Age Differences in Multiple Outcome Measures of Time-Based Prospective Memory

TIMO MÄNTYLÄ1, FABIO DEL MISSIER2, AND LARS-GÖRAN NILSSON3
1Department of Psychology, University of Umeå, Umeå, Sweden, 2Department of Psychology, University of Trieste, Trieste, Italy, and 3Department of Psychology, Stockholm University, Stockholm, Sweden

ABSTRACT

This study examined time-based prospective memory performance in relation to age, monitoring strategy, response accuracy, and dual-task demands. Young, middle-aged and older adults (N = 115) completed a prospective memory task, in which they indicated the passing of time every 5 min while listening to a short story (low task demands) or completing a series of cognitive tasks (high task demands). Young and older adults showed similar patterns of monitoring behavior, with low rates of clock checking during the early phase of each 5-min interval, followed by linearly accelerating monitoring functions. However, to obtain the same level of prospective memory performance older adults needed more frequent clock checks than young adults. Furthermore, older adults’ compensatory monitoring strategy was associated with an additional cost in primary task performance. Finally, increased primary task demands shifted age differences in prospective memory from monitoring frequency to response accuracy. These findings suggest that goal-directed behavior requires efficient task coordination and resource allocation, and that age-related differences in time-based prospective memory should be evaluated by using multiple outcome measures.

Keywords: Prospective memory; Monitoring strategy; Dual task; Multiple measures; Age effects.

INTRODUCTION

Most goal-directed tasks are completed in the context of other everyday activities, each of which requiring efficient task coordination and scheduling of processing resources. For example, being able to remember what one intends to do is a central aspect of goal-directed behavior, but everyday tasks...
of prospective memory are seldom completed in isolation. Instead, successful performance reflects a series of related activities, including (a) maintaining an intention in mind whilst completing and coordinating more urgent (primary task) activities, (b) occasionally thinking ahead and updating current intentions (task monitoring), and (c) interrupting primary task activity in order to initiate and execute the intention at the appropriate time (see Kliegel, McDaniel, & Einstein, 2008, for an overview). In most situations, these multiple ongoing activities demand attentional resources (e.g., listening to a conversation while paying attention to an approaching deadline) and they are associated with costs and tradeoffs (e.g., frequent clock checking facilitates prospective remembering but impairs primary task performance).

Although most goal-directed tasks, including prospective memory, reflect multiple task criteria, conclusions about individual, developmental and pathological differences are often assessed in isolation by using a single criterion (success vs. failure). These analyses may lead to premature or even incorrect conclusions about performance limitations and their underlying processes. For example, past research suggests that age differences in prospective memory vary substantially across studies. Some studies have reported significant age-related declines in prospective memory performance, whereas other studies found that older adults perform as well as their younger counterparts (see Henry, MacLeod, Phillips, & Crawford, 2004, for an overview). It is reasonable to assume that the magnitude of age difference in prospective memory is mediated by multiple factors (e.g., Einstein et al., 2005; Mäntylä & Nilsson, 1997; McDaniel & Einstein, 2000; Salthouse, Berish, & Siedlecki, 2004), and that a more complete evaluation of age differences should be based on multiple outcome measures of prospective memory (see also Glickson & Myslobodsky, 2006; Kliegel et al., 2008).

Most goal-directed activities, including prospective memory, require a monitoring process by which agents assess their environments (Atkin & Cohen, 1996; see Mäntylä & Carelli, 2006 for an overview). Furthermore, prospective memory tasks involve multiple performance criteria in that they are completed in the context of other everyday activities. In event-based prospective memory tasks, monitoring processes reflect covert activities (e.g., self-reminding) and cannot be easily studied. However, time-based tasks of prospective memory are more informative than event-based tasks in that the individual’s monitoring activities may reflect overt behaviors, and can be observed more directly, for example, by registering the frequency of clock checking in laboratory settings. This additional outcome measure (i.e., monitoring frequency) provides a window or a proxy to the participant’s strategic behavior during the interval that precedes the deadline (i.e., clock checking indicates withdrawal of attentional resources from the ongoing activity). Thus, a time-based prospective memory can (and should) be examined in terms multiple outcome measures, including response accuracy, monitoring frequency and primary task performance.
As an illustration of multiple outcome measures, Mäntylä, Carelli, and Forman (2007) examined developmental differences time-based prospective memory performance. In their study, school-aged children and young adults completed a prospective memory task, in which they indicated the passing of time every 5 min whilst watching a movie (see also Ceci & Bronfenbrenner, 1985). One interesting finding of their study was that time-based prospective memory performance was not related to participants’ age. Instead, the school-aged children (between 7 and 11 years of age) obtained the same level of response accuracy as did the adult participants. This result was somewhat surprising considering that age-related differences are typically accentuated in time-based tasks of prospective memory and that the adult participants constituted a selected group of university students.

