Journal of Memory and Language 140 (2025) 104583

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Individual differences in time-based prospective memory: The roles of working memory and time monitoring

Wiebke Hemming *, Kathrin Sadus, Jan Rummel

Department of Psychology, Heidelberg University, Germany

ARTICLE INFO

Keywords:
Time-based prospective memory
Working memory capacity
Binding
Updating
Strategic clock checking
Ongoing task slowing

ABSTRACT

Time-based prospective memory (PM) refers to the ability to remember to execute an intended action at a predefined future time. Previous research suggests that both general cognitive abilities (working memory) and task specific abilities (time monitoring) underly time-based PM performance. In three studies, we investigated the relevance of specific WM processes (binding, updating) for time-based PM and unravel their interplay with task specific abilities. In Experiment 1 ($N = 147$), we manipulated working-memory load, and found a greater influence of time monitoring on PM performance with increasing load. In Experiment 2 ($N = 132$), we found, in addition to time monitoring, specifically WM updating abilities to be associated with PM performance. In Experiment 3 ($N = 148$), we found PM performance to suffer when updating demands were increased but the effect vanished after controlling for time monitoring. These findings emphasize the complex interplay between general cognitive ability and task specific abilities in time-based PM.

Introduction

Prospective memory (PM) refers to the ability to remember and execute intended actions at a designated future moment. When studying PM, one can differentiate between event-based and time-based PM tasks. Event-based PM, involves executing an intended action in response to a specific event, such as posting a letter when passing the next mailbox or telling a colleague something important the next time one sees her. Time-based PM involves executing an intended action at a predetermined future time point, like remembering a dentist appointment at 11 am or remembering to meet a friend for coffee at 3 pm (Einstein & McDaniel, 1990; McDaniel & Einstein, 2007). From these everyday examples, it becomes evident that both types of PM tasks play an important role for goal-directed behavior in daily life (Ball et al., 2019; Rummel et al., 2023) and it is thus crucial to understand which cognitive demands PM tasks place on the individual and also which cognitive processes underlie the successful execution of these tasks. Research on PM, and especially its cognitive underpinnings, has flourished over the past three decades, with a steady increase in publications each year (Rummel & McDaniel, 2019, for a recent overview). However, many of these publications have been focusing on event-based PM and thus the cognitive underpinnings of time-based PM are currently less well understood. This discrepancy is remarkable considering that in everyday life most PM tasks are time-based (Kvavilashvili & Fisher, 2007). In the present article, we thus investigated which general cognitive and task specific abilities enable people to fulfill their time-based PM tasks.

A fundamental aspect of both event-based and time-based PM tasks is that they require the individual to maintain the intended action over time while being engaged in other activities or tasks (Ellis, 1996; Ellis & Kvavilashvili, 2000; Smith et al., 2017). In laboratory settings, the tasks participants perform while awaiting the right moment for executing the delayed intention are referred to as ongoing tasks. Oftentimes these ongoing tasks require continuous responding, like classifying stimuli according to certain rules (Einstein & McDaniel, 1990). Event-based PM tasks can be introduced in the setting by simply asking participants to press a special key when a specific event happens (e.g., when a predefined stimulus appears on the screen). Time-based PM tasks can be introduced similarly by asking participants to press a special key when a certain amount of time has elapsed (Einstein et al., 1995). In time-based PM tasks, participants are usually provided with opportunities to monitor the time via a clock that can be accessed on demand. A critical feature of both event-based and time-based laboratory PM settings is that the PM task has to be executed only infrequently, that is, on only 5 % – 10 % of the ongoing task trials (Ellis & Kvavilashvili, 2000). The rate of correct responses to the critical events or at the critical time points is usually considered a measure of PM performance (Einstein &

* Corresponding author at: Heidelberg University, Institute of Psychology, Hauptstr. 49, D-69117 Heidelberg, Germany.
E-mail address: wiebke.hemming@psychologie.uni-heidelberg.de (W. Hemming).

https://doi.org/10.1016/j.jml.2024.104583
Received 26 March 2024; Received in revised form 28 October 2024; Accepted 1 November 2024
Available online 13 November 2024
0749-596X/© 2024 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

McDaniel, 1990).

Whereas the exact nature of the cognitive processes underlying PM is still an issue of debate, there is a consensus that PM performance often relies on top-down processes (e.g., monitoring the environment for the right moment to fulfill the PM task), although there is some evidence that very salient and focal stimuli can also activate the PM intention in a bottom-up fashion (McDaniel et al., 2015; Shelton & Scullin, 2017; Smith et al., 2014; Strickland et al., 2018; Rummel & Kvavilashvili, 2023, for a discussion of current PM theories). General cognitive abilities like working-memory (WM) are sometimes considered a bottleneck for top-down monitoring processes (Smith & Bayen, 2005). In line with this notion, event-based PM performance regularly suffers under experimentally increased WM load (Einstein et al., 1997; Gonneaud et al., 2011; Kidder et al., 1997; Logie et al., 2004; Marsh & Hicks, 1998; Meier & Zimmermann, 2015; Park et al., 1997). The same pattern but with fewer studies can be found in time-based PM (Khan et al., 2008; Martin & Schumann-Hengsteler, 2001). Furthermore, it has been repeatedly shown that people with higher WM capacity also show better event-based PM performance (Kidder et al., 1997; Rose et al., 2010; Smith et al., 2011; Smith & Bayen, 2005). However, as said, it remains an open question which insights from event-based PM studies can directly be transferred to time-based PM. Indeed, it has been argued that the cognitive demands from event-based and time-based PM tasks differ in some respects. For instance, as there are no external events cuing the moment for intention execution in time-based tasks, their performance supposedly requires more self-initiated mental reactivation of the PM intention and a more continuous monitoring of the right moment to fulfill the intention. That is, one can assume that time-based PM relies on a general cognitive ability to maintain future task goals in WM and to monitor the time while performing some ongoing task (Kvavilashvili & Fisher, 2007). It is therefore crucial to understand which general cognitive abilities are associated with PM and how these processes influence task specific abilities, like monitoring behavior. Indeed, the few studies which have investigated the role of WM for time-based PM, found PM performance to decrease with increased WM load (Voigt et al., 2014) and time-based PM performance to be positively associated with WM capacity (Kretschmer et al., 2014; Mioni et al., 2017). Furthermore, there is empirical evidence that more efficient time monitoring goes along with better time-based PM performance (Joly-Burra et al., 2022).

In the present studies, we were primarily interested inasmuch WM processes are crucial for time-based PM and we used both experimental and individual difference approaches to address this issue. In Experiment 1, we manipulated WM capacity load during the ongoing task by systematically increasing it across three groups to examine the effects on time-based PM and on strategic clock checking behavior. In Experiment 2, we assessed WM capacity by measuring specific WM subprocesses. Specifically, drawing on previous research indicating their relevance for both time-based and event-based PM, we focused on updating (Kliegel et al., 2002; Mioni & Stablum, 2014; Zuber et al., 2019) and binding abilities (Lecouvey et al., 2017; Morand et al., 2021). In Experiment 3, we experimentally manipulated different types of WM load, increasing either capacity load or updating-specific load.

As said, there are likely general cognitive abilities (WM) and task-specific abilities (time monitoring) contributing to successful time-based PM performance. One open question from previous studies investigating the association between WM and time-based PM is whether WM and time monitoring contribute independently to time-based PM or whether people with better WM are simply better able to monitor the time and thus show better PM performance. Therefore, in the present studies, in addition to manipulating WM load or assessing WM capacity we also assessed time monitoring and tested whether eventually observed relationships between WM and time-based PM would remain stable after controlling for individual differences in time monitoring efficiency.

In addition to cognitive factors, non-cognitive factors like personality traits are believed to contribute to successful PM performance. However, these factors appear to be less strong predictors of PM (Arana et al., 2008) and the evidence surrounding them is somewhat contradictory. While some studies have found significant associations between certain personality traits and PM performance, others have not (Uttl et al., 2018). For instance, research has shown relationships between conscientiousness (Arana et al., 2008; Cuttler & Graf, 2007; Smith et al., 2011; Uttl et al., 2013), neuroticism (Cuttler & Graf, 2007), agreeableness, and openness (Uttl et al., 2013) and PM performance. However, a study with a large sample of 1170 participants found that personality traits did not significantly improve the prediction of PM performance beyond the variance that was already explained by cognitive factors (Uttl et al., 2018). In addition, self-reported PM abilities have been shown to be a reliable and valid way for assessing PM abilities, capturing individual differences (Rummel et al., 2019). Further investigation is needed to explore how these non-cognitive factors might interact with cognitive processes to influence PM outcomes. Consequently, we included personality traits and self-reported PM ability as additional individual difference variables in our studies.

To enable participants to monitor the time in laboratory time-based PM tasks, participants are typically provided with the opportunity to periodically check a screen clock display by pressing a key (Vanneste et al., 2016). This clock checking behavior is essential for determining the correct time for intention execution (Labelle et al., 2009; Mioni et al., 2020) and typically follows a J-shaped curve, with clock checking occurring more frequently when the right moment to execute the PM task gets closer (Einstein et al., 1995). In a time-based PM task, two different measures of clock checking behavior can be derived: absolute and strategic clock checking. Absolute clock checking is the absolute frequency with which the clock has been checked within the PM-critical time interval (i.e., if the PM task was to press a special key every five minutes, the critical time interval would be five minutes). Strategic clock checking considers that time monitoring is more efficient when it happens closer to the moment where the PM task shall be executed. It has been proposed to calculate a relative measure of clock checks by comparing the number of checks made closer to the PM execution time to the total number of checks in each block (Joly-Burra et al., 2022). Both absolute and strategic clock checking behaviors have been shown to be predictive of PM performance (Joly-Burra et al., 2022; Mioni & Stablum, 2014; Vanneste et al., 2016).

In event-based PM task, strategic slowing of ongoing task responding has been suggested as another task-specific strategic processes towards good PM performance. Specifically, slowing in ongoing task responding is often found when performances between conditions with and without an additional PM task are compared (Heathcote et al., 2015; Smith, 2003, 2010). Again, it might be plausible that any observed relationship between WM and time-based PM is due to the fact that people with higher WM capacity are more strategic in their ongoing task slowing and thus show better PM. Therefore, we also assessed and controlled for strategic slowing when investigating the relationship between time-based PM and WM. As a strategic slowing measure we used ongoing task response speed closer to the PM execution time and further compared it to response speed earlier in the block.

Open Science and data availability

All three experiments were approved by the Ethics Committee of Heidelberg University and conducted in compliance with the principles of the Declaration of Helsinki. All participants gave informed consent prior to their participation. All methods, hypotheses and analyses were preregistered prior to data collection on the OSF (https://osf.io/2yp4e/). Deviations from preregistered analyses are reported and explained when the analyses are introduced in each experiment. All materials, data and analysis code that we produced ourselves are provided via the OSF (link above). We do not provide program code for the WM tasks we used in Experiment 2 because it was programmed by Lewandowsky et al. (2010).

Experiment 1

In the first experiment, we examined the influence of WM capacity load imposed by the ongoing task on time-based PM performance. We realized three different load conditions and also assessed time monitoring in terms of clock checking during the ongoing task.

Method

Participants

A total of 153 young adults participated in this study, either for course credit or monetary compensation. Six participants were excluded because of software failure, because they did not adhere to the instructions, because they deliberately did not complete the PM task or withdrew their consent after participation, resulting in a sample of 147 participants (18–34 years, $M_{age} = 22.13$, $SD = 3.53$; 70.7 % female). Sample size was planned based on the recommendations for sufficient sample sizes for multilevel modeling provided by Maas and Hox (2005), indicating that a sample size of 50 participants per group can yield accurate estimates of regression coefficients and variance components. Additionally, the decision was informed by the simulation studies conducted by Paccagnella (2011).

