Prospective Memory Predictions in Aging: Increased Overconfidence in Older Adults

Stéphanie Cauvin

Faculté de Psychologie et des Sciences de l'Education, University of Geneva

Christopher J.A. Moulin

Laboratoire de Psychologie et Neurocognition, Université de Grenoble

Céline Souchay

Laboratoire de Psychologie et Neurocognition, Université de Grenoble

Matthias Kliegel

Faculté de Psychologie et des Sciences de l'Education, University of Geneva

Katharina M. Schnitzspahn

School of Psychology, University of Aberdeen

Author Notes:

We thank Charlotte Caparaos, Sophie Consigny, Clara Delaissiaz and Pauline Lopez for assistance with data collection. Preparation of this manuscript was funded by a joint grant from the Swiss National Science Foundation (SNSF), the Agence Nationale de Recherche (ANR; France) to MK, CM, KS and CS. CM gratefully acknowledges the support of the Institut Universitaire de France.

Correspondence concerning this article should be addressed to Stéphanie Cauvin, University of Geneva, Boulevard du Pont d'Arve 40, CH-1211 Genève. E-Mail: Stephanie.Cauvin@unige.ch

Abstract

This study investigated whether young and older adults can predict their future performance on an event-based prospective memory (PM) task. Metacognitive awareness was assessed by asking participants to give judgments-of-learning (JOLs) on an item-level for the prospective (remembering *that* something has to be done) and retrospective (remembering *what* to do) PM component. In addition, to explore possible age differences in the ability to adapt predictions to the difficulty of the task, encoding time and the relatedness between the prospective and the retrospective PM component were varied. Results revealed that both age groups were sensitive to our task manipulations and adapted their predictions appropriately. Moreover, item-level JOLs indicated that for the retrospective component, young and older adults were equally accurate and slightly overconfident. For the prospective component, predictions were fairly accurate in young adults, while older adults were overconfident. Thus, results suggest that general overconfidence is increased in older adults and concerns both components of PM. Findings regarding the conceptual differences between the prospective and retrospective components of a PM task, as well as the link between aging and metamemory in PM are discussed.

Keywords: prospective memory, metacognition, judgment-of-learning, aging

Introduction

Successful prospective memory (PM) is defined as the formation and delayed execution of planned intentions at the appropriate moment (Brandimonte, Einstein & McDaniel, 1996; Kliegel, McDaniel & Einstein, 2008). Examples of PM tasks in our everyday life are remembering to pass a friend a message when seeing him, or remembering to stop by the grocery store when passing it on the way home.

A characteristic of experimental PM tasks is that they are embedded in a separate activity, the so-called ongoing task (Einstein & McDaniel, 1990). Participants are occupied with the ongoing task and they have to interrupt it in order to perform the PM task. This reflects real-life PM situations, in which the intended action must be carried out while one is engaged in another activity (McDaniel & Einstein, 2000). In the daily example above, walking home after work thinking about the plans for the upcoming evening would be considered as the ongoing task and seeing the store would act as a cue to stop by. Typical laboratory paradigms mimic naturally occurring intentions and their context by asking participants to perform a computer-based task as an ongoing activity and to remember to press a particular key when a predefined target cue occurs.

Conceptually, PM has been described as a multi-phasic process that comprises two components: the prospective component and the retrospective component (Einstein & McDaniel, 1996). The first component refers to the PM cue detection and remembering *that* something has to be done. This component is thought to be supported by controlled attention and executive functions; such as switching one's attention between the ongoing and the PM task to monitor for the cue and inhibition of the ongoing task once this cue is detected (Schnitzspahn, Stahl, Zeintl, Kaller, & Kliegel, 2013). The second component describes the retrieval of the intention content from long-term memory and remembering *what* to do. This component is hypothesized to be related to episodic memory processes and represents the link to other traditionally studied memory processes (Kliegel, Altgassen, Hering, & Rose, 2011; McDaniel et al., 1999).

In terms of everyday relevance, PM tasks are ubiquitous in our private but also professional lives such as remembering friends' birthday, paying bills or attending scheduled meetings. They are of critical importance for efficiency and our wellbeing (Hering, Kliegel, Rendell, Craik, & Rose, 2018; Woods, Weinborn, Velnoweth, Rooney, & Bucks, 2012). This is even more so for older adults, because their PM tasks are often health-related: remembering to take medication or attending doctor's appointments (McDaniel & Einstein, 2008). Memory failures in this case can be harmful and threaten independent living. Thus, the effects of age on PM functioning have received increased attention in recent years (e.g., Kliegel et al., 2016).

Research into PM development in old age suggests a general decline in lab-based PM functioning (see Henry, MacLeod, Phillips, & Crawford, 2004; Ihle, Hering, Mahy, Bisiacchi, & Kliegel, 2013; Kliegel, Jäger, & Phillips, 2008 for meta-analyses). One potential explanation for this age effect are age-related changes in underlying cognitive processes: For example, working memory seems to partly explain age deficits in PM (Rose, Rendell, McDaniel, Aberle, & Kliegel, 2010). Further, there is evidence indicating larger age differences in PM when the PM tasks require strategic monitoring to detect the cue (Kliegel et al., 2008). Moreover, Schnitzspahn et al. (2013) investigated facets of executive control and found that shifting and inhibition predicted PM performance and also explained age differences in a sample of young and older adults.

Importantly, recent research suggests that the general cognitive decline in working memory or executive control does not inevitably lead to PM impairments in aging, but that older adults can perform as well as young adults if they direct their (possibly limited) attentional resources towards the PM task (Cherry et al., 2001; Ihle, Schnitzspahn, Rendell, Luong, & Kliegel, 2012). Specifically, a recent study manipulated the perceived importance of the PM task by instructing young and older adults to prioritize either the ongoing task or the PM task (Hering, Phillips, & Kliegel, 2013). When the importance of the ongoing task was emphasized, PM performance showed the expected age-related decline. By contrast, when the importance of the PM task was emphasized no age effect emerged. This finding suggests that older adults seem less likely to spontaneously allocate their attentional resources to the PM task, but a heightened PM task importance can improve older adults' performance.

What might influence resource allocation in PM tasks? It is the core assumption of the present paper that age effects in PM may be due to age differences in the beliefs about and evaluation of the PM task, such as its importance or its difficulty, encapsulated in the term *metamemory* (Dunlosky & Tauber, 2015). As an everyday example, one could decide to write down what to buy on a shopping list for a future trip to the grocery store instead of relying on one's memory alone, or to use a planner or reminders on a smartphone. In a lab-based task, one could decide to put more effort into a cognitive task based on the impression that the task is difficult. Accordingly, age differences in metamemory may significantly contribute to the PM decline in older adults. Surprisingly, the role of metamemory for PM age differences has not been addressed yet. It was therefore the goal of the present study to examine if young and older adults differ in their metacognitive awareness of their PM performance in the laboratory and if predictions are related to actual performance.

Metamemory has received substantial attention in the retrospective memory literature (for a review see Castel, Middlebrooks, & McGillivray, 2016). However, there is not a uniform picture of metacognitive decline in healthy aging, but a complex pattern of preservation and impairment according to task, methodology and memory domain. For instance, there seems to be a deficit in evaluations of working memory (Touron, Oransky, Meier, & Hines, 2009), with older adults being less accurate at predicting their performance, but in turn benefit more from the use of effective strategies. It is also mostly found that there are deficits in predicting the status of unrecalled material in episodic memory (e.g., Souchay, Moulin, Clarys, Taconnat, & Isingrini, 2007; but see Eakin, Hertzog, & Harris, 2014), whereas there is preservation of predictive accuracy in the very same judgements but for semantic materials (e.g., Morson, Moulin, & Souchay, 2015). The form of metamemory which bears closest resemblance to our interest in PM is the judgement of learning (JOL; Nelson & Narens, 1990) procedure, which we now review in more detail.

Experimentally, metamemory at encoding is usually studied by asking participants to predict their future performance with JOLs made after the encoding phase. JOLs vary according to the type of prediction: global (for the whole list, i.e. “How many items will you remember?”) or item-by-item (“How likely are you to recall this item later?”). They also vary according to when the predictions are made – either immediately after the encoding, or after a short delay.

In a seminal paper, Connor, Dunlosky, and Hertzog (1997) used JOLs in young and older adults to examine metamemory processes at encoding. Participants studied word pairs and either gave immediate or delayed item JOLs for each pair, and a global JOL before the test phase started. The authors found that older adults were overconfident with global JOLs, thus giving the same predictions as the young adults although they had a lower performance. However, with item-by-item JOLs, both age groups had greater accuracy with delayed JOLs compared to immediate JOLs. Moreover, older adults were as accurate as young adults in predicting their performance. More recent studies confirmed that monitoring at encoding evaluated with JOLs is not impaired with age, although older adults have lower memory performance than young adults (Hertzog, Sinclair, & Dunlosky, 2010; Souchay, Isingrini, Clarys, & Taconnat, 2004; Souchay & Isingrini, 2012).

So far, only one early study (Devolder, Brigham, & Pressley, 1990) explored possible age differences in the metacognitive awareness of PM performance. Devolder et al. used multiple retrospective memory lab-based tasks and one naturalistic PM task with global JOLs made before and after each task (pre- and post-dictions). Participants indicated how well they thought they would perform before starting the test phase. After the test, they estimated how well they had done. Results did not reveal a clear pattern of age differences in prediction accuracy. Rather, there were differences according to the task: older adults were more inaccurate than young adults on three out of eight retrospective memory tasks, but they were more accurate in the naturalistic PM task than young adults. In the remaining retrospective memory tasks, age differences in accuracies were not significant. Moreover, there were no age differences for post-dictions. Also, while young adults outperformed the older adults on all retrospective memory tasks, older adults performed better on the PM task than young adults.

