High-confidence memory errors in old age: The roles of monitoring and binding processes

Yana Fandakova, Yee Lee Shing & Ulman Lindenberger

To cite this article: Yana Fandakova, Yee Lee Shing & Ulman Lindenberger (2013) High-confidence memory errors in old age: The roles of monitoring and binding processes, Memory, 21:6, 732-750, DOI: 10.1080/09658211.2012.756038

To link to this article: https://doi.org/10.1080/09658211.2012.756038

Published online: 10 Jan 2013.

Submit your article to this journal

Article views: 325

View related articles

Citing articles: 8 View citing articles
High-confidence memory errors in old age: The roles of monitoring and binding processes

Yana Fandakova, Yee Lee Shing, and Ulman Lindenberger

Center for Lifespan Psychology, Max Planck Institute for Human Development, Berlin, Germany

Based on a two-component model of episodic memory development across the lifespan, we examined the contribution of memory monitoring and binding processes to older adults’ increased susceptibility to false memories with high subjective confidence. Younger and older adults worked on a modified version of the continuous recognition task (Schnider, von Daniken, & Gutbrod, 1996). Participants saw the same set of unrelated word pairs in three subsequent runs and had to identify pairs that were repeated within runs. Age group differences in the veridical recognition of repeated word pairs across the three runs were not reliable. False recognition of lure pairs increased across runs in older adults only, indicating increasing interference from previous encounters of the word pairs. Both age groups showed a decrease in false recognition of rearranged pairs across runs, but this decrease was more pronounced in younger adults. Older adults were more confident than younger adults when falsely recognising both lure and rearranged word pairs, indicating that overall memory-confidence calibration is less accurate in old age relative to early adulthood. Taken together, our results suggest that age-related decline in memory monitoring, in interaction with binding deficits, contribute to age differences in false memory for highly familiar events.

Keywords: Memory; Ageing; Confidence; Monitoring; Binding.

Episodic memory performance declines in late adulthood (Kausler, 1994; Rönnlund, Nyberg, Bäckman, & Nilsson, 2005). In particular, older adults have difficulties remembering the sources of different events; for example, where or when something happened in the past (Chalfonte & Johnson, 1996; Spencer & Raz, 1995). Several studies have further revealed that older adults show a higher tendency to falsely remember things that in fact did not take place in the past (e.g., Dodson, Koutstaal, & Schacter, 2000; Dywan & Jacoby, 1990; Jennings & Jacoby, 1997; Koutstaal & Schacter, 1997; McCabe, Roediger, McDaniel, & Balota, 2009; Schacter, Koutstaal, & Norman, 1997; Shing, Werkle-Bergner, Li, & Lindenberger, 2008). Age-related deficits in veridical memory are relatively well documented (Kausler, 1994; Verhaeghen, Marcoen, & Goossens, 1993; Zacks, Hasher, & Li, 2000), but adult age differences in false memories have not yet received the same degree of attention. Examining the mechanisms underlying false memory in old age would...
strengthen our understanding of age-related declines in episodic memory and help to further delineate its constructive nature (e.g., Schacter, Norman, & Koutstaal, 1998).

Pioneering studies on false memory used the Deese-Roediger-McDermott paradigm (DRM; Roediger & McDermott, 1995) to illustrate that under certain conditions participants show increased propensity to make false-memory responses. In this paradigm participants view a series of words that converge to a common non-presented theme (e.g., bed, tired, pillow, rest). After studying these items participants tend to falsely remember the non-presented theme sleep (Roediger & McDermott, 1995), a tendency that is magnified in older compared to younger adults (Norman & Schacter, 1997; Tun, Wingfield, Rosen, & Blanchard, 1998; but see Gallo & Roediger, 2003). This age difference has been attributed to older adults’ increased reliance on memory for “gist” information (Brainerd & Reyna, 1998) and deficits in source memory (Dodson & Schacter, 2002; Johnson, Hashtroudi, & Lindsay, 1993). Interestingly, several studies (Kensinger & Schacter, 1999; Watson, McDermott, & Balota, 2004; but see Budson, Daffner, Desikan, & Schacter, 2000) showed that over several repeated study-test trials in which the same DRM lists were presented, only younger but not older adults showed a decrease in recalling non-presented theme words. This result suggests that while younger adults can make use of the study-test repetition to better verify that words were actually presented during the study phase, older adults have difficulties in building up veridical representations of the actually studied items (Kensinger & Schacter, 1999). At the same time, as both studied and self-generated theme words become more familiar across repetitions, older adults may be more likely to base their judgement on the increasing strength of familiarity signals, possibly reflecting decreased control over familiarity responses (Gallo & Roediger, 2003; Watson et al., 2004).

