Author Accepted Manuscript

Prospective Memory Performance in Simulated Air Traffic Control: Robust to Interruptions but Impaired by Retention Interval

Micah K. Wilson [2 1], Luke Strickland[1], Simon Farrell[1], Troy A.W. Visser[1], & Shayne Loft[1]
[1] School of Psychological Science, University of Western Australia, Perth, Australia
[2] Future of Work Institute, Curtin University, Perth, Australia

Please note that this paper is published in *Human Factors: The Journal of the Human Factors and Ergonomics Society* and is reprinted here by permission of SAGE Publications. Please refer to this work using the complete citation shown below.

Wilson, M. K.., Strickland, L., Farrell, S., Visser, T. A. W., & Loft, S. (2019). Prospective Memory Performance in Simulated Air Traffic Control: Robust to Interruptions but Impaired by Retention Interval. *Human Factors*. https://doi.org/10.1177/0018720819875888

Prospective Memory Performance in Simulated Air Traffic Control: Robust to Interruptions but Impaired by Retention Interval

Micah K. Wilson [2 1], Luke Strickland[1], Simon Farrell[1], Troy A.W. Visser[1], & Shayne Loft[1] 1

School of Psychological Science, University of Western Australia, Perth, Australia

[2] Future of Work Institute, Curtin University, Perth, Australia

Manuscript Type: Research Article

Word Count: 5555

Author Note: This research was supported by DP12010311 awarded to Loft from the Australian Research Council and an Australian Postgraduate Award awarded to Wilson. Address correspondence to Micah K. Wilson (Micah.K..wilson@curtin.edu.au), Future of Work Institute, Curtin University, Australia. We would like to thank Mr. Thomas McKeesick for his assistance with data processing, and for refactoring and reviewing the processing code.

Abstract

Objective: To examine the effects of interruptions and retention interval on prospective memory for deferred tasks in simulated air traffic control. **Background:** In many safety-critical environments, operators need to remember to perform a deferred task, which requires prospective memory. Laboratory experiments suggest that extended prospective memory retention intervals, and interruptions in those retention intervals, could impair prospective memory performance. **Method:** Participants managed a simulated air traffic control sector. Participants were sometimes to perform a deferred handoff task, requiring them to deviate from a routine procedure. We manipulated whether an interruption occurred during the prospective memory retention interval or not, the length of the retention interval (37 s to 117 s), and the temporal proximity of the interruption to deferred task encoding and execution. We also measured performance on ongoing tasks. **Results:** Increasing retention intervals (37 s to 117 s) decreased the probability of remembering to perform the deferred task. Costs to ongoing conflict detection accuracy and routine handoff speed were observed when a prospective memory intention had to be maintained. Interruptions did not affect individuals speed or accuracy on the deferred task. **Conclusion:** Longer retention intervals increase risk of prospective memory error and of ongoing task performance being impaired by cognitive load; however, prospective memory can be robust to effects of interruptions when the task environment provides cuing and offloading. **Application:** To support operators in performing complex and dynamic tasks, prospective memory demands should be reduced, and the retention interval of deferred tasks should be kept as short as possible.

Keywords: deferred tasks, task interruptions, complex dynamic task, delay interval

Précis

A simulated air traffic control task was used to investigate how prospective memory for deferred tasks was affected by the prospective memory retention interval and interruptions. Prospective memory errors increased with longer retention intervals, and ongoing task performance decreased during the retention interval; but interruptions did not affect prospective memory.

Open Practices

The parsed experimental data and all data processing and analysis materials have been made publicly available via the Open Science Framework and can be accessed via https://osf.io/mz6fa/. The repository also contains materials required to replicate the study.

The cognitive processes involved in the maintenance, retrieval, and execution of deferred tasks are referred to as *prospective memory* (PM). Individuals often need to remember to perform deferred tasks in safety-critical work contexts, such as air traffic control (ATC), healthcare, piloting, and unmanned aerial vehicle control. For example, air traffic controllers sometimes must remember to deviate from a routine aircraft vectoring procedure and to instead hold an aircraft when it reaches a specific waypoint in the future because of heavy traffic. This requires the controller to defer the execution of the action and to remember to execute it at an appropriate time. Unfortunately, controllers can forget to complete deferred tasks, an outcome referred to as PM error (Shorrock, 2005). Such PM errors can, in turn, have serious safety implications (Dismukes, 2012; Loft, 2014; Loft, Dismukes, & Grundgeiger, 2019).

In order to reduce PM errors, it is important to understand the psychological processes underlying PM, and how task characteristics affect those processes. Previous laboratory research has identified two important factors likely to affect the performance of deferred tasks: (a) *retention interval* (see, Martin, Brown, & Hicks, 2011), which refers to the amount of time between the encoding of the PM intention and the opportunity to execute it; and (b) *interruptions* arising from competing task demands, which can occur frequently during PM retention intervals (e.g., Cook, Meeks, Clark-Foos, Merritt, & Marsh, 2014; Schaper & Grundgeiger, 2018). In the current study, we examine how the length of the PM retention interval, and the presence of interruptions during the retention interval, impact the probability and speed at which individuals remember to deviate from a routine aircraft handoff procedure in a simulated ATC task. Additionally, we measure performance on concurrent ongoing ATC tasks to examine performance costs associated with a PM load. The study aims to illuminate how individuals maintain deferred task goals and use situational cues to support PM in safety-critical work contexts such as ATC.

Theoretical Approach

The Dynamic Multiprocess View (DMPV) is a useful theoretical framework for understanding PM in applied dynamic multi-tasking contexts (Scullin, McDaniel, & Shelton, 2013). Its central tenet is that PM is supported by the dynamic interplay between top-down and bottom-up cognitive processes (Shelton & Scullin, 2017). Top-down processing involves deliberately maintaining the intention to perform the PM action in focal attention, or strategically monitoring the environment for PM cues (e.g., inspecting aircraft call signs). Evidence for such top-down processing in ATC has been demonstrated in several studies showing that maintaining PM is detrimental to performance on other ATC tasks (referred to as "PM costs"; for a review see Loft, 2014). However, PM can also be supported via bottom-up, cue-driven processes (Einstein & McDaniel, 2005). For instance, if events or environments become associated with a PM intention, attending to them can prompt retrieval of the PM action. A key feature of the DMPV is that bottom-up and top-down processes can interact. For instance, task context can trigger bottom-up processes that subsequently result in the engagement of strategic top-down monitoring processes (Scullin et al., 2013; see also, Smith, Hunt, & Murray, 2017). Thus, external cues can trigger an operator's intention to strategically monitor for PM events.

Controllers often report that PM tasks with long retention intervals are the most susceptible to PM error (Loft, Smith, & Remington, 2013). According to the DMPV, this would occur because limited-capacity top-down monitoring processes are difficult to sustain for extended durations. In line with this, studies of PM in laboratory-based tasks (e.g., lexical decision making) have shown that increasing retention interval decreases PM performance

(Martin, Brown, & Hicks, 2011; Scullin, McDaniel, Shelton, & Lee, 2010; Tierney, Bucks, Weinborn, Hodgson, & Woods, 2016; Zhang, Tang, & Liu, 2017). Furthermore, PM costs to other ongoing tasks have been found to decrease over longer retention intervals suggesting that individuals decrease top-down monitoring over time (Loft, Kearney, & Remington, 2008; McBride, Beckner, & Abney, 2011). As such, there is reason to suspect that in ATC, PM tasks with longer retention interval could be at high risk of not being completed. However, it is also possible that the continued presence of the PM relevant aircraft on the ATC display could act as a persistent contextual cue to prompt top-down monitoring over the retention interval (Todorov, Kubik, Carelli, Missier, & Mäntylä, 2018). In line with this, Stone, Dismukes, and Remington (2001) reported no effect on PM error when the PM retention interval increased from 1-min to 5-min in simulated ATC. To our knowledge, this is the only study that has manipulated retention interval in simulated ATC to date.