Even considering the overall PM literature, only very few studies have examined the role of metacognitive processes in PM in general. In one of the first studies, Meeks, Hicks and Marsh (2007) asked young participants to respond to particular PM cues embedded in an ongoing lexical-decision-task (LDT); either animal words in one condition or the syllable "tor" in another condition. After receiving the instructions, participants had to predict their future performance with a global JOL by estimating the percentage of the cues they expected to find. Results showed that participants were far from accurate and mostly demonstrated underconfidence in their PM. Moreover, predictions correlated positively with overall PM performance, but only in the "animal" condition. The authors concluded that people have some metacognitive awareness of their future PM performance, at least for some intentions. However, when participants made their predictions they did not know how many cues would appear in the task but only the general cue category (i.e. animal vs syllable "tor"), which could have distorted participant's predictions. And they only provided one global JOL, which limits interpretation of the results.

Further studies have assessed performance predictions in PM (Meier, von Wartburg, Matter, Rothen, & Reber, 2011; Rummel, Kuhlmann, & Touron, 2013). The main goal of these studies was to explore if predictions affect PM performance. Meier et al. (2011) found that making predictions prior to the task improved performance. Rummel et al. (2013) showed that making predictions on the PM task only, compared to additional predictions on the ongoing task, indirectly improved performance: allocation of attention was directed from the ongoing to the PM task and towards cue detection. Moreover, participants made different predictions according to the focal vs. non-focal characteristics of the task (focal PM tasks are those in which the ongoing task involves processing the defining features of the PM cue, while non-focal PM tasks are those in which the PM cue is not part of the information being extracted in the service of the ongoing task, Einstein & McDaniel, 2005). In the non-focal PM task, participants overestimated their performance, while in the focal task participants underestimated future performance.

To sum up, for global judgements, people appear inconsistent in their PM predictions. It seems that predicting PM performance is rather difficult and depends on the properties of the task. Besides, the relation between prediction and performance is not yet clear, and again, in these studies, participants only responded to global predictions. Thus, the present paper suggests a more fine-grained and process-based approach to study the role of metacognition for age differences in PM; using item-level predictions and disentangling different sub-components of PM.

The advantage of item-level predictions is that they allow a distinction of two aspects in predictions at encoding: absolute and relative accuracy. Absolute accuracy, also known as the *calibration* of the predictions to performance, is obtained by looking into the average JOLs and comparing it with the actual performance. It allows us to determine if participants are over- or underconfident. Relative accuracy, also known as *resolution*, is calculated by the Goodman-Kruskal correlation (referred thereafter as gamma correlation; Nelson, 1984). It indicates the degree to which an individual's JOLs correctly discriminate between recalled and non-recalled items.

Schnitzspahn, Zeintl, Jäger and Kliegel (2011) used such item predictions to investigate metacognitive factors in a PM task. More specifically, they wanted to explore two effects typically found in retrospective memory: the *delayed-JOL effect* and the *underconfidence-with-practice effect*. They asked participants to memorize cue-action pairs (i.e., cappuccino – clean glasses). For half of the pairs, participants had to give immediate JOLs and for the other half delayed JOLs. Further, half of the participants had a single encoding trial, while the other half had three encoding trials. For each pair, the participant first predicted performance for the *prospective* component as the probability to become aware *that something* has to be done upon encountering the PM target cue. Second, the participant predicted performance for the *retrospective* component and the probability to remember the specific *action* (e.g., clean glasses) related to the cue (e.g., cappuccino). The participants read a text on the computer as the ongoing task that contained the predefined words serving as PM target cues. PM responses could be made by clicking on the target cues and subsequently entering the associated action. In line with the delayed-JOL effect in retrospective memory, medium to high positive correlations between delayed but not immediate JOLs and performance for both the prospective and retrospective components were observed. The study also replicated the underconfidence-with-practice effect for the retrospective component; with participants being overconfident in the one trial condition, but showing underconfidence in the three trials condition. The prospective component, however, showed a different pattern, with general underconfidence regardless of the number of encoding trials, which is in line with previous studies (Knight, Harnett, & Titov, 2005; Meeks et al., 2007). This suggests that participants differentially assess the prospective and retrospective components of the task, but the difference between the two is not clear yet as regard to the underlying metacognitive processes.

The overall goal of the present study was to extend the existing research on PM and metacognition into healthy aging to test if this potentially important factor helps explaining age effects in PM performance. Item-by-item JOLs were used to obtain measures of calibration and resolution. In line with metacognition research in retrospective memory (Rhodes, 2016), performance and predictions were not only measured for one single task, but instead two experimental manipulations were used to vary task difficulty and former experience. This was done to examine whether older adults were able to adapt their PM performance predictions according to the situation (i.e. a sensitivity effect, e.g. Moulin, 2002).

Task difficulty was manipulated by adapting a paradigm commonly used in retrospective memory studies (e.g. Hertzog et al., 2010) in which participants are asked to study related (e.g., King-Crown) and unrelated word pairs (e.g., Turtle-Bean) and to give JOLs before the recall test phase. Similarly, the word association between the PM cues and the related words to be remembered in the PM response were manipulated in the present study to have one task condition with cue-action pairs that were highly related and a second task condition with low relatedness between cue and action.

Former task experience was manipulated by asking participants to perform each PM task (high vs low cue-action relatedness) twice. If people have insight into their PM performance, they should realize that a task was more or less difficult and accordingly adapt their predictions after performing the specific task for the first time.

In terms of hypotheses, first regarding PM performance, since our task required encoding of a high number of PM cues and constant monitoring during the ongoing task to detect the cues, we expected that PM would be highly prone to failure, especially for the older adults (Henry et al., 2004). Secondly, regarding PM predictions, based on previous research in retrospective memory (for a review, see Hertzog & Dunlosky, 2011) and the overlap of PM with retrospective memory processes (Zimmerman & Meier, 2006), we predicted on the one hand, poorer calibration in older adults compared to young adults; on the other hand, we predicted similar resolution in both age groups. Thirdly, concerning sensitivity effects, research on retrospective memory suggests that young and older adults are sensitive to a word relatedness manipulation at encoding and correctly predicted better performance for the related than unrelated pairs (Hertzog et al., 2010). Accordingly, we expected both age groups to be somewhat sensitive to word relatedness between PM cue and related word in the present study: related word pairs will lead to overall higher predictions than unrelated word pairs. Similarly, we predicted that young and older adults should adjust their judgments based on their previous task experience, with higher predictions and performance in the second block, especially in the easier task condition using highly related cue-word pairs. The finding that older adults seem less likely to spontaneously allocate their attentional resources to the PM task (Hering et al., 2013) suggests that they might be less aware of the difficulty of a PM task and the required focus. Thus, in addition to the described main effects, interactions between age and the manipulated factors seem plausible. Older adults may be less sensitive towards task difficulty and less able to adapt their behavior resulting in less accurate predictions and larger PM age effects in the more difficult task conditions. Finally, to further extend the current literature suggesting different patterns for the predictions of the prospective and retrospective components of the PM task (Schnitzspahn et al., 2011), we expected both young and older adults to be more accurate in predicting the retrospective than the prospective component, with no age differences. In contrast, older adults should be less accurate in predicting the prospective component.

Method

Participants and Design

Overall, 68 young (*M* = 24.53 years, *SD* = 3.6, age range = 19-35 years; 36 female) and 55 older adults (*M* = 69.83 years, *SD* = 8.45, age range = 60-89 years; 30 female) took part in the study. Young adults participated in exchange for course credits and older adults participated voluntarily. Older adults completed the Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975) in which they all scored above the cut-off score of 27 (*M* = 29.17, *SD* = .95) as well as a vocabulary test (Mill Hill; Raven, Court, & Raven, 1986) as a measure of verbal ability (*M* = 26.53; *SD* = 3.41). Young adults were all students at the University of Dijon, while older adults were community dwelling volunteers. All participants were native French speakers. Informed written consent was obtained from all participants prior to taking part in the experiment. The study conformed to local ethical procedures, the Helsinki Declaration, and Convention of the Council of Europe on Human Rights and Biomedicine.

The study followed a 2 X 2 X 2 mixed factorial design to investigate the effects of age, cue-word associativity and block. The between-subjects variable was *age* (young, older), and the within-subjects variables were *associativity* (high, low) and *block* (first versus second time participants studied the items).

Materials

Ongoing task. The paradigm conformed to the traditional Einstein and McDaniel (1990) laboratory assessment of PM, with the PM task being embedded in an ongoing task. Following previous studies (e.g., Cohen, Jaudas, & Gollwitzer, 2008), the ongoing task consisted of a lexical-decision task (LDT). 480 words from a normed set (Bonin et al., 2003) were used as verbal materials. The set comprises 866 French words in total and provides ratings on emotional valence and subjective frequency. On the basis of those ratings, we selected neutral words of medium frequency and similar length between one and two syllables. Non-words were constructed by randomly interchanging the letters of each word stimulus. This resulted in an equal number of valid French words (e.g., paume) and pronounceable non-words (e.g., maupe). At the beginning of each LDT trial, a fixation cross of variable duration (250-750 ms) was presented in the center of a black screen followed by a stimulus (i.e., either a word or non-word), which stayed on the screen until the participant responded. Instructions were given to press a "yes" key (i.e., j) with the right index finger if the presented letter string was a valid French word, otherwise a "no"' key (i.e., f) had to be pressed with the left index finger.

