Cognition 202 (2020) 104305

Contents lists available at ScienceDirect

Cognition

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

# Differences in time-based task characteristics help to explain the age-prospective memory paradox

Simon J. Haines[a,b,*], Susan E. Randall[a], Gill Terrett[a], Lucy Busija[c], Gemma Tatangelo[d], Skye N. McLennan[a], Nathan S. Rose[f], Matthias Kliegel[g], Julie D. Henry[e], Peter G. Rendell[a]

[a] Australian Catholic University, Melbourne, Australia
[b] La Trobe University, Melbourne, Australia
[c] Monash University, Melbourne, Australia
[d] Deakin University, Melbourne, Australia
[e] University of Queensland, Brisbane, Australia
[f] University of Notre Dame, South Bend, United States of America
[g] University of Geneva, Geneva, Switzerland

ARTICLE INFO

*Keywords:*
Prospective memory
Time cues
Cognitive aging
PM paradox
Ecological validity

ABSTRACT

Prior prospective memory (PM) research shows paradoxical findings—young adults outperform older adults in laboratory settings, but the reverse is found in naturalistic settings. Moreover, young-old outperform old-old adults in laboratory settings, but show no age differences in naturalistic settings. Here we highlight how time-based task characteristics have differed systematically between studies conducted in laboratory (time-interval cues) and naturalistic settings (time-of-day cues) and argue that this apparent paradox is a function of comparing disparate task types. In three experiments, we tested this hypothesis using analogous paradigms across settings, with event-based, time-of-day, and time-interval cued PM tasks. Experiment 1 compared young ($n$ = 40) and older ($n$ = 53) adults on a laboratory paradigm that measured PM tasks embedded in a virtual, daily life narrative; and on a conceptually parallel paradigm using a customized smartphone application (MEMO) in actual daily life. Results revealed that on the MEMO, older adults outperformed young adults on the time-of-day tasks but did not differ on the time-interval or event-based task. In contrast, older adults performed worse than young adults in the laboratory. Experiment 2 compared PM performance in young-old ($n$ = 64) and old-old ($n$ = 40) adults using the same paradigms. Young-old outperformed old-old adults in the laboratory; however, group differences were not evident in daily life. Experiment 3 compared young ($n$ = 42) and older ($n$ = 41) adults, and largely replicated the findings of Experiment 1 using a more demanding version of MEMO. These findings provide novel and important insights into the limiting conditions of the age-PM paradox and the need for a finer theoretical delineation of time-based tasks.

## 1. Introduction

Understanding the effects of normal aging on prospective memory (PM)—the ability to remember and execute delayed intentions at an appropriate time or event in the future (Ellis, 1996), such as taking medication—is important for providing appropriate support for independent living in older age (Haines et al., 2019; Hering, Kliegel, Rendell, Craik, & Rose, 2018; Kliegel et al., 2016; Lee, Ong, Pike, & Kinsella, 2018). However, a clear picture of the effects of normal aging on PM has not been established, largely because a number of apparently contradictory findings have been reported regarding the PM performance of young and older adults across laboratory and naturalistic-settings; a pattern that has been dubbed the age-PM paradox (Rendell & Thomson, 1999). The first aspect of the paradox is that young adults outperform older adults in laboratory-based studies, while the reverse occurs in naturalistic studies (Henry, Macleod, Phillips, & Crawford, 2004; Kliegel et al., 2016). The second aspect is that young-old adults outperform old-old adults in laboratory-based studies, but both groups

* Corresponding author at: Lincoln Centre for Research on Ageing, Health Science Building 1, La Trobe University, Plenty Rd & Kingsbury Dr, Bundoora, Victoria 3086, Australia.
*E-mail address:* s.haines@latrobe.edu.au (S.J. Haines).

https://doi.org/10.1016/j.cognition.2020.104305
Received 19 December 2019; Received in revised form 5 April 2020; Accepted 6 April 2020

perform comparably well in naturalistic studies (Kliegel, Rendell, & Altgassen, 2008; Rendell & Craik, 2000; Rendell & Thomson, 1999). This pattern is unique in cognitive psychology, as in many cognitive domains making a task more familiar (Kliegel, Martin, Mcdaniel, & Phillips, 2007), or measuring it in the context of daily life (Phillips, Henry, & Martin, 2008), attenuates age-related effects on performance—it does not lead to dramatic reversals of age-related performance patterns on cognitive measures found in the laboratory.

### 1.1. Age-PM paradox a function of incommensurate PM task types

Proposed explanations for these group level age-effects across settings include: variations in the cognitive demands of the ongoing task (i.e., whatever activity is being performed during the delay period and opportunity for intention execution; Schnitzspahn, Ihle, Henry, Rendell, & Kliegel, 2011); and use of external aids, such as diaries or alarms, in naturalistic settings (Ihle, Schnitzspahn, Rendell, Luong, & Kliegel, 2012; Maylor, 1990; Moscovitch, 1982). However, neither of these factors has yet been shown to account adequately for the paradoxical findings. In the present experiments we investigated another possible contributor: the use of conceptually distinct PM task types across laboratory and naturalistic settings. Could the paradox be reduced—or eliminated—by using conceptually similar PM tasks in both settings?

We suggest that PM task type is a key feature to consider when unpacking the age-PM paradox (Niedźwieńska & Barzykowski, 2012). According to the multiprocess framework (Mcdaniel & Einstein, 2000), many *event-based* tasks (e.g., giving a message when meeting a colleague) involve spontaneous cognitive processes spared with normal aging (Einstein & Mcdaniel, 1990), while *time-based* tasks (e.g., taking cakes out of the oven in 20 min) involve effortful cognitive processes (Hasher & Zacks, 1979), which are negatively affected by normal aging (Mcdaniel & Einstein, 2008). However, a further distinction can be made within the category of time-based tasks. *Time-of-day* tasks (e.g., phoning the doctor at 1 p.m.) lend themselves to being associated with a constellation of environmental cues (known as *conjunction cues*), which might be expected to occur in the relevant window of time, and support spontaneous retrieval of the previously formed intention (Maylor, 1990; Phillips et al., 2008; Rendell & Craik, 2000). In contrast, *time-interval* tasks (e.g., returning a phone call in 10 min; Rose et al., 2015) offer relatively few or no time relevant environmental cues to support spontaneous retrieval processes of the previously formed intention.

The distinction we are drawing here for time-based tasks in many ways parallels, and is of similar importance to, a distinction frequently made within event-based tasks (Haines et al., 2019; Kliegel, Jäger, & Phillips, 2008). This distinction is between event cues which are *focal* (where the cue is related to the cognitive processing involved in the ongoing task) and event cues which are *non-focal* (where the cue is unrelated to the cognitive processing involved in the ongoing task). Focal event-based tasks involve relatively spontaneous processes, less affected by aging (Einstein & Mcdaniel, 2005), while non-focal event-based tasks involve relatively effortful, strategic cognitive processes, which are markedly affected by aging (Kliegel, Jäger, & Phillips, 2008; Mcdaniel & Einstein, 2007). We propose that, within the category of time-based PM tasks, time-of-day tasks are analogous in terms of cognitive demand (and hence age-effects) to focal event-based tasks; while time-interval tasks (assuming no external aids are used) are analogous in terms of cognitive demand (and age-effects) to non-focal event-based tasks. This useful distinction, made explicit in the current series of studies, was adumbrated by McDaniel and Einstein (2008): “In *time-based* prospective memory tasks, the appropriate moment for executing the prospective memory intention is a particular *time-of-day* (a doctor's appointment) or *the passage of a particular amount of time* (taking cookies out of the oven in 10 minutes).” (Emphasis added.) Unfortunately, to date the significance of this distinction has received little research attention in the published age-PM literature (Haines et al., 2019).

### 1.2. Increasing the range of PM task types used in naturalistic-settings

Converging evidence for the age-PM paradox being driven by the use of disparate PM task types used across settings is beginning to emerge (Bailey, Henry, Rendell, Phillips, & Kliegel, 2010; Kliegel, Rendell, & Altgassen, 2008; Niedźwieńska & Barzykowski, 2012; Schnitzspahn, Kvavilashvili, & Altgassen, 2018). Historically, naturalistic measures of PM have been predominantly time-of-day tasks (Henry et al., 2004; Kvavilashvili & Fisher, 2007; Maylor, 1990; Schnitzspahn et al., 2011), while laboratory measures of PM have been primarily event-based tasks (e.g., press a button whenever an animal word appears; Henry et al., 2004). Recently, studies have explored event-based tasks in naturalistic settings (Bailey et al., 2010; Kvavilashvili, Cockburn, & Kornbrot, 2013; Schnitzspahn et al., 2018). Bailey et al. (2010) replicated older adults' poorer performance on event-based tasks in laboratory settings by embedding a ‘classic’, event-cued PM task in a questionnaire administered using mobile phone devices. Niedźwieńska and Barzykowski (2012) used features of the evening TV news as focal (first appearance of weather map) and non-focal (first mention of a politician's name) for event-based tasks in daily life, and found older adults performed at high levels on the former but not the latter PM task type. Kvavilashvili et al. (2013) used the completion of a take home questionnaire, for the event-based task of writing the date and time on the top of the form before sending it to experimenter, and found no difference between young, young-old, and old-old adults. Schnitzspahn et al. (2018) used the event-based tasks of sending a text message from a mobile phone when first seeing a form of public transport on the following day and mailing a postcard when seeing a mailbox two days later. Young and older adults did not differ on these naturalistic setting event-based tasks (though in the laboratory young outperformed older adults on the event-based tasks, while showing no difference on a time-interval task). Both Schnitzspahn et al. (2018) and Niedźwieńska and Barzykowski (2012) also used time-of-day tasks (scheduled phone calls), finding the common pattern of older adults outperforming young adults, though neither study investigated the age-related effects on time-interval tasks in a naturalistic setting.

The use of time-of-day tasks in naturalistic settings is common (even so far as to suggest that time-based tasks in daily life are largely homogenous in their cognitive demands). However, when time-based tasks are used in the laboratory, they are typically time-interval tasks, such as signaling when 10-minute intervals have elapsed (Niedźwieńska & Barzykowski, 2012; Rendell & Thomson, 1999; Schnitzspahn et al., 2018). The difference between typical time-based tasks used in the lab and those used in naturalistic settings is accentuated by the fact that in the latter, though not the former, many features of daily life indicate time-of-day (e.g., clocks, TV or radio programs, level of sunlight, foot and vehicle traffic, other people eating meals or snacks, etc.). Such features may facilitate performance on time-of-day tasks in naturalistic settings (see Henry et al., 2004), and provide reliable cues for older adults who lead more structured or routine lives compared to the socially less predictable lives of young adults (Carstensen, Isaacowitz, & Charles, 1999).