According to the DMPV, people are less likely to engage in top-down monitoring if they exit the environmental context that is associated with PM. This is consistent with many laboratory studies in which shifts in ongoing task context result in decreased evidence of top-down monitoring (Bowden, Smith, & Loft, 2017; Kuhlmann & Rummel, 2014; Marsh, Hicks, & Cook, 2006). In ATC, such situations are particularly likely to arise due to task interruptions, which can be defined as situations in which an individual must suspend a *primary task* (e.g., display monitoring) in order to perform a secondary *interrupting task* (e.g., answering a pilot communication), with the explicit intention to return to the primary task after the interruption (Trafton & Monk, 2007). Indeed, several laboratory studies have found that interruptions during PM retention intervals can impair PM performance (Cook et al., 2014; McDaniel, Einstein, Graham, & Rall, 2004; Schaper & Grundgeiger, 2018). However, in simulated ATC, Wilson, Farrell, Visser, and Loft (2018) found that interrupting participants for 27 s during the PM retention interval had no effect on the speed or accuracy of performing a PM task that required deviation from a routine aircraft handoff procedure.

The DMPV offers two explanations for why interruptions may not have impacted PM in simulated ATC. One is that PM retrieval in ATC may largely depend on bottom-up, cue-driven processes (Einstein & McDaniel, 2005), in which case any effect of interruptions on top-down monitoring processes would be irrelevant. However, previous ATC studies show that PM load (i.e., having an active PM intention) impairs concurrent air traffic management tasks, such as conflict detection (Loft, Chapman, & Smith, 2016; Loft et al., 2013; Loft, Finnerty, & Remington, 2011; Loft & Remington, 2010; Loft, Smith, & Bhaskara, 2011), indicating reliance on top-down monitoring. Another possibility is that top-down monitoring might have been reinstated in the interval between the interruption ending and the PM action being required. In Wilson et al. (2018) the PM task had to be performed approximately 1 min after the interruption had ended, thus permitting time to process contextual cues associated with the PM action, which in turn may have re-engaged top-down monitoring.

This second option suggests that is important to further consider how the temporal relationship between interruptions and the correct time for PM retrieval influences PM errors. In the interruptions literature, studies generally examine tasks where the resumption lag (interval between end of interruption and primary task resumption) is effectively zero - individuals must "resume" the intended primary task immediately after returning from the interruption. For PM tasks, however, the resumption lag (i.e., PM *execution-delay*) is likely to be heterogeneous, and this may impact how well individuals can orient to the updated visual scene and use contextual cues to trigger PM retrieval. Research on visual working memory has shown that memory representations are most volatile in the immediate moments following an interruption, resulting from insufficient time for attention to recover (Wang, Theeuwes, & Olivers, 2018). Similarly, PM can be improved by the provision of time *prior* to an interruption (i.e., increased PM *encoding-delays*). An increased encoding-delay allows more time to rehearse and consolidate intentions, and an opportunity to strengthen associations between intentions and contextual cues (Boehm-Davis & Remington, 2009). This can improve PM (Dismukes & Nowinski, 2006) and resumption time (Hodgetts & Jones, 2006) in simple tasks, and improve decision-making in complex dynamic tasks (Labonté, Tremblay, & Vachon, 2019). To be clear, this *interruption-delay hypothesis* specifies that increasing encoding-delay or execution-delay may improve PM when interrupted. However, increasing these delay intervals also increases the total retention interval, which may be expected to impair PM. In the current study, we compare these hypotheses by testing whether encoding-delays and execution-delays function in a different manner to what would be expected from an effect of retention interval when interrupted.

Current Study

We examined how PM in a simulated ATC task is affected by retention interval (as well as the separable contributions of differences in encoding and execution-delay to this interval), and by interruptions. Participants assumed the role of an air traffic controller responsible for accepting aircraft entering their air sector, detecting and resolving aircraft conflicts, and handing-off aircraft exiting their sector. Participants completed a number of trials, and in each, a PM task occurred that required them to acknowledge an instruction to perform an alternative action (e.g., press right arrow key) instead of a routine action (H key) when handing off a target aircraft (i.e., deferred handoff). During some PM retention intervals participants were interrupted by an additional ATC task (for 27 s) that comprised the same task objectives as the primary scenario, but with different aircraft and flight paths. The interruption and primary ATC tasks overlapped in visual appearance and processing modality, which has previously been linked to interference-based PM errors in ATC (Wilson et al., 2018).

Figure 1 shows how the timing of the PM task relative to the interruption yielded variations in encoding and execution delays. The square dots represent key stages in each trial (trial start, encode PM task, execute PM task, and trial end), while the two triangles indicate the point that the PM task was encoded, and need to be executed, relative to the interruption. The PM retention interval consisted of the encoding-delay, duration of the interruption, and the execution-delay. The combination of two encoding-delays (10 or 50 s), two execution-delays (0 or 40 s) and a 27 s interruption yielded three retention intervals (37 s, 77 s, 117 s).

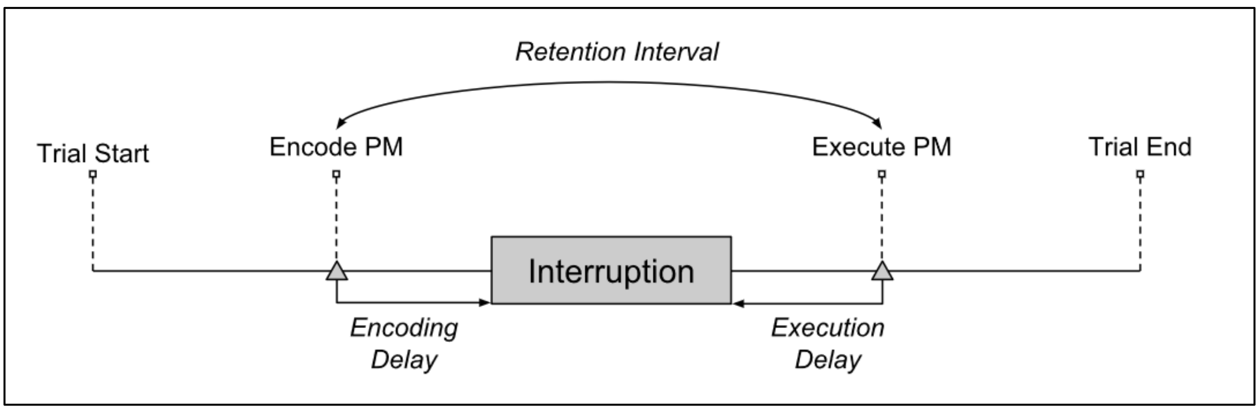

Figure 1. The time-course of the PM handoff task (not to scale). Italicized and bolded words indicate manipulated factors.

We examined: (1) whether the interruption decreased PM performance (i.e., decreased accuracy and increased PM RT); (2) whether longer encoding or execution delays improved PM performance; (3) whether longer PM retention intervals were associated with decreased PM performance; (4) and whether there were costs to aircraft acceptances or handoffs (non-response errors and RT) or conflict detection accuracy during the PM retention interval.

Method

Participants

78 undergraduate students (female = 31; median age = 20) from the University of Western Australia participated in the study in exchange for course credit or $50 AUD. This research complied with the American Psychological Association Code of Ethics and was approved by the Human Research Ethics Office at the University of Western Australia. Informed consent was obtained from each participant.

Design

The experiment used a 2 (interruption) x 2 (encoding-delay) x 2 (execution-delay) within-subjects design. The interruption manipulation was either ‘uninterrupted’ in which participants were not interrupted, or ‘interrupted’ in which they had to manage an additional ATC sector for 27 s. The timing of the encoding and execution-delay manipulations was anchored around the interruption. The encoding-delay was either *short* in which the deferred handoff PM task was encoded 10 s prior to the interruption; or *long* in which the PM task was encoded 50 s before interruption. Similarly, the execution-delay was either *short,* in which the PM aircraft flashed for handoff immediately after the interruption ended; or *long* in which the PM aircraft flashed for handoff 40 s after the interruption ended. These timings were identical on uninterrupted trials, except that there was no interruption. As shown in Table 1, this resulted in eight within-subjects conditions, associated with three retention interval durations (37 s, 77 s, 117 s).

Table 1

The experimental design along with the respective total retention interval.

