Original Articles

The influence of time on task on mind wandering and visual working memory

Marissa Krimsky, Daniel E. Forster, Maria M. Llabre, Amishi P. Jha

Department of Psychology, University of Miami, FL, USA

ARTICLE INFO

Keywords: Attention Mind wandering Working memory Time-on-task Executive resources Delayed recognition

ABSTRACT

Working memory relies on executive resources for successful task performance, with higher demands necessitating greater resource engagement. In addition to mnemonic demands, prior studies suggest that internal sources of distraction, such as mind wandering (i.e., having off-task thoughts) and greater time on task, may tax executive resources. Herein, the consequences of mnemonic demand, mind wandering, and time on task were investigated during a visual working memory task. Participants (N = 143) completed a delayed-recognition visual working memory task, with mnemonic load for visual objects manipulated across trials (1 item = low load; 2 items = high load) and subjective mind wandering assessed intermittently throughout the experiment using a self-report Likert-type scale (1 = on-task, 6 = off-task). Task performance (correct/incorrect response) and self-reported mind wandering data were evaluated by hierarchical linear modeling to track trial-by-trial fluctuations. Performance declined with greater time on task, and the rate of decline was steeper for high vs low load trials. Self-reported mind wandering increased over time, and significantly varied as a function of both load and time on task. Participants reported greater mind wandering at the beginning of the experiment for low vs. high load trials; however, with greater time on task, more mind wandering was reported during high vs. low load trials. These results suggest that the availability of executive resources in support of working memory maintenance processes fluctuates in a demand-sensitive manner with time on task, and may be commandeered by mind wandering.

1. Introduction

The capacity to use working memory, which is the ability to maintain and manipulate information over short intervals, can become derailed by task-unrelated thought, a phenomenon known as mind wandering (MW; Smallwood & Schooler, 2006). Although there is growing evidence that working memory and MW are related (e.g., Mrazek et al., 2012), their precise relationship is still poorly understood, thus limiting our ability to offer solutions for minimizing errors that may be driven by internally-generated distraction. One prominent model of MW, referred to as the executive-resource account (Smallwood & Schooler, 2006; see also Thomson, Smilek, & Besner, 2014), proposes that MW may compete with working memory processing demands for a limited pool of executive resources (Ram & Handy, 2014; Smallwood, Nind, & O’Connor, 2009; Teasdale et al., 1995).

A prediction of the executive-resource account is that the likelihood of MW’s occurrence will be tied to the resource requirements of the primary task at hand (Smallwood, McSpadden, & Schooler, 2007; Smallwood & Schooler, 2006). Support for this prediction comes from studies, such as Smallwood et al. (2009), in which less MW was reported by participants during a working memory task versus a choice reaction time task, two tasks differing in the amount of executive resources devoted to working memory processes. The working memory task required executive resources to be used in the service of encoding memoranda, as well as maintaining and updating information over short intervals, while the choice reaction time task did not (see also Smallwood & Schooler, 2015). In a related study by Forster and Lavie (2009), in which the level of demand was manipulated by varying perceptual load in a visual search task, greater MW was reported during low vs. high load trials; however, with greater time on task, more mind wandering was reported during high vs. low load trials. These results suggest that the availability of executive resources in support of working memory maintenance processes fluctuates in a demand-sensitive manner with time on task, and may be commandeered by mind wandering.

http://dx.doi.org/10.1016/j.cognition.2017.08.006

Received 24 January 2017; Received in revised form 14 August 2017; Accepted 16 August 2017

0010-0277/ © 2017 Elsevier B.V. All rights reserved.
encoding task that may be highly demanding when first encountered could become less demanding as it becomes familiar and automated. In task contexts in which such automation occurs, MW has been shown to increase with greater time on task (e.g., Smallwood et al., 2003; Mason et al., 2007). As such, when tasks require fewer executive resources for successful task performance, either due to low demand or practice-related automation, remaining resources may be commandeered by MW.

