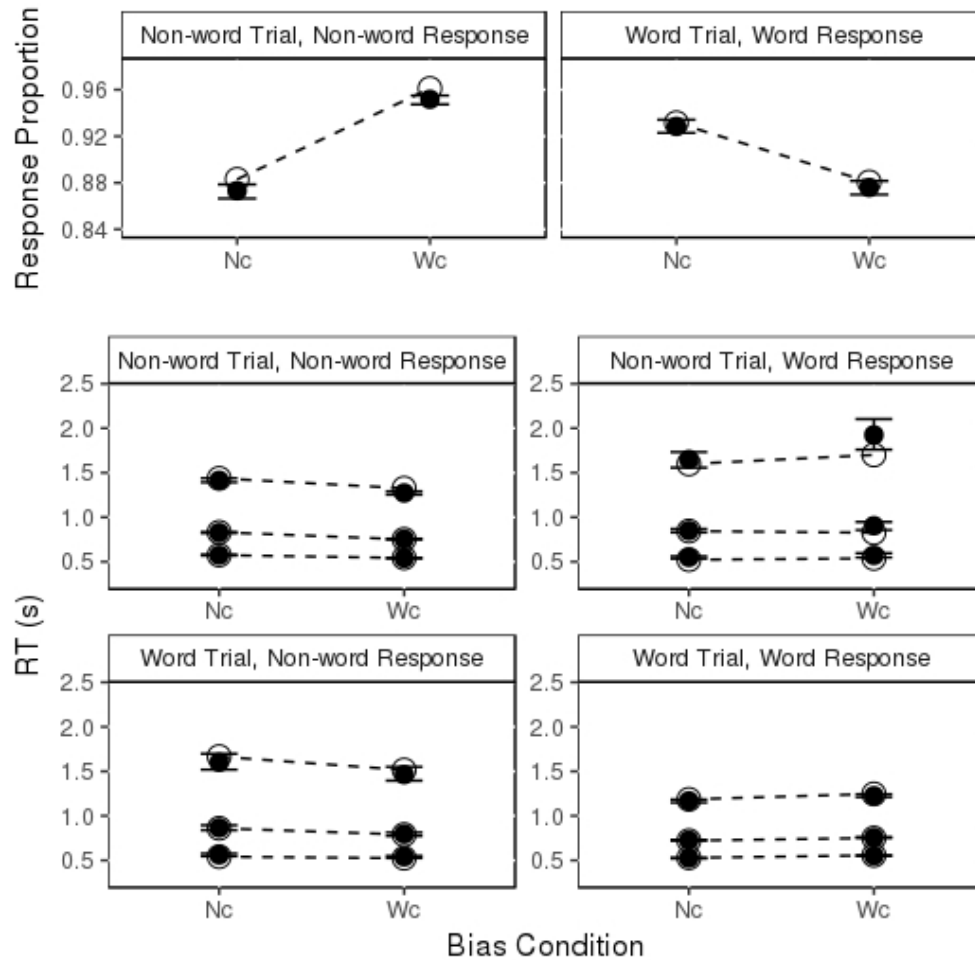


Figure 3. Model fits to the non-PM trial response data. The white dots indicate the observed data. The black dots indicate the mean prediction of the model. The black bars indicate the 95% credible intervals of the posterior predictions. For the purposes of these graphs, we concatenated all participants' 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.