Interruption	Encoding-Delay	Execution-Delay	Total Retention Interval [a]	*N* trials [b]
Interrupted	S (10 s)	S (0 s)	37 s	256
Interrupted	S (10 s)	L (40 s)	77 s	258
Interrupted	L (50 s)	S (0 s)	77 s	262
Interrupted	L (50 s)	L (40 s)	117 s	254
None	~~S (10 s)~~	~~S (0 s)~~	37 s	259
None	~~S (10 s)~~	~~L (40 s)~~	77 s	253
None	~~L (50 s)~~	~~S (0 s)~~	77 s	256
None	~~L (50 s)~~	~~L (40 s)~~	117 s	254

Note. The encoding-delay and execution-delay conditions are relative to the ‘interruption start point’ and ‘interruption end point’, respectively. The strike-through for the delay manipulations indicates that while timing was equivocal across interruption conditions, uninterrupted trials sum to the retention interval.
[a] The total retention interval includes the 27s of either continued ongoing air traffic management (uninterrupted trials) or the interrupting ATC task.
[b] The total number of observed PM trials per condition after the specified exclusion criteria (see results section).

ATC-Lab^{Advanced} Simulator

Figure 2 presents a screenshot of the ATC task (Fothergill, Loft, & Neal, 2009). The light grey polygon area is the flight control sector, whilst the dark grey area represents sectors outside the participants’ control. The black lines denote aircraft flight paths. Aircraft were represented by a circle with a leader-line indicating heading. The aircraft data-blocks specify call sign, speed, aircraft type, current/cleared altitude. Current altitude and cleared altitude were separated by an arrow that denotes whether the aircraft is climbing (^), descending (v), or cruising (>). Aircraft entered the sector from the edges of the display, cross sector boundaries, and then exit the sector. New aircraft entered throughout trials, with aircraft positions being updated every second (behavioral measures were recorded with millisecond precision). Time elapsed in each trial was displayed on the bottom of the display.

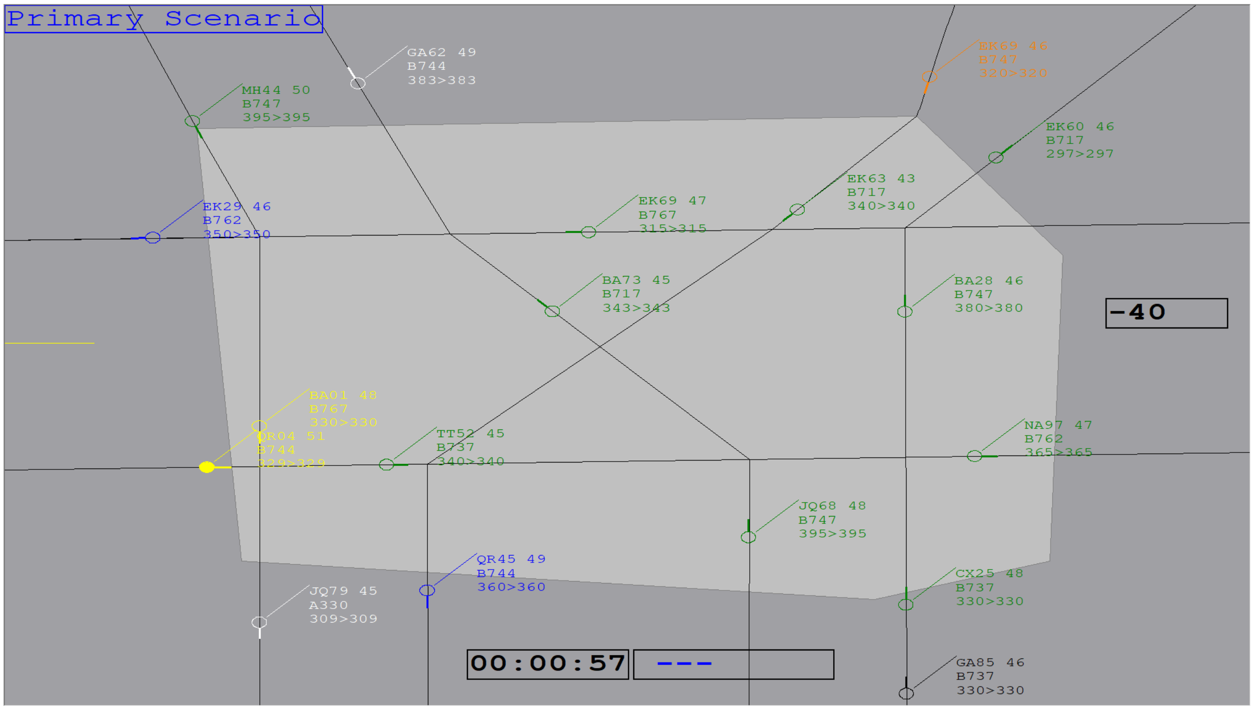

Figure 2. A screenshot of the ATC display. Inbound aircraft are black (GA85) as they approach the sector, and flash orange for acceptance (EK69) when they reached within 5 miles of the sector boundary. Aircraft turn green (e.g., MH44) when accepted. When outbound aircraft cross the sector boundary they flash blue (e.g., EK29), and then turn white (e.g., JQ79) when handed off. The example in Figure 2 shows that the individual is required to change the altitude of EK63 (cruising at 340) to an alternative altitude to resolve the conflict with JQ68 (cruising at 395). Aircraft turn yellow (e.g., QR04 & BA01) if they violate the minimum vertical and lateral separation. The running score (- 40 points) is presented in the middle right hand side of the display. Note the ‘primary scenario’ text box in the upper left to distinguish the primary from interrupting scenarios.

When aircraft approached the sector boundary, they flashed for acceptance and participants had to accept aircraft by clicking the aircraft and pressing the A key within 15 s. Similarly, as aircraft exited the sector boundary, they flashed for handoff and participants had to click the aircraft and press the H key within 15 s. Conflicts occurred when an aircraft pair violated both lateral (5 nautical miles) and vertical (1,000 feet) separation standards, and were indicated by the pair of aircraft turning yellow. Participants were required to prevent conflicts from occurring by clicking on one of the aircraft they believed to be in future conflict and changing the cleared altitude. Each trial comprised 3 conflicts, and 25 to 30 acceptances and handoffs. Participants were instructed to respond as quickly and accurately as possible.

Participants received points for successfully completing tasks and the current score was continuously updated on the right of the display. Ten points were awarded or deducted for a successful/failed handoff/acceptance. Between 10 and 40 points were awarded for resolving a conflict, depending on the speed of resolution, and 40 points were deducted for failing to resolve a conflict or for unnecessary interventions (i.e., altering the altitude of aircraft not in conflict).

Procedure

The experiment consisted of two 2.5 hour sessions. Session one comprised a training and the first test phase. Session two comprised the second test phase and a brief questionnaire.

Training Phase. Training started with completing an instructional website (~30 min). The website provided explanations of basic ATC concepts, instructions for completing ongoing tasks (handoffs, acceptances, and detecting/resolving conflicts), instructions regarding the deferred-handoff task, and information regarding the interruptions and point scoring. Participants then completed two 5 min practice trials, which followed the same structure as the test trials.

Test Phase. Each of the two test phase sessions comprised 16 5-min trials. One PM handoff was required in each trial, resulting in 32 PM task observations per participant (four per within-subject condition) across the study. The general order of events relating to the PM handoff (see Figure 1) was identical in every trial, but trials differed with respect to event timings and locations (e.g., conflicts occurred at differing times and locations; interruption onset times differed). The average overall proportion of handoffs that were PM handoffs (PM rate) was 7.62%, with a median of 13 routine handoffs required per trial. The same 16 trials were used for both sessions, however on the second session aircraft call-signs were randomized and experiment conditions were opposed (e.g., interrupted to uninterrupted; short to long delay). Experimental conditions were counterbalanced across trials and subjects with two 8 x 8 Latin square schemes (one for each session), the details of which are presented in Table 2, below. Final trial presentation order was randomized.

Table 2.

Counterbalancing scheme for the first experimental session.