### 1.3. Testing conceptually parallel PM tasks in the one study

Relatively few studies have tested PM in laboratory and naturalistic settings within the one study (Kvavilashvili et al., 2013; Niedźwieńska & Barzykowski, 2012; Rendell & Craik, 2000; Schnitzspahn et al., 2011; Schnitzspahn et al., 2018), and the use of parallel task types in each setting is even rarer. Work by Rendell and Craik (2000) provides one of the few attempts to systematically address the issue of whether the key features of the age-PM paradox are still apparent when PM task types are designed to be conceptually similar across laboratory and naturalistic settings. In their study, Rendell and Craik tested performance in the laboratory using Virtual Week, a laboratory paradigm in a board game format that simulates the typical structure and activities of daily

**Table 1**
A summary of key studies comparing young and older adults in laboratory and naturalistic settings on different PM task types.

| Study | Age groups | | Task types by setting with age-related differences in PM performance | | | |
|---|---|---|---|---|---|---|
| | Range (years) | *n*[a] | Laboratory | | Naturalistic | |
| Rendell and Thomson (1999) | 18–28 | 175/120 | Event | Y > YO > OO | Time-of-day | YO, O > Y |
| | 60–69 | 120/80 | Time-interval | Y > YO > OO | | |
| | 80–92 | 80/80 | | | | |
| Rendell and Craik (2000) | 19–24 | 20/16 | Event | Y > YO, OO | Event | OO, YO > Y |
| | 61–73 | 20/16 | 'Time-of-day' | Y > YO, OO | Time-of-day | OO, YO > Y |
| | 75–84 | 20/16 | Time-interval | Y > YO, OO | Time-interval | Y = YO = OO |
| Schnitzspahn et al. (2011) | 18–25 | 20/20 | Time-interval | Y > O | Time-of-day | O > Y |
| | 61–79 | 20/20 | | | | |
| Niedźwieńska and Barzykowski (2012) | 19–27 | 63/57 | Event (focal) | Y = M = O | Event (focal) | M, O > Y |
| | 42–51 | 50/46 | Event (non-focal) | Y, M > O | Event (non-focal) | Y = M = O |
| | 64–74 | 47/45 | Time-interval | Y = M = O | Time-of-day | O > M > Y |
| Kvavilashvili et al. (2013) | 18–30 | 72/61 | Event | YO, Y > OO | Event | Y = YO = OO |
| | 61–70 | 79/74 | | | | |
| | 71–80 | 72/68 | | | | |
| Schnitzspahn et al. (2018) | 20–29 | 31/31 | Event | Y > O | Event | Y = O |
| | 60–75 | 22/22 | Time-interval | Y = O | Time-of-day | O > Y |

*Note*. Y: young; YO: young-old; OO: old-old. The " > " sign indicates which group performed better (the group on the left-hand side of the sign); the " = " or a "," sign indicates no group differences in performance. For each study the same (or predominantly same) participants were tested across settings. Any difference in *n*s reflects attrition (except in the case of Rendell and Thomson (1999) where the naturalistic setting PM tasks were carried out before the laboratory setting tasks; and Rendell and Craik (2000) in which a different sample of young adults only was used in the naturalistic-setting).

[a] The numbers before and after the forward slash refers to the number of participants tested in the laboratory and naturalistic-setting, respectively.

life (for more recent versions see: Browning, Harris, Van Bergen, Barnier, & Rendell, 2018; Rendell & Henry, 2009; Terrett et al., 2019). A range of PM tasks (including virtual time-of-day tasks) are embedded in the game. The paradigm Rendell and Craik used for the parallel tasks in a naturalistic setting was Actual Week, which involved completing the same number and type of PM tasks per day as in Virtual Week (i.e., 10 tasks), at set times or in relation to set events within the participants' normal daily life routine.

The Rendell and Craik (2000) study largely replicated the general paradox: young were superior to older adults in the laboratory setting (Virtual Week) but inferior in the naturalistic setting (Actual Week); and young-old were superior to old-old adults in the laboratory but did not differ in a naturalistic setting. However, in one way the typical paradox was not identified in the naturalistic setting. Although results suggested an inverted U-shaped pattern of performance for the different age groups on the time-interval task in Actual Week (young-old performing better than both the young and old-old adults), there was no substantive difference between the three age groups on this PM task. The proportion correct in Actual Week for young, young-old, and old-old on the time-interval task was: 0.24, 0.46, and 0.26, respectively; which, for the older adults, was comparable to performance on the time-interval task in the lab (Virtual Week; young-old: 0.47; old-old: 0.34). Thus, the age-related decline seen in older adults in laboratory settings was replicated in the naturalistic setting for the time-interval task but not for the time-of day or event-based tasks.

An attempt to create conceptually parallel time-based PM tasks across settings was devised by Schnitzspahn et al. (2011). In that study the time-based tasks in the lab involved pressing the 'a' key on a computer keyboard at two set times within three 12-minute blocks of a naturalistic ongoing task. The duration of the time-based tasks in the lab were scaled up in the naturalistic-setting, with one set time cue in the first 12 h of the day, and one in the last 12 h, on three consecutive days. The time-based task in the naturalistic setting was to send two text messages, with 'a' as the content, in the morning and afternoon. However, a limitation of this approach is that the lab time-based task appears to be a time-interval task (few or no conjunction cues possible), while the apparent parallel task in the naturalistic setting appears to be a time-of-day task, and therefore amenable to a rich array of conjunction cues. Understandably practical limitations of the laboratory setting play a role in the challenge of developing conceptually parallel time-of-day tasks across settings.

The Virtual Week paradigm provides an elegant solution to the difficulty of separating time-of-day and time-interval tasks by providing, in a virtual format, some of the conjunction cues (e.g., regular meal times) that occur for time-of-day tasks in naturalistic-settings, while still including the more common laboratory time-interval task; involving breaking set with the ongoing task (i.e., switching attention from engaging in the virtual day with its various plausible commonplace scenarios). To date Virtual Week appears to be the only paradigm that explicitly distinguishes these two types of time-based tasks. Table 1 below summarizes the literature in which the same (or predominantly the same) sample of older and younger adults have been compared on PM tasks in each setting, illustrating the general lack of conceptually commensurable PM tasks across settings.

From Table 1 it is clear how time-interval tasks are dominant in the laboratory, while time-of-day tasks are dominant in naturalistic settings. Furthermore, the only study to investigate time-interval tasks in a naturalistic-setting, Rendell and Craik (2000), used delays (30 and 60 min) that were longer than those typically used in the laboratory (e.g., 10 min).

### 1.4. Clarity and verifiability of PM task types

There are two key limitations of the Rendell and Craik (2000) study that the present study seeks to address. First, the Rendell and Craik (2000) version of Virtual Week had times marked on the board which potentially compromised the validity of the "time-of-day" tasks (cf. Rose, Rendell, Mcdaniel, Aberle, & Kliegel, 2010). That is, while recent computer versions of Virtual Week have a virtual time clock calibrated to the token position on the game board (e.g., Mioni, Grondin, Mclennan, & Stablum, 2019; Terrett et al., 2019), the original board game version had the consecutive hours of the day marked on the squares on the board, which may have acted like event-based cues when the token passed the marked square corresponding to the set time. Second, in Actual Week successful performance of each age group on the event-based tasks could not be adequately verified. Specifically, while the recording device could verify when the time-based tasks were completed, it was not possible to verify whether the event-based cue had actually occurred, or whether the PM task had been completed as instructed (cf. Schnitzspahn et al., 2018). Related to this limitation is the fact that all participants were *given* the same event-based cues which may not have been equally suited to young and older adults' daily routines.

### *1.5. The current experiments*

Taking Rendell and Craik (2000) as a starting point, we addressed the limitations of Actual Week by developing a novel, naturalistic setting PM paradigm (MEMO); and the limitations of the original Virtual Week paradigm by using the innovation (e.g., Rendell et al., 2011) of a virtual time clock calibrated to the token position on the board, which replaced times marked on the board. Using refinements and manipulations of these two paradigms, we conducted three experiments to investigate the time-based task distinction and undertake the most comprehensive investigation to date of the extent to which the lack of conceptually similar PM tasks across laboratory and naturalistic settings might account for the two key aspects of the age-PM paradox.

## 2. Experiment 1

Experiment 1 used a mixed measures design with a sample of young and older adults. Based on a large body of previous laboratory research using Virtual Week (Henry et al., 2004; Rendell & Henry, 2009) it was hypothesized that, in the laboratory setting, older adults would show poorer performance than young adults on all three types of PM task (i.e., event-based, time-of-day, and time-interval), but that these group differences would be greatest for the (virtual) time-of-day and time-interval tasks. A different pattern of performance for each age group was predicted on the analogous PM tasks administered in a naturalistic setting. It was hypothesized that older adults would show superior performance than young adults on the event-based and time-of-day tasks (which lend themselves to environmental support). However, for the time-interval tasks, it was hypothesized that young and older adults would show more commensurate performance to that evidenced in the laboratory on the time-interval tasks, though with neither age group being superior to the other. That is, similar to Rendell and Craik (2000), who used 30 and 60 min delays, it was expected the two groups would show similar levels of performance on the time-interval task, and these levels of performance would be relatively poor compared to the event-based and time-of-day tasks. This hypothesis is based on the assumption from the multiprocess framework that the time-interval PM task type should be the most taxing of cognitive resources due to low environmental support and high strategic monitoring demands.

### *2.1. Method*

#### *2.1.1. Participants*

Ninety-three healthy, community dwelling volunteers—40 young (19–30 years; 75% women) and 53 older (65–86 years; 68% women) adults—were recruited via flyers at universities and a range of recreational facilities. The young adult sample consisted of both undergraduate students (61%; $n = 24$) and young professionals (39%; $n = 16$). Participants received $30 for their participation, with the exception of undergraduate students, who obtained partial course credit. Older adults were screened using the Mini-Mental State Examination (MMSE, $M = 28.9$; $SD = 1.3$) to determine eligibility (scores $\geq 24$, Folstein, Folstein, & Mchugh, 1975). Of the 53 eligible older adults enrolled into the study, 12 were subsequently excluded from analyses. Of these, five declined to complete the experiment, five completed less than one day of the naturalistic phase of the experiment, and two others reported neurological conditions, leaving a total of 40 young and 41 older adults whose data could be included in analyses.

**Table 2**
Characteristics of participants in Experiment 1.

| | Young adults | | Older adults | | *t-test* | | |
|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *t* | *p* | *d* |
| Age (in years) | 24.1 | 3.6 | 71.6 | 4.9 | | | |
| Education (in years) | 16.3 | 2.6 | 14.6 | 3.0 | 2.84 | .01 | 0.63 |
| Verbal IQ[a] | 102.2 | 8.7 | 110.0 | 6.3 | 4.66 | < .001 | 1.03 |

[a] Estimated using the National Adult Reading Test (NART; Nelson, 1982).

**Ethical approval was provided by the** Human Research Ethics Committee (HREC) **of the Australian Catholic University**. Characteristics of the participants included in the final analyses are reported in Table 2.

#### *2.1.2. Materials and procedure*

Participants were tested individually in a 2–3-hour laboratory testing session, followed by 6 days of naturalistic testing (3 days of event-based tasks and 3 days of time-based tasks, counterbalanced). Participants completed a demographic questionnaire, followed by the MMSE (Folstein et al., 1975; for older participants only) and the National Adult Reading Test (NART; Nelson, 1982). Participants were then instructed on how to use the naturalistic measure of PM (MEMO) before the laboratory PM measure (Virtual Week) was administered. Prior to leaving the laboratory, participants were again briefed on the MEMO.

*2.1.2.1. Laboratory PM measure*. Virtual Week (VW; Rendell & Craik, 2000; Rendell & Henry, 2009) is presented in a board-game format, recently adapted for computer automation (Browning et al., 2018; Henry, Rendell, Phillips, Dunlop, & Kliegel, 2012; Leitz, Morgan, Bisby, Rendell, & Curran, 2009; Mioni, Rendell, Stablum, Gamberini, & Bisiacchi, 2015; Niedźwieńska, Rendell, Barzykowski, & Leszczyńska, 2014; Rendell et al., 2011; Rendell, Jensen, & Henry, 2007; Terrett et al., 2015). Fig. 1 shows the automated board game with examples of PM tasks and dialogue boxes.