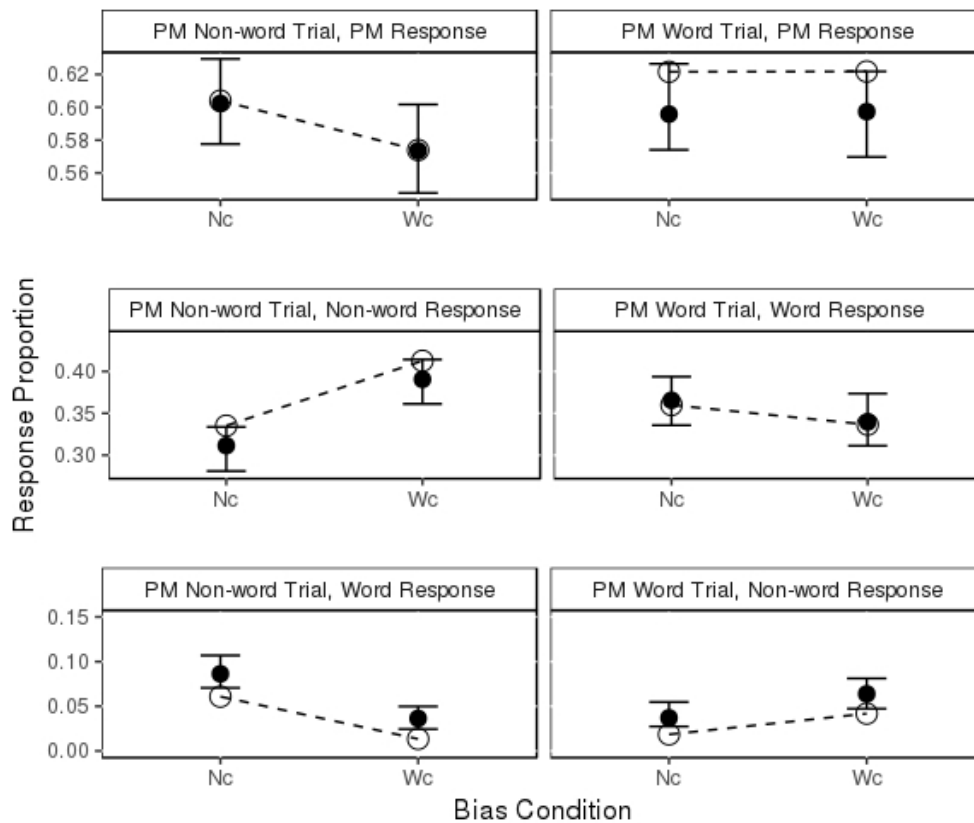


Figure 4. Model fits to PM trial response proportions. The white dots indicate the observed data. The black dots indicate the posterior mean prediction of the model. The black bars indicate the 95% credible intervals of the posterior predictions. For the purposes of these graphs, we concatenated all 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.

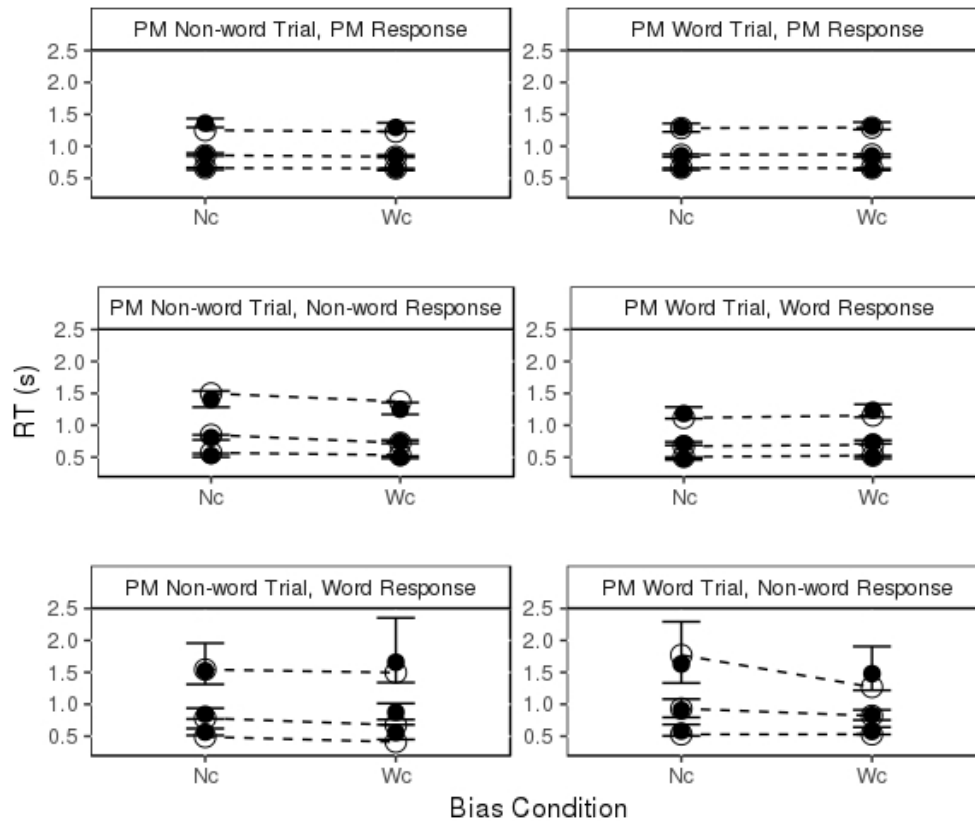


Figure 5. Model fits to PM trial response times. The white dots indicate the observed data. The black dots indicate the mean prediction of the model. The black bars indicate the 95% credible intervals of the posterior predictions. For the purposes of these graphs, we concatenated all participants' data 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.

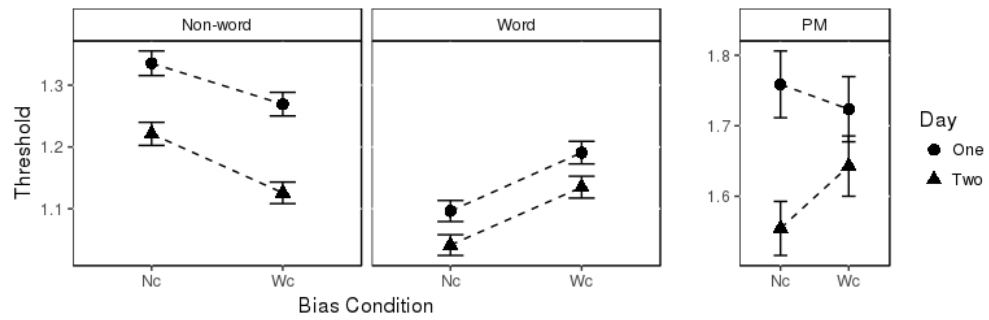


Figure 6. Participant-averaged thresholds. The symbols correspond to the posterior mean, and the error bars correspond to plus and minus one standard deviation from the mean. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against word decisions.