	ATC-SS	ATC-SL	ATC-LS	ATC-LL	None-SS	None-SL	None-LS	None-LL
Group 1	(1,9)	(2,10)	(3,11)	(4,12)	(5,13)	(6,14)	(7,15)	(8,16)
Group 2	(2,10)	(3,11)	(4,12)	(5,13)	(6,14)	(7,15)	(8,16)	(1,9)
Group 3	(3,11)	(4,12)	(5,13)	(6,14)	(7,15)	(8,16)	(1,9)	(2,10)
Group 4	(4,12)	(5,13)	(6,14)	(7,15)	(8,16)	(1,9)	(2,10)	(3,11)
Group 5	(5,13)	(6,14)	(7,15)	(8,16)	(1,9)	(2,10)	(3,11)	(4,12)
Group 6	(6,14)	(7,15)	(8,16)	(1,9)	(2,10)	(3,11)	(4,12)	(5,13)
Group 7	(7,15)	(8,16)	(1,9)	(2,10)	(3,11)	(4,12)	(5,13)	(6,14)
Group 8	(8,16)	(1,9)	(2,10)	(3,11)	(4,12)	(5,13)	(6,14)	(7,15)

Note. Participants were distributed equally across the counterbalancing groups (i.e., rows). ATC scenarios were grouped in pairs of two (shown in parentheses) and distributed across the counterbalancing scheme. Experimental conditions were assigned column-wise. For the second session, the column order was reversed (e.g., Group 1 would complete scenarios 8 and 16 under ATC-SS, and scenarios 2 and 10 under None-LS).

The interruption start time was fixed for each of the 16 trials (between 1m 30s and 2m 30s). On each trial either 50 s (long encode-delay) or 10 s (short encode-delay) before the interruption, a message box would appear adjacent to one aircraft instructing participants to handoff that aircraft with an arrow key that corresponded to the aircraft heading (e.g., ↑), instead of the routine ‘H’ key. This message was displayed for 10 s, and participants had to acknowledge it by clicking a button marked "Acknowledge" that became active after 3s to prevent accidental acknowledgement. Messages disappeared if not acknowledged within 10 s. Encoding-delay was manipulated by changing the onset of this message.

After the interruption, the deferred handoff target aircraft would immediately flash for handoff (i.e., cue for performing the PM action), or participants resumed the primary ATC tasks for 40s, and then the PM aircraft flashed for handoff. This was our execution-delay manipulation. No aircraft flashed for acceptance or handoff within 10 s of the PM aircraft flashing for handoff. After the PM aircraft was handed-off (or recorded as missed), participants continued ongoing ATC tasks until the trial ended. Event timings were identical for uninterrupted trials, but participants only performed the primary ATC task.

The interrupting ATC task required monitoring a different sector that was displayed in place of the primary sector, with task objectives identical to the primary task. Each interrupting trial comprised two or three aircraft acceptances, two or three aircraft handoffs, and two conflicts. The interruption began with a 1.7 cm crosshair presented in the center of the display for 2500 ms, a 24s ATC scenario, and a black visual buffer for approximately 500 ms. Participants were instructed that the interrupting ATC task was equally important as the primary ATC, but that no “special handoff aircraft task” (i.e., deferred handoff) would occur during the interruption. The timer was removed from the interrupting sector display. There were eight unique interruption trials which varied with regards to event timing and locations. The presentation order was randomized for each participant per session.

There was a 15 s break after trials, except for the 8th trial which was followed by a 180s break. Participants could then begin the following trial by pressing spacebar and could take a longer rest break if required by pausing before pressing spacebar.

Results

Four participants who did not complete the second experimental session were excluded, as were four participants who only correctly perform the deferred handoff task on less than 10% of trials (final n = 70). Hypothesis testing was conducted using mixed-effects modelling, implemented in the *lme4* R package (Bates, Mächler, Bolker, & Walker, 2015) for the R programming language (R Core Team, 2017). Continuous dependent variables (e.g. mean RTs) were analyzed with linear mixed models, and binary dependent variables (e.g., PM errors) with generalized linear mixed models using a logistic link function. Mixed effects modelling enables control of variance associated with random factors (e.g. participant) without data-aggregation (Baayen, Davidson, & Bates, 2008). Models were compared with likelihood-ratio tests. Specifically, for each dependent variable, a null model was specified that included only the dependent variable of interest and a random intercept across participants. The impact of each experimental factor was evaluated by comparing a model that included the fixed effect of interest to the null model. Interaction effects were tested by comparing a full model specifying the interactions with a model containing identical predictors but no interaction. Reported p values were obtained with the Satterthwaite approximation by conducting chi-square tests (χ^2) on the log-likelihoods of the respective models (Kuznetsova, Brockhoff, & Christensen, 2017). Coefficients (β) and standard errors (SE) for each effect in question are presented in-text.

PM Task Accuracy

To assess whether the number of excluded trials based on error type differed across conditions, Pearsons Chi -squared tests for count data were conducted. PM response execution errors (remembering to press an arrow key but pressing the incorrect arrow key) were made on 3.17% (*n trials* = 74) of trials and did not significantly differ between conditions, $\chi^2(7) = 3.84$, $p = 0.8$. Non-response errors occurred on 0.56% of trials (*n trials* = 13) and did not significantly differ between conditions, $\chi^2(5) = 1.31$, $p = 0.93$. PM false alarms (pressing the arrow key on non-PM aircraft) were made on 3.17% of trials (*n trials* = 74) and did not significantly differ between the conditions, $\chi^2(7) = 6.65$, $p = 0.47$. PM task acknowledgement errors (failing to acknowledge the PM encoding message *and* making a PM error) occurred on 1.67% of trials (*n* *trials* = 39) and did not significantly differ between conditions, $\chi^2(7) = 5.51$, $p = 0.6$. All these trial types were excluded from final analysis (*final n observations* = 2052).

PM errors were defined as pressing the routine handoff key (H) instead of the instructed key when handing off PM aircraft. PM error rates by condition are presented in Figure 3. First, we examined whether PM errors increased as a function of retention interval in the uninterrupted condition. There was a significant main effect of retention interval, $\beta = 0.64$, *SE* = 0.17, $\chi^2(1) = 14.75$, $p = < 0.001$, and this was associated with a significant polynomial linear contrast, z = 3.85, $p = < 0.001$, indicating PM errors increased over longer retention intervals. This analysis was repeated with both interruption conditions which revealed a similar pattern of results.

PM error rates did not significantly vary by interruption condition, $\beta = 0.01$, *SE* = 0.11, $\chi^2(1) = 0.01$, $p = 0.94$. To test whether delays prior to, or after, an interruption affected PM errors, we examined the interaction between interruption condition and the encoding and execution delay conditions, respectively. A non-significant interaction would indicate that the effect of encoding or execution delay is equivalent between uninterrupted and interrupted conditions, indicating that only retention interval impacts PM task performance. A significant interaction would indicate the effect of delay differs for each interruption condition, suggesting a unique role of delay for interrupted trials. There was no significant interaction between encoding-delay and interruption condition, $\beta = -0.16$, *SE* = 0.22, $\chi^2(1) = 0.51$, $p = 0.47$, or between execution-delay and interruption condition, $\beta = -0.32$, *SE* = 0.22, $\chi^2(1) = 2.07$, $p = 0.15$.

Table 3.

Model comparison table for all deferred handoff PM model comparisons (PM errors and RT).