To complete Virtual Week, participants must click on the image of the die (a random number generator of 1 to 6), and move a token around the board according to roll of the die, with each two squares traversed on the board representing "15 min" of a virtual day. A virtual day is represented by a complete circuit of the board (60 squares), beginning at "7 a.m." and finishing at "10 p.m.". A virtual clock, which changes as the token is moved, and a stop clock, counting up by seconds and minutes in real time, are displayed in the center and at the top of the screen. These clocks are relevant for the time-of-day and time-interval PM tasks, respectively. Ten event squares are marked on the board with an "E". When the token passes one of these, participants are required to click on the button labelled "Event Card" to reveal the image of an event card. The event cards form a plausible daily narrative, which acts as the ongoing task (e.g., "you visit the doctor") and participants are required to make choices at each event (e.g. "while waiting for the doctor you read: a) a magazine; b) a book you brought with you; or c) some brochures"). A button marked "Perform Task" is always present in the top right-hand corner of the board and at a similar location on event cards. Participants can click this button at any time to select a PM task to perform from a menu of 10 PM tasks. Participants are advised that they can still perform a task even if they realize it is late or the appropriate opportunity has passed.

At the beginning of the first day participants are given "regular" event- and time-based tasks, meaning the same tasks are to be performed each virtual day in response to the same cues. Additionally, at the start of each day, before rolling the die, participants are given two "irregular" PM tasks unique to that day (one time-of-day—e.g., "haircut at 1 p.m."—and one event-based—e.g., "return book when at the


**Fig. 1.** Screen shots from Virtual Week computerized version. From left-right, top row: board game interface; event-based task (“take antibiotics at breakfast and dinner”); time-of-day time-based task “use asthma inhaler at 11a.m. and 9 p.m.”); middle row: time-interval task (“test lung capacity”); irregular event task start of day (“pick up … at swimming pool”); irregular time-of-day task, start of day (“hair cut at 1 p.m.”); bottom row: ongoing task event card (“At the library”); additional irregular time-of-day task during day (“You make an appointment … 3 p.m.”); perform task menu.


library”). Two additional irregular tasks (one time-of-day, one event-based) are presented to participants after particular event-cards have been picked up during the virtual day (e.g., after interacting with an event-card about dropping off photocopying, a task card pops up with the following new PM task “remember to call your partner at 4 p.m. to collect photocopies” [time-of-day]; after being informed in an event card that their friend had a baby girl, a task card pops up with the following new PM task “when you see Margaret remember to tell her about Jane's new baby” [event-based]). As illustrated with the preceding examples, these PM tasks that “pop up” during the virtual day follow on from the events described in the event cards.

The first circuit of the board is a practice day, in which participants are introduced to features of the game. Following the practice day, participants completed 2 virtual days with a total of 20 PM tasks: 8 event-based tasks (4 regular, 4 irregular); 8 time-of-day tasks (4 regular, 4 irregular), and 4 time-interval tasks (all regular; i.e., “check lung capacity at 2 and 4 minutes on the stop clock”).

*2.1.2.2. Scoring criteria.* The proportion of correct responses for each of the PM task types (regular/irregular event-based, time-of-day, and time-interval tasks) was used as an indicator of PM performance. A correct response was defined as one where the participant completed the appropriate PM task in response to the relevant cue. More specifically, responses were categorized as correct if they were performed when the token arrived at (or just passed) the target position on the board, and before the next die roll. In regard to time-interval tasks, responses were categorized as correct if the task was completed within 10 s of the target time.

*2.1.2.3. Naturalistic PM measure.* Developed as a parallel measure for the PM task type data captured by Virtual Week, MEMO involves a customized smart phone application—for time-of-day and time-interval tasks—and the use of a smartphone's camera function for event-based tasks (see Fig. 2). A smartphone (Optus L3 II, model LG-E425f; operated on the Android™ 4.1 operating system; dimensions: 6.11 cm (W) × 10.26 cm (H) × 1.19 cm (D)) was provided to each participant for the duration of the experiment. Participants were provided with detailed instructions about how to use MEMO before they left the laboratory and were requested not to use any external aids to help remember the tasks (e.g., writing notes). In Experiment 1, a 3 day block was assigned just for event based tasks, and a separate 3 day block just for time-based tasks (PM task type blocks were counterbalanced). For the 3 day period

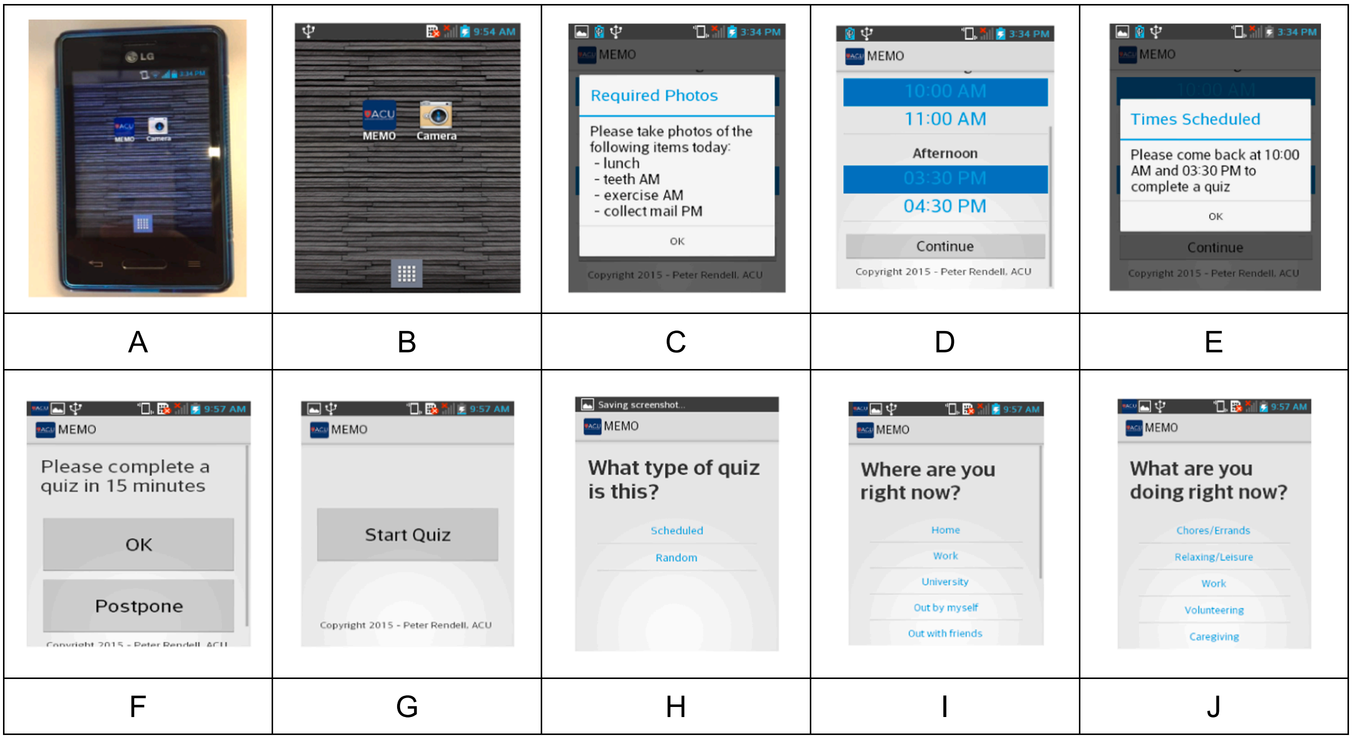


**Fig. 2.** Screen shots from MEMO assessment of PM in naturalistic setting. A: smartphone participants used. B: home screen showing MEMO quiz app and camera app. C: morning notification of the day's event-based tasks (require photos of events expected to occur later in the day). D: times selected for that day's time-of-day quizzes. E: final reminder of time-of-day tasks (shown immediately after "continue" button pressed in D). F: example of a random quiz notification. G: screen shown when participants remember to open app for "scheduled" (i.e., time-of-day) or "random" (i.e., time-interval) quiz. H: first quiz question (test of retrospective memory). I: first context question (location). J: second context question (activity).


that was exclusively for event-based tasks, smartphones were programmed to briefly display the selected target events each morning after participants opened a notification on the phone. For the 3 day period that was exclusively for time-based tasks, participants also received a notification in the morning, but in this case it requested the participant to select two time-of-day tasks—one in the morning (either 10 a.m. or 11 a.m.) and one in the afternoon (either 3:30 p.m. or 4:30 p.m.)—at which times participants were required to open the app to complete the scheduled time-of-day tasks (surveys).

For the days allocated to event-based tasks, the smartphones were programmed prior to participants being given the smartphone to take home. The smartphones were programmed to prompt participants each day at 9 a.m., using an auditory notification, to complete four event-based PM tasks each day (a total of 12 tasks over 3 days). After participants acknowledged the notification each morning, by opening the app, a list of four events that they were required to photograph that day were displayed (the act of photographing tokens of these events served as the event-based PM task). The list disappeared once participants pressed a button that read "OK". Half of the events were selected by participants from a larger list of common activities and events (e.g., collect mail, feed pet, water garden). Participants were instructed to choose events that were "extremely likely" to occur during the testing days. The remaining events were selected by the research assistant. The four events for each day differed along the following dimensions: regularity; experimenter given vs self-selected (from list of options); and duration of time-window for execution (i.e., short vs long). The first two events selected were regular (i.e., required on each of the 3 days): one was set by the research assistant (i.e., lunch each day) and one was self-selected (either: brushing teeth, taking medication, or passing a daily landmark). The other two photos chosen were irregular (i.e. unique to each day): one had a short (e.g., "putting on the dishwasher") and one had a long (e.g., "collect mail") window of opportunity.[1]

Both types of time-based tasks (time-of-day and time-interval) were intermixed and administered concurrently during the three day period for time-based PM tasks. Participants were again requested not to use any external aids (e.g., alarm clocks). For the time-of-day tasks, participants again received a notification on the smartphone at 9 am each day, but this time they were required to choose two times (one time-of-day in the morning and one in the afternoon) to come back and complete a brief quiz (quiz completion served as the time-of-day PM task). This selection of times (time-of-day quizzes) was made each morning in response to the smartphone notification. The times offered were 10 a.m. or 11 a.m. for the morning quiz and 3:30 or 4:30 p.m. for the afternoon quiz. At these times participants were required to click on the MEMO icon on the home screen (the only icon displayed, apart from the camera app icon) to start their quiz. For the time-interval task, participants received two "pop-up quiz" notifications (and accompanying auditory alerts) at times unbeknown to the participant (day one: 1:30

[1] As the dimensions of regularity, self-selection, and duration of window of opportunity for event-cued tasks were not of primary interest, they were not included in the analyses. However, as an aside it can be noted that there were no differences on these event task dimensions, except on the regularity dimension in Experiment 1 where both young and older adults showed better performance on self-selected regular tasks ($M$ = 0.79; $SD$ = 29) compared to experimenter-given regular tasks ($M$ = 0.69; $SD$ = 34), $F$ (3, 234) = 3.27, $p$ = .022, $\eta_p^2$ = 04, no other dimensions showed any substantive differences. ANOVAs with event-cued task dimension as a within subject variable found no difference for experiment 2: $F$ (3, 303) = 1.07, $p$ = .362, $\eta_p^2$ = 0.01; and experiment 3: $F$ (3, 213) = 0.64, $p$ = .590, $\eta_p^2$ = 0.01.

and 5:30 p.m.; day two: 12:30 and 6:00 p.m.; and day three: 1:00 p.m. and 5:30 p.m.). These pop up notifications instructed the participant to open the MEMO app after a specified time-interval (either 10, 15 or 20 min) to complete another brief quiz; thus participants had to carefully monitor the time to correctly perform the time-interval task. If the participant was unable to respond to the pop up quiz notification when it was sent, then the time-interval task commenced from the time when the participant was able to acknowledge receiving the notification. In Experiment 1 the mean delay between when the six pop up quiz notifications were sent and when participant acknowledged the time-interval task (i.e., began stop clock) ranged from 14 to 38 min, with a standard deviation range of 29 to 67 min, for each of the six time-interval quiz acknowledgements.

