The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 2025, **80**(6), gbaf020
https://doi.org/10.1093/geronb/gbaf020
Advance access publication: 11 February 2025
Research Article

Factors Explaining Age-Related Prospective Memory Performance Differences: A Meta-analysis

Marta Menéndez-Granda, MSc,[1,2] Nadine Schmidt, PhD,[1,2] Gianvito Laera, PhD,[3,4] Annick Clenin, MSc,[1] Matthias Kliegel, PhD,[3,5] Michael Orth, MD, PhD,[1] and Jessica Peter, PhD[1,*]

[1]University Hospital of Old Age Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland.
[2]Graduate School for Health Sciences, University of Bern, Bern, Switzerland.
[3]Cognitive Aging Lab (CAL), Faculty of Psychology and Educational Sciences, University of Geneva, Geneva, Switzerland.
[4]University of Applied Sciences and Arts Western Switzerland HES-SO, Geneva School of Health Sciences, Geneva, Switzerland.
[5]Centre for the Interdisciplinary Study of Gerontology and Vulnerability, University of Geneva, Geneva, Switzerland.

*Address correspondence to: Jessica Peter, PhD. E-mail: jessica.peter@unibe.ch

Decision Editor: Gali Weissberger, PhD (Psychological Sciences Section)

Abstract

Objectives: The age-prospective memory paradox states that younger adults perform better than older adults in laboratory tasks, whereas the opposite has been observed for naturalistic tasks. These terms insufficiently characterize tasks and task settings. We therefore revisited the age-prospective memory paradox using a newly developed taxonomy to better understand how task characteristics or task settings contribute to age-related differences in performance.

Methods: We conducted a meta-analysis of 138 studies, classifying prospective memory tasks according to our newly developed taxonomy. The taxonomy included 9 categories that considered how close any task or task setting was to daily life.

Results: When categorizing relevant studies with this taxonomy, we found that older adults did better than younger adults in "close to real-life" tasks done at home and, particularly, in to-do lists and diary tasks. However, they did worse in "far from real-life" tasks done in naturalistic environments or in simulations of real-life tasks in a laboratory.

Discussion: Results of this meta-analysis suggest that the level of abstraction of a task and familiarity with the environment in which the task is taken can explain some of the differences between the performances of younger and older people. This is relevant for the choice of task settings and task properties to experimentally address any prospective memory research questions that are being asked.

Keywords: Age-prospective memory paradox, Ecological validity, Focality, Future intentions, Old age

Prospective memory is the ability to look ahead in time, plan when to do what, and then do it (Einstein & McDaniel, 1990; Kliegel et al., 2002). It involves remembering to do something at a specific time (i.e., time-based), when a specific event occurs (i.e., event-based), or after a certain activity has been finished (i.e., activity-based). Prospective memory is key to living an independent life at all ages (Zuber & Kliegel, 2020). However, the challenges in life may change as people age. More older people for instance may have health issues and must remember to take their medication and keep their health-related appointments. Younger people need to use the same prospective memory abilities but the challenges they need to rise to may differ and be more related to their work environment and family life. Given its relevance for autonomy insight into what influences the different aspects of prospective memory function, and how this varies with age, are key to developing strategies for its preservation and the identification of differential vulnerabilities with aging.

What matters most is prospective memory functioning in tasks of daily living. However, the assessment of prospective memory performance experimentally involves computer-based tasks, tasks that simulate daily living, board games, observations in daily life, and diaries. These tasks can more (or less) closely mimic activities of daily living. Different theoretical frameworks exist to categorize methods with which prospective memory can be assessed. Phillips et al. (2008) proposed five categories ranging from low to high ecological validity. Their categorization considered where the task took place (i.e., task setting), whether the task was artificial or natural, and whether it was novel or familiar. Guynn et al. (2018) distinguished between everyday-life methods (e.g., experience sampling and diaries), naturalistic methods (tasks performed outside a laboratory), laboratory methods (e.g., artificial laboratory tasks), and assessments in clinical populations. Jones et al. (2021) created five categories based on ecological validity from one (self-report/questionnaires) to five (tasks selected from participants' daily activities performed at home). Rummel and Kvavilashvili (2019) proposed a two-dimensional approach that considered both the ecological validity of the ongoing task and the prospective memory task. Using some of these assessments, several studies reported that younger adults do better in prospective memory tasks in a

Received: August 15 2024; Editorial Decision Date: January 7 2025.

laboratory, whereas older adults do better in naturalistic environments. This so-called age-prospective memory paradox has been confirmed in different meta-analyses (Henry et al., 2004; Kliegel et al., 2008; Uttl, 2008), which mainly categorized tasks as either "laboratory" or "naturalistic" (Rummel et al., 2023). It is, however, difficult to categorize paradigms such as virtual reality tasks or online computer tasks into one of these categories as they are far from activities of daily living, but they can be done in naturalistic environments (Laera, Hering, et al., 2023). In addition, the terms "laboratory" and "naturalistic" are somewhat imprecise as they include computer-based tasks done in naturalistic environments but also tasks that closely mimic daily living in a laboratory. To tease apart what exactly leads to worse or better performance of older adults (i.e., the task, the environment, or both) it is important to widen the categorization beyond "laboratory" and "naturalistic." Hence, a taxonomy is needed that considers the distance between a given task and real-life activity both for what is done and where. Therefore, we developed a taxonomy that consists of nine categories. These categories considered characteristics of the ongoing task, characteristics of the prospective memory task, and the context in which the tasks were conducted.