However, the observed pattern of results was less surprising when both outcome measures of prospective memory were taken into consideration. Specifically, there were no age-related differences in response accuracy (i.e., on-time performance), but children monitored more frequently than adults. Thus, the absence of age differences in response accuracy had a cost in terms of monitoring frequency. One implication of this study and related work (e.g., Cicogna, Nigro, Occhionero, & Ésposito, 2005; Harris & Wilkins, 1982; Einstein, McDaniel, Richardson, Guynn, & Cunfer, 1985; Kerns, 2000; Maylor, Smith, Della Sala, & Logie, 2002; Park, Hertzog, Kidder, Morrell, & Mayhorn, 1997; see also Mäntylä & Carelli, 2006, for an overview) is that individual and age-related differences in prospective memory performance should be evaluated in relation to patterns of outcome measure, rather than relying on a single index of response accuracy.

Although it is reasonable to argue that age-related differences in prospective memory should be evaluated in terms of multiple outcome measures, past studies have not included more than one or two criterion measures. Specifically, studies involving event-based tasks (which provide minimal information about monitoring behavior) have examined performance in terms of accuracy and ongoing task performance, (e.g., Einstein, Smith, McDaniel, & Shaw, 1997; Kliegel, Martin, McDaniel, & Einstein, 2001; Marsh & Hicks, 1998; McDaniel & Einstein, 2000; Park et al., 1997; Smith, 2003; Smith, Hunt, McVay, & McConnell, 2007). As illustrated above, studies involving time-based tasks have virtually ignored ongoing task performance, while measuring response accuracy and, in some cases, also monitoring frequency.

The aim of the present study was to extend past work by examining age-related differences in prospective memory by considering the three measures within a single task. Thus, our starting point was the idea that age differences in prospective memory (and other complex goal-directed tasks) should be examined in terms of patterns of performance by using multiple outcome measures and varying task demands (cf. Craik, Byrd, & Swanson,
Specifically, we examined time-based prospective memory performance by involving young, middle-aged and older adults as participants, by varying ongoing task demands (high vs. low task demands) and by examining age-related differences in a pattern of performance that was based on three measures, namely, monitoring frequency, response accuracy, and ongoing task performance, respectively. It should be noted that our main aim was to examine how older adults allocate their limited cognitive resources by adjusting monitoring strategies in a complex goal-directed task, as reflected by patterns of multiple measures, rather than making absolute comparisons between primary and secondary task demands (cf. Marsh, Hicks, & Cook, 2005; McDaniel & Einstein, 2000; Smith, 2003; Smith et al., 2007).

Following the reasoning outlined above (see also Mäntylä et al., 2007), we expected marginal or nonsignificant age differences in time-based prospective memory performance under low concurrent task demands, because older adults would be able to adopt a more conservative (i.e., more frequent) monitoring strategy in order to compensate for their reduced attentional resources. Although increased monitoring frequency was expected to facilitate response accuracy (i.e., more frequent clock checking would increase on-time responding), this compensatory strategy was also expected to be associated with a task cost, as measured by ongoing task performance. Following this line of reasoning, an increased concurrent task demand (i.e., a more difficult ongoing task) was expected to accentuate age-related difference in time-based prospective memory performance, because older adults would not have sufficient attentional resources to maintain the same compensatory strategy as under less demanding task conditions. In other words, we expected that older adults would reduce their monitoring frequency, but that these limitations in compensatory strategies would have a cost in terms of prospective memory accuracy.