Dependent Variable (*y*)	Model Specification	*k*	AIC	BIC	Deviance	*p*
PM Errors (Full Dataset)						
	$y \sim \beta_0$	1	2073.71	2084.96	2069.71	—
	$y \sim \beta_0 + \beta_1(\textit{Interruption})$	2	2075.70	2092.58	2069.70	.94
	$y \sim \beta_0 + \beta_1(\textit{Retention})$	**3**	**2060.53**	**2083.04**	**2052.53**	**< .001**
	$y \sim \beta_0 + \beta_1(\textit{Encoding-Delay}) + \beta_2(\textit{Interruption})$	3	2074.21	2096.71	2066.21	—
	$y \sim \beta_0 + \beta_1(\textit{Execution-Delay}) + \beta_2(\textit{Interruption})$	3	2064.22	2086.72	2056.22	—
	$y \sim \beta_0 + \beta_1 + \beta_2 + \beta_3(\textit{Encoding-Delay} \times \textit{Interruption})$	4	2074.14	2102.27	2064.14	.47
	$y \sim \beta_0 + \beta_1 + \beta_2 + \beta_3(\textit{Execution-Delay} \times \textit{Interruption})$	4	2065.70	2093.84	2055.70	.15
PM Errors (Uninterrupted)						
	$y \sim \beta_0$	1	1032.68	1042.54	1028.68	—
	$y \sim \beta_0 + \beta_1(\textit{Retention})$	**3**	**1021.93**	**1041.65**	**1013.93**	**< .001**
PM RT (Full Dataset)						
	$y \sim \beta_0$	**1**	**8599.44**	**8612.26**	**8593.44**	—
	$y \sim \beta_0 + \beta_1(\textit{Interruption})$	2	8598.99	8616.08	8590.99	.12
	$y \sim \beta_0 + \beta_1(\textit{Resumption})$	2	8598.44	8615.53	8590.44	.08
	$y \sim \beta_0 + \beta_1(\textit{Encoding-Delay}) + \beta_2(\textit{Interruption})$	3	8598.76	8620.13	8588.76	—
	$y \sim \beta_0 + \beta_1(\textit{Execution-Delay}) + \beta_2(\textit{Interruption})$	3	8600.99	8622.35	8590.99	—
	$y \sim \beta_0 + \beta_1 + \beta_2 + \beta_3(\textit{Encoding-Delay} \times \textit{Interruption})$	4	8600.72	8626.36	8588.72	.83
	$y \sim \beta_0 + \beta_1 + \beta_2 + \beta_3(\textit{Execution-Delay} \times \textit{Interruption})$	4	8600.83	8626.47	8588.83	.14
PM RT (Uninterrupted)						
	$y \sim \beta_0$	**1**	**4338.89**	**4349.63**	**4332.89**	—
	$y \sim \beta_0 + \beta_1(\textit{Retention})$	3	4340.65	4358.55	4330.65	.33

Note. Bolded model names indicate selected models. k = number of fixed effect parameters. β_0 = intercept. All models included a participant-level random intercept term. Interaction models included main effects for both factors in the interaction, indicated by the unlabeled terms ($\beta_1 + \beta_2$). Reported values were obtained from Chi-square tests on the log-likelihoods via Satterthwaite approximation.

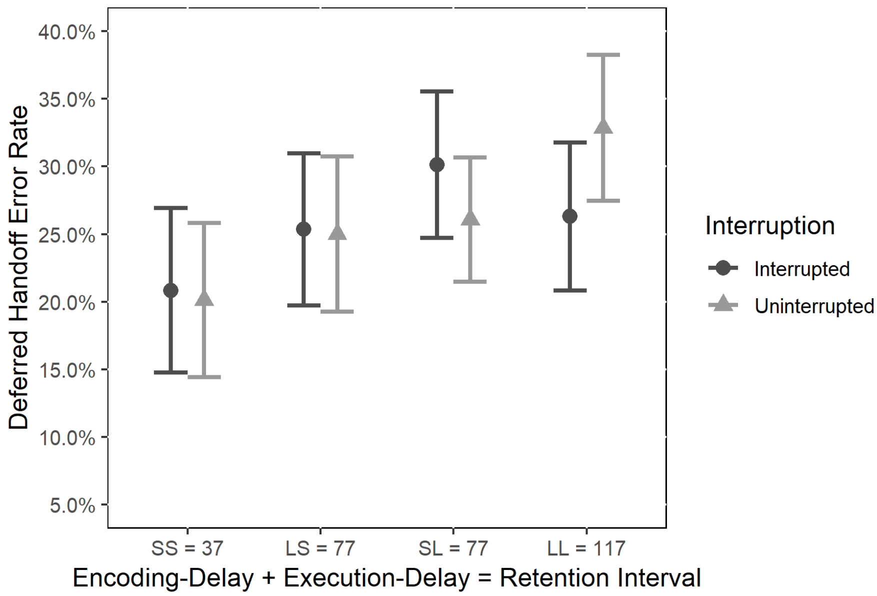

Figure 3. Mean deferred handoff error rate across the four timing conditions and the two interruption conditions for the deferred handoff task. Error bars represent 95% within-subjects confidence intervals (Cousineau, 2005).

PM Task Response Time

PM task response time (RT) was defined as the time taken to correctly handoff the PM aircraft after it flashed for handoff. Trials with RTs more than 3 SDs from a participant's grand mean were excluded from analysis (1.43% of RTs). Mean RTs are presented in Figure 4, separated by retention interval. There was no significant effect of retention interval for the uninterrupted condition, β = 134.64, *SE* = 89.52, $\chi^2(1)$ = 2.24, *p* = 0.33. There was also no significant main effect of interruption, β = -97.25, *SE* = 62.11, $\chi^2(1)$ = 2.45, *p* = 0.12, and no significant interaction between encoding-delay and interruption condition, β = -25.74, *SE* = 123.48, $\chi^2(1)$ = 0.04, *p* = 0.83, or between execution-delay and interruption condition, β = -181.86, *SE* = 123.62, $\chi^2(1)$ = 2.16, *p* = 0.14. Finally, we conducted a planned contrast to determine whether RT was slower for PM aircraft which had to be responded to immediately following an interruption (i.e., the ATC conditions with short execution-delay (SS & LS), relative to the SL and LL conditions), but this was not significant, β = 123.09, *SE* = 71.01, $\chi^2(1)$ = 3, *p* =

0.08. Thus, we did not find that retention interval or interruption impacted PM RT, and we did not find a resumption time effect.

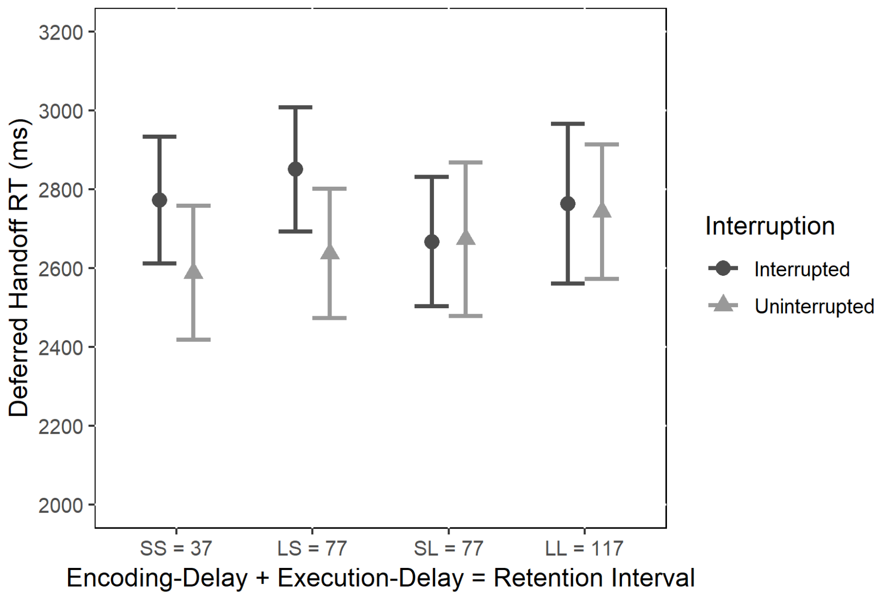

Figure 4. Mean deferred handoff RT across the four timing conditions and the two interruption conditions for the deferred handoff task. Error bars represent 95% within-subjects confidence intervals (Cousineau, 2005).