Another factor that influences how much age matters for differences in performance is the prospective memory task itself. In event-based tasks, there is usually an external cue that helps retrieve an intention. However, the amount of attention that is required to detect external cues while performing other tasks (i.e., cue focality) matters. The preparatory attention and memory processes (PAM) model (Smith & Bayen, 2004) argues that a certain degree of attention is *always* required to monitor for cues in the environment (Smith & Bayen, 2004). Age effects then largely depend on declining attention or working memory capacity. In contrast, the multi-process theory (McDaniel & Einstein, 2000) postulates that one factor that influences age differences in event-based performance is whether strategic or automated processes support retrieval (McDaniel & Einstein, 2007). Focal event-based tasks, in which cues are salient and easily detectable, mainly involve spontaneous or automated retrieval - hence older adults perform well. For example, while watching the news, a person may intend to make a phone call when commercials begin (focal cue). The focal cue is part of what they do at that moment (i.e., watching the news), hence, they will remember with little additional effort. If cues are not salient or cannot be processed as part of ongoing activities, strategic monitoring is essential (McDaniel & Einstein, 2000). In the above example, a non-focal cue would be the intention to call someone at dusk. This is not part of what they do at that moment (watching the news) and requires attentional control as it must be processed in addition to the current activity. Since older adults' attentional control often declines, they tend to perform worse than younger adults in non-focal tasks (Ihle et al., 2013; Kliegel et al., 2008; Uttl, 2011). It may be different in activity-based tasks, where an intention is retrieved when an activity has been finished (Kvavilashvili et al., 2009; Yang et al., 2013). Age effects tend to be smaller in such tasks as older adults perform similarly or even better than younger adults (Kvavilashvili et al., 2009; Yang et al., 2013).

For time-based tasks, the setting where the task takes place seems to be important for age differences in performance. In a laboratory, older adults usually perform worse than younger adults, possibly because no external cues remind them of an intention (Henry et al., 2004). They must rely on internal cues, and they are less well able than younger adults to do so. In naturalistic environments, however, older adults often perform better than younger adults, for example, when asked to send text messages to an experimenter at a specific time (Schnitzspahn et al., 2011). The degree of difficulty of time-based tasks may vary in terms of involvement of more or fewer degrees of freedom in using routines, reminders, and links to environmental cues such as time of day events or time of day characteristics (e.g., daylight or darkness). Naturalistic time-based tasks are mostly appointment-like tasks (e.g., sending a text message to an experimenter). External memory aids or environmental cues may facilitate retrieval. For example, when asked to send a text message at 12 p.m., having lunch at this time may serve as a reminder to send the text message. Laboratory time-based tasks are usually time-interval tasks (e.g., press a key every 2 min) and thus provide few or no time-relevant environmental cues (and external memory aids are usually not allowed). Hence, they require effortful cognitive processes (Haines et al., 2020; Henry et al., 2021; Marsh et al., 2006).

Previous studies mostly examined the influence of cue type (and its focality) on tasks in a laboratory. It is important to complement insights gained from laboratory tasks with what happens to task performance outside of a laboratory. This may help to better understand which cues (or what kind of task structure) could be helpful for older adults to improve their prospective memory performance.

The aims of our study were, therefore, to better understand what drives age differences in prospective memory performance and which prospective memory cues help older adults to perform better. To this end, we conducted a meta-analysis of studies that assessed prospective memory in younger and older healthy adults. We hypothesized that younger adults do better than older adults in "far from real-life" tasks in a laboratory. In contrast, we hypothesize that older adults do better in tasks that closely mimic real life, regardless of where they take place. Finally, we expected that certain prospective memory cue types would be helpful for older adults also outside of a laboratory.

Material and methods

Search Strategy

We conducted a meta-analysis following PRISMA guidelines (Moher et al., 2015) and registered the study before data coding (Open Science Framework pre-registration DOI: https://osf.io/ew2f9/?view_only=6dd78ea76bb04f65b9550846a3771495). First, we searched PyscInfo and Pubmed databases for studies from the earliest available date to the 15th of December 2022. We used the following search terms: (a) "Prospective memory" OR "memory for intentions" OR "delayed intention" OR "intentional memory" OR "future memory," (b) "aging" OR "ageing" OR "age," (c) "old" OR "older" OR "elderly," (d) "young" OR 'younger', combined with the AND operator. Second, we screened references from previous meta-analyses (i.e., Henry et al., 2004; Uttl, 2008, 2011) for relevant studies. The search yielded 480 results.

Inclusion and Exclusion Criteria

Figure 1 (PRISMA flowchart) provides an overview of the applied search and screening steps. We included studies if (a) at least one prospective memory task and/or questionnaire

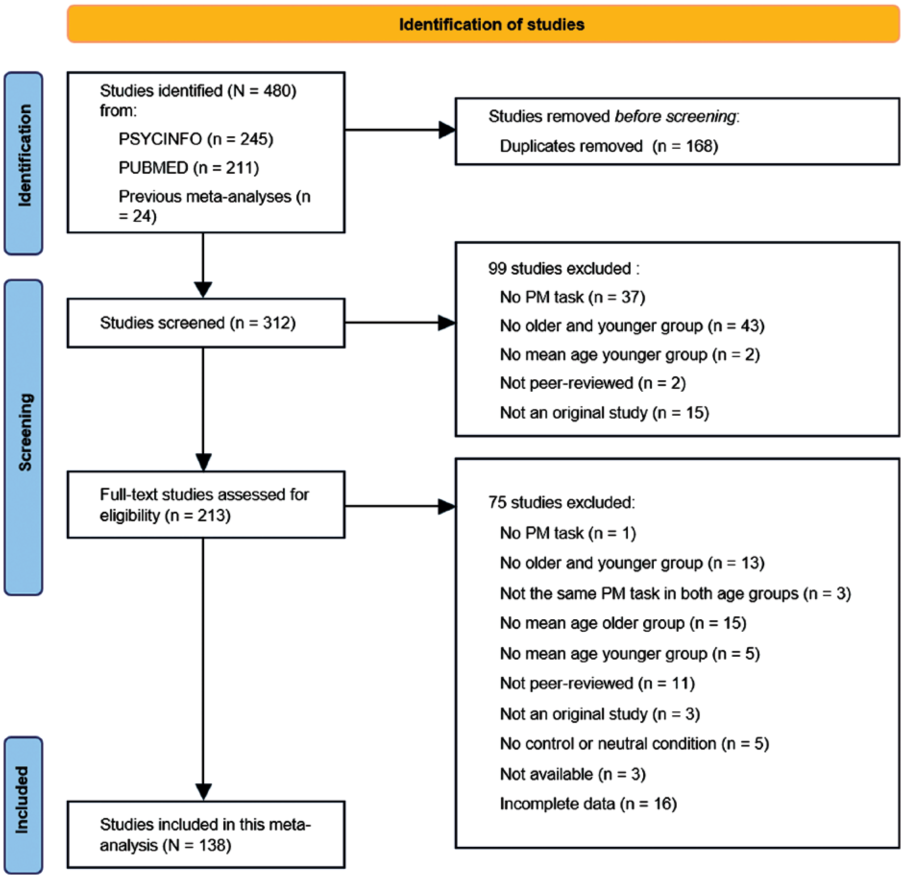

Figure 1. PRISMA flow chart of the current meta-analysis. PM = prospective memory.