In line with these considerations, recent studies have found that older adults’ propensity for false recognition is not restricted to the DRM paradigm, but also present in tasks that manipulate the familiarity of studied material (Jacoby, 1999; Jacoby, Bishara, Hessels, & Toth, 2005; Jennings & Jacoby, 1997; Pierce, Sullivan, Schacter, & Budson, 2005) or in associative recognition paradigms (Castel & Craik, 2003; Cohn, Emrich, & Moscovitch, 2008). For example, Old and Naveh-Benjamin (2008) tested memory for people and their performance of everyday actions in younger and older adults. They found larger age-related decline for associative than for item memory, in agreement with the associative deficit hypothesis of ageing (Naveh-Benjamin, 2000). According to this hypothesis, an important cause of poor episodic memory performance in older adults is their deficiency in creating and retrieving associations between single information units, including associations between different items, item and its context or item features (Naveh-Benjamin, Brav, & Levy, 2007; Naveh-Benjamin, Guez, Kilb, & Reedy, 2004; Naveh-Benjamin, Hussain, Guez, & Bar-On, 2003). However, when separately examining hits (i.e., correct response to studied pairs) and false alarms to rearranged pairs (i.e., wrongly accepting a recombined pair in which both of the presented items were studied but not together), the authors found that age differences in associative memory were mainly driven by age differences in false alarms. In another study, Shing, Werkle-Bergner, and Lindenberger (2008) examined associative recognition before and after extensive training on an elaborative memory strategy with a lifespan sample. The authors found no differences between older adults and the remaining age groups with respect to hits. However, older adults showed consistently more false alarms to rearranged pairs, particularly in an experimental condition with study materials of high associative demand. Moreover, older adults’ tendency to commit more false alarms was not eliminated by the instruction and extensive practice of the elaborative memory strategy. Thus, while ageing is associated with general reductions in correct recognition of studied information, the magnitude of age-related increase in false recognition seems to be greater (Bender, Naveh-Benjamin, & Raz, 2010).

Older adults are not only more likely to commit memory errors, but also to express higher subjective confidence in their false memories than younger adults. For example, older adults in the study by Shing and colleagues (2008) were disproportionately more likely than the other age groups to indicate high confidence (on a 3-point scale) following a false alarm on a rearranged pair (see Shing, Werkle-Bergner, & Lindenberger, 2009). Similarly, using an eyewitness misinformation paradigm, Dodson and Krueger (2006) found that older adults provide high-confidence ratings to a greater proportion of falsely recognised items.
than younger adults. High-confidence errors in old age may indicate that the ability to calibrate subjective confidence to memory accuracy (e.g., Jacoby & Rhodes, 2006) decreases with advancing adult age. Accordingly, Hertzog and Touron (2011) found substantial age differences in the accuracy of confidence judgements in an incidental associative learning task with interspersed associative recognition memory probes. This age effect was especially pronounced for rearranged pairs, presumably due to age-related deficits in monitoring of associative retrieval. However, age differences in confidence calibration are not observed under all circumstances. For example, while Kelley and Sahakyan (2003) found greater confidence miscalibration in older adults than in younger adults for misleading lures in a cued-recall procedure, Hines, Touron, and Hertzog (2009) detected only small age differences in confidence judgement calibration during an associative recognition task.

Taken together, the available literature on confidence judgement calibration suggests substantial variability in age-related effects depending on the task and its characteristics, underscoring the need to examine the specific contexts and mechanisms under which high-confidence memory errors occur in old age.

**MECHANISMS OF FALSE MEMORY IN AGING**

A variety of mechanisms may contribute to older adults’ increased tendency to falsely endorse familiar information. Foremost among them, available evidence indicates that recollection is more strongly affected by senescent changes than familiarity (Jacoby & Hay, 1998; Light, Prull, La Voie, & Healy, 2000). Given that correct rejection of novel configurations among familiar stimuli requires recollection of the specific context of the original pairs, age-related increases in memory errors during associative recognition might result from deficits in recollection and increased reliance on familiarity during the retrieval of episodic memories (Castel & Craik, 2003; Old & Naveh-Benjamin, 2008; Shing et al., 2008). Furthermore, strategic processes, in particular memory monitoring engaged to evaluate the source of retrieved information (Mitchell & Johnson, 2009), might play an important role in the efficacy with which false memories are rejected (Schacter & Dodson, 2001; Schacter, Koutstaal, Norman, 1997; Schacter et al., 1998). Memory monitoring is a central process in established models of memory retrieval (Burgess & Shallice, 1996; Mitchell & Johnson, 2009; Moscovitch & Winocur, 2002). In particular, whenever retrieval cues are very specific, monitoring processes need to be instantiated to a greater degree (Johnson, Hashtroudi, & Lindsay, 1993; Moscovitch & Winocur, 2002). Hence the need for memory monitoring is increased when strong familiarity signals need to be overcome or put on hold in order to engage in recollecting specific information.