The quizzes for both the time-of-day and time-interval PM tasks consisted of two multiple-choice questions: 1. "where are you at the moment?" options included "at home", "at work", "out by myself", "out with friends", "friend/family's home", "work", and "university"; and 2. "what are you doing at the moment?" options included "relaxing/leisure", "doing chores/errands", "commuting", "volunteering", "work", "eating", and "caregiving". For frequency analyses, "eating" (selected by both young and older adults for < 9% of quizzes) were pooled with "relaxing/leisure" into the more general category of "leisure" activity; while commuting, volunteering, and caregiving (reported < 7%,4%, 3%, respectively for both groups) were pooled into "other" activity. Experiment 2 below also included data from an additional question: "is this a scheduled (i.e., chosen by participant in the morning) or random quiz (i.e., pop up notification during the day)?" to test retrospective memory.

*2.1.2.4. Scoring criteria*. The time-of-day tasks were classified as correct if the quiz was opened within ± 5 min of the required time (e.g., 3:30 p.m.). The time-interval tasks were classified as correct if the quiz was opened within ± 2 min of the required time (e.g., opening the app after 11 min for a requested 10 minute time-interval would be classified as correct). Photos (event-based task) were scored using the following categories: *correct* (photo taken as intended), *missed* (no photo), *remembered forgetfulness* (participant remembered at a later time, and took a photo of a note indicating this), *reminded by another task* (photo taken 1 min after another required photo), *contrived* (judged by researcher to be an unlikely photo of the true event, either due to time-stamp or appearance of event being staged), and *unable to take photo* (photo of note indicating event did not occur—e.g., "doctor's appointment rescheduled"—or that it was impractical to take photo at time of event). In the current study correct photos were scored as 1 with all other categories (except "unable to take photo") scored as 0.[2] The proportion of correct scores out of the total possible for all event-based (photo) tasks was the dependent variable. To assess the reliability of the scoring, a second independent researcher classified the photos taken by a subset of 25 randomly selected participants. Inter-rater reliability was computed using Siegel and Castellan (1988) variant of Cohen's kappa (1960) as the scoring was nominal. Analyses for each individual MEMO photo task (e.g., 'regular' photo of lunch required on day 1) yielded kappa values between 0.84 (Day 1 photo of 'regular' other—non-lunch—photo, e.g., medicine) and 1.00, indicating excellent inter-rater reliability (Hallgren, 2012).

*2.1.2.5. Design*. Experiment 1 was analyzed with a 2 × 3 mixed factorial design analysis of variance (ANOVA). *Age group* (young, older) was the between-subjects variable and *task type* (event, time-of-day, time-interval) was the within-subjects variable. The dependent variable was mean proportion of correct responses. There were separate analyses for each setting (laboratory or naturalistic), as there is no claim that the tasks are identical across settings, rather there is conceptual parity across settings. Also, the main interest was in the pattern of results between age groups within each setting and between different task types. Partial eta-squared was used for effect sizes, with 0.010 classified as small, 0.059 as medium, and 0.138 as large following Cohen (1988). All statistical analyses were conducted using IBM SPSS version 22 software.

### 2.2. Results

Fig. 3 shows the proportion of correct responses for each PM task type as a function of *age group* (young, older), *task type* (event, time-of-day, time-interval) and *setting* (laboratory, naturalistic).

#### 2.2.1. Laboratory

In the laboratory, age group and task type did not interact, $F$ (2, 176) = 1.27, $p$ = .283, $\eta_p^2$ = 0.01, but there was a main effect for age group, $F$ (1, 88) = 31.24, $p$ < .001, $\eta_p^2$ = 0.26: with young adults ($M$ = 0.69, $SD$ = 0.25) performing better than older adults ($M$ = 0.39, $SD$ = 0.25). There was also a main effect of task type, $F$ (2, 176) = 21.57, $p$ < .001, $\eta_p^2$ = 0.20. Bonferroni post hoc tests showed that for all participants (across young and older adults), performance was better on event-based ($M$ = 0.66, $SD$ = 0.29) compared to both time-of-day tasks ($M$ = 0.49, $SD$ = 0.27; $p$ < .001) and time-interval tasks ($M$ = 0.47, $SD$ = 0.35; $p$ < .001), while there was no difference between performance on the time-of-day and time-interval tasks ($p$ = 1.000).

#### 2.2.2. Naturalistic setting

In the naturalistic-setting, there was a main effect for age group, $F$ (1, 75) = 8.97, $p$ = .004, $\eta_p^2$ = 0.11, and a main effect of task type, $F$ (2, 150) = 60.15, $p$ < .001, $\eta_p^2$ = 0.45, but there was also an interaction between age group and task type, $F$ (2, 150) = 5.02, $p$ = .008, $\eta_p^2$ = 0.06. Follow up analysis of the interaction (separated by the task type) revealed a difference between age groups for the time-of-day task, $F$ (1, 75) =17.61, $p$ < .001, $\eta_p^2$ = 0.19. Specifically, older ($M$ = 0.65; $SD$ = 0.29) outperformed young ($M$ = 0.38; $SD$ = 0.28) adults. There was no age group difference for both the event (young: $M$ = 0.68; $SD$ = 0.22; older: $M$ = 0.75; $SD$ = 0.21), $F$ (1, 75) =1.82, $p$ = .181, $\eta_p^2$ = 0.02, and time-interval tasks (young: $M$ = 0.27; $SD$ = 0.31; older: $M$ = 0.36; $SD$ = 0.32), $F$ (1, 75) =1.31, $p$ = .257, $\eta_p^2$ = 0.02. When considering task type performance for each age group separately, there was a main effect for both young, $F$ (2, 74) = 33.27, $p$ < .001, $\eta_p^2$ = 0.47, and for older adults $F$ (2, 74) = 31.42, $p$ < .001, $\eta_p^2$ = 0.46. For both young and older adults Bonferroni post hoc tests showed performance on event-based to be better than performance on time-interval tasks (both $p$s < .001); and performance on time-of-day tasks was also better than performance on time-interval tasks for both young ($p$ = .05) and older ($p$ < .001) adults.

*2.2.2.1. Context of correct time-based tasks performance*. The brief quizzes completed by participants when performing the time-of-day (young group quiz completion range: 78 to 98%; older group quiz completion range: 85 to 95%) and time-interval tasks (young group quiz completion range: 67 to 90%; older group quiz completion range 60 to 83%), offer some insight into the context within which successful

[2] The number of photos coded as "unable to take photo" (i.e., where participant *explicitly* informed researcher that they were prevented from taking the photo due to external circumstances beyond their control; in other words, the event simply did not or could not occur) was extremely rare for all three experiments. For all 12 event cued *trials* the total frequency across each age-group was as follows: Experiment 1: young: $n$ = 4 (< 1%), older: $n$ = 3 (< 1%); Experiment 2: young-old: $n$ = 3 (< 1%), old-old: $n$ = 0 (0%); Experiment 3: young: $n$ = 0 (0%), older: $n$ = 0 (0%). The number of "Reminded by another task photo" (i.e., photo taken within 1 min of a preceding event-cued photo), which was coded as "incorrect" (largely as it could plausibly be construed as a strategy related to *contriving* a photo), was also very low for all three experiments (range 2.5–7.9%); as was the number of contrived (range 0.2–4.6%) and "remembered forgetfulness" photos (range from 0 to 6%).

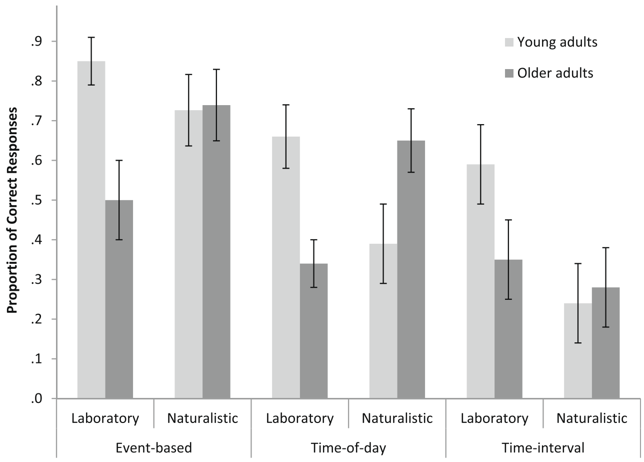


**Fig. 3.** Mean proportion correct on the PM tasks as function of task type (event-based, time-of-day, time-interval), setting (laboratory, naturalistic), and age (young, older). Error bars depict 2 standard error of the mean (corresponding to the 95% confidence interval for the mean).

performance by young and older adults occurred. Older adults who completed the time-of-day quiz were more likely to report being at home (62%[3]) compared to young adults (42%). Older and young adults reported interrupting leisure activities no more than half the time (50% vs 37%, respectively). Young adults reported "working" (44%) more than older adults (10%) but less "chores/errands" (13% vs 34%). Volunteering, caregiving, and commuting made up approximately 5% of interrupted ongoing activities for young adults and 10% for older adults.

For the time-interval task, older adults were more likely to be at home (58%) compared to young adults (39%). A similar pattern of activities as that found for the time-of-day tasks emerged—young and older adults switched from leisure activities 45% and 50% of the time, work 34% and 10%, and chores/errands 13% and 26%, respectively.

### 2.3. Discussion

The findings of Experiment 1 give both a broader and clearer picture of age-effects on PM in naturalistic settings in contrast to previous studies that used disparate task types across settings. The apparent paradox that has emerged from previous studies was replicated when just focusing on the results for tasks that previous studies have used in each setting. That is, on the time-interval tasks in the laboratory (Virtual Week) older adults were inferior to young adults, whereas on the time-of-day tasks in the naturalistic setting (MEMO) older adults were superior to young adults. The classic paradox of older adults being inferior to young adults in laboratory settings and superior in naturalistic settings was also replicated when just focusing on the other typical comparison of complex event-based tasks in the laboratory (event-based tasks in Virtual Week; complex because participants had to store multiple event-based tasks at the same time), with simpler time-of-day tasks in the naturalistic setting (simpler because participants only had to store two time-of-day tasks each day; selected each morning, with a much longer delay, and allowing for potential use of conjunction cues).

However, our inclusion of event-based and time-interval tasks in the naturalistic setting (tasks which have previously almost solely been confined to the laboratory) produced novel findings that are not consistent with the PM paradox. Specifically, the time-interval task findings demonstrated that older adults may not display a universal superiority to young adults in naturalistic-settings, nor that older adults' performance in naturalistic settings is always better than in the laboratory. For the event-based tasks a similar level of performance was found for both age groups in the naturalistic setting and, consistent with the greater environmental support afforded by event-based cues, both groups performed better on this task type compared to the time-interval PM tasks. Contrary to Rendell and Craik (2000), young and older adults performed comparably on the event-based tasks. This may reflect the more refined and controlled event-cued task paradigm (MEMO), or possibly the larger and more diverse young sample in the present study (Rendell and Craik only included 16 undergraduates in the young group). These findings of parity between both age groups, and the relative performance between PM task types within a naturalistic-setting, shows the critical importance of using conceptually parallel PM tasks across settings in order to elucidate the main factors contributing to the age-PM paradox.

The results of Experiment 1 also raise the question of whether a similar pattern of results would emerge when using parallel tasks in each setting to investigate the second key feature of the age-PM paradox: young-old being superior to old-old adults in the lab but there being no age differences in naturalistic settings. Thus, in Experiment 2, we focused on differences within older adults, at a period of the life span when most cognitive decline occurs, but when more established routines and lifestyle are assumed to provide a framework of environmental support to compensate for older adults' decline in PM found in the laboratory.