In a between-subjects design, we manipulated WM load in the ongoing task load across three groups: a high load ($N = 50$), a medium load ($N = 49$), and a low load group ($N = 48$). Participants were randomly assigned to a group.

Tasks

Ongoing Task. A word-picture matching task was used as an ongoing task in the present study (see Fig. 1). Participants performing this task had to determine whether a presented picture matched a highlighted word from a previous set of two, four or six words (depending on the assigned group). Ongoing task stimuli were derived from the ecological stimulus set by Moreno-Martínez and Montoro (2012). This set contains 360 pictures and matching words. Stimuli from the categories body parts, insects and weapons were excluded in anticipation of interference of arousal. We further excluded stimuli from the food and tree categories that were specific to Spain and possibly unknown to German participants leaving a sample of 293 words and matching pictures. On each trial, a white screen was presented for variable amount of time (i.e., either 300, 400, 500, 600 or 700 ms) first. Then participants were shown a set of words for 1200 ms in black in front of a white background. The set size varied depending on the condition between two words (low load group); four words (medium load group); or six words (high load group). For all groups, after the set was presented, the location of one of the previously presented words was marked with a red cross, indicating that this word should be evaluated against the picture shown afterwards. Half of the pictures matched the cued word, while the remaining half either matched a different word shown earlier in the current trial but wasn't marked (25 % of trials) or matched a word that was not presented in the current trial (25 % of trials). Both words and pictures were selected randomly from a list in each trial. The picture to match the previously shown word was on display until a response key was pressed. Participants were instructed to press the J-key for a match or the F-key for a non-match as quickly and accurately as possible.

PM Task. Participants were instructed to remember to press the Enter key and type in a word every 5 min during the ongoing task. They were instructed that, if the keypress for the PM task was on time, a response window would open and they would have to enter a target word (i.e.,"medal"). By adding this task we hoped to separate the PM component (remember the keypress) from the retrospective component (remember target word) (Cohen et al., 2003). The response window for the PM task was +/- 30 s around the target time point but participants were unaware of that. If the Enter key was pressed at any time outside the response window, the ongoing task was interrupted for 2 s to indicate that the keypress was registered but occurred outside the response window.

To monitor the time, participants could press the spacebar at any time during the ongoing task in order to display a screen clock ("00:00"). The number of screen clock showings was not limited but the ongoing task was interrupted while the screen clock was on display for 2 s.

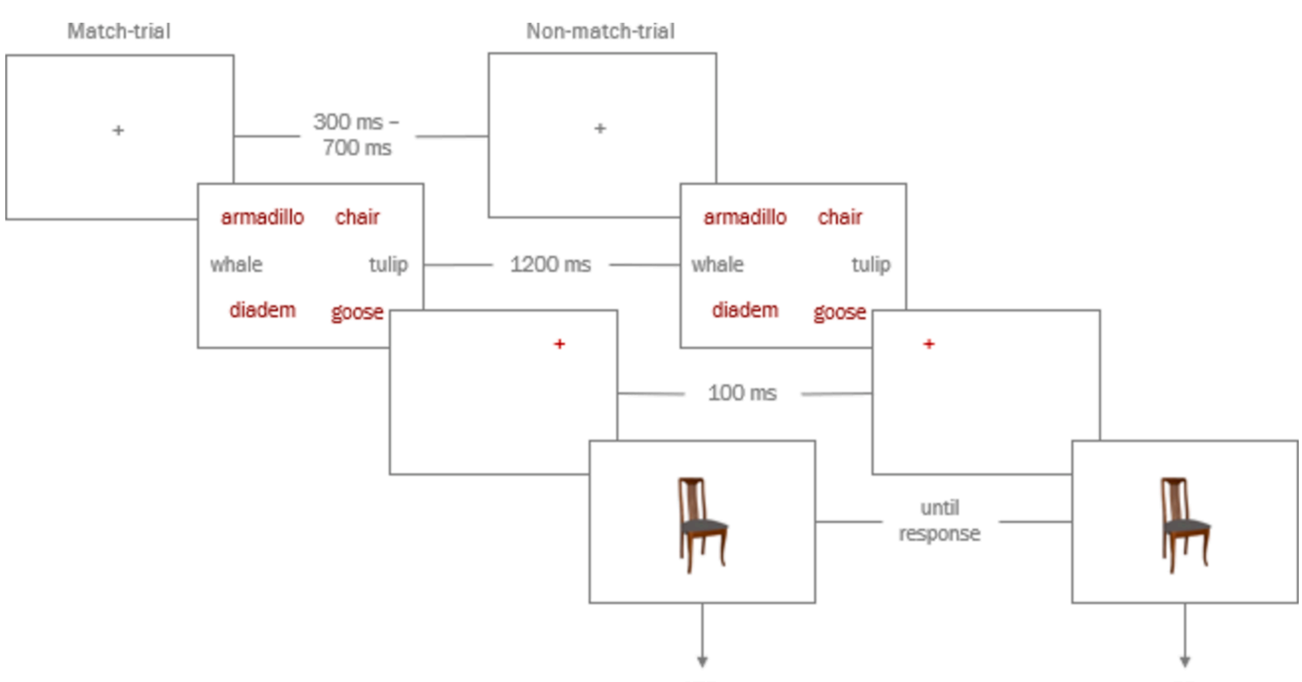

Fig. 1. Ongoing Task with Example of a Match- and Non-Match Trial. *Note.* Positions of the presented words in the different groups are indicated by different colors: Group 1 (low load): grey, Group 2 (medium load): red, Group 3 (high load): all, a 100-ms-interval was inserted between frames to prevent visual overlap.

Procedure

The study was advertised as a WM study without giving details about the PM task beforehand to avoid a primary focus on the PM task. The ongoing task and the PM task were introduced as equally important in the laboratory. The experiment was programmed in PsychoPy 2022.2.5 (Peirce, 2007). Participants were randomly assigned to one of three experimental groups. For all participants, the experiment started with instructions for the ongoing task and a practice block of 5 trials. Then participants continued with the ongoing task. The ongoing task consisted of 9 blocks each lasting 5 min making the experiments duration about 50 min including instructions. The first was a baseline block to familiarize participants with the ongoing task (not analyzed in the present study), the following eight consecutive blocks included the PM task. At the end of the PM experiment, participants were asked to rate how difficult they found it to remember the PM task on a 7-point Likert scale (from "not difficult" to "very difficult"). Additionally, they were asked to select a reason why they did not remember to execute the PM task every time (remembered PM task too late, forgotten, not executed on purpose, never forgotten to execute PM task).

Statistical analyses & dependent variables

We used the software R, version 4.2.2 (R Core Team, 2022) for data cleaning and analyses. We applied an alpha level of.05 for all statistical tests. We used accuracy rates and response times (RTs) as dependent variables in these analyses. For the RT analyses, only correct responses were analyzed and responses faster than 150 ms and those exceeding 2.5 standard deviations from participants' mean RT were excluded. For the analyses of ongoing task performance, we conducted one-way analyses of variance (ANOVAs) with load group as between-subjects factor and accuracies and RTs as dependent variables. In addition to preregistered analyses, whenever there was an overall effect of load, we followed up with a trend analysis to interpret the direction of the effect. This analysis tested whether performance would linearly change with increasing load from low to medium to high load, $\Psi_{\text{linear}} = (0.5, 0, -0.5)$, and whether the medium load group would deviate from the linear trend, $\Psi_{\text{quadratic}} = (-0.5, 1, -0.5)$.

For clock checking analyses, we also conducted one-way ANOVAs with load group as between-subjects factor and clock checking frequency and strategicness as dependent variables. Clock checking frequency was assessed as the number of spacebar presses per block. Contrary to preregistered analyses, instead of considering clock checks per minute as strategicness measures, we used a measure suggested by Joly-Burra et al. (2022). To this end, we further divided each 5-minute block into four segments of equal duration (S1- S4) and calculated strategicness by dividing the number of clock checks in the last segment before the PM timeframe (S4) by the total number of clock checks within the block and multiplied it by 100. This results in a measure ranging from 0 to 100 which we considered as a predictor in the PM analyses. In doing so, we are able to account for individual differences in clock checking frequency. Furthermore, this measure better captures the temporal dynamics of clock checking than traditional measures, especially in relation to proximity to the PM task (see also Joly-Burra et al., (2022) for additional advantages of this measure) (Joly-Burra et al., 2022; Laera et al., 2024).

We conducted exploratory analyses to test for a relationship between PM performance and strategic ongoing task slowing. The analysis focused on blocks with successful PM performances because PM performance was generally very high in the low and medium load groups and we thus did not have enough observations of PM misses. In each block, we considered mean RTs in S3 and S4 as a dependent variable and conducted a 3 x 2 repeated measures ANOVA with load group (low, medium and high load) as between-subjects factor and RTs in S3 and S4 as within-subject factors.

For further analyses of PM performance, we conducted a Fisher's exact test to assess whether there was a difference between remembering the PM task to press the Enter key within the right time window (prospective component) and remembering the word correctly (retrospective component). The result of this exploratory analysis suggested that there was no difference in prospective and retrospective performance, that is, that if participants remembered the PM task, they also remembered to type in the correct word. One participant appeared to have misunderstood the instructions and wrote random words as the PM task instead of the target word every time. One participant wrote no word instead of the target word twice and another participant wrote no word once, but in every other successful PM trial the target word was remembered successfully. Therefore, contrary to the preregistered analysis, we used only the prospective instead of the retrospective component as a measure of PM performance, because it directly corresponds to the primary focus of our research question, which revolves around understanding and predicting the prospective component of PM performance. Generalized linear mixed-effect models (GLMM) were utilized, to account for random effects and the binary nature of PM performance with a random intercept for participants (1 | participant), and the binary-coded PM performance (1 = successful intention fulfillment, 0 = unsuccessful intention fulfillment) as dependent variable. We used the lme4 package, version 1.1–33 (Bates et al., 2015). As we found a general trend in improvement in PM performance throughout the experiment, as preregistered, block number was introduced as a control variable (coded numerically and centered, −3.5, −2.5, −1.5, −0.5, 0.5, 1.5, 2.5, 3.5 for blocks 1–8, respectively).

To assess load effects on the PM performance, we applied the same contrasts as for the ongoing tasks analyses—namely, $\Psi_{\text{linear}} = (0.5, 0, -0.5)$ and $\Psi_{\text{quadratic}} = (-0.5, 1, -0.5)$. We used the z-test of regression weights to evaluate significance of model predictors. To test whether strategic clock checking was functionally related to PM performance, we included z-standardized strategic clock checking as an additional block-level predictor in the PM performance analysis in a stepwise fashion and compared the models using the chi-square test for model comparison to identify the model with the best fit. We calculated McFadden's R^2_{pseudo} as a measure for the variance in the dichotomous criterion explained by the predictors (McFadden, 1973). This measure can be interpreted similar to R^2. According to McFadden (1977) values above 0.2 to 0.4 indicate an excellent model fit. To maintain methodological consistency, we deviated from the preregistration and performed frequentist analyses instead of a Bayesian approach across all experiments.