Yet, not all tasks are amenable to practice-related automation. As Smallwood et al. (2002) found, attention-demanding tasks fail to demonstrate improvements with greater time on task. In their study, MW was probed while participants performed a verbal fluency task, in which both task performance and self-reported MW remained stable despite block length. Furthermore, in continuous performance tasks emphasizing sustained attention, performance has been found to wane over time, a pattern referred to as the vigilance decrement phenomenon (see Mackworth, 1948). One prominent theoretical explanation for vigilance decrement is the resource-depletion hypothesis, in which greater time on task is proposed to deplete a limited pool of executive resources, resulting in fewer resources available to successfully perform the task (Caggiano & Parasuraman, 2004).

Recently, Thomson, Besner, and Smilek (2015) suggested that MW may play a role in vigilance decrements. According to their resource-control account, motivational factors may lead to a reduction in task engagement over time, causing executive resources to shift away from the task at hand toward MW. In line with this view, studies that used vigilance tasks in which MW was indexed, report that performance decreases and MW increases with greater time on task (Thomson, Sel, Besner, & Smilek, 2014; McVay & Kane, 2012; Cunningham, Scerbo, & Freeman, 2000). Prior studies have found that the rate of performance decline over time is greater in tasks with high vs. low demand (Helton & Russell, 2011; Smit, Eling, & Coenen, 2004). Thomson et al. (2015) hypothesized that if these decrements are driven by task disengagement in the service of MW, there should be greater MW over time for high vs. low demand tasks. Testing this hypothesis would require an experimental paradigm in which: (1) executive demands are varied over trials so that the effects of high vs. low demand on performance and MW can be evaluated; (2) MW is probed at regular intervals over the course of the experiment; and (3) performance degradation is observed with greater time-on task (i.e. vigilance decrement).

Motivated by Thomson et al. (2015), the current study employed a paradigm to satisfy all three of these requirements, in order to investigate this hypothesis in the context of a visual working memory task. Here, we assessed participant performance during a delayed-recognition visual working memory task, in which demand was manipulated by varying mnemonic load and MW was probed throughout the experiment. Although prior studies have investigated MW during complex span tasks of WM (Mracek et al., 2012), we used a visual delayed-recognition task to understand maintenance-related processes for visual information over short intervals (Ranganath, DeGutis, & D’Esposito, 2004; Luck & Vogel, 2013; Fuster & Bressler, 2012; D’esposito & Postle, 2015; see Cowan (2016) for review); this is in contrast to working memory span tasks, which have multiple demand-sensitive task components tied to verbal information (e.g., maintenance, task-switching, retrieval). Our manipulation of mnemonic demand, on the other hand, could be better constrained to maintenance processes, allowing us to examine the influence of time on task and MW on working memory. Our key question of interest was to determine if working memory task performance and MW fluctuate with greater time on task in a demand-sensitive manner. To answer this question, task performance and MW data were analyzed using hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002).

2. Methods

2.1. Participants

Undergraduate students (N = 143, females = 88, Mage = 19.09, SDage = 1.37) were recruited from the University of Miami psychology subject pool. Participants received course credit for their participation. All participants provided informed consent in accordance with the Institutional Review Board of the University of Miami.

Before beginning the working memory delayed-recognition task (described below), participants were instructed to emphasize the accuracy of their response over speed. Participants first received instructions for the working memory task and completed a practice session of 10 trials. Following this practice, participants received instructions about the MW probes and concrete examples of each question. Participants then practiced 10 trials of the working memory task with thought probes, as described below.

2.2. Procedure

To measure working memory across 2 levels of cognitive load, we used a modified version of the delayed-recognition task from Jha and Kiyonaga (2010). All presented stimuli were displayed as grayscale images, centrally located on the computer screen on a gray background. Each trial began with the presentation of a memory array, consisting of either two faces (high mnemonic load) or one face and a noise mask (low mnemonic load), appearing side by side for 3000 ms. The memory item was followed by a 3500 ms delay period with a fixation cross, after which a test item was presented centrally for 2500 ms (depicted in Fig. 1).

Two levels of load were selected (low vs. high) based on results of prior studies (Jha, Fabian, & Aguirre, 2004; Jha & McCarthy, 2000), which indicated that participants’ performance was significantly better for 1 face vs. 2 faces; and performance was at near-chance levels when participants were required to remember 3 faces. In addition, past studies have found larger differences in activation in prefrontal regions between one vs. two faces, but not two vs. three faces (Jha & McCarthy, 2000). Stimuli were presented using E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2001).