was used in younger and older healthy adults, (b) task performance was reported as sum or proportion of correctly retrieved intentions, prospective memory failures (i.e., forgotten intentions), or as scores for self-reports, (c) the same task was used in both age groups, (d) older adults were 60 to 75 years old and younger adults 18–40 years, (e) the study was published in English in a peer-reviewed journal. We excluded studies if they (a) manipulated any experimental factor or condition (e.g., motivation, stimulus valence) without presenting control group data, (b) included only clinical populations, (c) included an intervention, or (d) reported results that were already considered in another study.

Study Selection

We screened abstracts of 312 studies and excluded 99. Then, we screened 213 full texts and excluded another 74 (Figure 1). In several studies, data was not reported but could be obtained by figure digitalization using DigitizeIt (version 2.5). This software has been used in a previous meta-analysis (Laera, Hering, et al., 2023) and values obtained with that software did not differ from real data (Wojtyniak et al., 2020). In other cases, we checked whether data were reported in previous meta-analyses, or we contacted the corresponding author. If nothing of the above was possible, we excluded the study (Figure 1). Screening and study selection were done by NS, MMG, and AC. We calculated inter-rater reliability using kappa statistics (McHugh, 2012) and found substantial to almost perfect agreement (Cohen‘s kappa > 0.61; see Table S1 in Supplementary Material 1; Landis & Koch, 1977).

Taxonomy: Classification of Prospective Memory Tasks

We developed a taxonomy based on previous studies (Guynn et al., 2018; Jones et al., 2021; Rummel & Kvavilashvili, 2019). The taxonomy consisted of nine categories that included abstract tasks "far from real life" (i.e., computerized tasks) or tasks very close to daily life (i.e., to-do lists or observational studies (see Figure 2 and Table 1 for a detailed description). We then categorized all included prospective memory studies based on our taxonomy. Whenever possible, we rated the type and focality of prospective memory cues as focal event-based, non-focal event-based, time-based, or activity-based. If no information was provided to categorize focality of the cue, we rated the task as “event-based unspecified.”

Statistical Analyses

We extracted the task performance of older/younger adults from each study and calculated Hedges’ g (i.e., estimated effect sizes that consider each study’s sample size; see Supplementary Material 1). We then calculated the percentage of studies reporting better performance in younger adults than in older adults (effect size = positive), no difference between younger and older adults’ task performance (effect size = zero), or better performance in older adults (effect size = negative). This gave us a descriptive impression of how many studies report better/worse task performance in older adults. Next, to examine whether age differences were positive, zero, or negative, we used a three-level random effects model, which accounts for interdependencies (Assink &

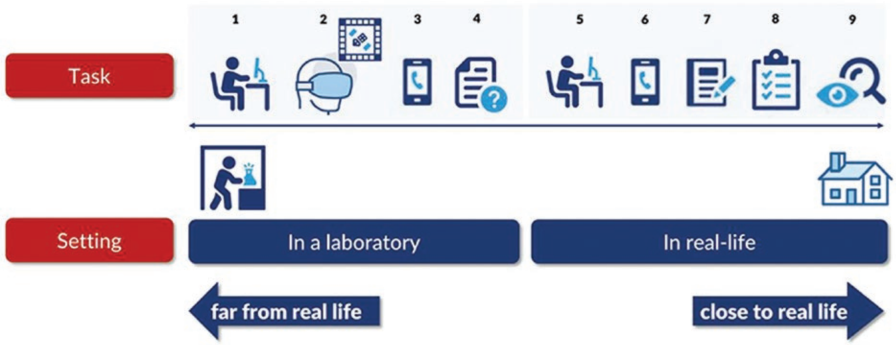

Figure 2. Taxonomy used in this study. Each prospective memory task (and each ongoing task) was classified according to its distance from real life. The taxonomy included computer-based tasks in a laboratory (1), everyday activities (e.g., a phone call) in a laboratory (2), simulations of daily activities or board games in a laboratory (3), questionnaires answered in a laboratory (4), computer-based tasks at home (5), diaries at home (6), everyday activities at home (7), to-do lists at home (8), and observations of daily activities at home (9).

Table 1. Taxonomy used in our meta-analysis.

Task properties	Proximity to daily life			Description
	Task setting	Ongoing task	Intention	
Far from real-life	Laboratory	Low-middle	Low	Participants need to do paper-pencil or computer tasks and, simultaneously, remember intentions (e.g., press a button). Tasks and intentions are determined by an experimenter.
Close to real-life	Laboratory	Low-middle	Middle	Participants need to solve a task or fill-in a questionnaire and to remember intentions that resemble their daily live (e.g., make a phone call). Tasks and intentions are determined by an experimenter.
Simulation	Laboratory	Middle	Middle	Ongoing and prospective memory task simulate real-life tasks (e.g., virtual week task). Tasks and intentions are determined by an experimenter.
Questionnaire	Laboratory	High	High	Participants rate the number of prospective memory failures in their daily live using questionnaires
Far from real-life or simulation	Home/lab	Low-middle	Low-middle	Participants perform “far from real-life” tasks or simulations at home. This includes studies that assessed older adults at home and younger adults in a lab. Tasks and intentions are determined by an experimenter.
Diary	Real-life	High	High	Participants report prospective memory failures retrospectively for each day (usually in the evening).
Close to real-life	Real-life	Middle-high	Middle	Participants need to retrieve intentions that resemble daily live activities (e.g., make a phone call) during their usual daily activities. Intentions are determined by an experimenter.
To-do list	Real-life	High	High	Participant first list intentions for the coming days. Then, they report which of those intentions were carried out. Activities and intentions are determined by participants.
Observation	Real-life	High	High	Prospective memory failures are observed in daily lives. Participant-generated activities and intentions.