Results & discussion

Mean accuracies, RTs of the ongoing task, clock checking strategicness, clock checking frequencies, and frequencies of correct responses in the PM task, are displayed in Table 1. We found a significant load effect on ongoing task accuracy, $F(2, 144) = 91.28$, $p < .001$, $\eta^2 = .56$, with a the linear trend, $t(144) = 13.44$, $p < .001$, whereas the quadratic trend was not significant, $t(144) = 1.57$, $p = .12$, suggesting a linear decline in accuracy with increasing ongoing task difficulty. We found a load effect on RTs, $F(2, 144) = 12.31$, $p < .001$, $\eta^2 = .15$, displaying both a significant linear trend, $t(144) = -4.43$, $p < .001$, and quadratic trend, $t(144) = 2.18$, $p = .03$, indicating a curve-linear effect with a stronger increase from low to medium than from medium to high load group. These results suggest that load has detrimental effects on RTs, and the curve-linear trend implies that while RTs generally increase with load, this effect becomes less pronounced as load reaches higher levels. Clock checking strategicness did not differ between load groups, $F(2, 144) = 0.54$, $p = .58$. Clock checking frequency varied with load group, $F(2, 144) = 5.27$, $p = .006$, $\eta^2 = .07$, with a significant linear trend, $t(144) = 3.08$, $p = .002$, and a non-significant quadratic trend, $t(144) = -0.99$, $p = .32$. These results suggest that participants check the clock less frequently, when the ongoing task load is increased while remaining equally strategic in their clock checking behavior across groups. Thus, it is possible that ongoing task load only influences the number of clock checks while having no effect on the strategicness with which participants check the clock.

Table 1
Mean RTs, Accuracies, PM Performance and Clock Checking by Group.

Load	Mean Acc (%)	SD Acc	Mean RT (ms)	SD RT	Mean CC strat.	SD CC strat.	Mean CC freq.	SD CC freq.	Mean PM freq.	SD PM freq.
Low	95.60	3.10	721.94	180.32	54.53	13.42	5.44	3.59	7.54	1.42
Medium	88.76	8.05	891.34	229.75	51.60	13.36	4.31	2.36	7.33	1.53
High	78.46	6.62	904.70	194.03	52.39	16.25	3.98	2.65	6.75	1.70

Note. N (low load) = 50, N (medium load) = 49, N (high load) = 48, Acc = accuracy, CC strat. = clock checking strategicness, CC freq. = clock checking frequency, PM freq. = frequency of correct responses in the PM task with a maximum of 8.

When analyzing ongoing task slowing, we found a significant effect of group, $F(2, 142) = 11.47$, $p < .001$, $\eta^2_G = .14$, indicating that there were differences in RTs in S3 and S4 among the three load groups. These findings are in line with the differences in RTs between groups reported earlier. Similarly, a significant effect of segment was found, $F(1, 142) = 15.84$, $p < .001$, $\eta^2_G = .003$. Post-hoc analyses suggest that there were differences in RTs between segment "S3" and segment "S4" which is immediately prior the PM intention fulfillment. The interaction between group and segment was not significant, $F(2, 142) = 0.67$, $p = .514$, suggesting that the amount of slowing prior to the moment of PM execution did not vary significantly between groups (Fig. 2).

In a next step, we used GLMMs to analyze the dependent variable PM performance. Our baseline model included the random intercept for participant and the detrending factor of block number, which was significant for all reported models. For the following models, we added the predictors load group, strategic clock checking and their interaction in a stepwise fashion. First, we specified an experimental effects model with the factor load group and found a significant linear trend, $b = 0.97$, $SE = 0.33$, $z = 2.91$, $p = .004$, but no quadratic trend, $b = -0.13$, $SE = 0.56$, $z = -0.23$, $p = .819$. The comparison to the baseline model showed a significant difference in model fit in favor of our experimental effects model, $\chi^2(2) = 8.93$, $p = .01$; $R^2_{\text{pseudo}} = .04$. These results indicate an influence of load group on PM performance, namely that our load manipulation results in a decrease in PM performance when ongoing task load increases. However, the low model fit also indicates that large proportions of the variance in PM remained unexplained. Second, we included the additional block-level predictor strategic clock checking. The linear effect of load group remained significant, $b = 0.730$, $SE = 0.364$, $z = 2.003$, $p = .045$. Strategic clock checking had a significant effect as well, $b = 1.988$, $SE = 0.214$, $z = 9.296$, $p < .001$. We compared this model to the experimental effects model and found a significant difference in model fit in favor of our model including the block level predictor, $\chi^2(1) = 168.46$, $p < .001$; $R^2_{\text{pseudo}} = .32$. These results suggest that strategicness in clock checking during the ongoing task has an influence on PM performance in addition to the load manipulation. Finally, we also allowed for the interactions between the significant predictors and tested whether this model would fit the data better than the model without interactions. The chi-square test showed a significant difference in model fit, $\chi^2(2) = 12.09$, $p = .002$; $R^2_{\text{pseudo}} = .34$, in favor of our final model including block level predictor and interactions. We found significant effects for clock checking strategicness, for the interaction between clock checking strategicness and the linear trend in the load group and for the detrending variable block number, while the linear trend alone was no longer significant (Table 2) with still excellent model fit.

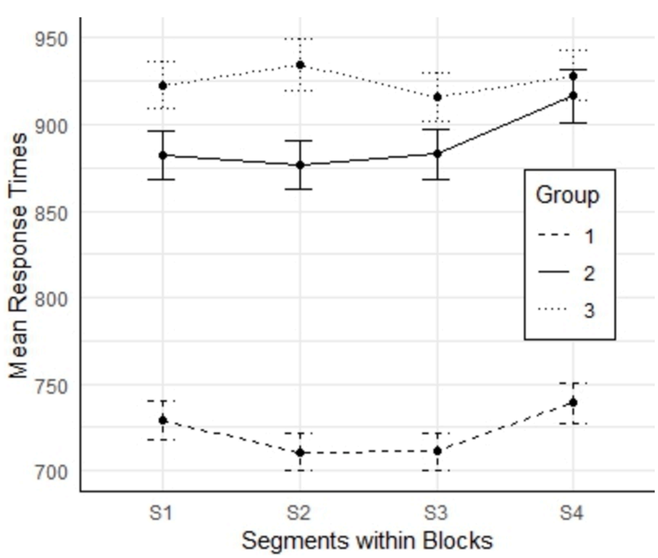

Fig. 2. Mean Response Times in the Ongoing Task in each Segment depending on Load Group. *Note.* Mean RT in milliseconds, only blocks with successful PM performance, low load (1) = dashed line, medium load (2) = solid line, high load (3) = dotted line, including standard errors.

The significant interaction between load group and clock checking strategicness indicates a different influence of strategic clock checking on PM performance for the different load groups. As evident from Fig. 3 the influence of clock checking strategicness became stronger when ongoing task load increased. This suggests that participants may rely more on strategic clock checking to manage their time more effectively when facing high ongoing task load. Although a similar pattern can be seen for absolute clock checking (Fig. 4), where the influence of absolute clock checking again became stronger with increasing load, absolute clock checking lacks the differentiated temporal focus seen in strategic time monitoring. Strategic clock checking not only reflects when participants check the clock relative to the target time but also accounts for individual differences in time monitoring behavior. Given the variability in clock checks across participants (ranging from 0 to 28 clock checks per block, for means for each group: see Table 1), strategic clock checking provides a better interpretable and standardized measure, allowing for a better comparison between participants.

Experiment 2

In Experiment 1, we found that time-based PM performance decreases when WM load increases. Despite the fact that we used a rather unspecific load manipulation, McFadden's pseudo R^2 for the final model suggested a good model fit, $R^2_{\text{pseudo}} = .340$. However, a further improvement in model fit may be achieved by considering individual differences in different WM sub-processes. To shed light on the question of which WM sub-processes contribute to this effect, we conducted a second study in which we adopted an individual difference approach. It has been shown, that WM capacity in general (Ball et al., 2019; Smith & Bayen, 2005), and specifically better binding (Lecouvey et al., 2017; Morand et al., 2021) and updating abilities (Kliegel et al., 2002; Mäntylä et al., 2007; Mioni & Stablum, 2014; Voigt et al., 2014; Yang et al., 2011), are associated with time-based and event-based PM performance (but see also Schnitzspahn et al., 2013). Based on this previous research, one can assume that the WM-load-induced time-based PM decrements we observed in Experiment 1, may have been either due to a load effect on binding, updating, or both binding and updating processes. Therefore, in Experiment 2, we had participants perform the high-load time-

Table 2
Fixed Effects of the Best Fitting Model with Block-Level Indicators and Interactions.

	b	*SE*	*z*	*OR*	*CI*	*p*	
Intercept	4.755	0.469	10.138	116.20	46.34 – 291.38	<.001	***
Linear trend	0.254	0.363	0.701	1.29	0.63 – 2.63	.483	
Quadratic trend	−0.285	0.633	−0.451	0.75	0.22 – 2.60	.652	
CC strat.	1.756	0.214	8.217	5.79	3.81 – 8.80	<.001	***
Detrended block number	0.238	0.073	3.243	1.27	1.10 – 1.46	.001	**
Linear trend: CC strat.	−0.851	0.240	−3.550	0.43	0.27 – 0.68	<.001	***
Quadratic trend: CC strat.	−0.372	0.432	−0.861	0.69	0.30 – 1.61	.389	

Note. Linear trend = linear relationship of PM performance for load groups, quadratic trend = quadratic relationship of PM performance for load groups, CC strat. = z-standardized measure of strategic clock checks in each block, detrended block number = block number as control variable.

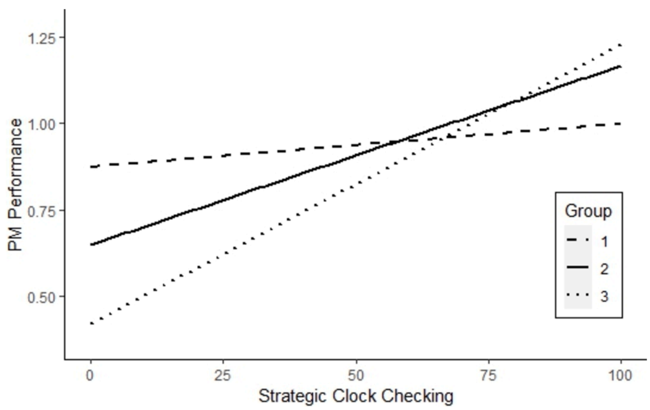

Fig. 3. Regression Lines for PM Performance Depending on Strategic Clock Checking and Ongoing Task Load. *Note.* Strategic CC in percent, binary PM performance (1 = successful intention fulfillment, 0 = unsuccessful intention fulfillment), low load (1) = dashed line, medium load (2) = solid line, high load (3) = dotted line.

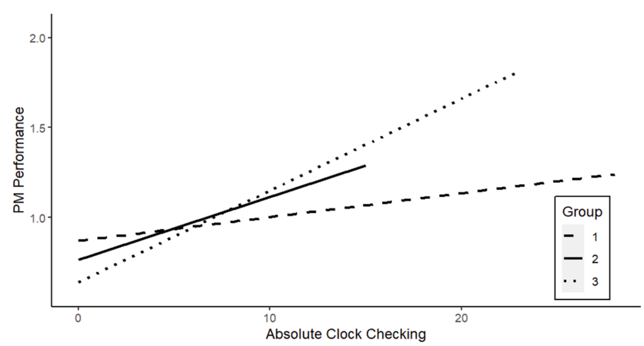

Fig. 4. Regression Lines for PM Performance Depending on Absolute Clock Checking and Ongoing Task Load. *Note.* Absolute number of clock checks, binary PM performance (1 = successful intention fulfillment, 0 = unsuccessful intention fulfillment), low load (1) = dashed line, medium load (2) = solid line, high load (3) = dotted line.

based PM task from Experiment 1 and also assessed WM capacity of these participants with multiple tasks. Using multiple tasks to assess WM capacity can reduce the influence of task specific variance and error variance, thereby increasing the reliability and validity of the measure. The variance observed in performance on one task may be more reflective of the task specific demands rather than the underlying construct of WM capacity. By using multiple tasks, we aim to capture a broader range of cognitive processes involved in WM and thereby gain a more nuanced understanding of an individual's WM abilities (Lewandowsky et al., 2010). To investigate the specific WM processes relevant for PM, we used a binding and an updating task. Binding tasks and updating tasks load high on the general WM capacity factor (Oberauer, 2005; Wilhelm et al., 2013). So, we also employed an operation span task as a widely utilized as a reference task to measure WM capacity (Lewandowsky et al., 2010).