Cost of PM Retention to Ongoing ATC tasks

To determine the impact of PM on ongoing task performance, we examined aircraft handoffs, acceptances, and conflict detection. Descriptive statistics for these tasks are presented in Table 5 and full model comparisons are presented in Table 4. Handoff RT was slower for handoffs that occurred during the PM retention interval relative to outside it, β = 62.05, *SE* = 24.41, $\chi^2(1)$ = 6.36, *p* = 0.012, but there was no significant difference in non-response errors, β = 0.27, *SE* = 0.16, $\chi^2(1)$ = 2.86, *p* = 0.09. Aircraft acceptance RT did not significantly differ for acceptances occurring during versus outside the PM retention interval, β = 46.21, *SE* = 29.9, $\chi^2(1)$ = 2.38, *p* = 0.12, nor was there a significant difference in non-response errors, β = -0.27, *SE* = 0.15, $\chi^2(1)$ = 3.27, *p* = 0.071.

A conflict detection failure occurred when two aircraft violated minimum separation. Because conflicts evolve over time, their degree of overlap with the PM retention interval differed. To examine the cost of the PM load to conflict detection accuracy, we calculated an 'overlap proportion' measure that indexed the proportion of time that the PM retention interval overlapped with the time aircraft pairs involved in a conflict were in the sector. An overlap proportion of 0% indicates that the PM retention interval did not overlap with the evolving conflict. An overlap proportion of 100% indicates that the entire time the conflict pair was evolving occurred during the PM retention interval. Figure 5 shows predicted detection probability by overlap proportion. There was a significant effect of overlap proportion, β = -0.94, *SE* = 0.13, $\chi^2(1)$ = 51.78, *p* = < 0.001, with higher overlap being associated with poorer conflict detection accuracy. Conflict response time was not examined as it varied systematically as a function of conflict duration (i.e. how long it takes the aircraft pair to violate separation from when they were first both on the display), and thus did not allow unconfounded comparison.

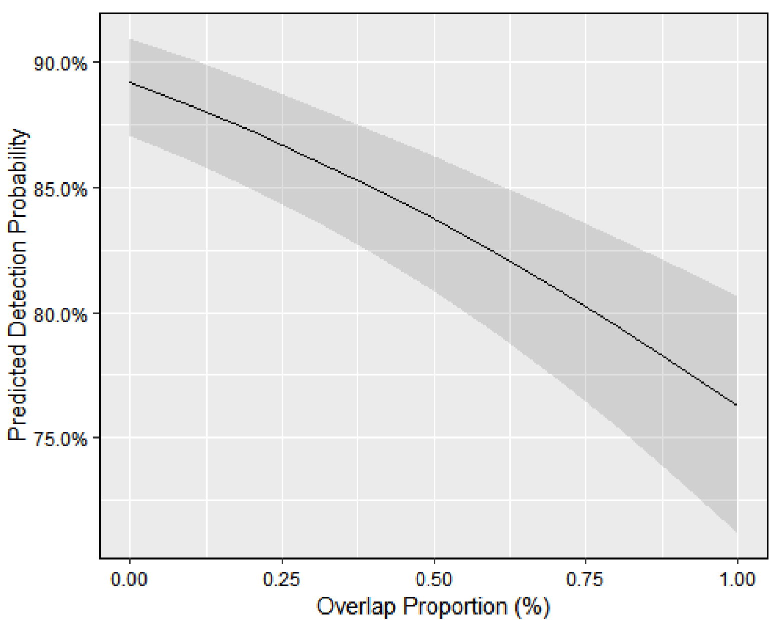

Figure 5. Effect display plots for the PM overlap proportion on the X axes and predicted conflict detection probability on the Y axis. The line represents the predicted detection probability means for each overlap proportion value. Error bars reflect 95% confidence intervals for the fixed effect.

Table 4.

Model comparison table for all ongoing task models.

Dependent Variable (*y*)	Model Specification	*k*	AIC	BIC	Deviance	*p*
Handoff RT						
	$y \sim \beta_0$	1	4050.85	4061.76	4044.85	—
	$y \sim \beta_0 + \beta_1$(*During PM Retention Interval*)	**2**	**4046.49**	**4061.03**	**4038.49**	**.012**
Handoff Errors						
	$y \sim \beta_0$	**1**	**2630.39**	**2646.58**	**2626.39**	—
	$y \sim \beta_0 + \beta_1$(*During PM Retention Interval*)	2	2629.53	2653.82	2623.53	.09
Acceptance RT						
	$y \sim \beta_0$	**1**	**4093.17**	**4104.07**	**4087.17**	—
	$y \sim \beta_0 + \beta_1$(*During PM Retention Interval*)	2	4092.79	4107.33	4084.79	.12
Acceptance Errors						
	$y \sim \beta_0$	**1**	**3421.29**	**3437.87**	**3417.29**	—
	$y \sim \beta_0 + \beta_1$(*During PM Retention Interval*)	2	3420.02	3444.9	3414.02	.07
Conflict Detection Accuracy						
	$y \sim \beta_0$	1	5184.92	5198.44	5180.92	—
	$y \sim \beta_0 + \beta_1$(*PM Overlap Proportion*)	**2**	**5135.15**	**5155.42**	**5129.15**	**< .001**

Note. Bolded model names indicate selected models. k = number of fixed effect parameters. β_0 = intercept. All models included a participant-level random intercept term. Reported values were obtained from Chi-square tests on the log-likelihoods via Satterthwaite approximation.

Table 5

Means and standard deviations for the three ongoing task performance measured by PM cost condition (i.e., whether the PM intention was to be maintained or not).

Measure	Type	Mean	*SD*	*N*
Accept Response Time	No PM	2818 ms	552 ms	140
Accept Response Time	PM	2864 ms	567 ms	140
Correct Accept Proportion	No PM	96.9%	17%	24178
Correct Accept Proportion	PM	97.4%	16%	5887
Handoff Response Time	No PM	3044 ms	658 ms	140
Handoff Response Time	PM	3106 ms	746 ms	140
Correct Handoff Proportion	No PM	98.9%	11%	18678
Correct Handoff Proportion	PM	99.13%	9.3%	5608
Conflict Detection Proportion	No PM	88%	32%	2720
Conflict Detection Proportion	PM	82%	39%	3634

Note. N = number of observations. The conflict detection proportion type includes any aircraft with any degree of overlap with the PM aircraft, and is reported here for descriptive purposes only.

Discussion

PM errors can have serious safety implications in complex work domains such as ATC (Dismukes, 2012; Loft, 2014). The current study examined how PM in simulated ATC was affected by the PM retention interval and sought to identify any contributions of encoding and execution delays under conditions where participants were interrupted. Our choice of experimental manipulations was motivated by previous research and ecological concerns, and our predictions motivated by the DMPV of PM (Scullin et al., 2013) which suggests that PM is supported by top-down strategic monitoring and maintenance, bottom-up cue-driven processes, and their interaction. We found that PM errors increased with longer retention intervals, but were not affected by the presence of an interruption during the retention interval. Further, there was no evidence that encoding or execution delays influenced PM error or RT. Conflict detection accuracy and routine aircraft handoff RT were both impaired during the PM retention interval, suggesting that participants relied on top-down PM maintenance and monitoring processes.

The finding that PM error rates and PM RTs were unaffected by interruptions sits in contrast to findings in basic tasks, where interruptions have been found to negatively impact PM (Cook et al., 2014; McDaniel et al., 2004; Schaper & Grundgeiger, 2018). Further, there was no increase in resumption time for handoffs immediately after the interruption. Interruptions may have failed to affect PM because shifting back to the pre-interrupted context cued participants to reinstate top-down monitoring, as suggested by DMPV (Martin et al., 2011; McDaniel et al., 2004). Participants may have done so by utilizing a meta-cognitive 'offloading strategy' (Risko & Gilbert, 2016), associating the PM intention with spatial and contextual features of the ATC display (Todorov et al., 2018). Indeed, offloading strategies can eliminate the costs of interruption on PM tasks in basic paradigms (Gilbert, 2015). Interestingly however, 70% of our subjects reported in a post-experiment questionnaire that the deferred handoff task was the task made most difficult by the interruption. This discrepancy between subjective reports and our findings highlights the importance of conducting empirical studies: subjective intuition obtained through qualitative methods may not align with objective performance data.