In metamemory frameworks, monitoring processes play an important role for subjective confidence calibration (Koriat & Goldsmith, 1996; Nelson & Narrens, 1990). An effective monitoring mechanism is needed to differentiate between correct and incorrect responses so that the subjective confidence related to a candidate response is matched to its objective correctness (Koriat & Goldsmith, 1996). A few studies have examined the effectiveness of metamemory monitoring and control by comparing confidence judgements in a forced cued-recall test (where participants are required to give an answer followed by a confidence rating) to a free cued-recall test where participants could volunteer to withhold a response (Koriat & Goldsmith, 1996). However, the results with regard to the pattern of age differences are contradictory. While Pansky, Goldsmith, Koriat, and Pearlman-Avnion (2009) reported age-related deficits in the decision to withhold a response, Kelley and Sahakyan (2003) demonstrated poorer confidence calibration in the forced cued-recall test, but did not find age differences in volunteering to withhold a response. Using feeling-of-knowing judgements that capture episodic memory monitoring after cued recall has failed, some studies reported decline in old age (e.g., Souchay, Moulin, Clarys, Taconnat, & Isingrini, 2007), whereas others did not find such developmental effects after equating encoding quality and memory performance between younger and older adults (Hertzog, Sinclair, & Dunlosky, 2010).

Neurally, patients with frontal lobe (PFC) lesions do not show as strong memory deficits as patients with lesions to the medio-temporal lobe (MTL), but they are more likely than those patients to falsely claim having experienced events that did not happen in the past with high subjective confidence (Baldo & Shimamura, 2002; Schacter & Slotnick, 2004; Schnider, 2003). Thus
the propensity to confidently endorse novel events as experienced before is likely to reflect failures in executive and control functions implemented by frontal regions (Gilboa et al., 2006). In ageing, these brain regions are among the areas that undergo the strongest changes (Raz, Ghisletta, Rodrigue, Kennedy, & Lindenberger, 2010; Raz et al., 2005). Thus age-related increases in the propensity to false recognition with high subjective confidence may be related to deficits in the ability to carry out executive and monitoring operations confined to the PFC (Jacoby & Rhodes, 2006; Schacter et al., 1997). Correspondingly, older adults are more prone to proactive interference than younger adults (Hay & Jacoby, 1999; Kliegl & Lindenberger, 1993; May, Hasher, Kane, & Roediger, 1999) and show deficits in the ability to inhibit irrelevant information (Hasher, Zacks, & May, 1999). Furthermore, lower levels of executive functioning in old age have been associated with increased false memory in the DRM paradigm (Butler, McDaniel, Dornburg, Price, & Roediger, 2004; Plancher, Guyard, Nicolas, & Piolino, 2009), as well as with declines in episodic feeling-of-knowing judgements (Perrotin, Tournelle, & Isingrini, 2008) and volunteering to withhold a response in a free cued-recall test (Pansky et al., 2009). However, no study to date examined age differences in the ability to distinguish currently relevant from irrelevant memories in episodic memory. Moreover, it is still unclear to what extent deficits in executive functioning are related to older adults’ deficits in confidence calibration.

Associative binding deficits during encoding of episodic information might also contribute to the increased false memory in old age (Castel & Craik, 2003; Naveh-Benjamin, 2000). Neurocomputational models suggest that less distinctive encoding of new information (Wilson, Gallagher, Eichenbaum, & Tanila, 2006) results in less distinctive memory representations (Li, Naveh-Benjamin, & Lindenberger, 2005), which in turn may increase the likelihood of false recognitions. A recent account suggests that high-confidence memory errors in old age may also be ascribed to associative deficits. For example, Dodson and colleagues (Dodson, Bawa, & Krueger, 2007) used a source-monitoring task to investigate high-confidence errors in older adults. Participants heard sentences spoken by a male or a female voice and were later asked to report the speaker for recognised sentences, followed by a confidence rating after each response. The results indicated that even if memory for source was equated between younger and older adults, older participants committed more high-confidence errors. By contrast there were no age differences in confidence judgements for item memory. Hence, Dodson and colleagues (2007) suggested that older adults are especially prone to making high-confidence errors only when recollection of specific details is necessary (i.e., the misrecollection account). However, the exact contribution of binding deficits to age differences in false memory remains largely unexplored.

In sum, the existing literature suggests that control processes in episodic memory, especially monitoring of retrieved information in service of current goals and demands, might play a major role in older adults’ propensity to remember past events that actually did not happen, in particular when familiarity of memory representations is high. In addition, associative binding deficits may lead to weaker and less distinctive memory representations that may also contribute to older adults’ difficulties in rejecting familiar information. Taken together, the extent to which adult age differences in high-confidence memory errors are related to failures of monitoring mechanisms, binding deficits, or both, remains to be explored.