With this study design, we aimed to test whether these abilities would be associated with time-based PM and, if so, whether their association would differ in magnitude. In addition to WM capacity, previous studies suggest that conscientiousness (Arana et al., 2008) and self-reported PM abilities (Rummel et al., 2019) can explain differences in event-based PM performance. To test whether this would also hold for time-based PM and to potentially increase the proportion of explainable variance in time-based PM performance, we assessed these variables as well.

Method

Participants

Either for course credit or monetary compensation, 145 young adults who had not participated in Experiment 1, participated in this study. Five participants were excluded prior to analyses because of software failure or because they withdrew consent after participation. Two participants were excluded because they deliberately ignored the PM task while performing the ongoing tasks. Following the exclusion criteria proposed by Lewandowsky et al. (2010), six participants scoring zero percent in the WM tasks in either the updating or operation span tasks were also excluded. Additionally, one participant's performance in the operation span task was more than 3 standard deviations below the sample mean. However, as their performance in other WM tasks was satisfactory, all observations from this participant were retained. This left a sample of 132 participants (18 – 34 years, M_{age} = 22.52, *SD* = 3.14; 74.2 % female).

Sample size was determined based on the data from Experiment 1 using the shiny app by Murayama, Usami and Sakaki (2022). The objective was to achieve a statistical power of.80 for predictors with *z*-values equal to or exceeding 3.00, with a significance level of.05. This analysis resulted in a recommended sample size of 133 participants.

Tasks

All WM tasks were programmed and run in MATLAB R2018b (The MathWorks Inc, 2018). Data from the questionnaires for conscientiousness and self-reported PM performance were collected online in the lab with SoSci Survey (Leiner, 2023). The PM task was again programmed in PsychoPy 2022.2.5 (Peirce, 2007).

Updating Task. This task is one of the four tasks included in the WM test battery by Lewandowsky et al. (2010). In this task, participants had to remember an initial set of digits displayed in separate frames on a screen and update them using arithmetic operations. On each trial, a set of frames of different set sizes (three, four or five frames) each with two, three, four, five, and six updating operations was presented. Resulting in a total of 15 test trials preceded by two practice trials. The first digits were displayed in frames for 1 s each. Then, cues for arithmetic operations such as "- 3″ or "+ 5″ appeared for 1300 ms each, followed by a 250 ms interval. Participants were to apply the shown operations to the previously displayed digits in the same frame and to replace the content mentally with the result. After the updating operations, question marks appeared in the frames indicating the final recall. Operations ranged from − 7 to + 7, excluding 0, with intermediate and final results between 1 and 9. The set of initial digits, operations, updating, recall prompt, and trial orders were randomly generated but held constant for all participants to control for between-subjects error variance (Goodhew & Edwards, 2019).

Operation Span Task. This task is also one of the four tasks included in the WM test battery by Lewandowsky et al. (2010). In the operation span task, developed by Turner and Engle (1989), participants alternately were to evaluate the correctness of arithmetic equations (e.g. 4 + 2 = 6) and to remember the subsequent consonants (except Y and Q). Trials began with a 1.5-second fixation cross followed by an equation. Participants were to judge the equation by responding within 3 s with the right arrow key for correct and the left arrow key for incorrect. After the equation disappeared, a consonant appeared for 1 s and had to be encoded for later recall. After 4 to 8 consonants were presented, participants were to recall the letters in the presented order. Trials were separated by a 500 ms interval, with pauses every 3 trials. There were 15 trials: 3 per trial list length (number of letters to be remembered). The equations featured numbers from −9 to 10, except 0, with results from 1 to 20. Half of the equations were correct. Three practice trials (trial list length 3, 4, 5 letters) preceded the experimental trials. Letter sequences, equations and trial order were in a fixed random order for all participants. The consonant lists had no repetitions.

Binding Task. We used the location-letter binding task, where participants remembered the positions of letters sequentially presented in a 3 × 3 grid. This task was derived from Wilhelm et al. (2013). During recall, participants were either presented with a letter and had to indicated the letter's location within the grid or were presented with a location and had to selected the correct letter for this specific grid location. The complexity varied from 2 to 6 location-letter pairs, with a total of 14 trials. Stimuli appeared for 1500 ms with a 500 ms inter stimulus interval. There was no time limit for the response.

Conscientiousness. To measure conscientiousness, we used the eight respective items of the German version of the International Personality Item Pool (IPIP40). It is an open access measure that has high convergent validities with the German version of the NEO-FFI (Hartig et al., 2003).

Self-Reported PM Abilities. We measured self-reported PM performance by using the Short Version of the Metacognitive Prospective Memory Inventory (MPMI-s). This questionnaire measures individual differences in self-reported PM abilities as well as the utilization of both internal and external strategic approaches to PM tasks in everyday life. (Rummel et al., 2019).

Procedure

The study was advertised as a WM study without giving details about the PM task beforehand to avoid a primary focus on the PM task. Participants first performed the WM tasks explained above in the order in which they were described. Next, participants performed the ongoing task with the embedded PM task which was very similar to the difficult condition in Experiment 1 with a few changes. In contrast to Experiment 1, the picture corresponding to the previously shown word remained visible for three seconds rather than indefinitely. This adjustment was made to prevent participants from taking breaks during the ongoing task. Trials in which participants did not respond within the three second time window were considered as incorrect. To increase motivation, participants could earn an extra 2 € if their accuracy in the ongoing task exceeded 75 %. This time, instead of the Enter key, we used the "a" key as the PM response key after the 5-minute intervals, to reduce the prominence of the PM task. The retrospective component in the PM task to enter a target word was no longer included as the analyses in Experiment 1 did not reveal a difference between remembering the intended action and executing it correctly. Instead, we assessed PM performance, as commonly done in the Einstein-McDaniel paradigm, by asking participants to simply press a special key at designated times. For correct PM responses, participants received feedback on the screen, that their keypress was successful and automatically returned to the next block of the ongoing task. The ongoing task with the embedded PM task lasted 45 min and consisted of a 5-minute baseline block followed by eight 5-minute PM blocks again. Lastly, participants completed the IPIP-40 and the MPMI-s questionnaires.

Statistical analyses & dependent variables

We used the software R, version 4.2.2 (R Core Team, 2022) for data cleaning and analyses and followed the data cleaning protocol established in Experiment 1. Performance in the WM tasks was scored as the proportion of correctly remembered items using partial credit scoring. For this purpose, the accuracy in each trial was used and then task performance was calculated as the mean value of these partial scores for each task individually. As a measure of WM capacity, we calculated a WM composite score by using the mean performance across all three WM tasks. This results in a measure ranging from 0 to 100 with higher values reflecting higher WM capacity. We also considered binding and updating performance individually.

RTs and accuracy in the ongoing task and clock checking were calculated as in Experiment 1. Instead of the preregistered linear mixed models, we used Spearman's rank correlation test to analyze the strength and direction in the relationship between WM capacity and ongoing task performance, strategic clock checking and clock checking frequency. This decision was based on the non-normal distribution of residuals in the LMM analysis. To analyze strategic slowing we used a similar approach as in Experiment 1. A paired-sample *t*-test using only blocks with successful PM performances was conducted. Within each block, mean RTs in the ongoing task in S3 and S4 were compared. Additionally, because of sufficient data with unsuccessful PM performance, we compared RTs observed in S4 between cases where PM performance was successful and cases where it was unsuccessful. The Wilcoxon signed-rank test was employed to address the remaining imbalance between the numbers of successful and unsuccessful trials.

To analyze PM performance, we utilized GLMMS, to allow for interactions and account for the binary nature of PM performance with a random intercept for participants, and the binary-coded PM performance (1 = successful intention fulfillment, 0 = unsuccessful intention fulfillment) as dependent variable. Again, we found a general trend with an improvement in PM performance throughout the experiment. Consequently, we introduced block number as a detrending variable like we did in Experiment 1. The same strategies as in Experiment 1 to evaluate significance of model predictors and identify the model with the best fit were used. To test which predictors are related to PM performance, we added the predictors WM composite score, the block-level predictor strategic clock checking, the conscientiousness measure, self-reported PM abilities and, in addition to the preregistered analyses, log RTs in S4 in a stepwise fashion to our model. Like in Experiment 1, we evaluated model fit based on McFadden's pseudo R^2_{pseudo}.

Results & discussion

Descriptive statistics for ongoing task performance, clock checking, PM performance, the WM measures, and non-cognitive measures are displayed in Table 3.

WM composite score and ongoing task accuracy showed a significant moderate positive correlation, $r_s(130) = .54$, $p < .001$. These results suggest that participants with higher WM capacity perform more accurately in the ongoing task. Conversely, WM composite scores and RTs in the ongoing task were not correlated, $r_s(130) = -.18$, $p = .983$, indicating that WM capacity was not associated with the speed of ongoing task performance. Furthermore, we found a significant positive correlation between WM composite score and both strategic clock checking, $r_s(130) = .19$, $p = .016$, and absolute clock checking $r_s(130) = .19$, $p = .016$. These results suggest that participants with higher WM capacity tend to

Table 3
Descriptive Statistics for all Measures.

	Mean	SD	Median	Min.	Max.	Skew	Kurtosis	SE
OT Accuracy	0.77	0.07	0.78	0.58	0.91	−0.53	−0.51	0.01
OT RT	844.30	167.54	807.07	541.49	1458.47	0.93	0.96	14.58
Strategic CC/block	54.97	33.00	50.00	0	100.00	−0.25	−0.87	1.02
Absolute CC/block	3.59	2.63	3.00	0	25.00	1.63	6.57	0.08
PM Performance	6.61	1.92	7.00	0	8.00	−1.81	2.70	0.17
Binding	0.90	0.09	0.93	0.60	1.00	−1.38	1.38	0.01
Updating	0.63	0.20	0.65	0.06	1.00	−0.25	−0.54	0.02
Operation Span	0.79	0.13	0.81	0.09	1.00	−1.67	5.41	0.01
WM Composite Score	0.77	0.12	0.79	0.45	0.99	−0.57	−0.11	0.01
Conscientiousness	3.44	0.67	3.38	1.50	4.88	−0.34	−0.17	0.06
Self-report PM Ability	3.89	0.62	4.00	1.88	5.00	−0.90	0.83	0.05

Note. OT accuracy = ongoing task accuracy, OT RT = ongoing task response time, strategic and absolute clock checking considered on a block level.

check the clock more often and more strategically. A full correlation matrix including reliability estimates for all measures can be found in Table 4.

The paired *t*-test to analyze slowing from S3 to S4 for successful PM blocks revealed no significant slowing effects, $t(872) = -1.89$, $p = .058$. Thus, while these results suggest a potential trend towards slower RTs in S4 compared to S3, these effects were not statistically significant in this sample. The Wilcoxon signed-rank test revealed a significant difference between RTs in S4 when PM fulfillment was successful, $M = 833.49$, $SD = 200.31$, compared to when it was forgotten, $M = 800.53$, $SD = 182.60$, $p < .001$ (Fig. 5). These findings indicate that RTs in segment S4 varied significantly between cases where PM performance was successful and those where it was unsuccessful, suggesting that people tended to slow down when approaching the time of intention execution.