## 3. Experiment 2

Experiment 2 addressed the issue of whether old-old adults are more liable to PM lapses in general than young-old adults, and whether performance on event and the two time-based PM task types show differential cognitive age-effects. For exploratory purposes, participants in Experiment 2 were permitted to use external aids (e.g., making notes or setting alarms).

A larger sample of young-old and old-old adults than Experiment 1 participated, but used the same laboratory paradigm, as well as a refined version of the MEMO. Specifically, the naturalistic event- and time-based PM tasks used in Experiment 1 were combined and administered in a single block over 3 days (rather than in two separate blocks of 3 days). Thus, the number of PM tasks in the same period of time was doubled; arguably increasing the cognitive load to a level more commensurate with that in Virtual Week, where both the event and time-based tasks are completed on each virtual day. Extending the logic and hypotheses from Experiment 1, it was hypothesized that the young-old adults would outperform the old-old adults in the laboratory, particularly on the time-based tasks, but that this pattern of differences would be eliminated or attenuated on all PM task types in the naturalistic setting with the exception of the time-interval task, on which it was expected that the young-old would outperform the old-old in both settings, in line with the findings of Rendell and Craik (2000; Study 2).

**Table 3**
Characteristics of participants in Experiment 2.

| | Young-old adults | | Old-old adults | | *t*-test | | |
|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *t* | *p* | *d* |
| Age (in years) | 68.6 | 4.0 | 79.2 | 3.2 | | | |
| Education (in years) | 15.1 | 3.4 | 14.6 | 4.7 | 0.60 | .55 | 0.12 |
| Verbal IQ[a] | 110.1 | 7.6 | 114.6 | 6.0 | 2.98 | < .01 | 0.66 |

[a] Estimated using the National Adult Reading Test (NART; Nelson, 1982).

[3] The percentages reported in the context sections of results are for the percentage of completed quiz responses for each category of quiz response across the entire age group. It should be noted that each participant had the opportunity to complete up to 6 time-of-day or 6 time-interval quizzes for each of the context questions.

## 3.1. Method

### 3.1.1. Participants

A new sample of 104 healthy older volunteers—64 young-old (60–74 years old; 70% female) and 40 old-old (75–87 years old; 63% female) adults—was recruited as part of a larger experiment on cognitive training. Ethical approval was provided by the HREC of Australian Catholic University. Participants were screened using the Telephone Interview Cognitive test (Jager, Budge, & Clarke, 2003) and all had an education adjusted cut-off score of ≥33. Characteristics of the participants included in the final analyses are reported in Table 3. Six young-old and three old-old participants were excluded from analyses due to self-reported technical errors (or misunderstanding the task) on > 50% of any of the PM task type trials in the MEMO. Thus, data for analyses was used from the remaining 58 young-old and 37 old-old participants.

### 3.1.2. Materials and procedure

The same version of Virtual Week used in Experiment 1 was used in Experiment 2,[4] however, the refined version of MEMO combined both event- and time-based tasks over the same 3-day period; thus there were eight tasks per day instead of four. The MEMO was also developed to include a "postpone" feature for the pop up quizzes used to measure time-interval PM, which meant that, if the postpone option was selected, the participant received another notification that a time-interval task was ready to be acknowledged (i.e., stop clock started) 1 h later. Each pop quiz was postponed no more than once; the percentage of participants who postponed each quiz (trial) ranged from 3.5% to 10.5%.

Participants were trained in and undertook the MEMO task after a session completing other cognitive measures as part of a baseline session for a larger cognitive training experiment. Approximately one week later participants completed a 2-day version of Virtual Week. A noteworthy difference in this induction in contrast to Experiment 1 was that participants were advised that they "could use whatever strategies they would normally use in daily life to remember a task they needed to perform".

Prior to participants commencing the MEMO, they were given a printed guide for use of the smartphone and a support contact phone number to reach a research assistant if they had any technical difficulties. The main technical difficulty reported was that no quiz notifications or photo instructions were provided for the third day; this glitch was subsequently rectified, but a substantial number of participants consequently were only able to provide data for 2 days (young-old: $n = 14$; old-old: $n = 10$). To explore the possible impact of this technical issue, an ANCOVA was conducted, comparing the mean proportion of correct responses in the two age groups, with the number of days of MEMO completed as a covariate. There was a main effect for number of days, $F(1, 103) = 11.80$, $p = .001$, possibly indicative of a practice effect, with those who completed 3 days having a higher proportion of correct responses for the combination of all MEMO tasks ($M = 0.88$, $SD = 0.17$) compared to those who completed only 2 days ($M = 0.74$, $SD = 0.23$). There was no main effect for age group, $p = .322$, or interaction between number of days and age group, $p = .117$. This pattern did not change when each PM task type was separately analyzed using separate ANCOVAs.

As with Experiment 1, data were analyzed with 2 (age groups) x 3 (task types) mixed ANOVAs, with separate analyses for each setting.

[4] There was one difference worth noting. As the data in Experiment 2 were collected as part of a larger cognitive training study, the version of Virtual Week also included a common measure of meta-memory, i.e., judgment of learning questions ("how likely are you to perform this task?" with slider bar for response, 0–100%) after each task was encoded. An additional task assessing memory of the event cards encountered and the decisions made was also included—presumably encouraging more focus on the ongoing task.

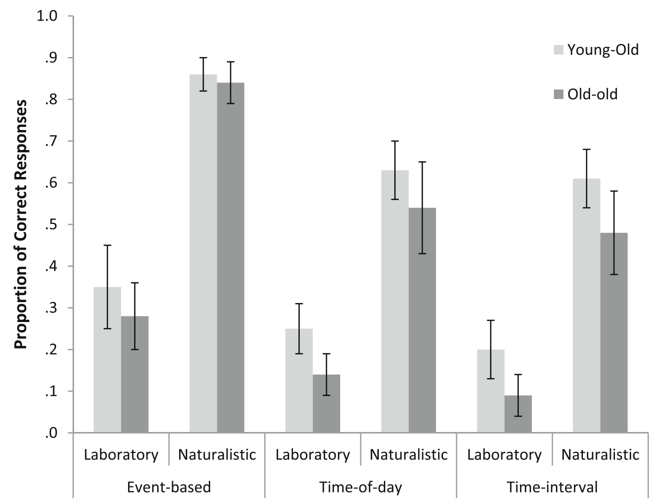


**Fig. 4.** PM performance as a function of PM cue type (event-based, time-of-day, time-interval), setting (laboratory, naturalistic), and age group (young-old, old-old). Error bars depict 2 standard error of the mean (corresponding to the 95% confidence interval for the mean).

## 3.2. Results

Fig. 4 shows the proportion of correct responses for each PM task type as a function of age group (young-old, old-old), task type (event, time-of-day, time-interval) and setting (laboratory, naturalistic).

### 3.2.1. Laboratory

There was no interaction between age group and task type in the laboratory, $F(2, 174) = 0.89$, $p = .413$, $\eta_p^2 = 0.01$, but there was a main effect for age group, $F(1, 87) = 6.40$, $p = .013$, $\eta_p^2 = 0.07$, with young-old ($M = 0.26$; $SD = 0.16$) performing better than old-old adults ($M = 0.17$, $SD = 0.16$) as shown in Fig. 4. There was also a main effect of task type, $F(2, 174) = 11.99$, $p < .001$, $\eta_p^2 = 0.12$. Bonferroni post hoc tests showed that for all participants (across young-old and old-old groups) performance was significantly better on event-based ($M = 0.30$, $SD = 0.27$) than time-interval ($M = 0.14$, $SD = 0.23$) tasks ($p < .001$), and performance on event-based tasks was also significantly better than on time-of-day tasks ($M = 0.20$, $SD = 0.21$; $p = .027$); but there was no difference between performance on time-of-day and time-interval tasks ($p = .178$).

### 3.2.2. Naturalistic setting

There was no main effect for age group $F(1, 93) = 3.51$, $p = .064$, $\eta_p^2 = 0.04$ and no interaction between age group and task type in the naturalistic setting $F(2, 186) = 1.19$, $p = .307$, $\eta_p^2 = 0.01$, however there was a main effect of task type, $F(2, 186) = 14.47$, $p < .001$, $\eta_p^2 = 0.38$. Bonferroni post hoc tests showed that for all participants (across young-old and old-old groups) performance was significantly better on event-based tasks ($M = 0.86$, $SD = 0.16$) than on both time-of-day tasks ($M = 0.60$, $SD = 0.32$) and time-interval tasks ($M = 0.54$, $SD = 0.30$; $p < .001$). There was no difference between performance on time-of-day and time-interval tasks ($p = .212$).

*3.2.2.1. Context of correct time-based task performance in naturalistic setting.* The completion of time-of-day quizzes ranged from 72 to 92% for the young-old group, and 63 to 73% for the old-old group. Recognition of time-of-day quiz type (i.e., "is this a scheduled or random quiz?") was comparably high for young-old (93%) and old-old (90%) participants. Home was a common location for time-of-day quizzes for young-old (65%) and old-old (70%) groups. Young-old and old-old groups reported leisure activity on 35% and 45% of time-of-day quizzes, respectively. Engagement in chores was comparable for young-old (36%) and old-old (36%) groups, while work was more common for young-old (13%) than old-old (6%) participants. The combined tasks of commuting, volunteering, and caregiving were infrequently reported by both young-old (6%) and old-old (10%) participants.

For participants who completed the time-interval quizzes (range for young-old: 52 to 86%; range for old-old: 70 to 97%), correct recognition of quiz type was 83% for young-old and 78% for old-old. Old-old adults were more often at home (79%) than young-old (66%), while young-old were engaged in leisure activity (57%) more than old-old (43%). For young-old and old-old, respectively, work was reported on 7% and 5%, and chores on 29% and 28%, of completed quizzes. Volunteering, caregiving, and commuting combined were reported least often as ongoing tasks at the time of time-interval task completion: 7% of quizzes for young-old and 10% for old-old.

### 3.3. Discussion

The pattern of results for Experiment 2 in both the laboratory and naturalistic settings essentially replicated the second aspect of the age-PM paradox (with young-old outperforming old-old in the laboratory, but groups performing equally well in a naturalistic-setting). The results were also consistent with what the multiprocess framework would predict (i.e., age-related declines in PM performance associated with lower levels of environmental support; Mcdaniel & Einstein, 2000; Mcdaniel, Umanath, Einstein, & Waldum, 2015). In particular, event-based tasks, which typically rely on relatively spontaneous cognitive processes, were the tasks performed best by both older age groups in each setting. In the laboratory, there was a pattern of decline in performance for both age groups as environmental support declined (i.e., performance was best on event-based tasks, followed by time-of-day tasks, and was worst on time-interval tasks). As predicted, young-old adults outperformed old-old adults in the laboratory on each task type, and age and task type did not interact.

Our findings for overall performance on the different task types in the naturalistic setting followed a similar pattern to those in the laboratory (i.e., performance was best on event-based, followed by time-of-day, and then time-interval PM tasks), albeit with generally better performance in the real life setting than in the laboratory setting, with reports of external aids and strategies used to compensate for PM task demands.[5] The pattern of better performance in the naturalistic setting compared to the laboratory setting was seen for both groups of older adults, for all task types, and most notably for the time-interval task. These findings clearly show that many healthy, older adults can compensate in a naturalistic setting for the detrimental effects of cognitive aging on PM performance as indexed in the laboratory.