The inter-trial interval was 500 ms, for a total of 9500 ms per trial. For half of the trials, the test item was a single face from the memory item array (match trials), and on the remaining trials, the test item was a novel face that had not previously appeared in the experiment (non-match trials). Stimuli were randomized prior to the experiment but appeared in the same order for all participants. Participants were instructed to determine whether a test item matched a face in the memory item array by pressing match or non-match designated buttons. They were again instructed to emphasize the accuracy of their response over speed. The experiment included an equal number of trials for each level of mnemonic load (low or high) with a total of 102 memory item trials. Item trials were divided into three equally sized blocks of 34 trials each with three self-timed breaks. Accuracy was calculated based on correct responses to the match and non-match trials. Failures to respond were coded as incorrect.

MW was assessed using probe questions presented throughout the task. There were 47 instances of thought probes throughout this experiment, with 15–17 MW probes in each of the three blocks. The thought probes, which were counterbalanced to follow an approximately equal number of high and low load trials, were presented after the test item during the inter-trial-interval. Probes were dispersed pseudo-randomly throughout the task and occurred after every 1–4 working memory trials. There were four questions presented one at a time as part of the probe, but only the first question (“Where was your attention focused on average during the last trial?”) was used to probe MW and is considered herein. The response to this question was presented on a 6-point likert scale with 1 as indicating being ‘on-task’ to 6.
being ‘off-task’. (See Fig. 1 for depiction of the MW probe). The other three questions, while outside the scope of the current report, asked, “How aware were you of where your attention was?”, “How dull was your mind over the last trial?”, “How much was your mind racing over the last trial?”. The purpose of these questions was to assess the qualitative nature of MW.

2.3. Statistical analysis

To account for the repeated measures within subjects, we used a hierarchical linear modeling (HLM) approach using the ‘lme4’ package (Bates et al., 2015) in R version 3.3.1 (R Core Team, 2016) to test our a priori predictions regarding the impact of time on task and mnemonic load on task performance and MW. Since the task instructions emphasized that participants should aim to be accurate over being as fast as possible, accuracy (correct/incorrect) was analyzed as the key measure of task performance.

To determine whether a linear approach was appropriate, we plotted average accuracy for each trial, separately for high and low load trials (see Figs. 2 and 3). We found a small number of trials that appeared to deviate from the rest around the mid-point of the experiment for low load. Since the overall pattern appeared linear, we pursued a linear model, since a non-linear model would likely result in over-fitting.

Subjects completed 102 working memory trials, and were probed pseudo-randomly across 47 of those trials (i.e., probe trials were randomly assigned once prior to the study and ascribed to all participants). Trials included in the models were limited to the 47 working memory trials that preceded MW probes. For all analyses, load was effect-coded (low load = −1, high load = 1) and trial numbers were mean-centered, then divided by 50 (to help with convergence) (raw data range: 2–102; mean-centered and scaled range: −1.01 to 1.02). We mean-centered and effect coded to reduce collinearity between our main effect and interaction estimates.

Our first model (Model 1) served to verify that our delayed recognition task behaved in a manner similar to previous findings, with lower accuracy for high vs. low load (Jha & Kiyonaga, 2010) and performance decrements with greater time on task (Smit et al., 2004; Helton & Russell, 2011). Specifically, in Model 1, we examined whether accuracy varied by mnemonic load and whether the effect of load fluctuated across trials. We only examined trials corresponding to the 47 working memory trials to maintain consistency between our models.

In Model 2, we asked whether mind wandering varied by mnemonic load and whether the effect of load fluctuated across trials. Models 1 and 2 alone did not provide direct evidence that MW is associated with accuracy with time on task. Therefore, with Model 3 we examined whether accuracy varied by trial, MW, load, and the interaction of MW and load, and whether the effect of MW on accuracy fluctuated across trials. For all models, there were no predictor variables at Level 2; instead, Level 2 allowed us to account for subject level variance in Level 1.
coefficients. We define the parameters for Model 1 below. For models predicting accuracy, which is a dichotomous variable, we used hierarchical logistic regression (Model 1 and Model 3; see Raudenbush & Bryk, 2002). In the case when the outcome variable was continuous (Model 2), we used hierarchical linear regression. Coefficients are interpreted in terms of log odds of accuracy.