In a next step, we used GLMMs to analyze the dependent variable PM performance. Our baseline model included the random intercept for participant and the detrending factor of block number. First, we specified a WM composite model and found a significant effect for WM composite score, $b = 4.836$, $SE = 1.553$, $z = 3.114$, $p = .002$. The comparison to the baseline model showed a significant difference in model fit in favor of our WM composite model, $\chi^2(1) = 9.576$, $p = .002$, $R^2_{\text{pseudo}} = .04$. These results suggest that PM performance is associated with the WM composite score. However, the specific component responsible for this association remains unknown. In the subsequent model, we replaced the WM composite score with the updating and binding scores as individual predictors. We found a significant effect of the updating component, $b = 3.716$, $SE = 1.094$, $z = 3.395$, $p < .001$, while the binding component was not significant, $b = -1.546$, $SE = 2.307$, $z = -0.670$, $p = .503$. The model comparison between the WM composite model and the individual WM components model showed a significant difference in model fit in favor of the individual WM components model, $\chi^2(1) = 4.030$, $p = .045$, $R^2_{\text{pseudo}} = .04$. These findings suggest that the updating component has a positive relationship with PM performance whereas the binding component has not. This pattern is in line with findings in previous studies highlighting the importance of updating abilities in PM (Kliegel et al., 2002; Mäntylä et al., 2007; Mioni & Stablum, 2014; Voigt et al., 2014; Yang et al., 2011; Zuber & Kliegel, 2020). As evident from Table 4, the reliability of the updating measure

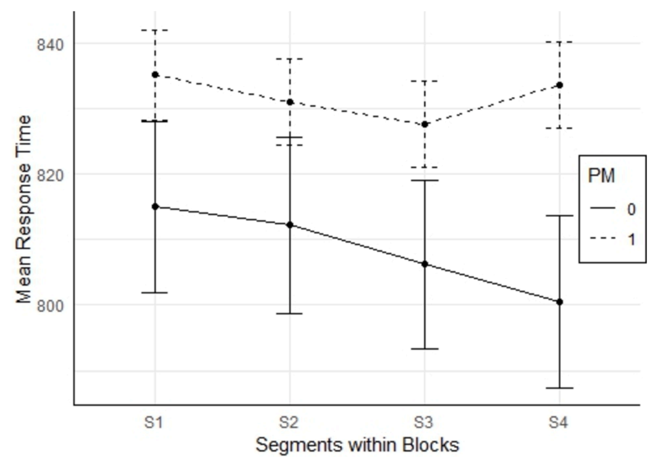

Fig. 5. Mean Response Times in the Ongoing Task in each Segment depending on PM Fulfillment. *Note.* Mean RT in milliseconds, separate for blocks with successful PM performance (PM = 1) = dashed line, and unsuccessful PM performance (PM = 0) = solid line, including standard errors.

Table 4
Full Correlation Matrix for all Measures including Reliabilities.

	OT Accuracy	OT RT	Strategic CC	Absolute CC	PM Performance	Binding	Updating	Operation Span	WM Composite Score	Conscien-tiousness	Self-report PM Ability
OT Accuracy	**.96**										
OT RT	.20	**.99**									
Strategic CC	.11	.06	**.76**								
Absolute CC	.06	−.03	.38	**.95**							
PM Performance	.17	.09	.60	.56	**.79**						
Binding	.47	−.07	.08	.07	.08	**.79**					
Updating	.52	−.13	.22	.20	.26	.55	**.94**				
Operation Span	.35	−.22	.06	.20	.19	.46	.66	**.85**			
WM Composite Score	.54	−.18	.19	.19	.21	.70	.94	.82	–		
Conscientiousness	.05	.00	−.11	−.17	−.20	.07	−.07	−.06	−.06	**.81**	
Self-report PM Ability	.06	−.02	−.04	−.06	.01	.09	.05	.08	.07	.39	**.76**

Note. OT accuracy = ongoing task accuracy, OT RT = ongoing task response time; the bold values on the diagonal represent Spearman-Brown corrected split half correlations as estimates of reliability, except for Cronbach's α for conscientiousness and self-reported PM ability. No reliability is reported for the WM composite score, as it is an average of all WM measures.

was higher than that of the binding measure (.94 vs .79). To address this difference, we applied an attenuation correction to the correlation of all three WM measures to estimate the correlation with PM performance assuming perfect reliability of the former measures. The corrected correlations increased from .26 to .30 for updating capacity, from .08 to.10 for binding capacity, and from .19 to.23 for operation span. Since the rank order of the correlations remained unchanged, we conclude that the differences in predicting PM performance between updating and binding are unlikely to be solely due to differences in reliability but nevertheless these differences should be taken into account when interpreting the present results.

In a block-level predictor model we added strategic clock checking and log RTs in S4 to our previous model. There was a significant effect of strategic clock checking,[1] $b = 2.112$, $SE = 0.168$, $z = 12.560$, $p < .001$, whereas RTs in S4 did not predict PM performance, $b = 1.072$, $SE = 0.637$, $z = 1.683$, $p = .092$. We found a significant increase in model fit compared to the individual WM components model, $\chi^2(2) = 246.98$, $p < .001$, $R^2_{\text{pseudo}} = .34$. Also, the overall model fit substantially improved. This replicates the clock checking results of Experiment 1, highlighting the importance of strategic clock checking as a predictor for PM performance. Even though we found significantly slower RTs in S4 if the PM task was remembered in the Wilcoxon signed-rank test reported above, it remains unclear whether these effects are functionally related to PM, as they are not a significant predictor of PM performance. Possibly, they are related to clock checking behavior which tends to increase towards the end of a block. In a final step, we introduced conscientiousness and self-reported PM ability as additional predictors to the previous model. Neither conscientiousness, $b = -0.079$, $SE = 0.255$, $z = -0.311$, $p = .756$, nor self-reported PM abilities, $b = -0.009$, $SE = 0.257$, $z = -0.034$, $p = .973$, were significant predictors. Model comparison revealed no significant improvement in model fit, $\chi^2(2) = 0.130$, $p = .937$, $R^2_{\text{pseudo}} = .34$, thus confirming the block level predictor model as our final model (Table 5). This is contradictory to previous research that suggests an association between conscientiousness and time-based and event-based PM (Arana et al., 2008; Cuttler & Graf, 2007; Smith et al., 2011). We did not find a significant effect of self-reported PM ability on PM.

Experiment 3

The results of Experiment 2 suggest that WM capacity, and specifically updating abilities, are positively correlated with PM performance, as are variations in clock checking strategicness. However, given the correlational nature of Experiment 2, it is unclear whether updating abilities are a prerequisite for time-based PM performance. For this reason, we ran a third experiment in which we experimentally manipulated updating demands in the ongoing task and tested the effect on PM performance. Similar to Experiment 1, we manipulated the WM load imposed by the ongoing task between participants, to test whether PM performance would suffer from an increased load. We created two different WM load groups, one updating-specific load group and a capacity-load group in addition to a control group. In Experiment 2, we did not observe a correlation between conscientiousness and PM performance. Nevertheless, previous studies report correlations between Big Five factors and PM performance (Uttl et al., 2013). We thus again assessed Big Five personality traits, this time conscientiousness, agreeableness, and neuroticism to test whether any of these factors would additionally explain variance in PM performance outcomes.

Table 5
Fixed Effects of the Best Fitting Model with Block-Level Indicators.

	b	*SE*	*z*	*OR*	*CI*	*p*	
Intercept	−5.054	4.471	−1.131	0.01	0.00 – 40.47	.258	
Updating	2.274	0.874	2.601	9.72	1.75 – 53.91	.009	**
Binding	−0.693	1.816	−0.382	0.50	0.01 – 17.57	.703	
Strategic clock checking	2.112	0.168	12.560	8.27	5.94 – 11.49	<.001	***
RT in S4	1.072	0.637	1.682	2.92	0.84 – 10.17	.092	.
Detrended block number	0.014	0.059	0.234	1.01	0.90 – 1.14	.815	

Note. Updating and binding as individual predictors, strategic clock checking as a z-standardized measure of strategic clock checks in each block, slowing (RT in S4) and detrended block number as block level predictors.

Method

Participants

A new sample of 155 young adults participated in this study, either for course credit or monetary compensation. Seven participants were excluded either due to scoring below chance in the ongoing task, because they were unable to follow instructions, or discontinuing their participation prematurely. This resulted in a sample of 148 participants (18 – 33 years, $M_{\text{age}} = 22.7$, $SD = 3.11$; 77 % female). Sample size was planned based on the recommendations for sufficient sample sizes for multilevel modeling provided by Maas and Hox (2005), indicating that a sample size of 50 participants per group can yield accurate estimates of regression coefficients and variance components. Additionally, the decision was informed by the simulation studies conducted by Paccagnella (2011).

In a between-subject design with three groups we manipulated the level of updating resources demanded by the ongoing task ($N = 49$), the WM capacity load imposed by the ongoing task ($N = 52$) and added a control group ($N = 47$). Participants were randomly assigned to one of the groups.

Tasks

The experimental tasks were programmed and run with PsychoPy 2022.2.5 (Peirce, 2007). Personality data was collected after the PM experiment using SoSci Survey (Leiner, 2023).

Ongoing Task. Participants completed the word-picture matching task as ongoing task, asking them to determine whether a presented picture matched a word cued by a red cross. This task featured eight different locations on which words were presented and the presentation pattern varied between the experimental groups (Fig. 6b). Trial duration was extended in this experiment, because, on each trial, participants were presented with two sets of stimuli instead of one. In the updating group, participants saw an initial set of five words at randomly chosen positions for 4000 ms. These words were followed by a second set of three words which were presented at the same positions as three of the previously presented words for 2400 ms. Participants were asked to update the three words from the first set with the words from the second that were presented at the same positions. So, for the final test participants had to memorize a total of five words, two from the initial set and three from the updating set, while removing three outdated words (Fig. 6a, G1).

In the load group, participants were also presented with a first set of five words at five different locations and then with a second set of three words at new locations. Participants had to remember all eight words. For both the updating and load groups, half of the pictures matched the word (50 % first set, 50 % second set), 25 % of them did not match but showed another word presented in the trial (50 % first set, 50 % second set) and 25 % of them matched a word that was not presented in the current trial (Fig. 6a, G2). Participants in the control group were presented with an initial set of five words in five random locations. The

[1] The interaction between updating abilities and strategic clock checking was not significant.