PM task. For the PM task, the participants were further asked to remember to carry out a specific action when they identified one of the pre-specified PM cues. These PM cues consisted of 20 words that the participants had to learn prior to the LDT. During encoding, each PM cue word was shown as the first word of a pair. Participants were instructed to remember those word pairs: the cue words and the related words. When they later detected one of the cue words during the LDT, they were instructed to press the "white key" (i.e. a key marked in white) instead of giving their ongoing task response. If they did so, a text box appeared and they were asked to type in the second word of the respective word pair by using the computer keyboard and press Enter to continue with the task. This approach enabled us to examine the two different PM components within PM performance: (a) performance in the prospective component, estimated by the accuracy of remembering to press the PM key in response to encountering the PM cues, and (b) the performance in the retrospective component, evaluated by the accuracy of retrieving the second word related to the cue (see Schnitzspahn et al., 2011 for a similar approach). Note that participants could press the white key in response to any word in the LDT and the text box would always open. They were asked to press the white key whenever they detected a PM cue, even when they forgot the second word related to the cue (when the text box appeared they could press Enter to continue the task without writing anything) or when they already gave the ongoing task response first.

Two different versions of the PM task were administered in a counterbalanced order. In the *high association* condition, the cues and the related words were highly associated (e.g., *Classeur - Bureau*), while in the *low association* condition the words forming the pairs were not associated (e.g., *Escalier - Carafe*). Participants had to work on two blocks of each condition (high vs. low) that varied in their encoding phase of the 20 word pairs. In the first encoding block each word pair was presented for 4 seconds followed by a fixation cross of 500ms. In the second encoding block, participants were asked to learn the same words pairs as before, but this time they could decide themselves how much time they wanted to spend for each word pair. In the PM task, each of the four blocks comprised 200 word/non-word trials. The same stimuli were used as cues in each of the high and low conditions.

JOLs. Participants estimated the probability of successful future PM performance before each task block on a scale from 0%-100% (0%, 20%, 40%, 60%, 80%, or 100% correct) (e.g. Nelson & Dunlosky, 1991). All JOLs were delayed, in that participants were asked to rate each of the cue words in a separate block having completed the encoding of all 20 pairs. In the same order as in the encoding phase, for each PM word pair, two types of JOLs were asked (see Figure 2 for an example). First, a pJOL for the *prospective* component defined by the probability of becoming aware that something had to be done upon encountering a PM cue; for example, "How probable is it that you will be aware to press the white key upon encountering the word "*Classeur*"?". Second, a rJOL for the *retrospective* component defined as the probability of retrospectively recalling the word associated with the cue; for example, "How probable is it that you will remember the specific word related to the word "*Classeur*"?".

Procedure

Participants were tested individually in a single session lasting approximately two hours including breaks. After informed consent was given, participants filled in a socio-demographic questionnaire. Thereafter, the computer task was introduced.

First, the ongoing LDT was introduced followed by a LDT training session composed of ten items with a feedback on performance accuracy. All participants scored above 80% correct in the training. Then, the PM task was introduced. Participants were instructed to memorize a list of pairs of words and to remember during the LDT to press the white key instead of the "yes" key when encountering the cue word of one of these pairs, then type in the second word of the pair. For both the ongoing task and the PM task, participants first read the instructions on the computer monitor. Upon completion, the experimenter cleared the screen and asked the participant to repeat the instructions in order to ensure correct understanding of each task.

Participants were informed that they would perform four blocks of the task. Half of the participants started with the two blocks of high association and the other half started with the two blocks of low association. For each block, after memorization of the 20 word pairs, participants were informed that we were interested in their ability to predict their memory abilities and asked to give JOLs on their future PM performance.

After the JOL assessment, participants worked on an arithmetical filler task. Ten simple equations were presented (e.g., 35 – 17 = 22?) and participants had to indicate if they were correctly solved or not by pressing yes/no. Then, the first PM task block started. This procedure (encoding, JOLs, delay task, PM task) was repeated for each block. There was a break between the high and low condition.

Results

Performance outcomes

As a first analytical step addressing the first hypothesis of reduced cognitive performance in older compared to young adults, mean performance in the ongoing task, the prospective component and the retrospective component were analyzed with 2 *age* (young, older) x 2 *association* (high, low) x 2 *block* (first, second) mixed repeated measures ANOVAs. Interaction effects were further explored using simple pairwise comparisons. Accuracy of PM performance is reported as a percentage of correct responses to facilitate subsequent comparison with performance predictions (see Table 1 for all inferential statistics). The alpha level was set to .05 unless stated otherwise.

Ongoing task performance

Young (M = .956, SE = .003) and older adults (M = .952, SE = .003) performed equally well, $F(1, 121) = .701$, $p = .404$, $\eta_p^2 = .006$.

PM task performance

Prospective component. Performance in the prospective component (see Table 1 for F-values, p-values and effect sizes) was calculated as the percentage of targets that were correctly identified (i.e., white key presses upon encountering PM cues). First, young adults (M = 70.3, *SE* = 2.6) outperformed the older adults (*M* = 55.6, *SE* = 2.9). Second, performance was better in the high (M = 68.5, *SE* = 2.2) compared to the low association condition (M = 57.4, *SE* = 2.3). Further, performance was better in the second block (*M* = 74.8, *SE* = 1.9) compared to the first block (*M* = 51.1, *SE* = 2.3). Furthermore, results show a significant interaction between block and age qualified by a significant three-way interaction.

Exploring the three-way interaction (see Figure 3 for means), subsequent follow-up tests conducted separately for each age group revealed that in the first block, young adults performed better in the high compared to the low association condition, $F(1, 121) = 10.090$, $p = .002$, $\eta_p^2 = .077$; while in the second block, young adults' performance did not differ anymore between high and low association, $F(1, 121) = 1.582$, $p = .211$, $\eta_p^2 = .013$.

As for the older adults, their performance in the high association was better than the low association condition in the first block, $F(1, 121) = 4.433$, $p = .037$, $\eta_p^2 = .035$, and in the second block, $F(1, 121) = 40.127$, $p < .001$, $\eta_p^2 = .249$. Further, with unlimited encoding time (i.e. in the second block) in the high association condition, older adults' performance reached young adults level of performance, $F(1,121) < 1$.

Retrospective component. The retrospective component of PM performance was the recall of the correct word in relation to the PM target cue (see Table 1 for F-values, p-values and effect sizes). As expected, young adults recalled more items (*M* = 51.7, *SE* = 2.5) than older adults (*M* = 27.0, *SE* = 2.7). In both age groups, highly associated words were better recalled (*M* = 52.3, *SE* = 2.2) than words in the low association condition (*M* = 26.5, *SE* = 2.1), and recall was better in the second block (M = 52.2, SE = 2.2) than the first one (M = 26.5, SE = 1.8). Further, the interactions between block and age and association and block were significant. Importantly, these interactions were qualified by a three-way interaction.

To explore the three-way interaction, follow-up tests were conducted separately for the two age groups (see Figure 4 for means). Young adults performed better in the second block compared to the first block for the high association pairs, $F(1, 121) = 123.784$, $p < .001$, $\eta_p^2 = .506$. They also performed better in the second block compared to the first block for the low association pairs, $F(1, 121) = 102.330$, $p < .001$, $\eta_p^2 = .458$.

Older adults show the same general pattern, performance in the retrospective component was better on the second block compared to the first one in the high association condition, $F(1, 121) = 128.413$, $p < .001$, $\eta_p^2 = .515$, and in the low association condition, $F(1, 121) = 14.287$, $p < .001$, $\eta_p^2 = .106$.

Metacognitive judgments

In a second analytical step, two critical measures of metacognition were considered as dependent variables to address the second hypothesis predicting that young adults should show better calibration than older adults, while no age differences were expected for resolution. As a brief reminder calibration describes the absolute accuracy of predictions, while resolution describes the relative accuracy of predictions. We will first report the results on calibration and examine whether the mean JOLs show the same pattern as the behavioral results. Mean JOL values for the prospective (pJOL) and the retrospective component (rJOL) were therefore computed and analyzed with two separate 2 *age* (young, older) x 2 *association* (high, low) x 2 *block* (first, second) ANOVAs (see Table 1 for F-values). These analyses will also address our third hypothesis that both age groups can adjust their predictions to the task requirements (also called sensitivity).

JOL calibration

Prospective component. For *pJOL*, older adults (M = 67.4, *SE* = 2.6) gave similar mean predictions as the young adults (M = 72.9, *SE* = 2.3). As expected, predictions were higher in the high association (M = 77.3, *SE* = 1.6) compared to the low association condition (M = 63.0, *SE* = 2.1), and predictions were higher for the second block (M = 79.6, *SE* = 1.8) compared to the first one (M = 60.7, *SE* = 1.9). The interaction between association and age was significant. Subsequent follow-up tests were conducted separately for the two age groups (see Figure 5). For the young adults, high association JOLs (*M* = 77.8, *SE* = 2.1) were higher than low association JOLs (*M* = 68.0, *SE* = 2.8), $F(1, 121) = 30.926$, $\eta_p^2 = .204$. The older adults showed the same pattern, with higher JOLs in the high association condition (*M* = 76.8, *SE* = 2.4) than for low association (*M* = 58.0, *SE* = 3.1), $F(1, 121) = 90.573$, $\eta_p^2 = .428$. The interaction between block and age was significant as well, $F(1, 121) = 18.718$, $\eta_p^2 = .134$. Again, planned comparisons were significant for both age groups. Young adults, $F(1, 121) = 85.331$, $p < .001$, $\eta_p^2 = .414$ as well as older adults, $F(1, 121) = 199.556$, $p < .001$, $\eta_p^2 = .623$ predicted higher performance in the second block (young: *M* = 79.9, *SE* = 2.4; older: *M* = 79.4, *SE* = 2.6), compared to the first (young: *M* = 65.9, *SE* = 2.5; older: *M* = 55.4, *SE* = 2.8).

Concerning calibration, we can also consider whether the mean values given in the JOLs are well matched to the actual levels of performance by calculating the discrepancy between the two (difference score). We used this measure for further analyses examining if a discrepancy between predicted and actual performance significantly differed from zero in young and older adults and if such discrepancies are truly stronger in older compared to young adults as predicted in our second hypothesis.