Wibbelink, 2016; Cheung, 2014). Level 1 contains the participants tested in studies, who are nested in individual effect sizes (level 2), which in turn, were part of several studies (level 3). Then, we examined whether age differences would depend on how close any task is to real life by including our taxonomy as a moderator in the same random effects model. Finally, we analyzed whether age differences would depend on prospective memory cue type (and its focality) by including this variable as a moderator. In case of a significant main effect, we first evaluated beta-values to examine in which category of our taxonomy older adults perform significantly better (or worse) than younger adults. Second, we conducted pairwise comparisons (Bonferroni corrected) to determine whether age differences were particularly pronounced in a category of our taxonomy. Third, we re-ran the three-level random effects model for each category of our taxonomy separately (only for the moderator cue type). For each model, we de-composed heterogeneity to account for within- and between-study heterogeneity (τ^2; Cheung, 2014). In addition, we used Q-statistic (Cochran, 1954) to evaluate whether heterogeneity was evident or not and I^2 statistic (Higgins & Thompson, 2002) to quantify heterogeneity within- and between studies (Huedo-Medina et al., 2006). If heterogeneity was present (i.e., $\tau^2 > 0$, regardless of the results of the Q-test), we provided prediction intervals for the true outcome (Riley et al., 2011). We performed an outlier analysis to detect other factors that may account for heterogeneity within and between studies (Viechtbauer & Cheung, 2010). If we found outliers, we used sensitivity analysis to examine the contribution of each outlier to heterogeneity and then we tested whether these outliers have common characteristics that may explain heterogeneity. Hence, we performed additional three-level random-effect models for those outliers only. To control for publication bias, we used Egger regression and a funnel plot (Borenstein et al., 2009, Chapter 30).

We used R with Rstudio (version 4.2.1; R Core Team, 2022) and the packages metafor (Viechtbauer, 2010) and meta (Balduzzi et al., 2019). We considered $p < .05$ statistically significant. Data is available within the Open Science Framework (Open Science Framework pre-registration DOI: https://osf.io/ew2f9/?view_only=6dd78ea76bb04f65b9550846a3771495).

Results

We included 138 studies, in which younger adults were on average 22.7 years old (range 18–35) and older adults were on average 69.7 years old (range 61–75).

Younger Adults Perform Better Than Older Adults in Almost 80% of Studies

We included $k = 301$ effect sizes from 138 studies (see Supplementary Material 2 for an overview of the included studies). The effect sizes ranged from −1.52 to 4.83 (Figure 3). 78.7% of them were positive, indicating better task performance in younger adults. 1.7% of them were zero, indicating similar performance of older and younger adults. 19.6% of them were negative, indicating better task performance in older adults. The pooled effect size (i.e., $g = 0.615$; 95% CI: 0.497–0.733) was significantly different from zero ($z = 10.29$, $p < .001$) which indicates better task performance in younger adults (Figure 3). Effect sizes were highly heterogeneous ($Q_{(300)} = 2812.05$, $p < .001$), particularly within-studies ($\tau^2 = 0.363$, $I^2 = 54.65$ %), whereas between-studies heterogeneity was moderate ($\tau^2 = 0.238$, $I^2 = 35.73\%$). The 95% prediction interval for the true outcome (−0.922 to 2.152) included 0. Hence, it cannot be expected in future studies that younger adults perform better than older adults. Egger statistic was significant ($z = 4.347$, $p < .001$), but the distribution of effect sizes in the funnel plot was symmetrical (Figure S1 in Supplementary Material 1). With p-curve analysis (Simonsohn et al., 2014), we found no evidence to suggest publication bias or p-hacking (test for right-skewness $p_{\text{half}} < .001$, $p_{\text{full}} < .001$; test for flatness $p_{\text{half}} > .99$, $p_{\text{full}} > .99$).

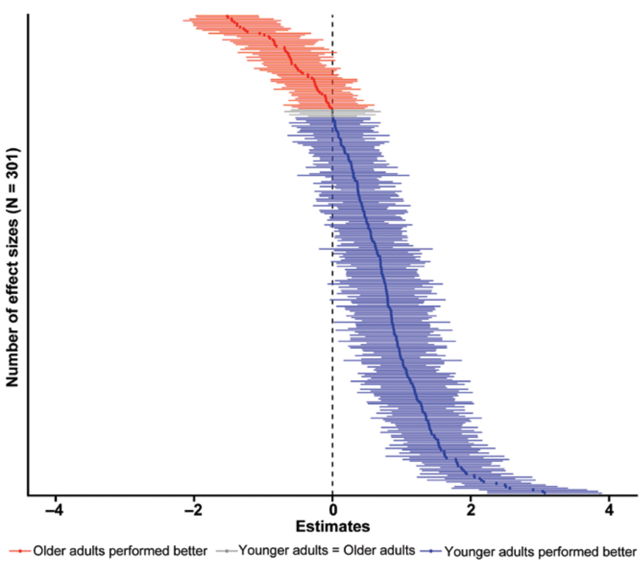

Figure 3. Forest plot of estimated effect sizes (and their confidence intervals) in a meta-analysis of prospective memory task performance in younger and older adults. Notes. Positive values indicate better performance of older adults than younger adults, negative values indicate the opposite.