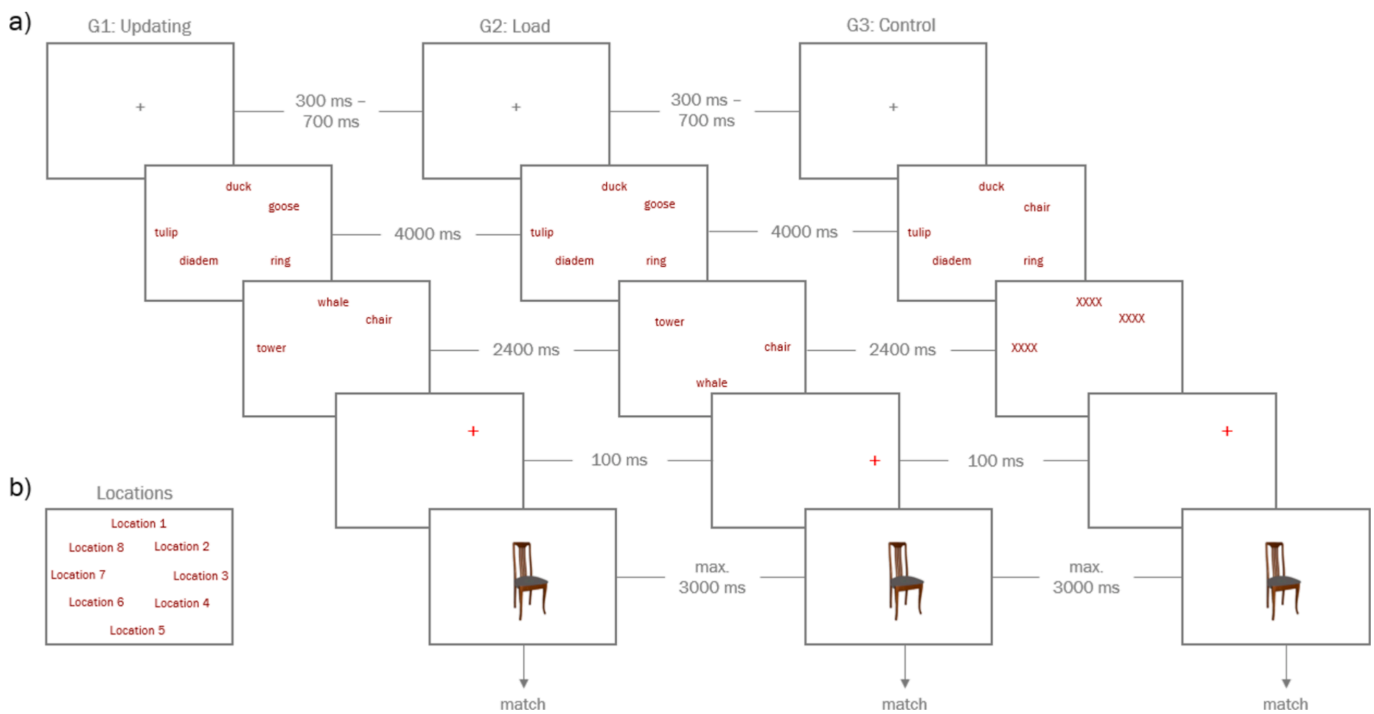

Fig. 6. Ongoing Task with Example of a Match Trial for each Group and Locations of Words. *Note.* a) In the updating group the three words in the second set replaced the ones presented before, in the load group more words were added in the second set, in the control group the "XXXX" stimuli were supposed to be ignored; all examples show match-trials, 100 ms interval between every frame to prevent visual overlap; b) Illustration of the eight locations for word presentation, locations are randomly selected in the first set.

second set featured three to be ignored ("XXXX") stimuli. These were either presented in locations of the previous words (conceptually matching the updating group) or in the locations that were not used in the initial set (conceptually matching the load group). Half of the pictures matched the word, 25 % did not match but showed a picture corresponding to another word in that set and 25 % matched a word that was not presented in the current trial (Fig. 6a, G3).

Personality assessment. To measure conscientiousness, agreeableness and neuroticism, we used the respective items of the German version of the IPIP40. This is an open access measure that has high convergent validities with the German version of the NEO-FFI (Hartig et al., 2003).

Procedure

The study was advertised as a WM study without giving details about the PM task beforehand to avoid a primary focus on the PM task. The ongoing task and the PM task were introduced as being equally important. Participants who already participated in Experiment 2 were not allowed to participate in this study. Participants had to perform the ongoing task for which we applied the changes described above. The PM task was embedded in this ongoing task, resulting in a 5-minute baseline block and 8 consecutive PM blocks. Duration and structure of the PM task remained the same as in the previous experiment. Participants were to press the "a" key every five minutes (i.e., at the end of each block).

Statistical analyses & dependent variables

We used the software R, version 4.2.2 (R Core Team, 2022) for data cleaning and analyses and, once more, followed the data cleaning protocol established in Experiments 1 and 2. For the analysis of strategic clock checking we used the same measure as in Experiment 1 and 2. In addition to preregistered analyses, we used ANOVAs to compare ongoing task performance, strategic and absolute clock checking with group as between-subjects factor, similar to the approach adopted for Experiment 1. Both the updating and the load capacity groups were presented with eight words in total. In the updating group, however, three words from the initial set were instructed to be replaced and forgotten. There is strong evidence that item replacement is highly effective in WM, meaning that they can be replaced often in a way the first presented words are fully removed from WM. The updating group should therefore be confronted with less capacity load (Dames & Oberauer, 2022). Whenever there was an overall effect of group, we followed up with a contrast analysis first testing for an updating specific effect by comparing the updating and the control groups, $\Psi_{\text{updating}} = (-0.5, 0, 0.5)$, and for a capacity load effect by comparing the updating and control condition against the load condition $\Psi_{\text{load}} = (0.5, -1, 0.5)$. To analyze strategic slowing, we implemented the same analysis as in Experiment 1, comparing mean RTs in the ongoing task in S3 and S4 with a 3 x 2 repeated measures ANOVA.

Again, GLMMs were utilized, to account for the binary nature of PM performance with a random intercept for participants, and the binary-coded PM performance (1 = successful intention fulfillment, 0 = unsuccessful intention fulfillment) as dependent variable. Again, we found general trend with an improvement in PM performance with time-on-task and thus a centered block variable was introduced as a detrending variable (see Experiment 1). To assess group effects on PM performance, we applied the same contrasts as for the ongoing tasks analyses—namely, $\Psi_{\text{updating}} = (-0.5, 0, 0.5)$ for an updating specific effect and $\Psi_{\text{load}} = (0.5, -1, 0.5)$ for a capacity load effect. To test which predictors are related to PM performance, we added the predictors contrasts for group effects, the block-level predictor strategic clock checking and the personality traits in a stepwise fashion. RTs in blocks with unsuccessful PM performances were not analyzed because of high PM performances in the capacity load and control groups, therefore leaving too little data to meaningfully compare the RTs in unsuccessful blocks. Measures to evaluate significance of the predictors, model fit and explained variance were used like in Experiment 1.

Results & discussion

Mean accuracies, RTs of the ongoing task, strategic slowing in the ongoing task, clock checking strategicness, clock checking frequencies, and frequencies of correct responses in the PM task, are displayed in Table 6. We found a significant effect of group on ongoing task accuracy, $F(2, 145) = 12.24$, $p < .001$, $\eta^2 = .14$. Contrast analyses showed that the accuracy in the groups with increased WM demands (updating and load condition) was lower than in the control group but that the other two experimental groups did not differ. We found a group effect on ongoing task RTs, $F(2, 145) = 4.68$, $p = .01$, $\eta^2 = .06$, which was due to a particularly fast responding in the updating group which was unexpected. Clock checking strategicness differed significantly between groups, $F(2, 145) = 6.84$, $p = .001$, $\eta^2 = .09$. Contrast analyses suggested that the updating group showed less strategic clock checking, $t(145) = 3.69$, $p < .001$. Absolute clock checking frequency did not differ significantly between groups, $F(2, 145) = 0.43$, $p = .65$.

When analyzing slowing in the ongoing task, we found a significant effect for group, $F(2, 143) = 3.57$, $p = .030$, $\eta^2_G = .05$, indicating that there were differences in RTs in S3 and S4 among the three groups. Similarly, a significant effect of segment was found, $F(1, 143) = 8.12$, $p = .005$, $\eta^2_G = .003$. These results suggest that there were differences in RTs between segment S3 and segment S4 which is immediately prior the PM intention fulfillment. The interaction between group and segment was not significant, $F(2, 143) = 0.96$, $p = .383$, suggesting that the effect of segment on RTs did not vary significantly between groups (Fig. 7). In sum, there was some evidence that participants slowed down strategically when time comes closer to fulfill the PM task but not differently so in the three experimental groups.

In a next step, we implemented GLMMs to analyze the dependent variable PM performance. The baseline model included a random intercept for participant and the detrending factor of block number. For the following models, we added the predictors group, strategic clock checking and the personality traits in a stepwise fashion. First, we specified an experimental effects model with the factor group and found a significant effect for the updating group, $b = 0.677$, $SE = 0.253$, $z = 2.680$, $p = .007$, but not for the load group, $b = -0.249$, $SE = 0.425$, $z = -0.586$, $p = .558$. The comparison to the baseline model showed a significant difference in model fit in favor of our experimental effects model, $\chi^2(2) = 7.83$, $p = .020$, $R^2{}_{\text{pseudo}} = .03$. These results indicate that our manipulation of updating demands negatively affected PM performance, while the load manipulation did not affect PM performance as compared to the control group. However, the low model fit also indicates that a large proportion of the variance in PM remains unexplained. Second, we included the additional block-level predictor of strategic clock checking. The effect of group was no longer significant, $b = 0.262$, $SE = 0.192$, $z = 1.365$, $p = .172$, while strategic clock checking had a significant effect, $b = 1.961$, $SE = 0.162$, $z = 12.109$, $p < .001$. We compared this model to the experimental effects model and found a significant difference in model fit in favor of the block-level predictor model, $\chi^2(1) = 205.88$, $p < .001$, $R^2_{\text{pseudo}} = .31$. These results suggest that strategic clock checking plays a crucial role for PM performance. Also, the overall model fit substantially improved. Finally, we introduced the personality traits as predictors to the model and found no significant effect for conscientiousness, $b = 0.167$, $SE = 0.223$, $z = 0.750$, $p = .454$, neuroticism, $b = 0.162$, $SE = 0.223$, $z = 0.729$, $p = .466$, or agreeableness, $b = -0.018$, $SE = 0.295$, $z = -0.061$, $p = .952$. The chi-square test

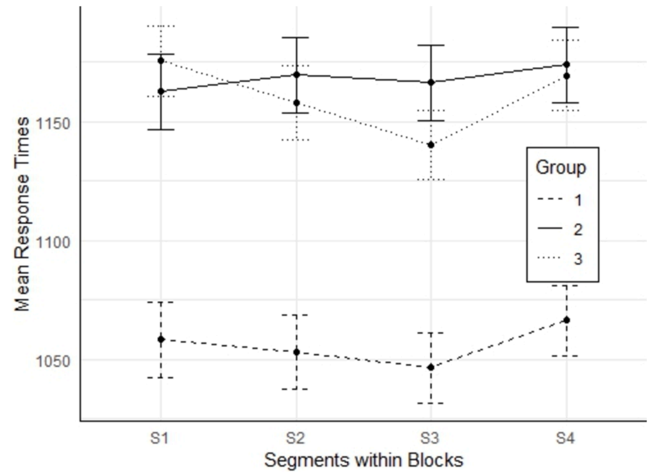

Fig. 7. Mean Response Times in the Ongoing Task in each Segment depending on Group. *Note.* Mean RT in milliseconds, only blocks with successful PM performance (PM = 1), updating group (1) = dashed line, load group (2) = solid line, control group (3) = dotted line, including standard errors.

showed no significant difference improvement in model fit, $\chi^2(3) = 0.945$, $p = .815$, $R^2_{\text{pseudo}} = .31$, confirming the block-level predictor model as the final model (Table 7).

The most variance in PM was explained by the clock checking model and in this model the group factor was longer significant. This result suggests, that as participants encounter increasingly demanding ongoing tasks, they seem to become less strategic in their clock checking behavior (although the absolute clock checking remains comparable with the other conditions). Thus, general cognitive ability and task-specific abilities are not independent of each other, despite the fact that they both contribute uniquely to overall PM performance as results of Experiment 2 suggested.

Table 7
Fixed Effects of the Best Fitting Model with Block-Level Predictor Clock Checking.

	b	*SE*	*z*	*OR*	*CI*	*p*	
Intercept	3.506	0.238	14.711	33.32	20.89 – 53.17	<.001	***
Updating contrast	0.262	0.192	1.365	1.30	0.89 – 1.89	0.172	
Capacity load contrast	−0.359	0.322	−1.113	0.70	0.37 – 1.31	0.266	
Strategic clock checking	1.961	0.162	12.109	7.10	5.17 – 9.76	<.001	***
Detrended block number	0.043	0.061	0.716	1.04	0.93 – 1.18	0.474	

Note. Updating contrast = updating specific effect, load contrast = capacity load specific effect, strategic clock checking = z-standardized measure of strategic clock checks in each block, detrended block number = block number as control variable.

Table 6
Mean RTs, Accuracies, PM Performance and Clock Checking by Group.