Although the lack of effect of interruptions on PM handoff errors is surprising, this is consistent with recent finding from Wilson et al. (2018). Wilson et al. found interruptions increased RT and resumption errors on a deferred conflict detection task that required a response immediately following an interruption, but did not find that interruptions impacted PM handoff errors. They reasoned that the effects of interruptions on PM might vary depending on the temporal proximity of PM encoding and execution to the interruption, whereby PM is improved by consolidation and recovery, respectively. However, we found no evidence that the effects of encoding or execution delay differed between our interruption conditions, indicating that the magnitude of the retention interval alone underpinned the observed PM error rate. Perhaps the differences between the deferred conflict task and handoff task can be attributed to the complexity of conflict detection, or the reliability of cuing of the deferred handoff task (i.e., aircraft flashes blue for handoff at a predictable future time). It would be valuable for future research to examine what properties of deferred tasks promote robustness to interruptions in applied contexts. The results here indicate that the timing of the deferred task relative to an interruption was not an important factor determining PM performance in simulated ATC. This contrasts with research suggesting that time for consolidation before an interruption, and time for recovery afterwards, may benefit PM (Dismukes & Nowinski, 2006; Hodgetts & Jones, 2006; Wang et al., 2018; Labonté et al., 2019). One possibility is that that encoding and execution delays may simply not impact PM when contextual cues can be quickly reinstated. However, because PM performance was unaffected by the interruptions in the current study, it is possible that individuals did not have to mitigate any disruptive effects through preparation or recovery strategies. Thus, future research needs to examine the effect of encoding and execution delays on PM under conditions where interruptions negatively impact PM.

In line with subjective reports from experienced controllers (Loft et al., 2013), and with the DMPV, we found PM errors increased over longer retention intervals, reflecting the challenge associated with sustaining top-down monitoring over long durations. This result was not guaranteed, given that the continued presence of the PM aircraft on the display could potentially have overcome the negative effects of PM retention interval. Assuming individuals were engaging in strategic offloading as suggested above, the negative effect of retention interval might indicate that the associations between cues and the intended action diminished over time.

This would also explain why Stone et al. (2001) found no effect of retention interval on PM. Participants in Stone's study performed multiple PM tasks with overlapping retention intervals, thus each consecutive PM instruction may have facilitated both recall of the remaining PM tasks, and strengthened the PM cue-intention associations.

Holding a PM intention also imposed costs to ongoing ATC tasks. During the PM retention interval, participants were slower to handoff routine aircraft and had poorer conflict detection accuracy. Costs to routine handoff RT likely may indicate that individuals were inspecting aircraft for potential PM features when handing them off (e.g., callsign of aircraft; relative spatial location). By contrast, there were no costs to aircraft acceptances as they were unrelated to PM and thus there was no PM features to inspect. However, conflict detection was also unrelated to PM, but was susceptible to PM load. Conflict detection is likely sensitive to shifts in allocated resources due to the high degree of attentional and cognitive demand posed by the conflict detection task (see Loft & Remington, 2010). The acceptance task is unlikely to be sensitive to such an attentional burden, because acceptance events were perceptually salient and did not require complex decision making. Thus, conflict detection PM costs may have occurred because participants were engaging in some form of active maintenance (e.g., PM rehearsal) that consumed limited cognitive resources (70% of participants reported using a rehearsal strategy in the post-experiment questionnaire), or because they were spending increased time attending to the PM aircraft (i.e., triggering bottom up cues).

Limitations, Future Directions and Practical Implications

The use of a student sample with limited training does constrain our ability to generalize the results to expert controllers. In addition to the differences in cognitive skill and motivation between experts and students, controllers learn to recognize specific events that occur routinely at certain sector locations (Bowden & Loft, 2016; Stein, Garland, & Muller, 2009), greatly reducing demands on their executive processing. The results of the current study may hold greatest relevance to situations where there aren't predictable patterns that controllers could rely upon for automatic processing (e.g., the sector is not highly familiar to controllers). Another limitation was that we were unable to examine how PM load impacted performance on the interrupting ATC task (as a PM intention was active in all interruption scenarios). Future research could also manipulate the retrospective memory demands, that is *what* must be remembered (i.e., the action to-be perform), and *when* it must be performed (i.e., PM cue features).

There are several practical implications of the current research. First, the presence of PM tasks and their respective retention intervals should be considered in cognitive work design. We demonstrated that longer retention intervals can not only lead to higher rates of PM error, but increasing the retention interval of the PM task also increases the risk that performance on other ongoing tasks will be contaminated by the PM load in simulated ATC. Thus, in applied contexts where operators are required to monitor dynamic displays and make complex cognitive decisions, it is crucial to attempt to minimize demands on PM and the retention interval of contaminant PM tasks. This recommendation is supported by our finding of a dose-dependent between PM overlap and proportion of conflicts missed.

In conclusion, the current study showed that longer retention intervals caused PM deficits and increased the risk of costs to ongoing conflict detection accuracy. As automation solutions emerge in ATC and other complex dynamic work tasks, it will be critical to continue to evaluate the nature of the memory load placed on human operators to prevent excessive memory demands competing with the overall safety of the human-machine system. Practitioners must examine whether automation solutions themselves inadvertently increase PM demands, for instance, by increasing the interleaved monitoring of greater numbers of concurrent tasks (Loft et al., 2019).

Key Points

- Individuals often need to remember to perform deferred tasks in safety-critical work contexts, requiring prospective memory.
- In the field, the retention interval of prospective memory tasks, and interruptions that occur during retention intervals, have been implicated as sources of prospective memory error.
- In a simulated ATC task, prospective memory errors increased with longer retention intervals.
- Performance on other ongoing ATC tasks decreased during the prospective memory retention interval, indicative of prospective memory costs.
- Prospective memory performance was not significant affected by task interruptions.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Boehm-Davis, D. A., & Remington, R. (2009). Reducing the disruptive effects of interruption: A cognitive framework for analysing the costs and benefits of intervention strategies. *Accident Analysis & Prevention*, *41*(5), 1124–1129. https://doi.org/10.1016/j.aap.2009.06.029

Bowden, V. K., & Loft, S. (2016). Using memory for prior aircraft events to detect conflicts under conditions of proactive air traffic control and with concurrent task requirements. *Journal of Experimental Psychology: Applied*, *22*(2), 211–224. https://doi.org/10.1037/xap0000085

Bowden, V. K., Smith, R. E., & Loft, S. (2017). Eye movements provide insights into the conscious use of context in prospective memory. *Consciousness and Cognition*, *52*, 68–74. https://doi.org/10.1016/j.concog.2017.04.003

Cook, G. I., Meeks, J. T., Clark-Foos, A., Merritt, P. S., & Marsh, R. L. (2014). The role of interruptions and contextual associations in delayed-execute prospective memory. *Applied Cognitive Psychology*, *28*(1), 91–103. https://doi.org/10.1002/acp.2960

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Dismukes, K., & Nowinski, J. (2006). Prospective Memory, Concurrent Task Management, and Pilot Error. *Attention: From Theory to Practice*, 225–236. https://doi.org/10.1093/acprof:oso/9780195305722.003.0016

Dismukes, R. K. (2012). Prospective Memory in Workplace and Everyday Situations. *Current Directions in Psychological Science*, *21*(4), 215–220. https://doi.org/10.1177/0963721412447621

Einstein, G. O., & McDaniel, M. A. (2005). Prospective Memory: Multiple Retrieval Processes. *Current Directions in Psychological Science*, *14*(6), 286–290. https://doi.org/10.1111/j.0963-7214.2005.00382.x

Fothergill, S., Loft, S., & Neal, A. (2009). ATC-lab Advanced: An air traffic control simulator with realism and control. *Behavior Research Methods*, *41*(1), 118–127. https://doi.org/10.3758/brm.41.1.118

Gilbert, S. J. (2015). Strategic offloading of delayed intentions into the external environment. *The Quarterly Journal of Experimental Psychology*, *68*(5), 971–992. https://doi.org/10.1080/17470218.2014.972963

Hodgetts, H. M., & Jones, D. M. (2006). Contextual cues aid recovery from interruption: The role of associative activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1120–1132. https://doi.org/10.1037/0278-7393.32.5.1120