Results show that the discrepancy between prediction and performance differed significantly from zero in the older adults, $t(54) = 4.132$, $p < .001$, but not in the young adults, $t(67) = 1.021$, $p = .311$. Accordingly, directly comparing young and older adults, results show that the discrepancy is significantly larger in older adults (*M* = 11.7, *SE* = 2.8) than young adults (*M* = 2.6, *SE* = 2.6), $t(121) = -2.388$, $p = .018$. This pattern of overconfidence in the older but not in the young adults is visualized when comparing results displayed in Figure 3 showing the mean levels of PM performance for the prospective component, with Figure 5 showing the mean levels of pJOLs.

Retrospective component. For *rJOL*, young adults gave higher predictions (*M* = 68.5, *SE* = 2.2) than older adults (*M* = 37.0, *SE* = 2.5) for the retrospective component. As expected, predictions were higher in the high association (*M* = 63.0, *SE* = 1.9) than the low association condition (*M* = 32.5, *SE* = 1.9), and they were higher for the second block (*M* = 53.8, *SE* = 1.9) compared to the first one (*M* = 41.7, *SE* = 1.6). Moreover, significant interactions between association and age, and association and block, were found. These interactions were qualified in a 3-way interaction.

Further follow-up analyses were performed separately for the two age groups (see Figure 6 for means). Within young adults, there was no difference between the predictions for the first and second block in the high association condition, $F(1, 121) = .861$, $p = .355$, $\eta_p^2 = .007$. While in the low association condition predictions were higher on the second block compared to the first, $F(1, 121) = 92.490$, $p < .001$, $\eta_p^2 = .433$. Within older adults, predictions were higher on the second block compared to the first block in both high, $F(1, 121) = 5.022$, $p = .027$, $\eta_p^2 = .040$, and low association conditions, the effect being larger in the low association condition, $F(1, 121) = 39.293$, $p < .001$, $\eta_p^2 = .245$.

Next, we calculated a difference score to investigate the discrepancy between prediction and performance for the retrospective component. Results show that the discrepancy between prediction and performance differed significantly from zero in the young, $t(67) = 3.098$, $p = .003$, and the older adults, $t(54) = 5.430$, $p < .001$. Comparing the mean values given in the rJOLs (Figure 6) to the actual performance levels in the retrospective component of PM (Figure 4) shows a general *overconfidence* in our participants. Comparing the discrepancy scores between young and older adults shows no difference between young (M = 6.7, SE = 2.2) and older adults (M = 10.0, SE = 1.8), $t(121) = -1.111$, p = 269.

JOL resolution

Resolution was operationalized using gamma correlations between each participant's JOLs and PM performance (Nelson, 1984). Gamma is a continuous variable that ranges from -1 to +1. A large positive value means a high degree of accuracy; a value of zero means chance-level accuracy, and a negative value means less than chance-level accuracy. To examine age differences on JOL resolution, 2 *age* (young, older) x 2 *association* (high, low) x 2 *block* (first, second) ANOVAs were performed separately on prospective and retrospective gamma correlations. All *gamma* values reported in this section were reliably greater than zero, *ps* < .001, indicating above-chance resolution of JOLs.

Prospective component. The *pJOL gamma* correlations were significantly lower in older adults (M = .343, SE = .042) than in young adults (M = .456, SE = .038), $F(1, 121) = 4.040$, p = .047, η_p^2 = .032. The mean *gamma* correlation was higher in the low association (M = .444, SE = .033) compared to the high association condition (M = .355, SE = .035), $F(1, 121) = 5.690$, p = .019, η_p^2 = .045. There was no difference between the first block (M = .386, SE = .034) compared to the second block (M = .412, SE = .034), $F(1, 121) = .495$, p = .483, η_p^2 = .004. None of the interactions reached significance, all *p*s > .165.

Retrospective component. For the *rJOL gamma* correlations, older adults (M = .689, SE = .041) were as accurate as young adults (M = .643, SE = .036), $F(1, 121) = .715$, p = .399, η_p^2 = .006. Again, *gamma* correlations were higher for the low association (M = .716, SE = .030) compared to the high association condition (M = .616, SE = .033), $F(1, 121) = 9.848$, p = .002, η_p^2 = .075. And participants were as accurate on the first (M = .670, SE = .031) as on the second block (*M* = .662, *SE* = .029), $F(1, 121) = .104$, $p = .758$, $\eta_p^2 = .001$. No interaction was significant, all *p*s > .272.

Prospective vs. Retrospective component

Comparing results from the above analyses on the prospective and the retrospective component of PM already suggests differences in performance and prediction between the two components. However, to directly address our fourth hypothesis that young and older adults predict their performance in the retrospective component more accurately than their performance in the prospective component of PM, we directly compared the gamma correlations for the two components. In young adults the mean rJOL *gamma* correlation is significantly higher (*M* = .643, *SE* = .039) than the mean pJOL *gamma* correlation (*M* = .456, *SE* = .039), $t(67) = -5.454$, $p < .001$. In older adults, the mean rJOL *gamma* correlation is also significantly higher (*M* = .690, *SE* = .036) than mean pJOL *gamma* correlation (*M* = .343, *SE* = .042), $t(54) = 6.721$, $p < .001$.

Discussion

The overall goal of the study was to examine metacognition as a possible factor in explaining PM age declines. In particular, we examined age-related effects on actual performance and item-level predictions of future PM performance and the relation between prediction and performance. Further, we tested for sensitivity effects in PM in young and older adults by manipulating PM task difficulty and previous task experience within participants. Possible effects on the prediction and performance of the prospective versus the retrospective components of the PM task were disentangled by considering them separately in all analyses and directly comparing the accuracy of the predictions for both.

Performance and predictions of future PM performance

In line with our first hypothesis predicting lower PM performance in older compared to young adults, the main pattern showed an age impairment in both retrospective and prospective components of PM. Our second hypothesis and main research question concerned possible age differences in individuals' predictions about their future PM performance. The age impairment observed in performance was not fully reflected in the predictions: the magnitude of the JOLs did not differ for young and older adults for the prospective component of the task, despite the older adults' poorer performance. This is indicative of a failure to appreciate their difficulties in this kind of laboratory PM task, and suggests an overconfidence in the older adults. Further analyses examining the discrepancy between predictions and performance support these conclusions, as results show that only the older adults were significantly overconfident concerning their performance in the prospective component. However, qualifying this general conclusion, for the retrospective component, the magnitude of predictions was significantly lower for the older adults, and appropriately so, since they performed worse on that part of the task. Comparing predictions and performance levels suggests that both age groups were overconfident regarding the retrospective component.

The critical measure in our study, was the accuracy of the item-by-item predictions (i.e. the JOL resolution operationalized using gamma correlations), which we have measured here for the first time in older adults. In line with our findings on calibration of the predictions reported above, we found that older adults were disproportionately inaccurate for the prospective component, whereas there were no age differences for the retrospective component. As such, older adults were less able to accurately predict cue detection, whereas they were as able as young participants in predicting the recall of the cued item. First, the lack of an effect of age for the retrospective component maps onto the repeatedly observed age equivalence on delayed JOLs on more standard episodic memory tasks (e.g. Hertzog et al., 2010). Second, the comparative deficit for cue detection is a novel finding, and one which we discuss below.

Task manipulation effects on PM performance and exploration of sensitivity effects

The third aim of the present study was to test whether young and older adults' JOLs are sensitive to factors which influence their PM function (sensitivity effects). To that end, PM task difficulty was experimentally manipulated. One condition was considered as easier, in which participants had to memorize and give JOLs for high associated words; and the other condition was considered as more difficult, in which participants had to memorize and give JOLs for low associated words. We also measured predictions' adjustments regarding encoding time in two blocks with the same word materials: In the first block study time was fixed, whereas in the second block study time was unlimited.

The findings on PM confirm that our task manipulations were successful and influenced performance as expected. In particular, the general age deficit in PM was qualified by interaction effects of our experimental manipulations. When participants could decide how long they wanted to study the highly associated target cues and related words, older adults improved their performance more strongly and reached the same level of performance on the prospective component as the young adults (up to 80%). Thus, self-paced encoding and a reduction of the retrospective memory load of the PM task can greatly improve older adults' performance and thereby eliminate age differences. These differences in performance were correctly reflected in young and older adults' responses for pJOLs and rJOLs which differed according to association and whether they were learning the items for a first or second time. Thus, the present study is the first one showing sensitivity effects in PM and more importantly that these are preserved in aging.

Differences between pJOLs and rJOLs

The final aim of the present study was to disentangle the predictions regarding the two components of PM. As described above, in line with the retrospective metamemory literature for

JOLs, older and young adults showed a similar calibration and resolution for the retrospective component of the PM task. The overlap between our PM task and a typical delayed JOL task is clear. In both tasks, participants study a set of word pairs, and after a delay are given a cue word and are asked to predict the likelihood of recalling the second word. It may be unsurprising that given that they detect the cue word, their prediction accuracy in the retrospective component is as unimpaired as it is in a standard delayed JOL task. Yet, our research adds to the body of work on episodic memory and delayed JOLs also from a PM perspective, and shows that older adults are able to accurately gauge which information will later be available for recall.

Conceptually importantly, the same age equivalence was not found for the prospective part of the PM component. To recap, with the pJOL, participants were judging whether they would later detect a cue word when it was embedded in an ongoing task. Comparing these predictions with actual performance showed significantly stronger differences in the older adults compared to the young controls. Accordingly, young adults showed a significantly better resolution for the prospective component compared to older adults. The mean values again point to overconfidence in the older adults, this inaccuracy stems from not appreciating how difficult the task will be at the point at which they make their judgements.