Model 1:

\[
\begin{align*}
WMAij &= \beta_0i + \beta_1i*(TRIALij) + \beta_2i*(LOADij) + \beta_3i*(TRIALij \times LOADij) + tij \\
\beta_1i &= \gamma_{00} + u_1i \\
\beta_2i &= \gamma_{20} + u_2i \\
\beta_3i &= \gamma_{30} + u_3i
\end{align*}
\]

where WMAij represents the accuracy of trial i for person j; \(\beta_0i\) represents the participant’s accuracy across trials and load; \(\beta_1i\) represents the difference in accuracy as trial numbers increase for person i averaged across loads, or the main effect of trial; \(\beta_2i\) represents the difference in accuracy for person j between the grand-mean and the mean for a given load, or the main effect for load; \(\beta_3i\), the interaction term, represents the difference in change in accuracy across trials between high and low loads; and \(tij\) is a residual term that reflects the difference between each participant’s observed and predicted accuracy.

In the Level-2 models, each participant’s intercept is modeled as the grand mean intercept (\(\gamma_{00}\), plus a residual term that reflects deviations in participants’ accuracy scores about the grand mean (\(u_0i\)). Each participant’s slope across trials (\(\beta_1i\)) is modeled as grand mean slope (\(\gamma_{20}\) plus a residual term that reflects individual participants’ slope differences about the grand mean (\(u_1i\)). The magnitude and statistical significance of the grand mean slope (\(\gamma_{20}\)) enabled us to evaluate whether accuracy varied as a function of trial number. The magnitude and statistical significance of the mean slope coefficient (\(\gamma_{30}\)) enabled us to evaluate whether accuracy varied as a function of demand while trial number increased.

Model 2 was similar to Model 1, but the outcome variable, which was self-reported MW, was a continuous variable.

Model 3 built on Model 1 and sought to explore the relationship between accuracy and MW with time on task. It included MW, load and the interaction of MW and load, and the interaction of MW and trial as predictors of accuracy.

Graphs for Models 1 and 2 were made using the ‘ggplot2’ package (Wickham, 2009) in R v. 3.3.1 (R Core Team, 2016). For coefficients that predict accuracy, log odds are reported for coefficients, as well as odds ratios (ORs) and their 95% confidence intervals. Coefficients from Model 1 are from the population-average model; coefficients from both models are reported with robust standard errors.

3. Results

Observed and model-predicted averages for MW and accuracy can be found in Table 1.

Results from the three models can be found in Tables 2–4.

Model 1. The results from Model 1 (accuracy predicted by trial, load, and their interaction) revealed a significant main effect of load, \(b = -0.341, SE = 0.04, p < 0.001, OR = 0.757\), which indicated that people were less likely to be accurate in high load trials than in any given trial, regardless of load. The main effect of trial was marginally significant, \(b = -0.174, SE = 0.089, p = 0.051\). However, there was also a significant interaction between load and trial, \(b = -0.212, SE = 0.071, p = 0.003, OR = 0.809\), which indicated that participants’ accuracy decreased across trials, more so for high load than low load trials (Fig. 2). It is worth noting that, especially for high load trials, observed accuracies appeared to deviate from model-predicted accuracies as Blocks 2 and 3 (see Table 1); suggesting that, after an initial decline in performance, performance may have increased somewhat, which is not reflected in our model. However, the general interpretation of the linear interaction is supported by the data, which shows that accuracy was generally lower for later than earlier high load trials, whereas accuracy was stable across time for low load trials.

Model 2. Results from Model 2 revealed a significant main effect of load on self-reported MW, \(b = 0.029, SE = 0.011, p = 0.008\), which indicated that participants reported more MW during high vs. low load trials. There was also a significant main effect of trial, \(b = 0.439, SE = 0.174, p = 0.008\), which indicated that participants reported more MW during high vs. low load trials. 95% confidence intervals are for the coefficient.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Observed and predicted averages of MW and accuracy by block.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mind wandering (M, SD)</td>
</tr>
<tr>
<td></td>
<td>Block 1</td>
</tr>
<tr>
<td>Low load</td>
<td>Observed</td>
</tr>
<tr>
<td>Predicted</td>
<td>1.72</td>
</tr>
<tr>
<td>High load</td>
<td>Observed</td>
</tr>
<tr>
<td>Predicted</td>
<td>1.55</td>
</tr>
</tbody>
</table>