Our Taxonomy Explains Which Tasks Are Difficult to Solve for Older Adults

Next, we repeated the random effects model and included our taxonomy as a moderator. The moderator reached significance ($Q_{M\ (9)} = 366.48$, $p < .001$) and reduced heterogeneity that was, however, still significant between and within studies ($Q_{(292)} = 1554.07$, $p < .001$; between: $\tau^2 = 0.128$, $I^2 = 32.62\%$; within: $\tau^2 = 0.201$, $I^2 = 51.11\%$). Beta estimates indicated that older adults performed better than younger adults in diary tasks ($\beta = -0.895$, $SE = 0.446$, $z = -2.004$, $p = .045$), to-do lists ($\beta = -0.532$, $SE = 0.270$, z $= -1.970$, $p = .049$), and in "close to real-life" tasks done in naturalistic environments ($\beta = -0.555$, $SE = 0.120$, $z = -4.625$, $p < .001$; Figure 4; Table S3 in Supplementary Material 1). In contrast, they performed worse than younger adults in "far from real-life" tasks ($\beta = 0.712$, SE $= 0.061$, $z = 11.755$, $p < .001$), "close to real-life" tasks ($\beta = 0.700$, $SE = 0.121$, $z = 5.777$, $p < .001$), and simulations ($\beta = 1.056$, $SE = 0.100$, $z = 10.511$, $p < .001$) done in a laboratory as well as in "far from real-life" tasks or simulations done by half of the sample at home and the other half in a laboratory ($\beta = 1.010$, $SE = 0.209$, $z = 4.842$, $p < .001$; Figure 4; Table S3 in Supplementary Material 1). For questionnaires and everyday life observations, no difference in task performance was observed. Pairwise comparisons revealed that age effects were smaller in tasks that were closer to real-life (Table S4 in Supplementary Material 1).

Cue Type (and its Focality) Matter for Age Differences Only in the Laboratory

As a third step, we included cue type (and its focality) as a moderator. Again, the moderator reached significance ($Q_{M\ (5)} = 125.56$, $p < .001$) and reduced heterogeneity, which was still significant between- and within-studies ($Q_{(276)} = 2352.47$, $p < .001$; between: $\tau^2 = .212$, $I^2 = 34.32\%$; within: $\tau^2 = .342$, $I^2 = 55.26\%$). Beta estimates revealed a significant effect for each cue type (Table S5 in Supplementary Material 1), indicating that younger adults perform better than older adults regardless of cue type (all β ~.6, all $p < .05$; Table S5 in Supplementary Material 1). Pairwise comparisons between the different levels of the moderator indicated that age effects did not differ significantly between event-based non-focal, event-based focal, event-based unspecified, time-based, or activity-based prospective memory tasks (Table S6 in Supplementary Material 1).

To further specify which cue type (or focality) may be helpful in each category of our taxonomy, we repeated the moderator analysis for each category (Table S7-S16 in Supplementary Material 1). The results indicate that the closer a task was to real life, the smaller the effect of cue type on age effects. Hence, we found no evidence to suggest that cue type matters for differences in performance in older and younger adults—only for tasks that were done in a laboratory. All results are summarized in Table S3–S16 (in Supplement 1).

Outlier and sensitivity analysis

So far, we found that our taxonomy reduced heterogeneity in our statistical model. However, heterogeneity was still moderate to high. In an exploratory analysis, we therefore examined whether some of this heterogeneity was

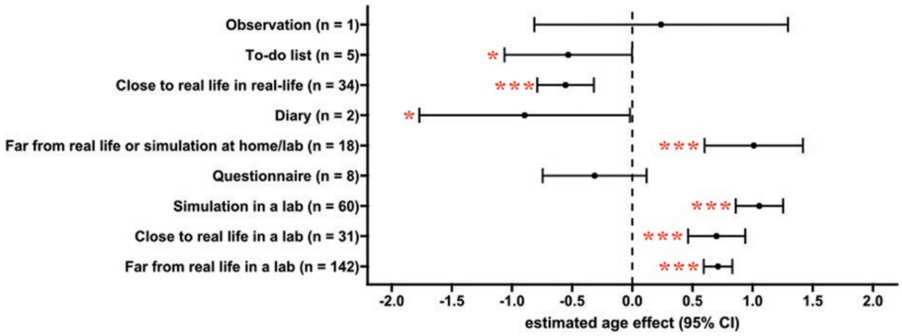

Figure 4. Estimated effect sizes and their confidence intervals when including our taxonomy as a moderator in a three-level random-effects model comparing younger and older adults' prospective memory performance.Positive values indicate better performance in younger adults, negative values indicate better performance in older adults. Asterisks indicate significant post-hoc tests (older adults vs. younger adults). Significant at $p < .05$*, or $p < .001$***.

explained by outliers and whether those outliers have something in common (i.e., whether factors beyond our taxonomy reduce heterogeneity). We found that 125 effect sizes were outliers. When excluding them, between-studies heterogeneity was reduced by 14.35% (from $I^2 = 35.73\%$ to $I^2 = 21.38\%$) and within-study heterogeneity by 45.11% (from $I^2 = 54.65\%$ to $I^2 = 9.54\%$; $Q_{(175)} = 257.565$, $p < .001$). In a sensitivity analysis, we found that three other factors reduced heterogeneity within studies (see Supplementary Material 1): The first factor, "repetition of an intention" (i.e., whether an intention had to be retrieved several times) reduced heterogeneity by 7.4%. The second factor, closeness of the ongoing task to everyday activities, reduced heterogeneity by 4.0%. The third factor, whether participants could decide what they wanted to remember, reduced heterogenity by 2.4%. Between-studies, three factors significantly reduced heterogeneity. Our taxonomy reduced heterogeneity by 25.2%. Other factors were "length of the delay between instruction and retrieval of intentions" (24% reduction) and where the task took place (14% reduction).

Hence, factors beyond our taxonomy influenced how pronounced performance differences were between younger and older adults. These factors included how close to real-life ongoing activities were, whether participants were able to decide what to remember, the delay between instruction and retrieval of an intention, whether external reminders were allowed, and whether intentions were to be retrieved several times.