Young and older adults were more accurate in predicting their performance in the retrospective component compared to their performance in the prospective component. This result is in line with Schnitzspahn et al. (2011) who also found higher gamma correlations for the retrospective than the prospective component on their PM task carried out only with young adults. The present aging data further supports that these two aspects of metamemory are dissociable by showing that the aging process affects predictions of prospective function more than retrospective function. At face value, this is an important finding both for aging and for PM research. From a conceptual perspective, it highlights the importance to distinguish the two components within PM. From an applied perspective, it gives crucial insight on which processes possible interventions aiming to improve PM in general and especially in aging should focus on.

Why should metamemory be accurate for retrospective but not prospective components of the same task in older adults? This could be due to task demands which are different for the two components. The prospective component could be operationalized as a recognition task, and the retrospective component as a recall task. Souchay and Isingrini (2012) showed that even within the same group of older adults, predictions of upcoming recall (JOLs) were accurate, whereas predictions of upcoming recognition (Feeling-of-Knowing, FOK) were not.

Besides episodic memory, successful PM function draws on several other cognitive abilities, including attention and executive function (Kliegel, Altgassen, Hering, & Rose, 2011). Here we propose that for PM predictions to be accurate, one needs to take on board the current state of all these domains in order to make an accurate assessment of PM. That is, the older adult has to estimate the ease with which they can encode the information and recognize it later (just as in any episodic FOK task) but additionally, they need to consider whether they will be overly engrossed in the ongoing task, unable to switch between one task or another for example. That is, the predictions have to take on board several different stages of planning and execution (e.g. Kliegel, Martin, McDaniel, & Einstein, 2002) and not just the simple retention and recognition of a stimulus. We propose here that the older adults may not be able to integrate and use all these sources of information in making their predictions. This is an idea which needs further experimentation. We point out, however, that the only other paradigm whereby metamemory deficits are routinely found in older adults is with the FOK. A key explanation for this deficit is that older adults lack the requisite recollection process to generate suitable information on which to make their metamemory evaluations (Daniels, Toth, & Hertzog, 2009; Souchay et al., 2007). Here we propose the same interpretation of the PM task could be made. Whereas young people can make a complex evaluation of the upcoming recognition of a PM cue, older adults fail to appreciate their particular difficulty on this task, even though they are aware of their memory difficulties more generally.

Possible limitations and outlook

The design constraints of taking metamemory judgements in a PM task lead to a somewhat specific task, and this has perhaps partly contributed to the results found here. Specifically, a high number of PM cues and related words (i.e., twenty) was necessary to calculate gamma correlations, which made the task more difficult than common PM tasks that usually have less than half this number of cues (up to six different cues, e.g. Cohen, Jaudus & Gollwitzer, 2008; but see Albinski, Kliegel, & Gurynowicz, 2016). This difference might explain why in previous studies, young adults were usually underconfident concerning their PM performance (Meeks et al., 2007; Schnitzspahn et al., 2011), while in this specific task, they were generally overconfident.

While we acknowledge that the number of PM cues was relatively high in the present paradigm we also underline that this is not exceptional in the available PM literature. While in some PM tasks there are as few as one or two PM cues, in other PM tasks, such as ERP studies, there are up to 50 cues. To ensure that the task is still a PM task, a ratio of 10% PM items among the ongoing task trials is generally accepted to prevent overt sustained attention towards the PM task and sufficient attention capture of the ongoing task. As for neuroimaging studies, a high number of cues is also better to ensure stability of the metamemory judgments measure.

The finding of lower performance in the retrospective compared to the prospective component could also partly reflect the relationship between the two components as operationalized in our task. Specifically, the recall component was contingent on the correct identification and response to a PM cue. However, the task used in the present study follows previous work (Schnitzspahn et al., 2011).

For future research, the link between metamemory and actual function should be considered more closely. We observed that leaving the participants free to study for longer on the second block did reduce age differences in performance (although the older adults did remain inaccurate in terms of their pJOLs). In a next step, one could measure actual study time (a typical measure of strategic control) to test if control behavior actually mediates the observed relations between predictions and PM performance and age differences. One could argue that our paradigm, with a fixed encoding time in the first block and a self-paced encoding time in the second block, produces practice effects: performing the same task for a second time should make it easier and therefore more accurate to predict one's performance. Thus, the difference in encoding time may not be the only reason why young and older adults predicted higher performance in the second block than in the first one. However, we believe it is crucial for the examination of possible effects of previous task experience to ensure that the initial experience is comparable for all participants and therefore did not counterbalance blocks using different encoding times.

It would be interesting to test if the present findings generalize to time-based PM tasks (remembering to perform the PM task at a specific time or after a specified period of time has elapsed) in future studies on metamemory and PM in aging. Time-based PM tasks lack an external event indicating the right moment to initiate the PM task and require more self-initiated processing for monitoring the time (e.g., Einstein, McDaniel, Richardson, Guynn, & Cunfer, 1995; Kvavilashvili & Fisher, 2007). Thus, the prospective component in a time-based task is not relying on recognition and accordingly pJOLs and rJOLs might not differ from each other for time-based tasks.

While we observed general overconfidence in both age groups, the only earlier study on metamemory, PM and aging (Devolder et al., 1990) reported that older adults were more accurate in their performance predictions than young adults. An important difference between the present study and the one by Devolder et al. concerns the setting in which the PM task had to be performed. While we assessed PM and related predictions in the laboratory, Devolder et al. asked their participants to predict performance in a naturalistic task that had to be performed in their everyday environment. As older adults have more experience in naturalistic PM tasks they might evaluate their performance in those settings more precisely and might then initiate proper strategies to achieve an optimal level of performance. This might explain the common finding of age benefits in naturalistic PM tasks (Henry et al., 2004) which contrasts the usual age declines observed in the laboratory. Accordingly, this pattern has been described as the "age-PM-paradox". Future studies should test the role of metamemory to explain this paradox.

Conclusion

The main aim was to address whether a metamemory deficit might be found alongside the PM deficit in older adults. We replicated the standard PM deficits seen in older adults for the prospective and retrospective components. Alongside this we found that predictions of PM were impaired. Thus, taking a real-life example, we might imagine that older adults not only fail to remember to go to the pharmacy when they are at the shops, but they also are less aware that they will fail to go to the pharmacy, meaning that they possibly fail to strategically control PM in an appropriate fashion (such as by setting an alarm or asking for a reminder). However, for the retrospective information, should they remember to go into the pharmacy, even if they cannot remember why they were going there, they should at least be able to appreciate that this can happen and do something about it (such as make a list of things to buy), since they show an appropriate awareness of retrospective memory.

Compliance with Ethical Standards

Funding: This study was founded by a joint grant from the Swiss National Science Foundation (SNSF) and the Agence Nationale de Recherche (ANR; France).

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

References

Albiński, R., Kliegel, M., Gurynowicz, K. (2016). The influence of high and low cue-action association on prospective memory performance. *Journal of Cognitive Psychology*, *28*(6), 707-717.

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2001). Age-of acquisition and word frequency in the lexical decision task: Further evidence from the French language. *Current Psychology of Cognition*, *20*(6), 401-444.

Bonin, P., Méot, A., Aubert, L.-F., Malardier, N., Niedenthal, P. M., & Capelle-Toczek, M.-C. (2003). Normes de concrétude, de valeur d'imagerie, de fréquence subjective et de valence émotionnelle pour 866 mots. *L'année Psychologique*, *103*(4), 655-694.

Brandimonte, M. A., Einstein, G. O., & McDaniel, M. A. (Eds.). (2014). *Prospective memory: Theory and applications*. New York and London: Psychology Press.

Bialystok, E., Craik, F., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 859-873.

Castel, A. D., Middlebrooks, C. D., & McGillivray, S. (2016). Monitoring memory in old age: impaired, spared and aware. In Dunlosky, J., Tauber, S. U. K., & Tauber, P. D. R. S. (Eds.). (2016). *The Oxford Handbook of Metamemory* (pp. 519-535). Oxford University Press.

Cherry, K. E., Martin, R. C., Simmons-D'Gerolamo, S. S., Pinkston, J. B., Griffing, A., & Drew Gouvier, W. (2001). Prospective remembering in younger and older adults: Role of the prospective cue. *Memory*, *9*(3), 177-193.

Cohen, A.-L., Jaudas, A., & Gollwitzer, P. M. (2008). Number of cues influence the cost of remembering to remember. *Memory & Cognition*, *36*(1), 149-156.

Craik, F. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: mechanisms and performances* (pp. 409-422). Amsterdam: Elsevier Science Publishers.

Craik, F., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191-211). New York: Plenum.

Connor, L. T., Dunlosky, J., & Hertzog, C. (1997). Age-related differences in absolute but not relative metamemory accuracy. *Psychology and Aging*, *12*(1), 50-71.

Daniels, K. A., Toth, J. P., & Hertzog, C. (2009). Aging and recollection in the accuracy of judgments of learning. *Psychology and Aging*, *24*(2), 494-506.

Devolder, P. A., Brigham, M. C., & Pressley, M. (1990). Memory performance awareness in younger and older adults, *Psychology and Aging*, *5*(2), 291-303.

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition*, *25*(5), 691-700.

Dunlosky, J., & Tauber S. K. (2015). A brief history of metamemory research and handbook overview. In J. Dunlosky & S. K. Tauber (Eds.), *Oxford handbook of metamemory* (pp. 7-34). New York, NY: Oxford University Press.

Eakin, D. K., Hertzog, C., & Harris, W. (2014). Age invariance in semantic and episodic metamemory: Both younger and older adults provide accurate feeling-of-knowing for names of faces. *Aging, Neuropsychology, and Cognition*, *21*(1), 27-51.

Einstein, G. O., & McDaniel, M. A. (1990). Normal aging and prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 717-726.

Einstein, G. O., & McDaniel, M. A. (1996). Retrieval processes in prospective memory: Theoretical approaches and some new findings. In M. Brandimonte, G. O. Einstein, & M. A. McDaniel (Eds.), *Prospective memory: Theory and applications* (pp. 115-141). New Jersey: Erlbaum.