Note: Observed averages and standard deviations were based on the 47 trials with MW probes (scale: 1–6); model-predicted averages were computed with the 47 trial numbers used in the model. Since variance for proportions is defined by the proportion itself, we did not report standard deviations for accuracy here.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Unstandardized coefficients for Model 1: Accuracy.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level 1</td>
</tr>
<tr>
<td>Intercept, (\beta_0)</td>
<td>2.408</td>
</tr>
<tr>
<td>Trial, (\beta_1)</td>
<td>-0.174</td>
</tr>
<tr>
<td>Load, (\beta_2)</td>
<td>-0.341</td>
</tr>
<tr>
<td>Trial (\times) Load, (\beta_3)</td>
<td>-0.212</td>
</tr>
</tbody>
</table>

Note: The effect for trial is on the scale of mean-centered trial number divided by 50 (range: \(-1.01\) to 1.02), which helped with convergence when trial was estimated with load. The effect of load represents the difference between average accuracy regardless of load and average accuracy on high load trials. 95% confidence intervals are for the coefficient.
Table 3
Unstandardized coefficients for Model 2: Mind Wandering.

<table>
<thead>
<tr>
<th>Level 1 Coefficient</th>
<th>SE</th>
<th>p</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, $\beta_0$</td>
<td>1.959</td>
<td>0.067</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Trial, $\beta_1$</td>
<td>0.439</td>
<td>0.048</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>Load, $\beta_2$</td>
<td>0.029</td>
<td>0.011</td>
<td>0.008</td>
</tr>
<tr>
<td>Trial × Load, $\beta_3$</td>
<td>0.114</td>
<td>0.018</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Note: The effect for trial is on the scale of mean-centered trial number divided by 50 (range: −1.01 to 1.02), which helped with convergence when trial was estimated with load. The effect of load represents the difference between average mind wandering regardless of load and average mind wandering on high load trials. 95% confidence intervals are for the coefficient.

Table 4
Unstandardized coefficients for Model 3: Accuracy predicted by MW.

<table>
<thead>
<tr>
<th>Level 1 Coefficient</th>
<th>SE</th>
<th>p</th>
<th>Odds ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, $\beta_0$</td>
<td>2.435</td>
<td>0.088</td>
<td>0.001</td>
<td>11.414</td>
</tr>
<tr>
<td>MW, $\beta_1$</td>
<td>−0.376</td>
<td>0.037</td>
<td>&lt; 0.001</td>
<td>0.687</td>
</tr>
<tr>
<td>Trial, $\beta_2$</td>
<td>−0.042</td>
<td>0.097</td>
<td>0.667</td>
<td>0.959</td>
</tr>
<tr>
<td>Load, $\beta_3$</td>
<td>−0.340</td>
<td>0.043</td>
<td>&lt; 0.001</td>
<td>0.712</td>
</tr>
<tr>
<td>MW × Trial, $\beta_4$</td>
<td>0.204</td>
<td>0.059</td>
<td>&lt; 0.001</td>
<td>1.226</td>
</tr>
<tr>
<td>Trial × Load, $\beta_5$</td>
<td>−0.150</td>
<td>0.074</td>
<td>&lt; 0.001</td>
<td>0.860</td>
</tr>
<tr>
<td>MW × Load, $\beta_6$</td>
<td>−0.038</td>
<td>0.030</td>
<td>0.215</td>
<td>0.963</td>
</tr>
</tbody>
</table>

Note: The effect for trial is on the scale of mean-centered trial number divided by 50 (range: −1.01 to 1.02), which helped with convergence when trial was estimated with load. MW was also centered. The effect for load represents the difference between average accuracy regardless of load and average accuracy on high load trials. 95% confidence intervals are for the Odds Ratio.

SE = 0.048, p < 0.001, which indicated that mind wandering generally increased across trials. However, these main effects were qualified by a significant interaction, b = 0.114, SE = 0.018, p < .001, which revealed that over the span of the experiment, MW increased at a greater rate for high load trials than low load trials (Fig. 3). The results of Models 1 and 2, together, inform us that load level and trial are jointly influencing accuracy and MW effects. Specifically, accuracy decreased and MW increased with more time on task, and these effects are load-sensitive.