Discussion

The current study revisited the age–prospective memory paradox which states that younger adults perform better in laboratory tasks, whereas older adults do better in naturalistic tasks as these tasks are usually done in naturalistic environments or are closer to real-life. However, these terms confuse task properties and task setting. Hence, for our meta-analysis, we used a taxonomy that considered whether a task was distant or close to what people do in real life *and* where it was done, that is, in a familiar environment or in a laboratory. Our results indicate that older adults indeed perform better in "close to real-life" tasks done in naturalistic environments and, particularly, in to-do lists and diary tasks. However, they were not better when "far from real-life" tasks were done at participants' homes or in simulations of real-life tasks in a laboratory. We will discuss these findings and their implications below.

Task Setting or Task Properties Alone Do Not Predict Older Adults' Task Performance

In line with our hypothesis and previous studies (Schnitzspahn et al., 2020), "close to real-life" tasks were easier to solve for older adults in naturalistic environments than in a laboratory. In contrast, in "far from real-life" tasks, older adults performed worse than younger adults, regardless of where the task was done (i.e., in naturalistic environments or in a laboratory). Phillips and colleagues concluded in their theoretical framework in 2018 that "the most critical factor in determining the direction of age effects in prospective memory is the task setting". We add to their framework that the analysis of actual data shows that both task setting (i.e., familiarity of the environment) and level of abstraction of the task are important. Hence, older adults need a familiar environment *and* "close to real-life" tasks to perform at their best. However, if true, they should have performed well in tasks that simulate real-life activities and are done at participants' home, but that was not the case. Hence, factors other than task setting and task properties seem to play a role. In task simulations, for example, older adults may be less familiar with the virtual reality devices that are being used as part of the task set-up (Chen et al., 2017). Hence, it may be that the experimental set-up rather than task properties (or task setting) affect performance. In addition, how often an intention needs to be retrieved may be important (see also Laera, Borghese, et al., 2023). Intentions are usually retrieved several times a day as part of a well entrained daily routine (e.g., take medication each day at 9:00 a.m.) while simulations take place in shorter time periods and intentions are only retrieved once or a few times. Hence, a routine seems to be helpful for older adults (see also Blondelle et al., 2016; Rose et al., 2010). This assumption is supported by our outlier analysis which indicates that repetition of an intention matters for age differences in performance. The time lag between when instructions were given and when retrieval was required may also be important—as indicated by our outlier analysis (see also Laera, Borghese, et al., 2023). This is related to the concept of dividing time-based tasks into short-delay tasks (e.g., taking the cake out of the oven in 2 minutes), or longer-delay tasks (e.g., remembering to meet a friend in 5 days; Haines et al., 2020). In simulations, this lag is—at most—1 hour.

In "close to real-life" tasks or in to-do lists, there is usually a longer delay (e.g., from several hours to days), and older adults may have developed strategies to remind them of their intentions (such as agendas, see also Henry et al., 2012; Jones et al., 2021). Reminders are usually not possible (or even not allowed) in simulations. In addition, performance in simulations is time-sensitive, which means that one indicator of performance is the speed with which participants respond. The faster people need to respond the more they must rely on factors like processing speed or attention that are known to decline with age (Nettelbeck & Burns, 2010; von Krause et al., 2022). Processing speed is less relevant in "close to real-life" tasks or in to-do-lists, in which an accurate response depends on the completion of an intention rather than on speed. All these factors do also play a role in "far from real-life" tasks performed in a laboratory, which is why older adults seem to do worse in such tasks than younger adults.

Our findings raise the question of how we should continue to assess prospective memory in a laboratory as it may not resemble prospective memory abilities in everyday life, even if tasks in a laboratory are close to real life. It could be that lab-based tasks are too dependent on processing speed or attention, which declines in older adults. Hence, any deficit in these cognitive domains may have a disproportionate influence on prospective memory performance. It could also be that tasks in naturalistic environments are too easy for older adults; hence subtle cognitive deficits (i.e., that may be detected under pressure in a laboratory) cannot be found in a naturalistic environment. It could also be that subtle deficits can be compensated for in real-life as older adults can use strategies, reminders, or simply because they better prepare for it and take their time. On the other hand, there are several advantages of using "far from real-life" tasks in a laboratory. If one were, for example, interested in the neural substrates of prospective memory, this needs neuroimaging. Another advantage of laboratory-based studies is the systematic and standardized approach, which is difficult to implement in naturalistic environments. Perhaps sometimes a combination of tasks is useful, as well as complementary measures that can influence prospective memory performance, such as processing speed and attention, among others, that may help to discern whether participants have attentional deficits that may cause prospective memory problems in a laboratory task, but not a deficit in prospective memory itself.

An Effect of Prospective Memory Cue Type Can Only Be Found in Abstract, "Far From Real-Life" Tasks in a Laboratory

Based on studies using "far from real-life" tasks in a laboratory, we hypothesized that prospective memory cue types (and their focality) moderate age differences both in a laboratory and in naturalistic environments. We expected that age differences would be most pronounced in time-based tasks, followed by event-based nonfocal tasks, event-based focal tasks, and then activity-based tasks (Jäger & Kliegel, 2008; Yang et al., 2013). Contrary to our expectations, we did not find evidence to suggest that cue types have any effect in tasks other than "far from real-life" tasks in a laboratory. There are several possible explanations. In a laboratory, an ongoing task is defined by the experimenter and hence, whether an event-based prospective cue is focal or nonfocal is very clear and standardized. This is different outside of a laboratory. If an experimenter, for example, asks participants to call them when they see a bus, focality of the cue depends on what the participant is doing at that moment. In other words, here, any ongoing activity is determined by the participant (or the context/setting the person is in at that moment) and not the experimenter and hence, the task is less well standardized. Moreover, the use of external help is usually not allowed in "far from real-life" tasks in a laboratory. In contrast, participants can mostly use external help in "close to real-life" tasks or tasks done in a naturalistic environment. Whether or not external help is allowed changes the type of the prospective memory cue and its focality (see also Haines et al., 2020). An external reminder such as setting an alarm clock, for example, may transform a time-based task into an event-based task. In other words, cue type and focality may not be as distinct in "close to real-life" tasks or tasks performed in a naturalistic environment, because experimenters have limited control over the task. This is supported by our sensitivity analysis that shows that the use of external help is one factor that influences how much age matters for performance differences in prospective memory.