Einstein, G. O., & McDaniel, M. A. (2005). Prospective memory: Multiple retrieval processes. *Current Directions in Psychological Science*, *14*, 286 –290.

Folstein, M., Folstein, S., & McHugh, P. (1975). "Mini-mental state": A practical method for grading the cognition state of patients for the clinician. *Journal of Psychiatry Research*, *12*, 189-198.

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.

Hering, A., Phillips, L. H., & Kliegel, M. (2013). Importance effects on age differences in performance in event-based prospective memory. *Gerontology*, *60*(1), 73-78.

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,* 1-6.

Hertzog, C., & Dunlosky, J. (2011). Metacognition in later adulthood spared monitoring can benefit older adults' self-regulation. *Current Directions in Psychological Science*, *20*(3), 167-173.

Hertzog, C., Sinclair, S. M., & Dunlosky, J. (2010). Age differences in the monitoring of learning: cross-sectional evidence of spared resolution across the adult life span. *Developmental Psychology*, *46*(4), 939-948.

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.

Ihle, A., Kliegel, M., Hering, A., Ballhausen, N., Lagner, P., Benusch, J., … Schnitzspahn, K. M. (2014). Adult age differences in prospective memory in the laboratory: are they related to higher stress levels in the elderly? *Frontiers in Human Neuroscience*, *8*, 1021-1030.

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*(1-2), 84-10*1.*

Kliegel, M., Altgassen, M., Hering, A., & Rose, N. (2011). A process-model based approach to prospective memory impairment in Parkinson's disease. *Neuropsychologia, 49*(8), 2166-2177.

Kliegel, M., Ballhausen, N., Hering, A., Ihle, A., Schnitzspahn, K. M., & Zuber, S. (2016). Prospective memory in older adults: Where we are now and what is next. *Gerontology*, *62*(4), 459-466.

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.

Kliegel, M., Martin, M., McDaniel, M. A., & Einstein, G. O. (2002). Complex prospective memory and executive control of working memory: A process model. *Psychological Test and Assessment Modeling*, *44*(2), 303.

Kliegel, M., Schnitzspahn, K., Souchay, C. & Moulin, C. (2016) Prospective memory: New perspectives for geropsychological research. In NA Pachana (ed.), *Encyclopedia of Geropsychology*. Springer, Singapore.

Knight, R. G., Harnett, M., & Titov, N. (2005). The effects of traumatic brain injury on the predicted and actual performance of a test of prospective remembering. *Brain Injury*, *19*(1), 19-27.

Mattli, F., Schnitzspahn, K. M., Studerus-Germann, A., Brehmer, Y., & Zöllig, J. (2014). Prospective memory across the lifespan: Investigating the contribution of retrospective and prospective processes. *Aging, Neuropsychology, and Cognition, 21*(5), 515-543.

McDaniel, M. A., & Einstein, G. O. (2000). Strategic and automatic processes in prospective memory retrieval: A multiprocess framework. *Applied Cognitive Psychology*, *14*(7), 127-144.

McDaniel, M. & Einstein, G. (2008). Prospective memory and aging: old issues and new questions. In S. M. Hofer & D. F. Alwin (Eds.). *Handbook of cognitive aging: Interdisciplinary perspectives* (pp. 168-180). Thousand Oaks, CA: SAGE Publications Ltd.

McDaniel, M. A., Glisky, E. L., Guynn, M. J., & Routhieaux, B. C. (1999). Prospective memory: A neuropsychological study. *Neuropsychology*, *13*(1), 103-110.

Meeks, J. T., Hicks, J. L., & Marsh, R. L. (2007). Metacognitive awareness of event-based prospective memory. *Consciousness and Cognition*, *16*(4), 997-1004.

Meier, B., Wartburg, P. Von, Matter, S., Rothen, N., & Reber, R. (2011). Performance predictions improve prospective memory and influence retrieval experience. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *65*(1), 12-18.

Morson, S. M., Moulin, C. J. A., & Souchay, C. (2015). Selective deficits in episodic feeling of knowing in ageing: A novel use of the general knowledge task. *Acta Psychologica*, *157*, 85-92.

Moulin, C. J. A. (2002). Sense and sensitivity: Metacognition in Alzheimer's disease. In T. J. Perfect & B. L. Schwartz (Eds.). *Applied Metacognition* (pp. 197-223). Cambridge.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, *95*, 109-133.

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". *Psychological Science*, *2*, 267-270.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation*, *26*, 125-322.

Raven, J. C., Court, J. H., & Raven, J. (1986). Manual for Raven's progressive matrices and vocabulary scales. Oxford: Oxford Psychologists Press.

Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: A Meta-analytic review. *Psychological Bulletin*, *137*(1), 131-148.

Rhodes, M. G. (2016). Judgments of learning: Methods, data, and theory. In Dunlosky, J., Tauber, S. U. K., & Tauber, P. D. R. S. (Eds.). (2016). *The Oxford Handbook of Metamemory* (pp. 65-80). Oxford University Press.

Rose, N. S., Rendell, P., McDaniel, M., Aberle, I., & Kliegel, M. (2010). Age and individual differences in prospective memory during a "virtual week": The role of working memory, vigilance, task-regularity, and cue-focality. *Psychology and Aging, 25*(3)*,* 595-605.

Rummel, J., Kuhlmann, B. G., & Touron, D. R. (2013). Performance predictions affect attentional processes of event-based prospective memory. *Consciousness and Cognition*, *22*(3), 729-741.

Rummel, J., & Meiser, T. (2013). The role of metacognition in prospective memory: Anticipated task demands influence attention allocation strategies. *Consciousness and Cognition*, *22*(3), 931-943.

Schnitzspahn, K. M., Stahl, C., Zeintl, M., Kaller, C. P., & Kliegel, M. (2013). The role of shifting, updating, and inhibition in prospective memory performance in young and older adults. *Developmental Psychology*, *49*(8), 1544-1554.

Schnitzspahn, K. M., Zeintl, M., Jäger, T., & Kliegel, M. (2011). Metacognition in prospective memory: are performance predictions accurate? *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *65*(1), 19-26.

Souchay, C., Isingrini, M., Clarys, D., Taconnat, L., & Eustache, F. (2004). Executive functioning and judgment-of-learning versus feeling-of-knowing in older adults. *Experimental Aging Research*, *30*(1), 47-62.

Souchay, C., & Isingrini, M. (2012). Are feeling-of-knowing and judgment-of-learning different? Evidence from older adults. *Acta Psychologica*, *139*(3), 458-464.

Souchay, C., Moulin, C. J., Clarys, D., Taconnat, L., & Isingrini, M. (2007). Diminished episodic memory awareness in older adults: Evidence from feeling-of-knowing and recollection. *Consciousness and Cognition*, *16*(4), 769-784.

Tauber, S. U. K., & Dunlosky, J. (2016). A brief history of metamemory research and handbook overview. In Dunlosky, J., Tauber, S. U. K., & Tauber, P. D. R. S. (Eds.). (2016). *The Oxford Handbook of Metamemory* (pp. 7-21). Oxford University Press.

Thadeus Meeks, J., Hicks, J. L., & Marsh, R. L. (2007). Metacognitive awareness of event-based prospective memory. *Consciousness and Cognition*, *16*(4), 997-1004.

Touron, D. R., Oransky, N., Meier, M. E., & Hines, J. C. (2010). Metacognitive monitoring and strategic behaviour in working memory performance. *The Quarterly Journal of Experimental Psychology*, *63*(8), 1533-1551.

Woods, S. P., Weinborn, M., Velnoweth, A., Rooney, A., & Bucks, R. S. (2012). Memory for intentions is uniquely associated with instrumental activities of daily living in healthy older adults. *Journal of the International Neuropsychological Society*, *18*(1), 134-138.

Zimmermann, T. D., & Meier, B. (2006). The rise and decline of prospective memory performance across the lifespan. *The Quarterly Journal of Experimental Psychology*, *59*(12), 2040-2046.

Zöllig, J., West, R., Martin, M., Altgassen, M., Lemke, U., & Kliegel, M. (2007). Neural correlates of prospective memory across the lifespan. *Neuropsychologia*, *45*(14), 3299-3314.

Footnote

[1] Performance measures and predictions for the prospective and the retrospective component will be analyzed and reported separately. Given that previous studies suggest different patterns for the predictions of the prospective and retrospective components of the PM task (Schnitzspahn et al., 2011) and stronger age effects in the prospective compared to the retrospective component on a behavioral level (Mattli, Schnitzspahn, Studerus-Germann, Brehmer, & Zöllig, 2014; Zöllig, West, Martin, Altgassen, Lemke, & Kliegel, 2007), it seems important to distinguish them. Moreover, the measurement of the two components was not independent. Specifically, the recall component was contingent on the correct identification and response to a PM cue. Thus, performance in the two components cannot be compared directly from a conceptual, methodological and developmental point of view.

Table 1 F-values, Significance Levels and Partial Eta-square for the PM Performance and Predictions for the Prospective and Retrospective Components

	Prospective component				Retrospective component			
	Performance		Prediction		Performance		Prediction	
	F (df)	η_p^2	*F (df)*	η_p^2	*F (df)*	η_p^2	*F (df)*	η_p^2
Age effect (younger vs. older)	13.692 (1,121)***	.893	2.478 (1,121)	.020	44.381 (1,121)***	.268	41.230 (1,121)***	.254
Association effect (high vs. low)	29.455 (1,121)***	.196	116.530 (1,121)***	.491	137.702 (1,121)***	.532	302.811 (1,121)***	.714
Block effect (first vs. second)	240.913 (1,121)***	.666	278.241 (1,121)***	.697	292.659 (1,121)***	.707	126.487 (1,121)***	.511
Age x Association	2.990 (1,121)	.024	11.273 (1,121)**	.085	0.951 (1,121)	.008	20.529 (1,121)***	.145
Age x Block	5.387 (1,121)*	.043	18.718 (1,121)***	.134	5.479 (1,121)*	.043	0.019 (1,121)	.000
Association x Block	.768 (1,121)	.006	2.317 (1,121)	.019	8.524 (1,121)**	.066	56.342 (1,121)***	.318
Age x Association x Block	13.544 (1,121)***	.101	0.199 (1,121)	.002	14.130 (1,121)***	.105	6.059 (1,121)*	.048

*p < .05, **p < .01, *** p < .001.