Table 5
Percentages of accurate trials and percentages of endorsement for each scale of MW probes, separated by block, corresponding to Model 3: Accuracy predicted by MW and trial.

<table>
<thead>
<tr>
<th>MW probe</th>
<th>Block 1</th>
<th>Block 2</th>
<th>Block 3</th>
<th>Total average accuracy</th>
<th>Total endorsement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy</td>
<td>Endorsement</td>
<td>Accuracy</td>
<td>Endorsement</td>
<td>Accuracy</td>
</tr>
<tr>
<td>1</td>
<td>95</td>
<td>61</td>
<td>88</td>
<td>49</td>
<td>93</td>
</tr>
<tr>
<td>2</td>
<td>93</td>
<td>22</td>
<td>86</td>
<td>26</td>
<td>89</td>
</tr>
<tr>
<td>3</td>
<td>90</td>
<td>11</td>
<td>78</td>
<td>13</td>
<td>88</td>
</tr>
<tr>
<td>4</td>
<td>84</td>
<td>3</td>
<td>78</td>
<td>6</td>
<td>88</td>
</tr>
<tr>
<td>5</td>
<td>74</td>
<td>2</td>
<td>73</td>
<td>4</td>
<td>84</td>
</tr>
<tr>
<td>6</td>
<td>45</td>
<td>1</td>
<td>56</td>
<td>2</td>
<td>71</td>
</tr>
</tbody>
</table>

Note: Item trials were divided into three equally sized blocks of 34 trials each with three self-timed breaks. Trials included in the models were limited to the 47 working memory trials that preceded MW probes.
4. Discussion

Herein, we investigated the influence of time on task on MW and working memory maintenance processes. We employed a lengthy working memory delayed-recognition task in which maintenance demands were varied at two levels (high load = 2 faces; low load = 1 face) and MW was probed throughout the experiment. Our findings support a key prediction of the resource-control account, proposed by Thomson et al. (2015), that high demand (vs. low demand) results in greater increases in MW over time. Results from hierarchical linear modeling revealed that, similar to prior studies manipulating maintenance demands via memory load (see Jha & McCarthy, 2000; Jha & Kiyonaga, 2010), task accuracy was lower for high vs. low load trials. The present study also found that mind wandering predicted the change in accuracy, in-line with both the executive-resource account and the resource-control account. However, the relationship between MW and accuracy did not change as a function of whether subjects were completing a high vs low load trial.

A considerable limitation of the executive-resource account is the interpretation of why MW increases with time on task. Smallwood and Schooler (2006) posit that MW increases due to practice effects, suggesting that as people become more familiar with the task, fewer resources are required to perform the task and, therefore, more resources are available for MW. This interpretation is inconsistent with our results, in which we demonstrate that task performance decreases with time on task, as predicted by the resource-control account. In addition, task accuracy declined and MW increased with greater time on task, which is in line with results of prior vigilance decrement paradigms (Thomson et al., 2014). Interestingly, we found an interaction between MW and time on task, such that low levels of MW were associated with decreases in task performance over time, whereas high levels of MW were associated with increases in task performance over time. This interaction indicated that people, perhaps, were less accurate in their reporting of MW over time.

The rate of change as a function of time on task, in both accuracy and MW, analyzed separately, was greater for high vs. low load trials. An examination of the interaction between load and time on task revealed that, in line with prior research on MW (e.g., Forster & Lavie, 2009; Mazrek et al., 2012), greater MW was reported for low vs. high load trials for trials at the beginning of the experiment. In this case, it could be argued that more resources are available for MW in low levels of load. Yet, in contrast to prior research, the opposite pattern was found during the latter two-thirds of the experiment. Thus, for the final two-thirds of the experiment, the results were in line with the predictions of the resource-control account (Thomson et al., 2015), with greater MW reported for high vs. low load trials. Additionally, towards the end of the experiment the association between accuracy and MW weakened, with participants making more errors when still responding that they were on-task. These results suggest that the availability of executive resources, in support of successful working memory maintenance, fluctuates with time on task in a demand-sensitive manner, and that these resources may be commandeered by MW throughout the task. Specifically, time on task could be considered as an indicator of executive resource demand.