METHOD

Participants

One hundred and fourteen adults between 20 and 81 years of age participated in the study. The young participants were Umeå university undergraduates between 20 and 30 years. The middle-aged participants (between 36 and 56 years) and the older participants (between 64 and 81 years) were based on population-based sample of adults, recruited through the population register of Umeå (see also Nilsson et al., 1997, 2004). All participants had Swedish as their native language, and none of them suffered from dementia or severe sensory deficits. Background information describing each age group is summarized in Table 1.
Procedure and Tasks

Participants completed a time-based prospective memory tasks under varying ongoing task demands. In the low-demand condition, they listened to a short story (‘The hound’ by Lovecraft, and read by a Swedish actor). They were also informed that the experimenter would ask questions about the story content. In the high-demand condition, the ongoing task comprised a series of cognitive tasks (see also below). In both conditions, participants were instructed to indicate the passing of every 5 min, while completing the ongoing task. Specifically, a response box with a green and a red button was placed next to a computer monitor in front of the participant. The experimenter instructed the participant that the red button should be pressed when the clock on the monitor showed 05:00, 10:00, 15:00 and so on, without informing about the duration of the task. The experimenter also clarified that the red button should be used only for indicating the passing of every 5 min, but that the green button could be used to check the clock any time during the task. After pressing one of the buttons, the corresponding task time appeared (in red or green color) for 2 s on the computer screen. A practice phase was included in the beginning of the task to clarify the instructions and to familiarize participants with the task. The experimenter also demonstrated that the clock would start at 00:00 and that, for example, 10:00 would mean 10 min. Participants were instructed that good performance in this task meant a combination of ‘red’ and ‘green’ responses in that that each 5-min target response should be as accurate as possible (within ± 10 s) while attempting to minimize the number of clock checks during each task interval. The experimenter also emphasized that the primary and ongoing tasks were equally important. None of the participants indicated problems in understanding the task instructions.

Each participant was tested individually during two 40-min sessions. During the first session, participants completed a background inventory and the low-demand condition. Specifically, participants were seated in front of

<table>
<thead>
<tr>
<th>Variable</th>
<th>Age group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Young</td>
</tr>
<tr>
<td>n (male/female)</td>
<td>39 (18/21)</td>
</tr>
<tr>
<td>Mean age</td>
<td>23.3 (2.4)</td>
</tr>
<tr>
<td>Years of education</td>
<td>14.7 (3.2)</td>
</tr>
<tr>
<td>MMSE</td>
<td>–</td>
</tr>
<tr>
<td>Word comprehension¹</td>
<td>–</td>
</tr>
</tbody>
</table>

MMSE, Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975).
¹Word comprehension test, a 30-item multiple-choice synonym test.
computer screen and they were wearing headphones. After completing the task, participants’ ongoing task performance was assessed by means of a text recall task. Specifically, the task comprised a list of statements that referred to 10 distinct story events (e.g., ‘Mentions Abdul-Al-Hazred’). Participants were asked to rank the temporal order of each statement in the story by writing a number between 1 and 10 after each statement (e.g., ‘Mentions Abdul-Al-Hazred’ = 6). Participants were encouraged to guess if necessary, but the experimenter did not provide any feedback.

During the second session, about one week later, participants completed the high-demand condition. The procedure and instructions were the same as those during the first session, except that the ongoing task comprised a series of cognitive tasks. Specifically, participants completed a three task of response inhibition (Stroop, stop signal and flanker task) and three tasks of working memory (n-back, keep track and letter-memory; see also Mäntylä et al., 2007). The order of tasks was the same for all participants, and each task (or a subtask) took between 2 and 5 min to complete, including instructions. Participants were also informed that they could check the clock whenever they wanted while completing these tasks. The experimenter attempted to verify that the participant had understood the instructions, and in case of difficulties, the specific task instruction was further clarified. Furthermore, all the tasks included a practice phase, during which the experimenter illustrated the task instructions.

RESULTS

Prospective memory performance was examined in terms of monitoring frequency (i.e., ‘green’ responses) and response accuracy (i.e., ‘red’ responses). Figure 1 summarizes the monitoring data as a function of age and concurrent task demand (collapsed across the 5-min blocks).

As can be seen, the mean number of clock checks (per a 5-min interval) increased as a function of age, $F(2, 111) = 4.36$, $MSE = 8.17$, $\eta_p^2 = .08$, $p < .02$, and these effects interacted with tasks demands, $F(2, 111) = 3.03$, $MSE = 0.57$, $\eta_p^2 = .07$, $p < .05$. Tests of simple effects showed that the mean of old adults was greater than that of young adults in the low-demand condition ($p < .05$), but not in the high-demand condition.