Limitations and Future Directions

This meta-analysis has several limitations. First, we deliberately included younger adults (20–40 years old) and older adults (60–75 years old) to reveal differences between, and commonalities in, two homogenous age groups. This, however, limits the generalizability of our results to all other age groups. Second, ratings of task category, cue type (and focality) depend on the individual rater. Although there was sufficient agreement between raters, determining cue type and focality was sometimes difficult as a detailed description of task instructions was not provided. For example, authors of a study may have simply noted that participants needed to remember an intention at the end of a testing session. If they were asked to recall the intention when the experimenter said goodbye, this would be referred to as an event-based cue. If they were instead asked to remember the intention when opening a laboratory door, this would be referred to as an activity-based cue. Third, we did not further subdivide time-based tasks. Ellis (1996), for example, differentiated between step-intentions (carried out within a time-window, e.g., next week) and pulse-intentions (carried out at a specific time), whereas Haines and colleagues (2020) distinguished between time-of-day tasks (carried out at a specific time) and time-interval tasks (carried out after a time interval has elapsed). Finally, we included only cross-sectional studies. Hence, we cannot say anything about the dynamics of how prospective memory abilities change with age in an individual. This requires longitudinal studies, which would be an important addition to prospective memory research.

Conclusion

Our study revisiting the age-prospective memory paradox revealed several important factors that influence performance in prospective memory tasks. The level of abstraction of a task, from very close to real life to far from it, familiarity with the environment in which the task is taken, and what constitutes task performance, can all explain some of the differences between performance of younger and older people. It may be less important to know whether an age-prospective memory paradox exists than being mindful of choosing task setting and task properties to experimentally address the question that is being asked and control for instance for processing speed or attention. Performance assessments relevant to day-to-day functioning require tasks close to a naturalistic environment. Wanting to gain insight into mechanisms underlying prospective memory performance requires laboratory experiments. There does not seem to be a one-size-fits-all for experimental set-up or task properties. Perhaps sometimes a combination of tasks is useful, and the results of this meta-analysis can provide some guidance for what to choose.

Supplementary Material

Supplementary data are available at *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* online.

Funding

This work was funded by the Swiss National Science Foundation (185105, 218252 to J. Peter).

Conflict of Interest

There are no conflicts of interest.

Data Availability

Data is available within the Open Science Framework (Open Science Framework pre-registration DOI: https://osf.io/ew2f9/?view_only=6dd78ea76bb04f65b9550846a3771495).

Author Contributions

N. S., M. M.-G., G. L., J. P., and M. K. contributed to the study conception and design. Material preparation, data collection, and analysis were done by N. S., M. M., and A. C. Data interpretation was done by all authors. M. M.-G.and N. S. drafted the first version of the manuscript. All authors contributed to revising it critically for important intellectual content. All authors approved its final version.

References

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology*, 12(3), 154–174. https://doi.org/10.20982/tqmp.12.3.p154

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-based Mental Health*, 22(4), 153–160. https://doi.org/10.1136/ebmental-2019-300117

Blondelle, G., Hainselin, M., Gounden, Y., Heurley, L., Voisin, H., Megalakaki, O., Bressous, E., & Quaglino, V. (2016). Regularity effect in prospective memory during aging. *Socioaffective Neuroscience & Psychology*, 6, 31238. https://doi.org/10.3402/snp.v6.31238

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386

Chen, Y., Lian, R., Yang, L., Liu, J., & Meng, Y. (2017). Working memory load and reminder effect on event-based prospective memory of high- and low-achieving students in math. *Journal of Learning Disabilities*, 50(5), 602–608. https://doi.org/10.1177/0022219416668322

Cheung, M. W. -L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19(2), 211–229. https://doi.org/10.1037/a0032968

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10(1), 101. https://doi.org/10.2307/3001666

Ellis, J. (1996). Prospective memory or the realization of delayed intentions: A conceptual framework for research. *Prospective Memory: Theory and Applications*, 1–22.

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

Guynn, M. J., Einstein, G. O., & McDaniel, M. A. (2018). Methods of studying prospective memory. In H. Otani & B. L. Schwartz (Eds.), *Handbook of research methods in human memory* (1st ed., pp. 284–312). Routledge. https://doi.org/10.4324/9780429439957-17

Haines, S. J., Randall, S. E., Terrett, G., Busija, L., Tatangelo, G., McLennan, S. N., Rose, N. S., Kliegel, M., Henry, J. D., & Rendell, P. G. (2020). Differences in time-based task characteristics help to explain the age-prospective memory paradox. *Cognition*, 202, 104305. https://doi.org/10.1016/j.cognition.2020.104305

Henry, J. D., Hering, A., Haines, S., Grainger, S. A., Koleits, N., McLennan, S., Pelly, R., Doyle, C., Rose, N. S., Kliegel, M., & Rendell, P. G. (2021). Acting with the future in mind: Testing competing prospective memory interventions. *Psychology and Aging*, 36(4), 491–503. https://doi.org/10.1037/pag0000593

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(1), 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. *Quarterly Journal of Experimental Psychology (2006)*, 65(7), 1274–1287. https://doi.org/10.1080/17470218.2011.651091

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, 21(11), 1539–1558. https://doi.org/10.1002/sim.1186

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychological Methods*, 11(2), 193–206. https://doi.org/10.1037/1082-989X.11.2.193

Ihle, A., Hering, A., Mahy, C. E. V., Bisiacchi, P. S., & Kliegel, M. (2013). Adult age differences, response management, and cue focality in event-based prospective memory: A meta-analysis on the role of task order specificity. *Psychology and Aging*, 28(3), 714–720. https://doi.org/10.1037/a0033653

Jäger, T., & Kliegel, M. (2008). Time-based and event-based prospective memory across adulthood: Underlying mechanisms and differential costs on the ongoing task. *Journal of General Psychology*, 135(1), 4–22. https://doi.org/10.3200/GENP.135.1.4-22

Jones, W. E., Benge, J. F., & Scullin, M. K. (2021). Preserving prospective memory in daily life: A systematic review and meta-analysis of mnemonic strategy, cognitive training, external memory aid, and combination interventions. *Neuropsychology*, 35(1), 123–140. https://doi.org/10.1037/neu0000704

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

Kliegel, M., Martin, M., Mcdaniel, M., & Einstein, G. (2002). Complex prospective memory and executive control of working memory: A process model. *Psychology Science*, 44, 303–318.