[5] Ad hoc interviews with a subset of young-old ($n = 48$) and old-old ($n = 31$) adults were conducted after completing the MEMO. For time-based tasks more young-old reported using a compensatory strategy than old-old: time-of-day (73% vs 58%) and time-interval (75% vs 67%). The most common strategy for time-of-day tasks was “making a note” (young-old: 33% vs old-old: 29%) followed by “setting an alarm” (young-old: 23% vs old-old: 13%). In contrast for time-interval tasks the most common strategy was “setting an alarm” (young-old: 31% vs old-old: 20%) followed by “monitoring a clock” (young-old: 21% vs old-old: 13%). For event-based tasks, both groups had 90% of respondents report using a compensatory strategy; by far the most common strategy was “making a note” (young-old: 67% vs old-old: 80%). Only a small percentage of young-old (13%) and old-old (17%) reported their strategy use as being *dissimilar* to how they would normally remember PM tasks in daily life. As this data is from ad hoc interviews we suggest caution in their interpretation, and do not include them in the main analyses.

## 4. Experiment 3

A natural corollary of Experiments 1 and 2 is to investigate the pattern of performance between young and older adults when the MEMO tasks are all given in one 3 day block and all participants are permitted to use strategies. The aim of Experiment 3 was therefore to investigate whether the age-PM paradox would persist when both young and older adults were administered a 2-day version of Virtual Week, along with the MEMO, including the same parameters for MEMO as were used in Experiment 2. It was hypothesized that young participants would continue to outperform older adults on all Virtual Week PM task types in the laboratory setting, while on the MEMO, in the naturalistic-setting, young and older adults were predicted to show comparable performance on both event- and time-interval tasks, with older adults showing better performance compared to younger adults on time-of-day tasks.

### 4.1. Method

#### 4.1.1. Participants

A new sample of 83 healthy volunteers—42 young (18–34 years old; 59.5% female) and 41 older (60–83 years old; 68.3% female) adults—were recruited, screened, and compensated for their participation as in Experiment 2. The young adults were primarily university students. Characteristics of the participants included in the final analyses are reported in Table 4. Due to self-reported technical difficulties, misunderstanding on > 50% of PM tasks, or poor compliance with MEMO instructions (e.g., not acknowledging time-interval task notifications: 18 young adults and 9 older adults), the final sample for the analyses was 23 young adults and 31 older adults. **Ethical approval was provided by the HREC of the Australian Catholic University.**

#### 4.1.2. Materials and procedure

The version of Virtual Week and MEMO, and the procedure used in Experiment 3, was the same as in Experiment 2.

### 4.2. Results

Fig. 5 shows the proportion of correct responses for each PM task type as a function of age group (young, older), task type (event, time-of-day, time-interval) and setting (laboratory, naturalistic).

#### 4.2.1. Laboratory

There was no interaction between age group and task type in the laboratory, $F(2, 140) = 0.58$, $p = .561$, $\eta_p^2 = 0.01$, but there was a main effect for age, $F(1, 70) = 32.39$, $p < .001$, $\eta_p^2 = 0.32$, with young adults ($M = 0.76$; $SD = 0.20$) performing better than older adults ($M = 0.48$, $SD = 0.20$), as shown in Fig. 5. There was also a main effect of task type, $F(2, 140) = 21.20$, $p < .001$, $\eta_p^2 = 0.23$. Bonferroni post hoc tests showed that for all participants (across young and older groups) performance was better on event-based ($M = 0.70$, $SD = 0.25$) than time-of-day ($M = 0.48$, $SD = 0.24$) tasks ($p < .001$); but there was no difference between performance on event-based and time-interval tasks ($M = 0.68$, $SD = 0.32$; $p = 1.00$). Performance was also better on time-interval tasks compared to time-of-day tasks ($p < .001$).

#### 4.2.2. Naturalistic setting

In the naturalistic-setting, there was no main effect for age group, $F(1, 52) = 2.79$, $p = .101$, $\eta_p^2 = 0.05$. However, there was a main effect of task type, $F(2, 104) = 30.96$, $p < .001$, $\eta_p^2 = 0.37$. Bonferroni post hoc tests showed that performance on event-based tasks ($M = 0.82$, $SD = 0.15$) was superior to that on time-of-day ($M = 0.58$; $SD = 0.26$) and time-interval ($M = 0.56$; $SD = 0.26$) tasks ($p < .001$); performance on time-of-day and time-interval tasks did not substantially

**Table 4**
Characteristics of participants in Experiment 3.

| | Young adults | | Older adults | | *t-test* | | |
|---|---|---|---|---|---|---|---|
| | *M* | *SD* | *M* | *SD* | *t* | *p* | *d* |
| Age (in years) | 22.9 | 4.1 | 70.6 | 5.5 | | | |
| Education (in years) | 15.6 | 2.0 | 16.7 | 4.0 | 1.59 | .12 | 0.35 |
| Verbal IQ[a] | 108.9 | 4.8 | 119.2 | 5.4 | 9.02 | < .001 | 2.01 |

[a] Estimated using the National Adult Reading Test (NART; Nelson, 1982).

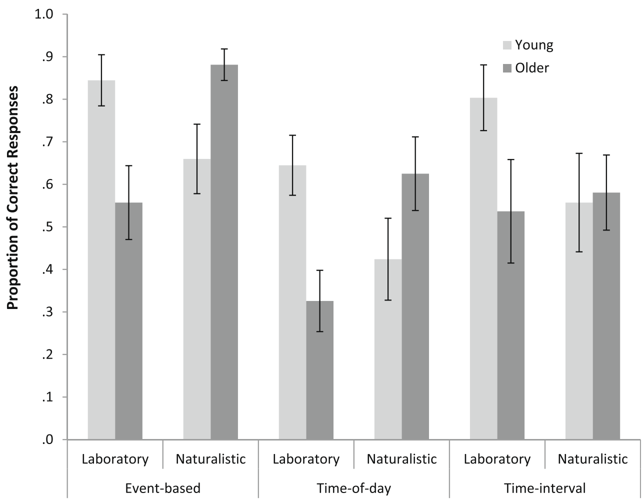


**Fig. 5.** PM performance as a function of PM cue type (event-based, time-of-day, time-interval), setting (laboratory, naturalistic), and age group (young, older). Error bars depict 2 standard errors of the mean (corresponding to the 95% confidence interval for the mean).

differ ($p = 1.000$). There was no interaction between age group and task type, $F(2, 104) = 1.16$, $p = .316$, $\eta_p^2 = 0.02$.

*4.2.2.1. Context of correct time-based task performance in naturalistic setting.* The completion of time-of-day quizzes ranged from 50 to 67% for the young and 71 to 88% for the older adult groups. Recognition of time-of-day quiz type was greater for young (98%) compared to older adults (93%). Being at home was frequently reported by young (59%) and older (41%) adults, while young adults (22%) were more often at work or university than older adults (12%), and older adults (12%) were more likely to be out with friends than young adults (5%) when responding to time-of-day quizzes.

In terms of activities reported on time-of-day quizzes, young adults tended to be more engaged in leisure activity (49%) than older adults (35%); while engagement in work for young and older adults was similar (22% vs 17%, respectively). Older adults (36%) reported being more frequently engaged in chores immediately prior to responding to time-of-day quizzes than young adults (20%).

For participants who completed the time-interval quizzes (range for young adults: 26 to 55%; range for old-old: 66 to 76%), recognition of quiz type was higher for young adults (96%) than older adults (88%). Older adults were more often at home (71%) than young adults (55%), while similar engagement in leisure activity was reported by young (44%) and older adults (49%). Work was reported as an ongoing task for 32% of completed time-interval tasks for young adults and 16% for older adults, while chores were 21% and 26%, respectively. Volunteering, caregiving, and commuting combined were reported least often: 3% of quizzes for young and 8% for older adults.

### *4.3. Discussion*

As with Experiment 1 and 2, young adults consistently outperformed older adults on the Virtual Week laboratory paradigm in Experiment 3. For both young and older adults, mean proportion correct on the time-interval tasks was comparable with levels of performance on event-based tasks; while performance on both task types was substantially better than performance on the time-of-day tasks. The generally better performance of both groups on Virtual Week in Experiment 3 compared to Experiment 2 is likely due to the removal of pop up messages designed to capture data on meta-memory (which was used in Experiment 2 as part of a larger study on cognitive training).

The lack of an age-effect on performance in the naturalistic setting was particularly striking as a contrast to the naturalistic setting pattern found in Experiment 1. This may reflect the pattern of higher levels of missing data for technical and compliance reasons. For example, for the time-interval task the sample number reduced to 22 young and 21 older adults; whereas in Experiment 1 there were 38 young and 39 older adults. While the lower sample numbers in Experiment 3 require cautious interpretations of the results, the lack of a difference between the two time-based tasks, in contrast to Experiment 1 in which external aids were prohibited, suggests that individuals in both age groups could effectively compensate for the low environmental support in these tasks.[6] In terms of location and activities reported for those who remembered to perform the time-based tasks, it was also noteworthy that being at home did not differ between groups on time-of-day tasks, but did differ (with older adults more likely to be at home) on time-interval tasks. As the time-interval tasks unexpectedly "popped up" during the course of the day, there was no opportunity for participants to strategically select less cognitively demanding environments when responding to these tasks. Contrary to common assumptions, younger adults *were more likely* to be engaged in leisure activities than older adults when remembering to perform time-of-day tasks; while in the case of time-interval tasks there was almost no difference in terms of leisure activities.

The better performance of older adults on event-based MEMO tasks in Experiment 3 may reflect older adults making more use of the opportunity to use external aids or reminders; for example, in ad hoc interviews 50% of older adults reported making a written note regarding event-based tasks compared to only 34% of young adults. It can also be noted that young adults performance on the event-based tasks is quite comparable between Experiment 1 and 3 (both around 0.7), whereas older adults show better performance compared to the older adults in experiment 1, consistent with the idea that they are using the opportunity for external aids to their advantage.

[6] As with Experiment 2, an ad hoc post-MEMO interview was completed by a subset of young ($n = 32$) and older ($n = 23$) participants. For time-of-day tasks 87% of young and 83% of older adults reported using a strategy. For young and older adults, respectively, these included: "making a note" (19% vs 38%), "setting an alarm" (25% vs 17%), "planning/scheduling into their day" (3% vs 13%), using a conjunction cue (9% vs 8%), "monitoring a clock" (3% vs 0%) and "other" strategies (28% vs 8%). For time-interval tasks 65% of young and 74% of older adults reported a strategy. For young and older adults, respectively, these included: "setting an alarm" (35% vs 25%), "making a note" (0% vs 33%), "monitoring a clock" (9% vs 4%), and "other" strategies (21% vs 12%). For event-based tasks 72% of young and 87% of older adults reported using a strategy. For young and older adults, respectively, these included: "making a note" (34% vs 50%), "planning/scheduling into their day" (6% vs 4%), using conjunction cues (3% vs 0%), and "other" strategies, including placing phone in prominent place, removing delay, mental note, and asking a spouse or partner to remind them (28% vs 33%). Most older adults (55%) reported the time-interval task as the "most difficult" task (compared to 19% of young adults). Importantly the majority of young (73%) and older adults (87%) claimed that their reported strategy use was "similar" to what they would normally do to complete PM tasks in their daily life.

**Table 5**
A summary of key findings from three experiments comparing young, older, and young-old and old-old adults in laboratory and naturalistic settings on different PM task types.

| | Age groups | | Task types by setting with age-related differences in PM performance | | | |
|---|---|---|---|---|---|---|
| | Range (years) | *n*[a] | Laboratory | | Naturalistic | |
| Experiment 1 | 19–30 | 40/40 | Event | Y > O | Event | Y = O |
| | 65–91 | 41/41 | 'Time-of-day' | Y > O | Time-of-day | Y < O |
| | | | Time-interval | Y > O | Time-interval | Y = O |
| Experiment 2 | 60–74 | 64/58 | Event | YO > OO | Event | YO = OO |
| | 75–87 | 40/37 | 'Time-of-day' | YO > OO | Time-of-day | YO = OO |
| | | | Time-interval | YO > OO | Time-interval | YO = OO |
| Experiment 3 | 18–34 | 42/23 | Event | Y > O | Event | Y = O |
| | 60–83 | 41/31 | 'Time-of-day' | Y > O | Time-of-day | Y = O |
| | | | Time-interval | Y > O | Time-interval | Y = O |

*Note*. Y: young; YO: young-old; OO: old-old. The " > " sign indicates which group performed better (the group on the left-hand side of the sign); the "=" or a "," sign indicates no group differences in performance.