Figure 2 shows a more detailed description of the clock checking data in the low-demand condition. Apart from differences in overall levels, young adults, middle-aged adults and old adults showed rather similar functions across the six 5-min blocks. Figure 2 also shows that the rate of clock checking was minimized immediately after each 5-min deadline (i.e., during the 6th, 12th and 15th min, respectively), producing a saw tooth-like monitoring function for both age groups (middle-aged are not included in Figure 2, but they showed a similar pattern).
Response accuracy was measured in terms of timing errors and proportion correct responses. The former measure reflects timing errors regardless of their direction. The latter measure is a more conventional measure of prospective memory performance, and was obtained by calculating the number of on-time responses (max 10 s after the target time). The two measures showed similar patterns of data, and Figure 3 shows response accuracy in terms of proportion correct responses. As can be seen, all three age groups showed similar patterns of performance in the low-demand condition in that more than 50% of the target responses were obtained within 10 s (the mean timing error was less than 18 s for all three groups). However, increased
ongoing task demand reduced response accuracy, and again these effects were accentuated for old adults.

Separate 2 (age) × 2 (task demand) analyses of variance (ANOVA)s confirmed these observations by yielding significant main effects of task demand, $F(1, 111) = 50.54, MSE = 0.05, \eta^2_p = .33, p < .01$, and its interaction with age $F(1, 111) = 3.21, MSE = 0.05, \eta^2_p = .06, p < .05$. The main effect of age was not significant, $F(1, 111) = 2.56, MSE = 0.11, \eta^2_p = .04, p < .10$. Subsequent contrast tests showed a significant difference between young and old adults in the high-demand, but not in the low-demand, condition. The mean of the middle-aged adults was not significantly different from those of the young and older age groups.

Table 2 summarizes ongoing task performance as a function of age and task condition. In the low-demand conditions, participants were given a story recall task in which they indicated the temporal order of 10 story statements. We obtained an index of task performance by calculating rank order correlation between the expected and observed ratings. The corresponding measure of the high-demand condition was based on a compound measure across the cognitive tasks, with a positive $z$-value indicating good performance. As shown in Table 2, older adults had greater difficulties in story recall than younger adults $F(2, 111) = 7.61, MSE = 0.074, p < .01$, and the effects of age on concurrent task performance were accentuated in the

<table>
<thead>
<tr>
<th>Task</th>
<th>Young</th>
<th>Middle-aged</th>
<th>Old</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low demand</td>
<td>0.73</td>
<td>0.63</td>
<td>0.49</td>
</tr>
<tr>
<td>High demand</td>
<td>0.79</td>
<td>-0.11</td>
<td>-0.87</td>
</tr>
</tbody>
</table>

*Note: Low demand = mean correlation; high demand = mean $z$-score.*
high-demand condition, $F(2, 103) = 38.62, MSE = 0.167, p < .01$. Tukey’s post-hoc tests showed that the mean of the young group was greater than that of old group in both task conditions.

Table 3 summarizes correlation coefficients among age, prospective memory and ongoing task performance, separately for the high- and low-demand conditions. For both conditions, age correlated significantly with prospective memory performance and ongoing task performance. In the high-demand condition, ongoing task performance (executive functioning) was related to variability in response accuracy, but not in monitoring frequency. The low-demand condition showed the opposite pattern with a significant correlation with monitoring frequency, but not with response accuracy.

In the absence of a baseline condition it is not possible to make strong conclusions about these data, but this pattern at least suggests that task requirements were greater in the high-demand condition than in the low-demand condition, and that older adults’ compensatory monitoring strategy was associated with a task cost in terms of impaired ongoing task performance. Additional regression analysis supported this conclusion in that older adults’ (but not young and middle-aged adults’) prospective memory performance was related to the ongoing task performance in the low-demand condition. We completed these analyses separately for the three age groups, with story recall scores as a regressor and age, monitoring frequency and responses accuracy, respectively, as predictors. These analyses showed that older adults’ monitoring frequency ($\beta = -0.35, t = 2.12, p < .05$), but not response accuracy ($\beta = -0.24$) or age ($\beta = -0.09$), was related to variability in story recall performance. For young and middle-aged participants, none of the predictors were related to story recall performance.