Kuhlmann, B. G., & Rummel, J. (2014). Context-specific prospective-memory processing: Evidence for flexible attention allocation adjustments after intention encoding. *Memory & Cognition*, *42*(6), 943–949. https://doi.org/10.3758/s13421-014-0405-2

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Labonté, K., Tremblay, S., & Vachon, F. (2019). Forewarning Interruptions in Dynamic Settings: Can Prevention Bolster Recovery? *Journal of Experimental Psychology: Applied*. https://doi.org/10.1037/xap0000219

Loft, S. (2014). Applying Psychological Science to Examine Prospective Memory in Simulated Air Traffic Control. *Current Directions in Psychological Science*, *23*(5), 326–331. https://doi.org/10.1177/0963721414545214

Loft, S., Chapman, M., & Smith, R. E. (2016). Reducing prospective memory error and costs in simulated air traffic control: External aids, extending practice, and removing perceived memory requirements. *Journal of Experimental Psychology: Applied*, *22*(3), 272–284. https://doi.org/10.1037/xap0000088

Loft, S., Dismukes, R. K., & Grundgeiger, T. (2019). Prospective memory in safety-critical work contexts. In Rummel, Jan (Ed.) & McDaniel. Mark (Ed.), *Current Issues in Memory: Prospective Memory*. (pp. 170-185). Taylor and Francis. Retrieved from https://doi.org/10.4324/9781351000154-11

Loft, S., Finnerty, D., & Remington, R. W. (2011). Using Spatial Context to Support Prospective Memory in Simulated Air Traffic Control. *Human Factors: The Journal of the Human*

Factors and Ergonomics Society, *53*(6), 662–671. https://doi.org/10.1177/0018720811421783

Loft, S., Kearney, R., & Remington, R. (2008). Is task interference in event-based prospective memory dependent on cue presentation? *Memory & Cognition*, *36*(1), 139–148. https://doi.org/10.3758/mc.36.1.139

Loft, S., & Remington, R. W. (2010). Prospective memory and task interference in a continuous monitoring dynamic display task. *Journal of Experimental Psychology: Applied*, *16*(2), 145–157. https://doi.org/10.1037/a0018900

Loft, S., Smith, R. E., & Bhaskara, A. (2011). Prospective memory in an air traffic control simulation: External aids that signal when to act. *Journal of Experimental Psychology: Applied*, *17*(1), 60–70. https://doi.org/10.1037/a0022845

Loft, S., Smith, R. E., & Remington, R. W. (2013). Minimizing the disruptive effects of prospective memory in simulated air traffic control. *Journal of Experimental Psychology: Applied*, *19*(3), 254–265. https://doi.org/10.1037/a0034141

Marsh, R. L., Hicks, J. L., & Cook, G. I. (2006). Task interference from prospective memories covaries with contextual associations of fulfilling them. *Memory & Cognition*, *34*(5), 1037–1045. https://doi.org/10.3758/BF03193250

Martin, B. A., Brown, N. L., & Hicks, J. L. (2011). Ongoing task delays affect prospective memory more powerfully than filler task delays. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *65*(1), 48–56. https://doi.org/10.1037/a0022872

McBride, D. M., Beckner, J. K., & Abney, D. H. (2011). Effects of delay of prospective memory cues in an ongoing task on prospective memory task performance. *Memory & Cognition*, *39*(7), 1222–1231. https://doi.org/10.3758/s13421-011-0105-0

McDaniel, M. A., Einstein, G. O., Graham, T., & Rall, E. (2004). Delaying execution of intentions: overcoming the costs of interruptions. *Applied Cognitive Psychology*, *18*(5), 533–547. https://doi.org/10.1002/acp.1002

R Core Team. (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Risko, E. F., & Gilbert, S. J. (2016). Cognitive Offloading. *Trends in Cognitive Sciences*, *20*(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002

Schaper, P., & Grundgeiger, T. (2018). The effect of different distractions on remembering delayed intentions. *Memory*, *26*(2), 154–170. https://doi.org/10.1080/09658211.2017.1339090

Scullin, M. K., McDaniel, M. A., & Shelton, J. T. (2013). The Dynamic Multiprocess Framework: Evidence from prospective memory with contextual variability. *Cognitive Psychology*, *67*(1–2), 55–71. https://doi.org/10.1016/j.cogpsych.2013.07.001

Scullin, M. K., McDaniel, M. A., Shelton, J. T., & Lee, J. H. (2010). Focal/nonfocal cue effects in prospective memory: Monitoring difficulty or different retrieval processes? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(3), 736–749. https://doi.org/10.1037/a0018971

Shelton, J. T., & Scullin, M. K. (2017). The Dynamic Interplay Between Bottom-Up and Top-Down Processes Supporting Prospective Remembering. *Current Directions in Psychological Science*, *26*(4), 352–358. https://doi.org/10.1177/0963721417700504

Shorrock, S. T. (2005). Errors of memory in air traffic control. *Safety Science*, *43*(8), 571–588. https://doi.org/10.1016/j.ssci.2005.04.001

Smith, R. E., Hunt, R. R., & Murray, A. E. (2017). Prospective memory in context: Moving through a familiar space. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*(2), 189–204. https://doi.org/10.1037/xlm0000303

Stein, E., Garland, D., & Muller, J. (2009). Air-Traffic Controller Memory. In J. Wise, V. Hopkin, & D. Garland (Eds.), *Handbook of Aviation Human Factors, Second Edition* (pp. 21–1–21–39). https://doi.org/10.1201/b10401-26

Stokes, D. E. (1997). *Pasteur's Quadrant: Basic Science and Technological Innovation*. Brookings Institution Press. Retrieved from https://books.google.com.au/books?id=9tnCDAEACAAJ

Stone, M., Dismukes, K., & Remington, R. (2001). Prospective memory in dynamic environments: Effects of load, delay, and phonological rehearsal. *Memory*, *9*(3), 165–176. https://doi.org/10.1080/09658210143000100

Tierney, S. M., Bucks, R. S., Weinborn, M., Hodgson, E., & Woods, S. P. (2016). Retrieval cue and delay interval influence the relationship between prospective memory and activities of daily living in older adults. *Journal of Clinical and Experimental Neuropsychology*, *38*(5), 572–584. https://doi.org/10.1080/13803395.2016.1141876

Todorov, I., Kubik, V., Carelli, M. G., Missier, F. D., & Mäntylä, T. (2018). Spatial offloading in multiple task monitoring. *Journal of Cognitive Psychology*, *30*(2), 230–241. https://doi.org/10.1080/20445911.2018.1436551

Trafton, J. G., & Monk, C. A. (2007). Task Interruptions. *Reviews of Human Factors and Ergonomics*, *3*(1), 111–126. https://doi.org/10.1518/155723408x299852

Wang, B., Theeuwes, J., & Olivers, C. N. L. (2018). When shorter delays lead to worse memories: Task disruption makes visual working memory temporarily vulnerable to test interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(5), 722–733. https://doi.org/10.1037/xlm0000468

Wilson, M. K.., Farrell, S., Visser, T. A. W., & Loft, S. (2018). Remembering to execute deferred tasks in simulated air traffic control: The impact of interruptions. *Journal of Experimental Psychology: Applied*, *24*(3), 360–379. https://doi.org/10.1037/xap0000171

Zhang, H., Tang, W., & Liu, X. (2017). The Effect of Task Duration on Event-Based Prospective Memory: A Multinomial Modeling Approach. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01895

Biographies

Micah K. Wilson is a PhD candidate at the University of Western Australia's and a Research Fellow at the Future of Work Institute, Curtin University. He received a Bachelor of Arts with First Class Honours in Psychology from the University of Western Australia in 2013.

Luke Strickland is a Research Associate at the University of Western Australia. He received his PhD in psychological science in 2018 from the University of Western Australia.

Troy A. W. Visser is an Associate Professor at The University of Western Australia. He received his PhD in cognitive systems in 2001 from the University of British Columbia.

Simon Farrell is a Professor at the University of Western Australia. He received his PhD in experimental psychology in 2001 from the University of Western Australia.

Shayne Loft is an Associate Professor at the University of Western Australia. He received his PhD in experimental psychology in 2004 from the University of Queensland.