Load	Mean Acc (%)	SD Acc	Mean RT (ms)	SD RT	Mean CC strat.	SD CC strat.	Mean CC freq.	SD CC freq.	Mean PM freq.	SD PM freq.
Updating	75.80	7.97	1045.10	244.95	53.52	31.70	4.57	3.86	6.55	1.98
C. Load	75.34	6.63	1176.38	252.92	58.11	28.64	4.92	4.60	7.21	1.31
Control	82.25	8.18	1172.64	214.04	64.30	26.37	4.30	2.72	7.43	1.11

Note. N (updating) = 49, *N* (capacity load) = 52, *N* (control) = 47, Acc = accuracy, CC strat. = clock checking strategicness, CC freq. = clock checking frequency, PM freq. = frequency of correct responses in the PM task with a maximum of 8.

General discussion

Ensuring timely execution of intended actions is essential for everyday life. We conducted three studies to better understand the contributions of general cognitive abilities (WM) and task specific abilities (time monitoring) to successful time-based PM and the interplay of these processes. In our first experiment, we manipulated WM capacity load within the ongoing task. Results revealed a significant decrease in PM performance with increased WM demands in the ongoing task and further showed that strategic clock checking plays a role for time-based PM. The interaction between the load manipulation and strategic clock checking indicated that clock checking disproportionally decreases when WM load increases. In Experiment 2, we took an individual differences approach, measuring WM capacity through the processes of updating and binding and participants engaged in the high capacity load PM task from Experiment 1. We observed a significant association between WM capacity, particularly updating capacity, and PM performance and found a substantial part of the variance in PM performance to be explained by strategic clock checking. Importantly, WM updating remained a significant predictor of PM, in addition to strategic clock checking. Experiment 3, again, employed an experimental approach, manipulating updating resources and capacity load in the ongoing task. Although an effect of the updating load was evident, it vanished when group differences in strategic clock checking were additionally considered. Notably, no personality trait or self-reported PM ability showed association with PM performance across all experiments. Our findings consistently underscored time monitoring in terms of strategic clock checking as the most powerful predictor of time-based PM performance.

Taken together, these findings suggest that general cognitive abilities (WM) and task-specific abilities (time monitoring) are not completely independent of each other. When the demands on the general cognitive abilities are particularly high in an ongoing task, the task specific strategies will also suffer. Nevertheless, individual differences in WM (updating or binding) abilities do not fully explain individual differences in task specific time monitoring abilities and their contribution to time-based PM performance. Future research is certainly necessary to better understand the complex interplay between task general and task specific time-based PM processes.

A couple of additional findings are worth discussing. In Experiment 1, we observed a significant effect of increased capacity load in the ongoing task on PM performance, which nevertheless disappeared once we allowed an interaction between these predictors. Surprisingly, in Experiment 3, we did not find a WM capacity load effect on PM performance. One reason might be that the control group in Experiment 3 was presented with five words (one more than the medium load condition in Experiment 1) but for a longer interval (1200 ms in Experiment 1 vs. 4000 ms in Experiment 3) and was required to maintain them for a longer duration (100 ms in Experiment 1 vs. 2600 ms in Experiment 3). Consequently, the WM demands may have substantially differed between the two experiments, making it possible to compensate task complexity with rehearsal strategies.

Previous research has demonstrated that implementing a delay strategy by slowing RTs in the ongoing task can enhance event-based PM performance (Heathcote et al., 2015). In our study, we sought to explore whether similar slowing effects occur in the ongoing task during a time-based PM task and whether they relate to PM performance. Our findings revealed a significant increase in RTs during the last quarter of each block compared to the previous one in Experiments 1 and 3, although this trend was absent in Experiment 2. However, in Experiment 2, a significant difference in RTs emerged between the last segment before successful completion of the PM task and the segment after which the PM task was forgotten. Surprisingly, this observed slowing effect was not a significant predictor of PM performance. It is conceivable that the detected slowing effects are not crucial for successful PM performance, or perhaps our measurement methodology was not well chosen. Employing a segmentation method similar to Joly-Burra et al. (2022), we divided each block into four segments, which may not fully capture the slopes of RT patterns. Understanding how time-based PM tasks influence ongoing task performance is crucial, as these tasks reflect the activities we routinely undertake, such as completing work assignments while awaiting meetings (Kvavilashvili & Fisher, 2007). Insights into this relationship can enhance our comprehension of cognitive processes in real-world contexts, shedding light on how maintaining an intention influences our current engagements.

While our studies provide valuable new insights into the factors influencing time-based PM performance, several limitations should be noted. One limitation is the generally relatively high PM performance observed across all three experiments. This fact may have limited the variability in PM performance and thus prevented the detection of subtle effects or individual differences that could have been revealed with a broader range of performance levels. Future research could benefit from recruiting participants with a wider range of PM abilities to better capture individual differences and their underlying mechanisms. Moreover, while our laboratory-based experiments shed light on the cognitive processes involved in time-based PM tasks, applying these findings to real-world settings may be challenging. The complexity and variability of everyday tasks may not fully align with the controlled conditions of laboratory experiments. Thus, future studies should explore PM performance in naturalistic settings to enhance ecological validity and better understand the practical implications of our findings (Rummel et al., 2023). Regarding generalizability of the present findings it is worth noting that age represents another significant factor contributing to variations in PM performance (Ball et al., 2019; Gonneaud et al., 2011), which was not covered in the current studies. Previous research has highlighted age-related declines in PM performance across both event-based (Kliegel et al., 2008) and time-based PM tasks, where these declines are often accompanied by a deficient use of strategic clock checking to monitor the time until PM intention fulfillment (Mäntylä et al., 2009; Vanneste et al., 2016). Similarly, age-related declines in WM performance have been well-documented (Salthouse, 1994; Salthouse et al., 1991), indicating a potential association between age-related changes in task-general WM and task-specific time monitoring processes (Gonneaud et al., 2011). Future research should test whether age-related declines PM performance depend on a deficit in general cognitive abilities or task specific abilities in PM. Furthermore, while the experiments reported here provide a more detailed analysis of the role of WM processes, particularly the updating component, they don't examine how other executive functions contribute to time-based PM performance. Future research should explore how different aspects of executive processes, such as distractor inhibition and dominant response inhibition, affect time-based PM and task specific abilities like clock checking. It would also be valuable to investigate whether the role of these executive functions differs between time-based and event-based PM tasks. Such research could provide a more detailed understanding of how various executive processes contribute to PM performance. Especially whether specific executive functions are differentially involved depending on the type of PM task.

Finally, present research contributes to our understanding of time-based PM by highlighting the role of top-down cognitive processes, particularly updating, in monitoring and executing future intentions. While previous studies have primarily focused on event-based PM, our findings demonstrate that time-based PM tasks place substantial demands on cognitive functions, such as time-monitoring strategies. Our studies extend the theoretical understanding of time-based PM by showing that successful PM performance is not merely a function of task-specific strategies but also relies on the availability of general cognitive resources, such as WM processes. In particular, under conditions of high cognitive load, the strategic time monitoring behavior becomes more critical for successful PM performance. Consequently, this finding underscores the need to understand PM as a dynamic process that combines general cognitive abilities with task-specific abilities in future research (Shelton & Scullin, 2017).

Conclusions

In summary, our research highlights the role of general cognitive abilities, particularly WM, in predicting time-based PM performance. Our results show that individuals are more likely to forget their PM intentions when WM resources, more specifically updating resources, are impacted by ongoing task demands. Strategic clock checking proofed to always be a powerful task specific predictor of PM performance which does not seem to be completely independent of general cognitive PM processes. These findings contribute to a deeper understanding of the cognitive mechanisms underlying time-based PM and highlight the importance of considering WM capacity in the context of PM tasks.

CRediT authorship contribution statement

Wiebke Hemming: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Project administration. **Kathrin Sadus:** Writing – review & editing, Conceptualization. **Jan Rummel:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data and analysis protocols are available in the Open Science Framework repository at https://osf.io/2yp4e/.

References

Arana, J. M., Meilan, J. J. G., & Perez, E. (2008). The effect of personality variables in the prediction of the execution of different prospective memory tasks in the laboratory. *Scandinavian Journal of Psychology, 49*(5), 403–411. https://doi.org/10.1111/j.1467-9450.2008.00671.x

Ball, H., Vogel, A., & Brewer, G. (2019). Individual differences in prospective memory. In J. Rummel, & M. A. McDaniel (Eds.), *Prospective memory*. Routledge.

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

Cohen, A.-L., Dixon, R. A., Lindsay, D. S., & Masson, M. E. J. (2003). The effect of perceptual distinctiveness on the prospective and retrospective components of prospective memory in young and old adults. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale, 57*(4), 274–289. https://doi.org/10.1037/h0087431

Cuttler, C., & Graf, P. (2007). Personality predicts prospective memory task performance: An adult lifespan study. *Scandinavian Journal of Psychology, 48*(3), 215–231. https://doi.org/10.1111/j.1467-9450.2007.00570.x

Dames, H., & Oberauer, K. (2022). Directed forgetting in working memory. *Journal of Experimental Psychology: General, 151*(12), 2990–3008. https://doi.org/10.1037/xge0001256

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 717–726. https://doi.org/10.1037/0278-7393.16.4.717

Einstein, G. O., McDaniel, M. A., Richardson, S. L., Guynn, M. J., & Cunfer, A. R. (1995). Aging and prospective memory: Examining the influences of self-initiated retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 996–1007. https://doi.org/10.1037/0278-7393.21.4.996

Einstein, G. O., Smith, R. E., McDaniel, M. A., & Shaw, P. (1997). Aging and prospective memory: The influence of increased task demands at encoding and retrieval. *Psychology and Aging, 12*(3), 479–488. https://doi.org/10.1037/0882-7974.12.3.479

Ellis, J. (1996). Prospective memory or the realisation of delayed intentions: A conceptual framework for research. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications*. Erlbaum.

Ellis, J., & Kvavilashvili, L. (2000). Prospective memory in 2000: Past, present, and future directions. *Applied Cognitive Psychology, 14*(7). https://doi.org/10.1002/acp.767

Gonneaud, J., Kalpouzos, G., Bon, L., Viader, F., Eustache, F., & Desgranges, B. (2011). Distinct and shared cognitive functions mediate event- and time-based prospective memory impairment in normal ageing. *Memory, 19*(4), 360–377. https://doi.org/10.1080/09658211.2011.570765

Goodhew, S. C., & Edwards, M. (2019). Translating experimental paradigms into individual-differences research: Contributions, challenges, and practical recommendations. *Consciousness and Cognition, 69*, 14–25. https://doi.org/10.1016/j.concog.2019.01.008

Hartig, J., Jude, N., & Rauch, W. (2003). *Entwicklung und Erprobung eines deutschen Big-Five-Fragebogens auf Basis des International Personality Item Pools (IPIP40)*. Arbeiten aus dem Institut für Psychologie der Johann Wolfgang Goethe-Universität, Heft 1.

Heathcote, A., Loft, S., & Remington, R. W. (2015). Slow down and remember to remember! A delay theory of prospective memory costs. *Psychological Review, 122*(2), 376–410. https://doi.org/10.1037/a0038952

Joly-Burra, E., Haas, M., Laera, G., Ghisletta, P., Kliegel, M., & Zuber, S. (2022). Frequency and strategicness of clock-checking explain detrimental age effects in time-based prospective memory. *Psychology and Aging, 37*(5), 637–648. https://doi.org/10.1037/pag0000693

Khan, A., Sharma, N. K., & Dixit, S. (2008). Cognitive load and task condition in event- and time-based prospective memory: An experimental investigation. *The Journal of Psychology, 142*(5), 517–532. https://doi.org/10.3200/JRLP.142.5.517-532

Kidder, D. P., Park, D. C., Hertzog, C., & Morrell, R. W. (1997). Prospective memory and aging: The effects of working memory and prospective memory task load. *Aging, Neuropsychology, and Cognition, 4*(2), 93–112. https://doi.org/10.1080/13825589708256639

Kliegel, M., Jäger, T., & Phillips, L. H. (2008). Adult age differences in event-based prospective memory: A meta-analysis on the role of focal versus nonfocal cues. *Psychology and Aging, 23*(1), 203–208. https://doi.org/10.1037/0882-7974.23.1.203

Kliegel, M., Martin, M., McDaniel, M. A., & Einstein, G. O. (2002). Complex prospective memory and executive control of working memory: A process model. *Psychologische Beitrage, 44*(2), 303–318.