The findings of the present study demonstrate that the rate of MW varies depending on contextual factors of time on task and demand. MW has been shown to be context dependent (Smallwood & Andrews-Hanna, 2013); yet, time on task as part of the parameter space for context has been largely understudied. Researchers have studied the demand-sensitive nature of MW in tasks while neglecting possible time on task effects (e.g., Forster & Lavie, 2009). Prior findings reporting that MW is greater during low vs high demand tasks have been used to advance the view that MW relies on executive resources but that these resources are privileged for use by the primary task to ensure its successful completion. From the perspective of the executive-resource account (see Smallwood et al., 2009), resources may become available for MW only secondarily, when the primary task demands are low. Yet, the results from the current study suggest that executive resource allocation may privilege engagement in the primary task in this fashion only when the time on task is relatively short. As the duration of the experiment is protracted, MW was found to be greater for high vs. low demand trials. Perhaps this uptick in MW over time is due to flagging motivation to remain engaged on the primary task (vs. in the service of MW). Decreases in motivation may result from an assessment of opportunity costs leading to prioritization of MW over task performance (see Kurzban, Duckworth, Kable, & Myers, 2013) or due to depletion of executive resources over time (Muraven & Baumeister, 2000), which may degrade goal maintenance for the task set. Recently, the depletion model of executive control has come under scrutiny, and it is unclear whether the phenomenon exists (Carter, Koffer, Forster, & McCullough, 2015; Hagger et al., 2016), or merely needs reinterpretation (Kurzban et al., 2013).

We acknowledge that our paradigm, too, is limited in a number of ways. First, our approach assumes that the influence of “trials” on task accuracy and MW is reflecting the influence of ‘time on task’. Yet, it is possible that the effects we see in the current study are being driven by the passing of time independent of the number of trials which occur over time. Future studies could disambiguate this by maintaining the total task duration while varying the inter-trial-intervals and number of trials. For example, the inter-trial-intervals could be lengthened and the number of trials could be cut, or the inter-trial-interval could be shortened and the number of trials increased. Manipulations that disentangle the effects of time and trials would allow for greater clarity on whether increases in MW are due to the passing of time, per se, or due to task engagement over repeated trials.

A second limitation of our study is that we did not assess whether participants were experiencing a greater sense of task-mastery with greater time on task. Smallwood and Schooler (2006) proposed that greater time on task may lead to task automation, freeing up executive resources for MW. While we found no evidence of task automation in the current study, it is possible that participants experienced a greater sense of mastery over the task which led them to strategically and erroneously withdraw executive resources away from the primary task in the service of MW. This is consistent with our finding that in later trials, accuracy decreased despite participants still reporting that they were on-task. This could be clarified with further research, in which performance feedback is provided and/or response confidence is probed. On the other hand, it is possible that participants’ insight of their performance impacted their MW response. This is a question that should be further explored in future studies. Future studies could also bolster our findings by examining the putative contribution of individual differences in both the cognitive (e.g., working memory capacity) and affective domains (e.g., depression and anxiety scores) that influence the relationship between accuracy and MW with time on task.

A third limitation of our experiment has to do with our method of randomizing trial order. Here, we randomized trial order once and applied the same order to all participants. Any trial-specific idiosyncrasies that may have resulted in differences in accuracy or mind wandering were not distributed evenly across trials. Therefore, it is possible that there were other trial difficulty parameters beyond time on task that contributed to our results. Future research should randomize stimuli across trials for each subject separately to reduce the influence of nuisance variables.

Despite these limitations, the results of this experiment do advance our understanding of how time on task influences mind-wandering and working memory task performance in a demand-sensitive fashion. Our findings suggest that theoretical models of MW should be expanded to account for its variability over time and as a function of executive demands. As well, cognitive neuroscience models of working memory, which have carefully considered the neurocognitive consequences of perturbations from external sources of interference (see Sreenivasan & Jha, 2007), must now formally account for the
contribution of internal sources of distraction (i.e., MW) on working memory maintenance processes.

Acknowledgements

This work was funded by United States Department of Army Grant #WX81XWH-11-2-0044 to APJ and NSF GRFP Award #145111 to MK. We thank Emily Brudner, Keith Chichester, Sydney Feldman, and Ariel Paz for their assistance with data collection.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2017.08.006.

References