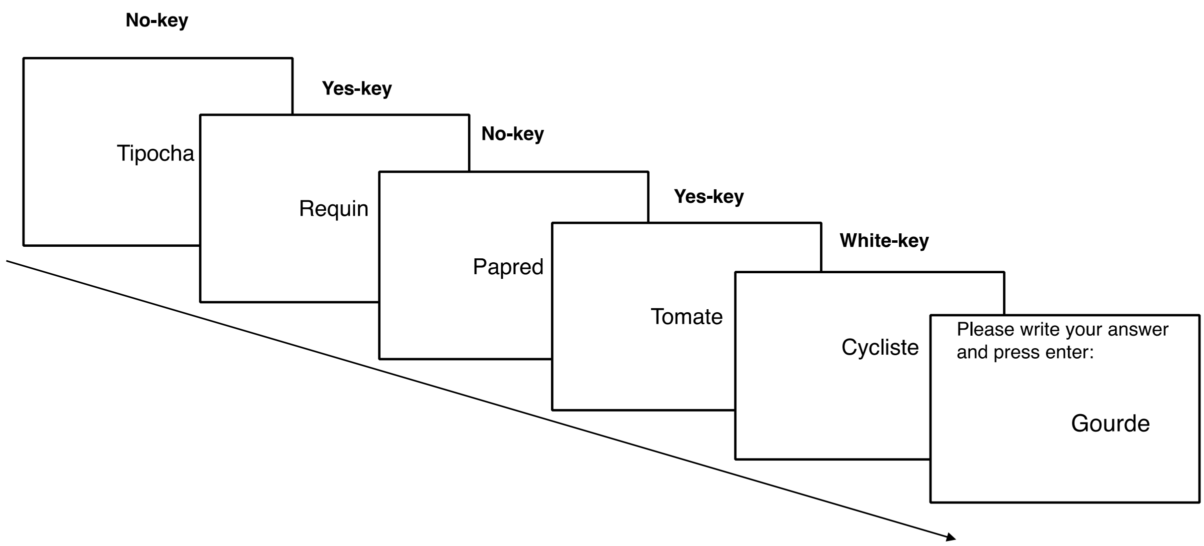

Figure 1. PM task: lexical decision task with the high associated word pair "Cycliste – Gourde" as PM cue and related action.

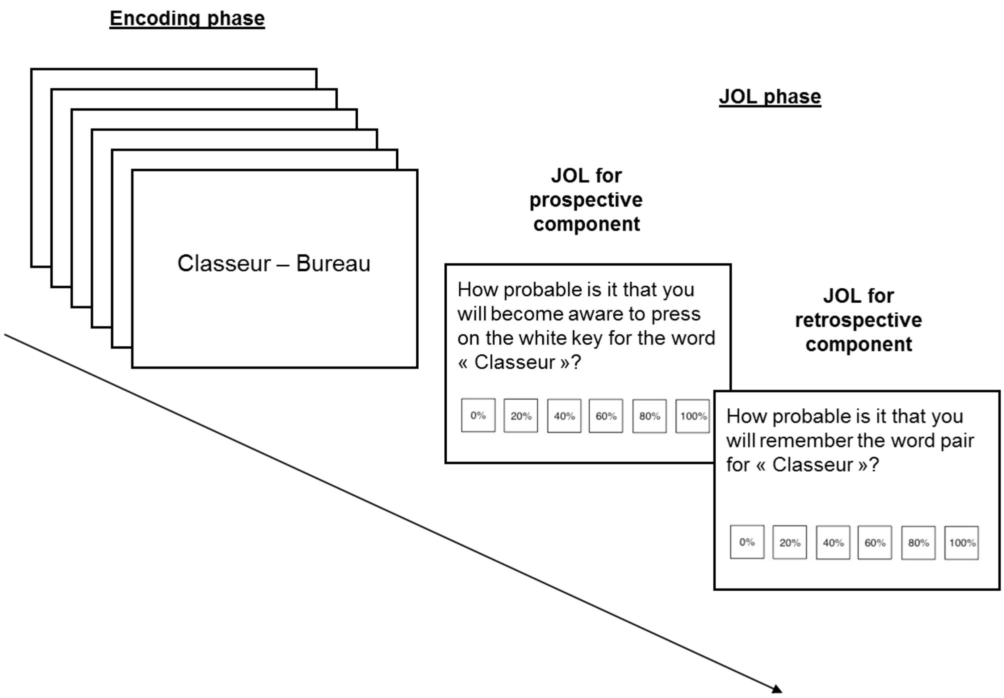

Figure 2. Encoding phase followed by the JOL phase with an example of a JOL for the high associated word pair "Classeur – Bureau".

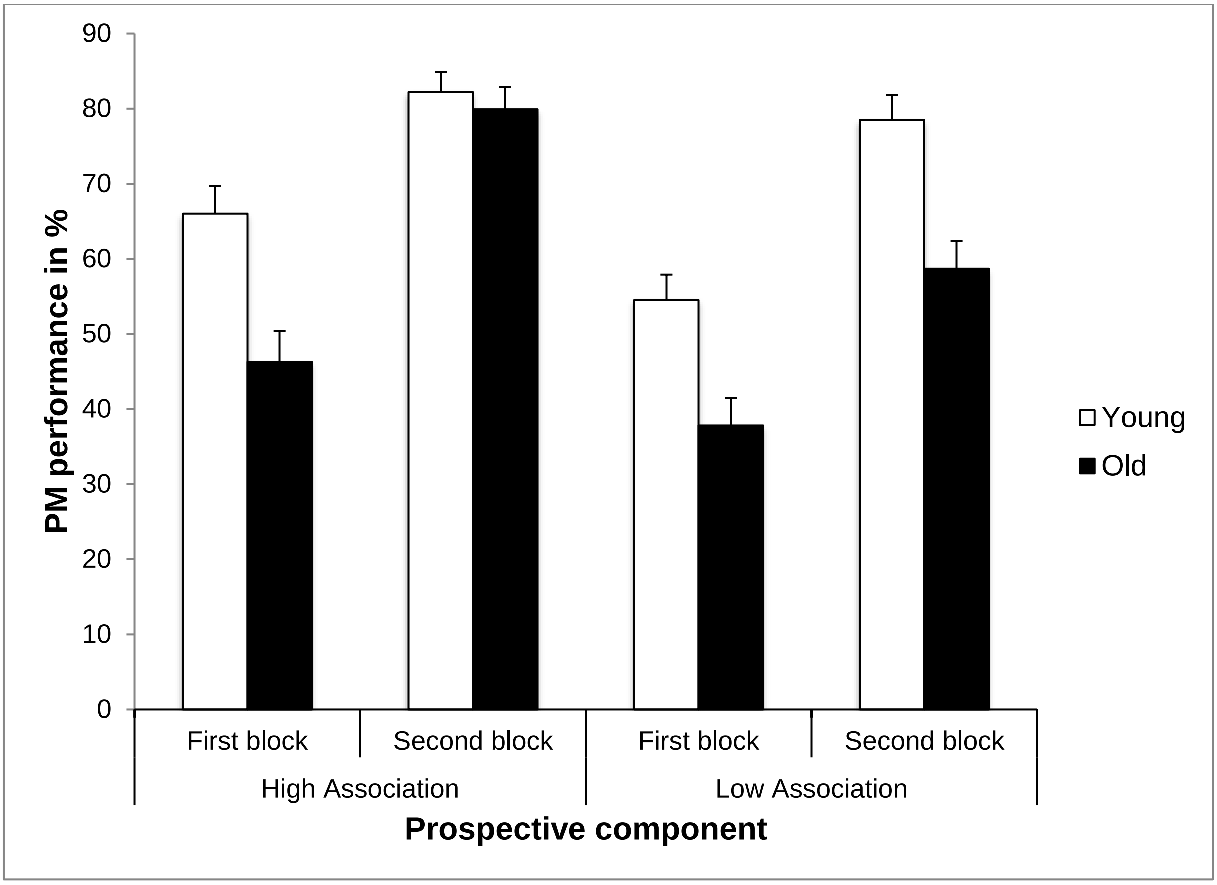

Figure 3. PM performance (proportion of correct PM responses) for the prospective component in both age groups as a function of association condition and blocks. Error bars represent the standard error (SE).