[a] The numbers before and after the forward slash refers to the number of participants tested in the laboratory and naturalistic-setting, respectively.

## 5. General discussion

The present investigation is the first to explicitly consider the different cognitive demands and expected age-effects associated with theoretically distinct time-based tasks that have typically been conflated in the cognitive aging and PM literature. By making and testing this distinction, these experiments contribute substantially to a growing body of literature on the nuances of the PM paradox (e.g., Aberle, Rendell, Rose, Mcdaniel, & Kliegel, 2010; Bailey et al., 2010; Hering, Cortez, Kliegel, & Altgassen, 2014; Kliegel, Rendell, & Altgassen, 2008; Niedźwieńska & Barzykowski, 2012; Schnitzspahn et al., 2011). The results show that the apparent pattern of age-effects differing across settings is partially a function of task type (e.g., time-based tasks) rather than setting per se. In particular, our results showed that in naturalistic-settings, older adults showed comparable performance on time-interval tasks. This was the case both under conditions when the use of external aids is permitted (Experiment 2 and 3) and when not permitted (Experiment 1). The difference in performance between tasks for both age groups was generally in the direction expected using the multiprocess framework (with the exception of substantially better performance on time-interval tasks compared to time-of-day tasks in the laboratory in Experiment 3), and thus consistent with the theory of more self-initiated operations being required as environmental support decreases (Craik, 1986; Mcdaniel & Einstein, 2000). Table 5 gives a schematic summary of the key findings of the three experiments.

It can be seen in the pattern of results presented in Table 5 that there was a general consistency within laboratory and naturalistic settings (cf. Table 1 of pattern of results of key studies on the PM paradox). In the more cognitively demanding context of the laboratory (where environmental support is typically less than in daily life), PM performance was worse with increasing age. However, the results for the naturalistic setting show that older adults can compensate for this decline, particularly when permitted to use their own strategies (Experiments 2 and 3). Thus, a key explanation for the PM paradox is a lack of parallel PM task types across settings, which systematically differ in the level of environmental support afforded, and how these task characteristics interact with age-related changes in cognitive processes (e.g., more reliance on automatic rather than effortful processes).

A key strength of the three experiments was the development and use of the MEMO, which addressed limitations of the Rendell and Craik (2000) naturalistic paradigm (Actual Week). The MEMO was relatively easy and a convenient means for participants to record their PM task performance, and the use of the mobile phone camera meant that the completion of event-based tasks could be more reliably verified. It is plausible to speculate that these improvements in the event-cued naturalistic task may account for the main divergence in the present study's findings from those of Rendell and Craik. That is, while Rendell and Craik found older adults performed better than young adults on event-based tasks in Actual Week, no difference emerged between young and older adults on event-based tasks in Experiment 1 and 3. In contrast, the similar performance for young-old and old-old in Experiment 2 on event-based tasks was consistent with the findings of Rendell and Craik. The time-interval task in the MEMO was also more similar to typical time-interval tasks in the laboratory (e.g., 10-minute delay) than the longer, 30- and 60-minute delays used in Rendell and Craik. This inclusion of more comparable time-interval tasks in laboratory and naturalistic settings makes a novel and important contribution to the cognitive aging and PM literature.

The experiments confirmed previous findings (e.g., Rendell & Craik, 2000; Rendell & Thomson, 1999) that older adults are very accurate in naturalistic settings on the type of time-based tasks used in previous naturalistic PM studies (time-of-day tasks). However, Experiment 1 showed that older adults perform very poorly in naturalistic-settings—when asked to refrain from the use of external aids and compensatory strategies—on the type of time-based tasks typically used in laboratory studies (time-interval tasks). As with Actual Week in Rendell and Craik (2000), Experiment 1 found the poorer performance of older adults in the laboratory on time-interval tasks in Virtual Week was also evident for the same older adult group on another set of time-interval tasks in a naturalistic setting (MEMO). When external aids and compensatory strategies were permitted, the generally poor performance of both age groups on time-interval tasks in the naturalistic setting improved greatly (from approximately 30% in Experiment 1 to 50–60% in Experiment 2 and 3).

The pattern of results for the time-based tasks in the laboratory raises the possibility that the time-of-day (and to a lesser extent the event-based) tasks are not as well matched across settings as the time-interval tasks. In Virtual Week, the time-of-day tasks were indicated by *virtual* time-of-day cues; that is, the visible virtual clock (calibrated to the position of the token on the board) and the content of event cards, which had descriptions of activities relevant to the virtual time-of-day (e.g., choosing what to eat at breakfast). However, the virtual time-of-day cues do not have the same intensity or richness of time-of-day cues in naturalistic settings that may support spontaneous retrieval of time-of-day PM tasks.[7] Moreover, the virtual time-of-day PM tasks (like the virtual event-based PM tasks) are more demanding given the tighter time window. Thus, more demanding "time" monitoring is probably required in Virtual Week than the real time monitoring in MEMO for the time-of-day PM task type.

The need for further research on temporal contexts for PM task performance was also highlighted in the pattern of findings on MEMO event-based tasks. For example, there were robust differences in the performance of young and older participants when presented with multiple event-based tasks spread over approximately an hour in the laboratory setting using the Virtual Week paradigm. However, there were no age-differences on MEMO event-based tasks in Experiment 1 when using the same number of event-based task trials spread over several days. This longer time span, plus a separate block of days in Experiment 1 exclusively for event-based tasks (four per day), may have made the PM tasks more like a series of 'single' targets. This would be consistent with the finding of Einstein, Holland, Mcdaniel, and Guynn (1992) on the complexity effect, in which age-effects in the laboratory were apparent when there were multiple, different target events, but eliminated when there was only a single, repeated target event. Similarly, older adults' better PM performance in naturalistic settings in general, compared to their own performance in laboratory settings, may reflect the reduced complexity and cognitive burden of the naturalistic setting PM tasks, in all the Experiments, by spreading the tasks over a longer time period—3 days in MEMO for 12 PM tasks in Experiment 1 (separate blocks for event and time-based tasks), and 24 PM tasks in Experiment 2 and 3 (all tasks combined in one block) compared to approximately an hour for 20 PM tasks in Virtual Week for each Experiment.

Related to temporal context is the potential for planning provided by the familiarity and possible salience of typically occurring, event targets available in naturalistic settings. The naturalistic-setting target events in the current study were pre-selected by participants from a list of options to ensure that they were very likely to occur. Being able to (strategically) select familiar target events that were "extremely likely to occur" may have provided substantial environmental support for older adults to complete these tasks successfully. Planning was also possible for the time-of-day tasks in the naturalistic-setting, which may have been reflected by the high percentage of young-old (65%) and old-old (70%) in Experiment 2, and young (59%) and older (41%), in Experiment 3 being at home (low cognitive demand environment) when completing this PM task. In contrast, for the time-interval tasks strategic situation-selection was less possible, and thus it was likely fortuitous whether the participant was at home or not. Future studies should systematically manipulate the use of external aids and strategies, between and within groups, to better elucidate how older adults compensate for time-interval (and other) PM tasks.

A limitation of the Experiments presented here is that we did not investigate the focal versus non-focal distinction for event-cued PM tasks. Exploring this event-based task dimension *alongside* the time-of-day versus time-interval task distinction would potentially strengthen the argument that time-based tasks may be usefully divided in terms of activating predominantly automatic (time-of-day) versus controlled (time-interval) processes. For example, it remains to be established whether age-effects would be equally apparent for non-focal event-based tasks as they are for time-interval tasks in a naturalistic setting. If such a pattern were to be identified, it would provide converging evidence that an extension of multiple processes to time-based tasks is a legitimate (and exciting) development of the multiprocess framework (McDaniel & Einstein, 2000).

In sum, these data provide an important theoretical advance in explicitly considering the differing cognitive demands and age-effects associated with distinct time-based tasks in laboratory and naturalistic settings. It also makes a novel contribution and adds important evidence about some of the neglected mechanisms contributing to the long-standing age-PM paradox. The experiments show that previous findings of superior performance by older adults in naturalistic settings appear to be largely a function of time-of-day PM tasks being predominantly used in this setting in past studies. Additionally, they indicate that for short time-interval tasks older adults' performance was low across both settings, while younger adults' performance was higher than older adults in the laboratory setting and at a similar low level to older adults' in the naturalistic setting. However, on the time-interval task in Experiment 2 and 3, which allowed the use of external aids, both young and older adults performed at substantially higher levels. This pattern suggests that older adults are able to use strategies to compensate for tasks that would otherwise have high monitoring demands (and not by strategically choosing less distracting locations, e.g., at home, or less demanding ongoing tasks, e.g., leisure activities, compared to younger adults). Thus, the pattern that previously seemed paradoxical is no longer apparent when parallel tasks are used across settings and a distinction is made between time-interval and time-of-day tasks. This time-based task distinction, particularly in naturalistic settings using the MEMO paradigm, may prove to be as theoretically fruitful for future research on cognitive aging and PM as that frequently made in the laboratory for focal (low in cognitive demand) and non-focal (high in cognitive demand) event-based tasks.

## Supplementary material

The data supporting all analyses presented in the results of each experiment are provided as supplementary online material. For each experiment, we provide each participant's demographic data; proportion correct score for event, time-of-day, and time-interval cued tasks for each setting (i.e., Virtual Week and MEMO); and contextual data (MEMO). The contextual data for Experiment 1 includes reported location and activity; for Experiment 2 and 3: location, activity, and retrospective memory test (recognition of quiz type). Note: odd quiz numbers (e.g., "1.1"; day 1 quiz 1] are always time-of-day quizzes; and even quiz numbers (e.g., "1.2"; day 1 quiz 2) are always time-interval quizzes. Supplementary data associated with this article can be found in the Mendeley Data archive at http://dx.doi.org/10.17632/8m84swkfyp.3

## Credit author statement

**Simon Haines**: Conceptualisation, Methodology, Investigation, Formal analysis, Writing - original and revised draft; **Susan Randall**: Methodology, Investigation, Formal analysis; **Gill Terrett**: Methodology, Investigation, Formal analysis, Supervision, Writing - review & editing; **Lucy Busija**: Formal analysis, Supervision; **Gemma Tatangelo**: Formal analysis, Supervision; **Skye N. McLennan**: Methodology, Supervision, Writing - review & editing; **Nathan S. Rose**: Methodology, Writing - review & editing; **Matthias Kliegel**: Methodology, Writing - review & editing; **Julie Henry**: Methodology, Formal analysis, Investigation, Writing - review & editing; **Peter Rendell**: Conceptualization, Methodology, Investigation, Formal analysis, Supervision, Writing - review & editing.

[7] It may appear that time-of-day tasks in naturalistic settings are an obscure or exotic species of event-cued PM tasks given the constellation of time-related cues that occur in such settings and the possibility of using conjunction cues as a strategy. However, we continue the convention of subsuming these appointment-like tasks under the category of time-based tasks, while arguing they are not the only type of time-based task. Routine events or activities that do occur at specific times of day might be *sufficient* to elicit spontaneous recall of time-of-day PM tasks, but there is no specific event or activity (or cluster of these) which is *necessary* for the environmental support that time-of-day tasks afford.