Kvavilashvili, L., Kornbrot, D. E., Mash, V., Cockburn, J., & Milne, A. (2009). Differential effects of age on prospective and retrospective memory tasks in young, young-old, and old-old adults. *Memory (Hove, England)*, 17(2), 180–196. https://doi.org/10.1080/09658210802194366

Laera, G., Borghese, F., Hering, A., Kliegel, M., & Mioni, G. (2023). Aging and time-based prospective memory in the laboratory: A meta-analysis on age-related differences and possible explanatory factors. *Memory*, 31(5), 747–766. https://doi.org/10.1080/09658211.2023.2191901

Laera, G., Hering, A., & Kliegel, M. (2023). Assessing time-based prospective memory online: A comparison study between laboratory-based and web-based testing. *Quarterly Journal of Experimental Psychology*, 77, 2214–2227. https://doi.org/10.1177/17470218231220578

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–74. https://doi.org/10.2307/2529310

Marsh, R. L., Hicks, J. L., & Cook, G. I. (2006). Task interference from prospective memories covaries with contextual associations of fulfilling them. *Memory & Cognition*, 34(5), 1037–1045. https://doi.org/10.3758/bf03193250

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology*, 14(7), 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*. Sage Publications, Inc.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22, 276–282. https://doi.org/10.1016/j.jocd.2012.03.005

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A.; PRISMA-P Group (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1. https://doi.org/10.1186/2046-4053-4-1

Nettelbeck, T., & Burns, N. R. (2010). Processing speed, working memory and reasoning ability from childhood to old age. *Personality and Individual Differences*, 48(4), 379–384. https://doi.org/10.1016/j.paid.2009.10.032

Phillips, L. H., Henry, J. D., & Martin, M. (2008). Adult aging and prospective memory: The importance of ecological validity. In M. Kliegel, M. A. McDaniel, & G. O. Einstein (Eds.), *Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives* (pp. 161–185). Taylor & Francis Group/Lawrence Erlbaum Associates.

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *BMJ*, 342(feb10 2), d549–d549. https://doi.org/10.1136/bmj.d549

Rose, N. S., Rendell, P. G., McDaniel, M. A., Aberle, I., & Kliegel, M. (2010). Age and individual differences in prospective memory during a "Virtual Week": The roles of working memory, vigilance, task regularity, and cue focality. *Psychology and Aging*, 25(3), 595–605. https://doi.org/10.1037/a0019771

Rummel, J., & Kvavilashvili, L. (2019). Take the field! Investigating prospective memory in naturalistic and real-life settings. In J. Rummel (Ed.), *Current issues in memory: Memory research in the public interest* (pp. 179–191). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9781003106715-10

Rummel, J., Snijder, J. -P., & Kvavilashvili, L. (2023). Prospective memories in the wild: Predicting memory for intentions in natural environments. *Memory & Cognition*, 51(5), 1061–1075. https://doi.org/10.3758/s13421-022-01379-y

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(4), 583–592. https://doi.org/10.1017/S1041610210001651

Schnitzspahn, K. M., Kvavilashvili, L., & Altgassen, M. (2020). Redefining the pattern of age-prospective memory-paradox: New insights on age effects in lab-based, naturalistic, and self-assigned tasks. *Psychological Research*, 84(5), 1370–1386. https://doi.org/10.1007/s00426-018-1140-2

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *p* -curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9(6), 666–681. https://doi.org/10.1177/1745691614553988

Smith, R. E., & Bayen, U. J. (2004). A multinomial model of event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4), 756–777. https://doi.org/10.1037/0278-7393.30.4.756

Uttl, B. (2008). Transparent meta-analysis of prospective memory and aging. *PLoS One*, 3(2), e1568. https://doi.org/10.1371/journal.pone.0001568

Uttl, B. (2011). Transparent meta-analysis: Does aging spare prospective memory with focal vs. non-focal cues? *PLoS One*, 6(2), e16618. https://doi.org/10.1371/journal.pone.0016618

Viechtbauer, W. (2010). Conducting meta-analyses in *R* with the **metafor** package. *Journal of Statistical Software*, 36(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Viechtbauer, W., & Cheung, M. W. -L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. https://doi.org/10.1002/jrsm.11

von Krause, M., Radev, S. T., & Voss, A. (2022). Mental speed is high until age 60 as revealed by analysis of over a million participants. *Nature Human Behaviour*, 6(5), 700–708. https://doi.org/10.1038/s41562-021-01282-7

Wojtyniak, J., Britz, H., Selzer, D., Schwab, M., & Lehr, T. (2020). Data digitizing: accurate and precise data extraction for quantitative systems pharmacology and physiologically-based pharmacokinetic modeling. *CPT: Pharmacometrics & Systems Pharmacology*, 9(6), 322–331. https://doi.org/10.1002/psp4.12511

Yang, T. -X., Wang, Y., Lin, H., Zheng, L. -N., & Chan, R. C. K. (2013). Impact of the aging process on event-, time-, and activity-based prospective memory. *PsyCh Journal*, 2(1), 63–73. https://doi.org/10.1002/pchj.19

Zuber, S., & Kliegel, M. (2020). Prospective memory development across the lifespan. *European Psychologist*, 25(3), 162–173. https://doi.org/10.1027/1016-9040/a000380