**DISCUSSION**

This study examined time-based prospective memory performance in relation to age, monitoring strategy and dual-task demands. The primary hypothesis of the study was that (time-based) prospective memory performance

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>–</td>
<td>.13</td>
<td>–.35**</td>
<td>–.69**</td>
</tr>
<tr>
<td>Monitoring frequency</td>
<td>.29**</td>
<td>–</td>
<td>.36**</td>
<td>.07</td>
</tr>
<tr>
<td>Response accuracy</td>
<td>–.09</td>
<td>.42**</td>
<td>–</td>
<td>.27**</td>
</tr>
<tr>
<td>Ongoing task performance</td>
<td>–.37**</td>
<td>–.23*</td>
<td>–.02</td>
<td>–</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01.
AGING AND MULTIPLE MEASURES OF PROSPECTIVE MEMORY

involves multiple outcome measures, including response accuracy, monitoring frequency and primary task performance, and that age-differences in prospective memory reflects a dynamic pattern of tradeoffs and tasks costs among these measures. From that perspective, it is not meaningful to determine the magnitude of age effect in prospective memory by relying on an isolated outcome measure.

The results of this study are consistent with this notion in that age and concurrent task demands produced a pattern of outcome measures, rather than showing an absolute presence or absence of age-related differences in time-based prospective memory performance. Specifically, age differences in prospective memory (responses accuracy) were not observed when the ongoing task demands were relatively low. However, the monitoring frequency data revealed a clear age difference in the low-task demand condition, suggesting that older adults compensated for their reduced cognitive resources by relying on more frequent monitoring strategy than young adults.

This pattern of results was consistent with the findings of Mäntylä et al. (2007) in that primary school children showed the same level of response accuracy as university students, but this absence of age differences in prospective memory performance had a cost in terms of monitoring frequency. Taken together, these studies suggest that individual and developmental differences in prospective memory performance should be evaluated in relation to patterns of outcome measure, rather than relying on a single index of response accuracy.

Extending the findings of Mäntylä et al. (2007), the present study also showed that increased monitoring frequency not only increase on-time responding, but this compensatory strategy was associated with a task cost, as measured by ongoing task performance. Specifically, older adults obtained the same level of response accuracy as young adults by checking the clock more frequently, but this monitoring strategy had indirect effects on primary task performance. Compared to young adults, older adults paid less attention to (the temporal structure of) the story, because they were more busy with time monitoring. However, it should be noted that, in the absence of a baseline condition, the observed age difference in story recall is only an indirect indicator of primary task cost. Yet, these data are at least consistent with the notion that older adults’ more frequent monitoring not only facilitated prospective memory (primary task) performance, but also had a cost on ongoing task performance. Furthermore, the regression analysis showed that those older adults who checked the clock frequently had greater difficulties in ongoing task performance.

Another central finding of this study was that increased concurrent tasks demands resulted in age-related difference in prospective memory, as measured by response accuracy. Similarly, the monitoring frequency data
showed that increased concurrent task demands reduced the frequency of clock checking in older adults. A reasonable explanation of this pattern of results is that older adults did not have sufficient attentional resources to maintain the same compensatory strategy as under the low concurrent task condition (i.e., reduced clock checking), which in turn led to significant age effects in response accuracy.

It should be mentioned that our primary aim was to examine patterns of performance (cf. Craik et al., 1987), rather than making any strong conclusions about resource allocation, including the contrasting views about automatic vs. strategic prospective memory retrieval (e.g., Kliegel et al., 2001; Marsh & Hicks, 1998; McDaniel & Einstein, 2000; Park et al., 1997; Smith, 2003; Smith et al., 2007). Although the task instructions emphasized that the primary and ongoing tasks were equally important, the two ongoing tasks differed both in qualitative and quantitative terms. Furthermore, it is reasonable to assume that the observed pattern of results reflected individual and age-related differences in a variety of factors, including cognitive resources and metacognitive strategies (e.g., perceived task difficulty). Yet, the present findings suggest that most goal-directed tasks, including prospective memory, involve a dynamic pattern of processes and that age-related differences in prospective memory should be evaluated in terms of multiple outcome measures.

ACKNOWLEDGEMENTS

This research was supported by a grant from the Swedish Research Council. Fabio Del Missier thanks Consorzio Universitario di Pordenone for financial support.

REFERENCES