Kretschmer, A., Voigt, B., Friedrich, S., Pfeiffer, K., & Kliegel, M. (2014). Time-based prospective memory in young children—Exploring executive functions as a developmental mechanism. *Child Neuropsychology, 20*(6), 662–676. https://doi.org/10.1080/09297049.2013.841881

Kvavilashvili, L., & Fisher, L. (2007). Is time-based prospective remembering mediated by self-initiated rehearsals? Role of incidental cues, ongoing activity, age, and motivation. *Journal of Experimental Psychology: General, 136*(1), 112–132. https://doi.org/10.1037/0096-3445.136.1.112

Labelle, M.-A., Graf, P., Grondin, S., & Gagné-Roy, L. (2009). Time-related processes in time-based prospective memory and in time-interval production. *European Journal of Cognitive Psychology, 21*(4), 501–521. https://doi.org/10.1080/09541440802031000

Laera, G., Brummer, J., Hering, A., Kliegel, M., & Horn, S. (2024). The cost of monitoring in time-based prospective memory. *Scientific Reports, 14*(1), 2279. https://doi.org/10.1038/s41598-024-52501-w

Lecouvey, G., Gonneaud, J., Piolino, P., Madeleine, S., Orriols, E., Fleury, P., Eustache, F., & Desgranges, B. (2017). Is binding decline the main source of the ageing effect on prospective memory? A ride in a virtual town. *Socioaffective Neuroscience & Psychology, 7*(1), 1304610. https://doi.org/10.1080/20009011.2017.1304610

Leiner, D. J. (2023). *SoSci Survey (Version 3.5.01) [Computer Software]* [Computer software]. https://www.soscisurvey.de.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MATLAB. *Behavior Research Methods, 42*(2), 571–585. https://doi.org/10.3758/BRM.42.2.571

Logie, R., Maylor, E., Della Sala, S., & Smith, G. (2004). Working memory in event- and time-based prospective memory tasks: Effects of secondary demand and age. *European Journal of Cognitive Psychology, 16*(3), 441–456. https://doi.org/10.1080/09541440340000114

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology, 1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

Mäntylä, T., Carelli, M. G., & Forman, H. (2007). Time monitoring and executive functioning in children and adults. *Journal of Experimental Child Psychology, 96*(1), 1–19. https://doi.org/10.1016/j.jecp.2006.08.003

Mäntylä, T., Missier, F. D., & Nilsson, L.-G. (2009). Age differences in multiple outcome measures of time-based prospective memory. *Aging, Neuropsychology, and Cognition, 16*(6), 708–720. https://doi.org/10.1080/13825580902912721

Marsh, R. L., & Hicks, J. L. (1998). Event-based prospective memory and executive control of working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(2), 336–349. https://doi.org/10.1037/0278-7393.24.2.336

Martin, M., & Schumann-Hengsteler, R. (2001). How task demands influence time-based prospective memory performance in young and older adults. *International Journal of Behavioral Development, 25*(4), 386–391. https://doi.org/10.1080/01650250042000302

McDaniel, M. A., & Einstein, G. O. (2007). Prospective memory: An overview and synthesis of an emerging field. *Sage*.

McDaniel, M. A., Umanath, S., Einstein, G. O., & Waldum, E. R. (2015). Dual pathways to prospective remembering. *Frontiers in Human Neuroscience, 9*. https://doi.org/10.3389/fnhum.2015.00392

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.

McFadden, D. (1977). Quantitative methods for analyzing travel behaviour of individuals: Some recent developments. *Cowles Foundation Discussion Papers, 707*. https://elischolar.library.yale.edu/cowles-discussion-paper-series/707.

Meier, B., & Zimmermann, T. D. (2015). Loads and loads and loads: The influence of prospective load, retrospective load, and ongoing task load in prospective memory. *Frontiers in Human Neuroscience, 9*. https://doi.org/10.3389/fnhum.2015.00322

Mioni, G., Grondin, S., McLennan, S. N., & Stablum, F. (2020). The role of time-monitoring behaviour in time-based prospective memory performance in younger and older adults. *Memory, 28*(1), 34–48. https://doi.org/10.1080/09658211.2019.1675711
Mioni, G., Santon, S., Stablum, F., & Cornoldi, C. (2017). Time-based prospective memory difficulties in children with ADHD and the role of time perception and working memory. *Child Neuropsychology, 23*(5), 588–608. https://doi.org/10.1080/09297049.2016.1172561
Mioni, G., & Stablum, F. (2014). Monitoring behaviour in a time-based prospective memory task: The involvement of executive functions and time perception. *Memory, 22*(5), 536–552. https://doi.org/10.1080/09658211.2013.801987
Morand, A., Segobin, S., Lecouvey, G., Gonneaud, J., Eustache, F., Rauchs, G., & Desgranges, B. (2021). Brain substrates of time-based prospective memory decline in aging: A voxel-based morphometry and diffusion tensor imaging study. *Cerebral Cortex, 31*(1), 396–409. https://doi.org/10.1093/cercor/bhaa232
Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS ONE, 7*(5), e37527.
Murayama, K., Usami, S., & Sakaki, M. (2022). Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modeling. *Psychological Methods*. https://doi.org/10.1037/met0000330
Oberauer, K. (2005). Binding and inhibition in working memory: individual and age differences in short-term recognition. *Journal of Experimental Psychology: General, 134*(3), 368–387. https://doi.org/10.1037/0096-3445.134.3.368
Paccagnella, O. (2011). Sample size and accuracy of estimates in multilevel models: New simulation results. *Methodology, 7*(3), 111–120. https://doi.org/10.1027/1614-2241/a000029
Park, D. C., Hertzog, C., Kidder, D. P., Morrell, R. W., & Mayhorn, C. B. (1997). Effect of age on event-based and time-based prospective memory. *Psychology and Aging, 12*(2), 314–327. https://doi.org/10.1037/0882-7974.12.2.314
Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1–2), 8–13. https://doi.org/10.1016/j.jneumeth.2006.11.017
R Core Team. (2022). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/.
Rose, N. S., Rendell, P. G., McDaniel, M. A., Aberle, I., & Kliegel, M. (2010). Age and individual differences in prospective memory during a "Virtual Week": The roles of working memory, vigilance, task regularity, and cue focality. *Psychology and Aging, 25*(3), 595–605. https://doi.org/10.1037/a0019771
Rummel, J., Danner, D., & Kuhlmann, B. G. (2019). The short version of the Metacognitive Prospective Memory Inventory (MPMI-s): Factor structure, reliability, validity, and reference data. *Measurement Instruments for the Social Sciences, 1*(1), 6. https://doi.org/10.1186/s42409-019-0008-6
Rummel, J., & Kvavilashvili, L. (2023). Current theories of prospective memory and new directions for theory development. *Nature Reviews Psychology, 2*(1), 40–54. https://doi.org/10.1038/s44159-022-00121-4
Rummel, J., & McDaniel, M. A. (Eds.). (2019). *Prospective Memory* ((1st ed.).). Routledge. https://doi.org/10.4324/9781351000154.
Rummel, J., Snijder, J.-P., & Kvavilashvili, L. (2023). Prospective memories in the wild: Predicting memory for intentions in natural environments. *Memory & Cognition, 51*(5), 1061–1075. https://doi.org/10.3758/s13421-022-01379-y
Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology, 8*(4), 535–543. https://doi.org/10.1037/0894-4105.8.4.535
Salthouse, T. A., Babcock, R. L., & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychology and Aging, 6*(1), 118–127. https://doi.org/10.1037/0882-7974.6.1.118
Schnitzspahn, K. M., Stahl, C., Zeintl, M., Kaller, C. P., & Kliegel, M. (2013). The role of shifting, updating, and inhibition in prospective memory performance in young and older adults. *Developmental Psychology, 49*(8), 1544–1553. https://doi.org/10.1037/a0030579
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
Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: Investigating the capacity demands of delayed intention performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(3), 347–361. https://doi.org/10.1037/0278-7393.29.3.347
Smith, R. E. (2010). What costs do reveal and moving beyond the cost debate: Reply to Einstein and McDaniel (2010). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(4), 1089–1095. https://doi.org/10.1037/a0019183
Smith, R. E., & Bayen, U. J. (2005). The effects of working memory resource availability on prospective memory: A formal modeling approach. *Experimental Psychology, 52*(4), 243–256. https://doi.org/10.1027/1618-3169.52.4.243
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
Smith, R. E., McConnell Rogers, M. D., McVay, J. C., Lopez, J. A., & Loft, S. (2014). Investigating how implementation intentions improve non-focal prospective memory tasks. *Consciousness and Cognition, 27*, 213–230. https://doi.org/10.1016/j.concog.2014.05.003
Smith, R. E., Persyn, D., & Butler, P. (2011). Prospective Memory, Personality, and Working Memory: A Formal Modeling Approach *It is journal policy that papers coauthored by an issue's editor are handled by a guest editor. We thank Edgar Erdfelder, who handled the current manuscript, for his efforts. *Zeitschrift Für Psychologie, 219*(2), 108–116. https://doi.org/10.1027/2151-2604/a000055
Strickland, L., Loft, S., Remington, R. W., & Heathcote, A. (2018). Racing to remember: A theory of decision control in event-based prospective memory. *Psychological Review, 125*(6), 851–887. https://doi.org/10.1037/rev0000113
The MathWorks Inc. (2018). *MATLAB version: 9.5 (R2018b)* [Computer software]. https://www.mathworks.com.
Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5
Uttl, B., White, C. A., Cnudde, K., & Grant, L. M. (2018). Prospective memory, retrospective memory, and individual differences in cognitive abilities, personality, and psychopathology. *PLoS One1, 13*(3), e0193806.
Uttl, B., White, C. A., Wong Gonzalez, D., McDouall, J., & Leonard, C. A. (2013). Prospective memory, personality, and individual differences. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00130
Vanneste, S., Baudouin, A., Bouazzaoui, B., & Taconnat, L. (2016). Age-related differences in time-based prospective memory: The role of time estimation in the clock monitoring strategy. *Memory, 24*(6), 812–825. https://doi.org/10.1080/09658211.2015.1054837
Voigt, B., Mahy, C. E. V., Ellis, J., Schnitzspahn, K., Krause, I., Altgassen, M., & Kliegel, M. (2014). The development of time-based prospective memory in childhood: The role of working memory updating. *Developmental Psychology, 50*(10), 2393–2404. https://doi.org/10.1037/a0037491
Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00433
Yang, T., Chan, R. C. K., & Shum, D. (2011). The development of prospective memory in typically developing children. *Neuropsychology, 25*(3), 342–352. https://doi.org/10.1037/a0022239
Zuber, S., & Kliegel, M. (2020). Prospective memory development across the lifespan: An integrative framework. *European Psychologist, 25*(3), 162–173. https://doi.org/10.1027/1016-9040/a000380
Zuber, S., Mahy, C. E. V., & Kliegel, M. (2019). How executive functions are associated with event-based and time-based prospective memory during childhood. *Cognitive Development, 50*, 66–79. https://doi.org/10.1016/j.cogdev.2019.03.001