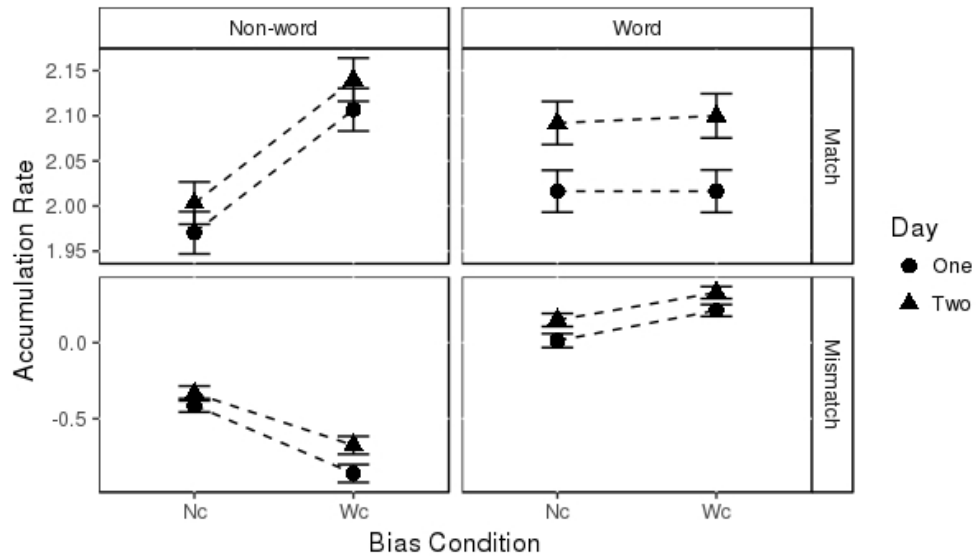


Figure 7. Participant-averaged non-PM trial accumulation rates for word and non-word stimuli (columns).

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.

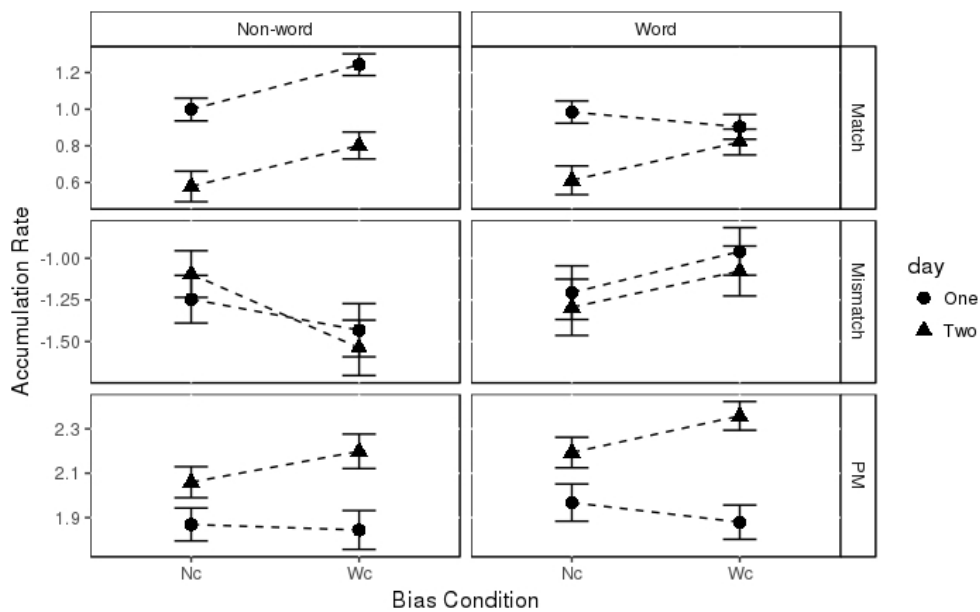


Figure 8. Participant-averaged PM trial accumulation rates. The symbols correspond to the posterior mean, and the error bars correspond to plus and minus one standard deviation from the mean. Nc refers to the condition with a bias against non-word decisions, and Wc to bias against word decisions.

Table 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.*

Stimulus Type	Wc		Nc	
	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)

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)

Table 3. *Prior distributions*

Model Parameter	<i>Distribution</i>	<i>Mean</i>	<i>SD</i>	<i>Lower</i>	<i>Upper</i>
<i>A</i>	Truncated Normal	1	1	0	10
<i>B</i>	Truncated Normal	1	1	0	None
ν (Correct Lexical Response)	Normal	1	2	None	None
ν (Incorrect Lexical Response)	Normal	0	2	None	None
ν (Correct PM Response)	Normal	1	2	None	None
ν (PM false alarm)	Normal	-1	2	None	None
s_ν	Truncated Normal	1	1	0	None
t_0	Uniform			0.1	1

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)