## Acknowledgements

This research was funded by an ARC Linkage Project Grant (LP150100140). We gratefully acknowledge the generous support of our linkage partner Villa Maria Catholic Homes. We would also like to acknowledge the support and help from Colleen Doyle, Nick Koleits, Rachel Pelly, Emma Lawrence, Aimee Brown, Gavriel Garrison and Kirra Liu for their important contribution to conducting the data collection and assistance with recruitment; Trevor Daniels for programming both Virtual Week and MEMO; and Yvonne Wells for proofreading the revised manuscript.

## References

Aberle, I., Rendell, P. G., Rose, N. S., McDaniel, M. A., & Kliegel, M. (2010). The age prospective memory paradox: Young adults may not give their best outside of the lab. *Developmental Psychology, 46*, 1444–1453. https://doi.org/10.1037/a0020718.

Bailey, P. E., Henry, J. D., Rendell, P. G., Phillips, L. H., & Kliegel, M. (2010). Dismantling the "age–prospective memory paradox": The classic laboratory paradigm simulated in a naturalistic setting. *The Quarterly Journal of Experimental Psychology, 63*, 646–652. https://doi.org/10.1080/17470210903521797.

Browning, C. A., Harris, C. B., Van Bergen, P., Barnier, A. J., & Rendell, P. G. (2018). Collaboration and prospective memory: Comparing nominal and collaborative group performance in strangers and couples. *Memory, 26*, 1206–1219. https://doi.org/10.1080/09658211.2018.1433215.

Carstensen, L. L., Isaacowitz, D. M., & Charles, S. T. (1999). Taking time seriously: A theory of socioemotional selectivity. *American Psychologist, 54*, 165–181. https://doi.org/10.1037/0003-066X.54.3.165.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, N.J.: L. Erlbaum Associates.

Craik, F. I. (1986). A functional account of age differences in memory. In F. Klix, & H. Heagendorf (Eds.). *Human memory and cognitive capabilities*. Amsterdam, Netherlands: Elsevier.

Einstein, G. O., Holland, L. J., McDaniel, M. A., & Guynn, M. J. (1992). Age-related deficits in prospective memory: The influence of task complexity. *Psychology and Aging, 7*, 471–478. https://doi.org/10.1037/0882-7974.7.3.471.

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

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

Ellis, J. A. (1996). Prospective memory or the realization of delayed intentions: A conceptual framework for research. In M. A. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.). *Prospective memory: Theory and applications* (pp. 1–22). Mahwah, NJ: Erlbaum.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198. https://doi.org/10.1016/0022-3956(75)90026-6.

Haines, S. J., Shelton, J. T., Henry, J. D., Terrett, G., Vorwerk, T., & Rendell, P. G. (2019). Prospective memory and cognitive aging. In B. Knight (Ed.). *Oxford research encyclopedia of psychology*Oxford University Presshttps://doi.org/10.1093/acrefore/9780190236557.013.381.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*, 23–34. https://doi.org/10.20982/tqmp.08.1.p023.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356–388. https://doi.org/10.1037/0096-3445.108.3.356.

Henry, J. D., MacLeod, M. S., Phillips, L. H., & Crawford, J. R. (2004). A meta-analytic review of prospective memory and aging. *Psychology and Aging, 19*, 27–39. https://doi.org/10.1037/0882-7974.19.1.27.

Henry, J. D., Rendell, P. G., Phillips, L. H., Dunlop, L., & Kliegel, M. (2012). Prospective memory reminders: A laboratory investigation of initiation source and age effects. *The Quarterly Journal of Experimental Psychology, 65*, 1274–1287. https://doi.org/10.1080/17470218.2011.651091.

Hering, A., Cortez, S. A., Kliegel, M., & Altgassen, M. (2014). Revisiting the age-prospective memory-paradox: The role of planning and task experience. *European Journal of Ageing, 11*, 99–106. https://doi.org/10.1007/s10433-013-0284-6.

Hering, A., Kliegel, M., Rendell, P. G., Craik, F. I. M., & Rose, N. S. (2018). Prospective memory is a key predictor of functional independence in older adults. *Journal of the International Neuropsychological Society, 24*, 640–645. https://doi.org/10.1017/S1355617718000152.

Ihle, A., Schnitzspahn, K., Rendell, P. G., Luong, C., & Kliegel, M. (2012). Age benefits in everyday prospective memory: The influence of personal task importance, use of reminders and everyday stress. *Aging, Neuropsychology, and Cognition, 19*, 84–101. https://doi.org/10.1080/13825585.2011.629288.

Jager, C., Budge, M. M., & Clarke, R. (2003). Utility of TICS-M for the assessment of cognitive function in older adults. *International Journal of Geriatric Psychiatry, 18*, 318–324. https://doi.org/10.1002/gps.830.

Kliegel, M., Ballhausen, N., Hering, A., Ihle, A., Schnitzspahn, K., & Zuber, S. (2016). Prospective memory in older adults: Where we are now and what is next. *Gerontology, 62*, 459–466. https://doi.org/10.1159/000443698.

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*, 203–208. https://doi.org/10.1037/0882-7974.23.1.203.

Kliegel, M., Martin, M., McDaniel, M. A., & Phillips, L. H. (2007). Adult age differences in errand planning: The role of task familiarity and cognitive resources. *Experimental Aging Research, 33*, 145–161. https://doi.org/10.1080/03610730601177395.

Kliegel, M., Rendell, P. G., & Altgassen, M. (2008). The added value of an applied perspective in cognitive gerontology. In S. M. Hofer, & D. F. Alwin (Eds.). *Handbook of cognitive aging: Interdisciplinary perspectives* (pp. 587–602). Thousand Oaks, California: Sage.

Kvavilashvili, L., Cockburn, J., & Kornbrot, D. E. (2013). Prospective memory and ageing paradox with event-based tasks: A study of young, young-old, and old-old participants. *The Quarterly Journal of Experimental Psychology, 66*, 864–875. https://doi.org/10.1080/17470218.2012.721379.

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*, 112. https://doi.org/10.1037/0096-3445.136.1.112.

Lee, S. D., Ong, B., Pike, K. E., & Kinsella, G. J. (2018). Prospective memory and subjective memory decline: A neuropsychological indicator of memory difficulties in community-dwelling older people. *Journal of Clinical and Experimental Neuropsychology, 40*, 183–197. https://doi.org/10.1080/13803395.2017.1326465.

Leitz, J. R., Morgan, C. J., Bisby, J. A., Rendell, P. G., & Curran, H. V. (2009). Global impairment of prospective memory following acute alcohol. *Psychopharmacology, 205*, 379–387. https://doi.org/10.1007/s00213-009-1546-z.

Maylor, E. A. (1990). Age and prospective memory. *The Quarterly Journal of Experimental Psychology Section A, 42*, 471–493. https://doi.org/10.1080/14640749008401233.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology, 14*, S127–S144. https://doi.org/10.1002/acp.775.

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

McDaniel, M. A., & Einstein, G. O. (2008). Prospective memory and aging: Old issues and new questions. In S. M. Hofer, & D. F. Alwin (Eds.). *Handbook of cognitive aging: Interdisciplinary perspectives*. Thousand Oaks, CA: 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.

Mioni, G., Grondin, S., McLennan, S. N., & Stablum, F. (2019). The role of time-monitoring behaviour in time-based prospective memory performance in younger and older adults. *Memory,* 1–15. https://doi.org/10.1080/09658211.2019.1675711.

Mioni, G., Rendell, P. G., Stablum, F., Gamberini, L., & Bisiacchi, P. S. (2015). Test–retest consistency of Virtual Week: A task to investigate prospective memory. *Neuropsychological Rehabilitation, 25*, 419–447. https://doi.org/10.1080/09602011.2014.941295.

Moscovitch, M. (1982). A neuropsychological approach to perception and memory in normal and pathological aging. In F. I. M. Craik, & S. Trehub (Eds.). *Aging and cognitive processes* (pp. 55–78). New York, NY: Plenum Press.

Nelson, H. E. (1982). *National Adult Reading Test (NART): For the assessment of premorbid intelligence in patients with dementia: Test manual: NFER-Nelson*.

Niedźwieńska, A., & Barzykowski, K. (2012). The age prospective memory paradox within the same sample in time-based and event-based tasks. *Aging, Neuropsychology, and Cognition, 19*, 58–83. https://doi.org/10.1080/13825585.2011.628374.

Niedźwieńska, A., Rendell, P. G., Barzykowski, K., & Leszczyńska, A. (2014). Only social feedback reduces age-related prospective memory deficits in "Virtual Week". *International Psychogeriatrics, 26*, 759–767. https://doi.org/10.1017/S1041610214000027.

Phillips, L. H., Henry, J. D., & Martin, M. (2008). In M. Kliegel, M. A. McDaniel, & G. O. Einstein (Eds.). *Adult aging and prospective memory: The importance of ecological validity* (pp. 161–181). New York: Erlbaum.

Rendell, P. G., & Craik, F. I. (2000). Virtual week and actual week: Age-related differences in prospective memory. *Applied Cognitive Psychology, 14*, 43–62. https://doi.org/10.1002/acp.770.

Rendell, P. G., & Henry, J. D. (2009). A review of Virtual Week for prospective memory assessment: Clinical implications. *Brain Impairment, 10*, 14–22. https://doi.org/10.1375/brim.10.1.14.

Rendell, P. G., Jensen, F., & Henry, J. D. (2007). Prospective memory in multiple sclerosis. *Journal of the International Neuropsychological Society, 13*, 410–416 (doi: 10.10170S1355617707070579).

Rendell, P. G., Phillips, L. H., Henry, J. D., Brumby-Rendell, T., de la Piedad Garcia, X., Altgassen, M., & Kliegel, M. (2011). Prospective memory, emotional valence and ageing. *Cognition & Emotion, 25*, 916–925. https://doi.org/10.1080/02699931.2010.508610.

Rendell, P. G., & Thomson, D. M. (1999). Aging and prospective memory: Differences between naturalistic and laboratory tasks. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 54*, 256–269. https://doi.org/10.1093/geronb/54B.4.P256.

Rose, N. S., Rendell, P. G., Hering, A., Kliegel, M., Bidelman, G. M., & Craik, F. I. (2015). Cognitive and neural plasticity in older adults' prospective memory following training with the Virtual Week computer game. *Frontiers in Human Neuroscience, 9*, 1–13. https://doi.org/10.3389/fnhum.2015.00592.

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*, 595. https://doi.org/10.1037/a0019771.

Schnitzspahn, K. M., Ihle, A., Henry, J. D., Rendell, P. G., & Kliegel, M. (2011). The age-prospective memory-paradox: An exploration of possible mechanisms. *International Psychogeriatrics, 23*, 583–592. https://doi.org/10.1017/S1041610210001651.

Schnitzspahn, K. M., Kvavilashvili, L., & Altgassen, M. (2018). Redefining the pattern of age-prospective memory-paradox: New insights on age effects in lab-based, naturalistic, and self-assigned tasks. *Psychological Research Psychologische Forschung*. https://doi.org/10.1007/s00426-018-1140-2.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York, NY: McGraw-Hill.

Terrett, G., Horner, K., White, R., Henry, J. D., Kliegel, M., Labuschagne, I., & Rendell, P. G. (2019). The relationship between episodic future thinking and prospective memory in middle childhood: Mechanisms depend on task type. *Journal of Experimental Child Psychology, 178*, 198–213. https://doi.org/10.1016/j.jecp.2018.10.003.

Terrett, G., Rose, N. S., Henry, J. D., Bailey, P. E., Altgassen, M., Phillips, L. H., ... Rendell, P. G. (2015). The relationship between prospective memory and episodic future thinking in younger and older adulthood. *The Quarterly Journal of Experimental Psychology,* 1–34. https://doi.org/10.1080/17470218.2015.1054294.