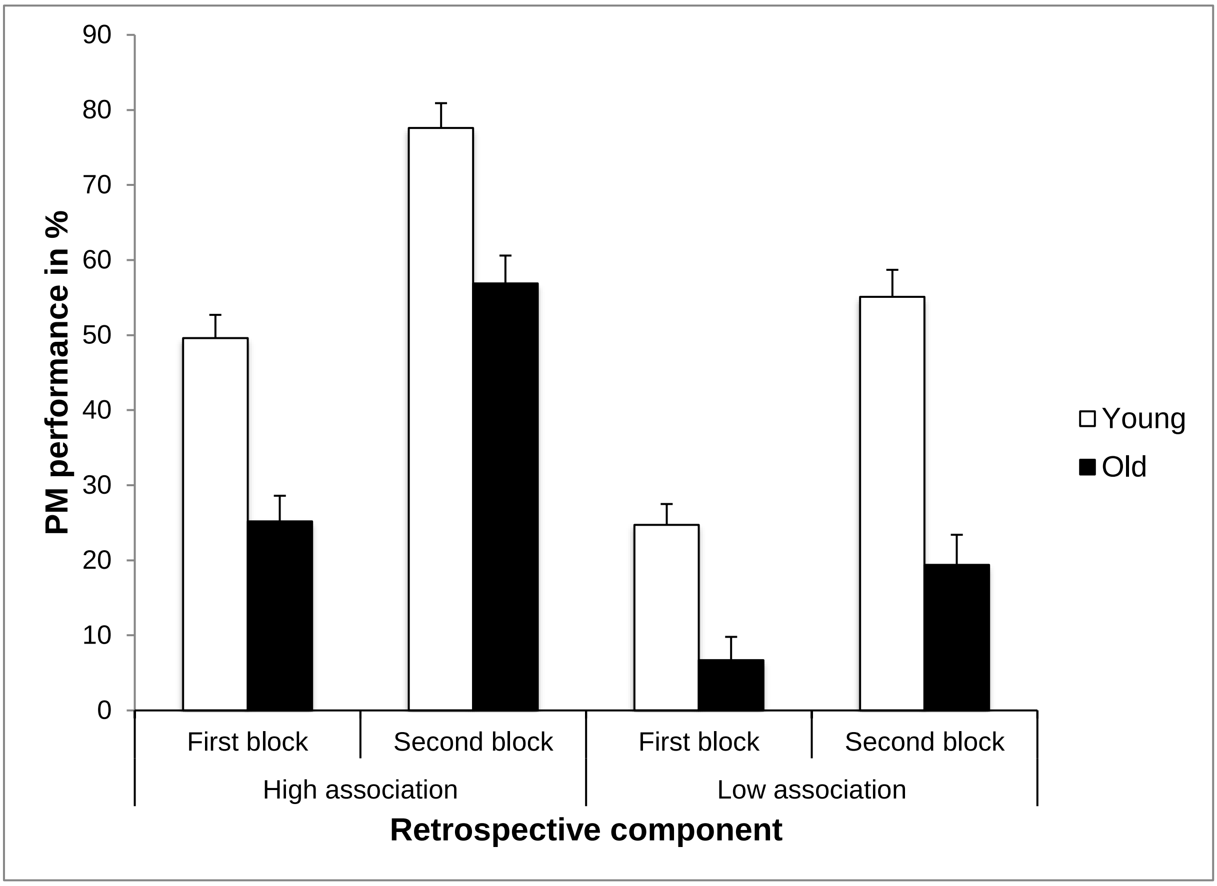

Figure 4. PM performance (proportion of correct PM responses) for the retrospective component in both age groups as a function of association condition and blocks. Error bars represent the standard error (SE).

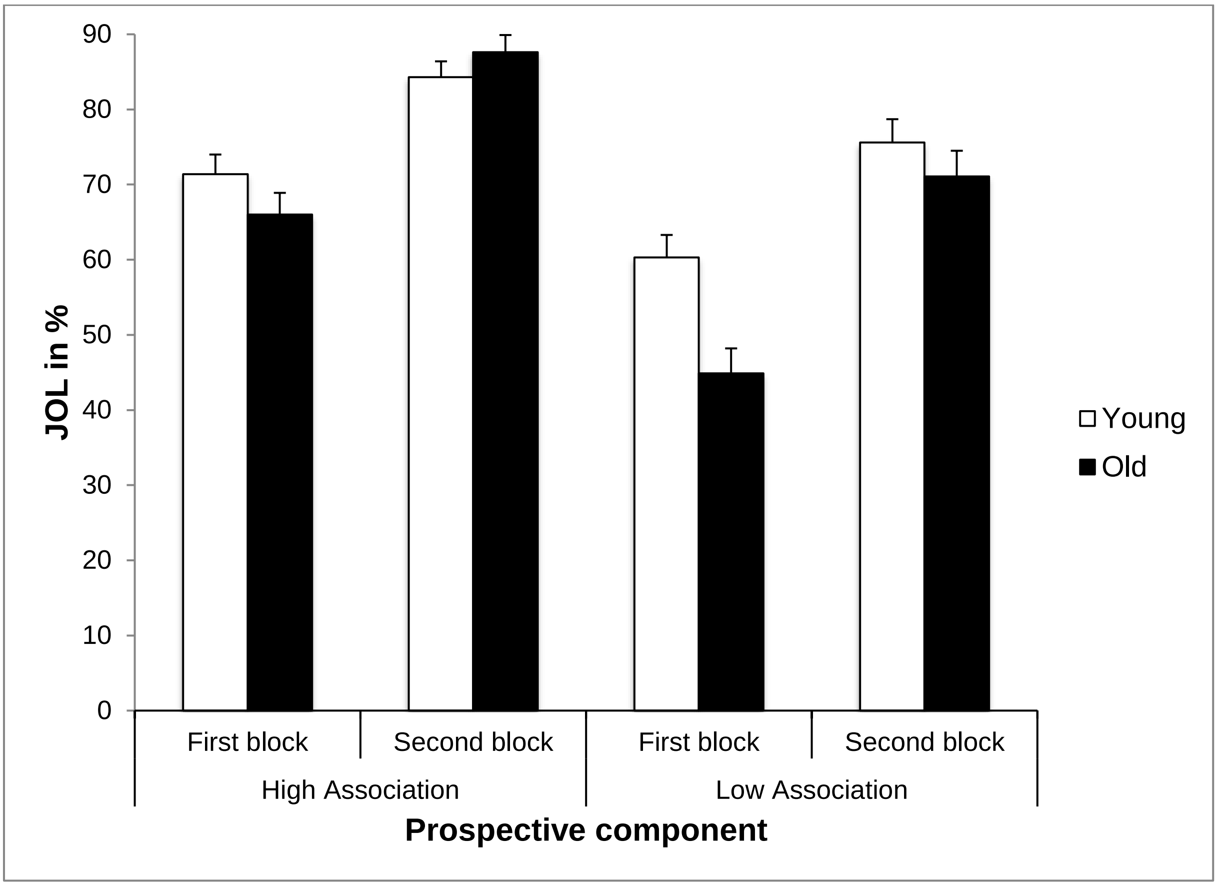

Figure 5. Predictions for the prospective component in both age groups as a function of association condition and blocks. Error bars represent the standard error (SE).

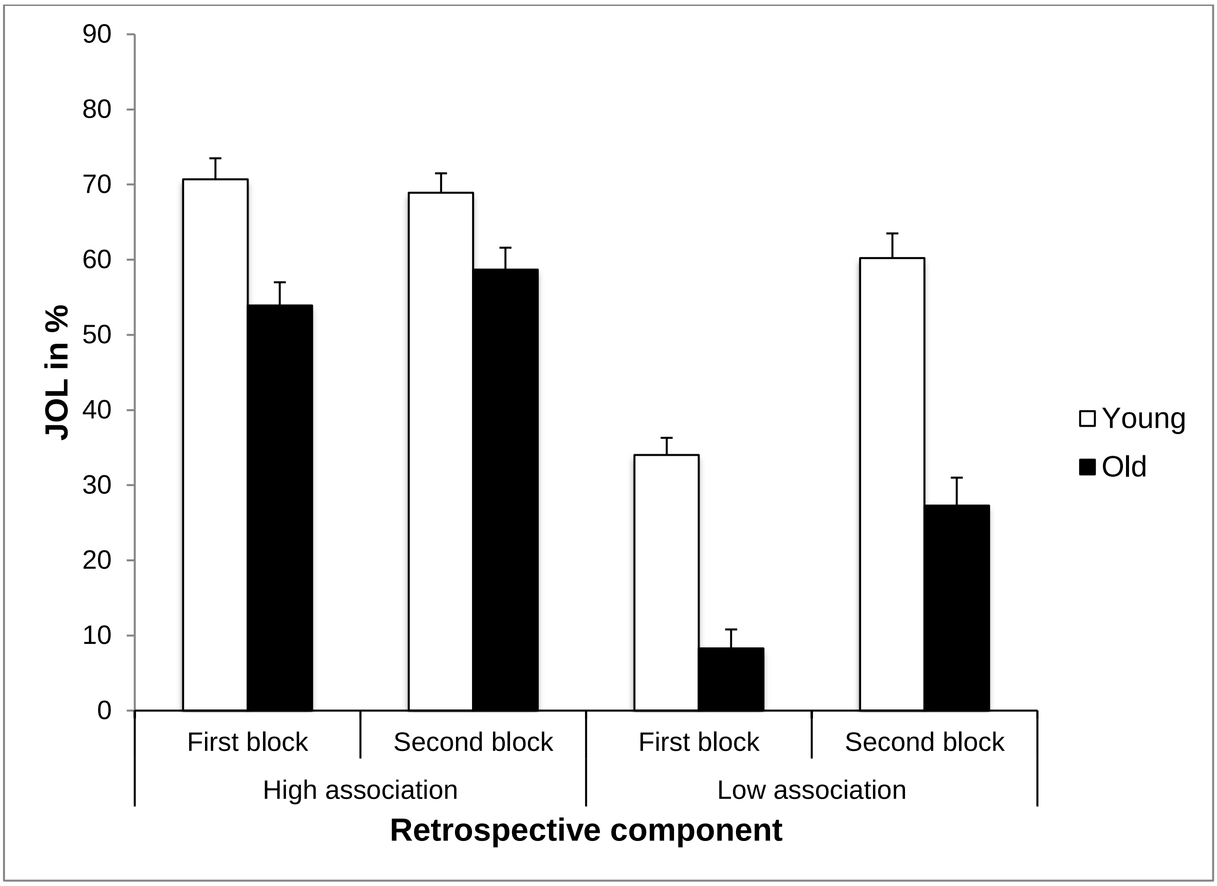

Figure 6. Predictions for the retrospective component in both age groups as a function of association condition and blocks. Error bars represent the standard error (SE).