RESEARCH HYPOTHESES

Old age is related to an increased propensity to make false-memory errors as well as to increased confidence in those false memories. However, not much is known about the different mechanisms contributing to these phenomena. The goal of the present study was to examine the contribution of deficits in strategic monitoring and associative binding to false memory in old age and the tendency to commit high-confidence errors.

We used a modified version of the continuous recognition task (CR task; see Figure 1) by Schnider and colleagues that was developed to investigate the deficits of confabulating patients (Schnider et al., 1996). In this task participants are presented with the same series of stimuli (e.g., object pictures) across repeated runs and have to indicate stimuli reoccurrence within each run by ignoring having seen the stimuli from previous runs (Schnider, 2003). The advantage of using repeating runs is that it allows the investigation of the ability to withstand interference from irrelevant but highly familiar memory traces.
Confabulating patients with specific prefrontal brain damage show a greater increase in false-positive responses as they endorse having already seen the stimuli that in fact only appeared for the first time in the actual run (Schnider & Ptak, 1999). We created an additional task condition that supposedly placed high demands on associative binding by presenting recombinations of stimuli across runs (Figure 1).

We assessed measures of binding (Face–Name task; Naveh-Benjamin et al., 2004) and executive-frontal functioning (Wisconsin Card Sorting Task; Heaton, Chelune, Talley, Kay, & Curtis, 1993) to help relate age differences in false memory in the CR task more directly to the contrasting mechanisms. We did not assume that these covariate tasks purely measure binding or executive functioning, respectively. Rather we assumed that the Face–Name task would rely more strongly on binding than on strategic processes, whereas the reverse would be true for the WCST.

Three main hypotheses were tested in the present study. First, we expected older adults to commit more false-memory errors than younger adults when distinguishing currently relevant from currently irrelevant information, reflecting their inability to withstand interference from previous exposures to the same events, due to age-related declines in memory monitoring (Gilboa et al., 2006; Schacter et al., 1997). Second, we hypothesised that deficits in binding process would contribute to the age-related increase in false memory. We expected that repeated exposure to the same word pairs across runs of the task would strengthen memory representations for these word pairs. If strengthening memory representations has beneficial effects on false memory, we expected that multiple repetitions of the original word pairs would lead to reduced false recognition of novel recombinations of familiar pairs across runs, in line with previous findings with the associative recognition paradigm in younger adults (Light, Patterson, Chung, & Healy, 2004). We hypothesised that the beneficial effect of repetition would be less pronounced in older adults compared to younger adults, due to associative binding deficits in old age (Jacoby, 1999; Light et al., 2004; Naveh-Benjamin, 2000). Finally, we tested whether older adults would be more likely to commit memory errors with high subjective confidence than younger adults. In line with the misrecollection account (Dodson et al., 2007) we expected that older adults would be particularly prone to commit high-confidence errors when recollection of specific details is needed, as in the case of rejecting novel recombinations of familiar word pairs. In this case we expected that individual differences in

**Figure 1.** Structure of the repeated continuous recognition task (CR task). FA = false alarm, CR = correct rejection.
high-confidence false recognition would be uniquely related to a measure of associative binding (i.e., Face–Name task). In contrast, if strategic monitoring deficits also contribute to failures of confidence calibration with age (Koriat & Goldsmith, 1996), older adults’ high-confidence memory errors would not be limited to the rejection of novel recombinations of familiar word pairs and would be related to an independent measure of executive functioning (i.e., WCST).

**METHOD**

**Participants**

Participants in this study were 28 younger adults aged 21 to 28 years ($M = 24.92, SD = 1.84$), and 30 older adults aged 69 to 75 years ($M = 72.28, SD = 2.01$). Older adults were community dwellers who reported no major health problems. They were screened for cognitive impairment with the Mini Mental State Examination (MMSE, $M = 29.6, SD = 0.73$, Range $= 28 – 30$). Both age groups were equated on levels of formal education ($p > .10$). The results on measures of fluid (Digit Symbol; cf. Wechsler, 1955) and crystallised intelligence (verbal knowledge; cf. Lehrl, 1977) revealed the typical age-related increase in verbal knowledge, $t(56) = 7.02, p < .05$, as well as age-related decline in processing speed, $t(56) = 3.53, p < .05$ (see Table 1).

**Materials**

A pool of 414 highly imaginable words was selected to form 207 unrelated word pairs for the CR task. A set of 132 pairs was randomly selected to constitute the task set for younger adults. Out of the task set for younger adults, 102 word pairs were randomly selected to constitute the task set of older adults. List length differed between age groups to render the task manageable for both age groups. The remaining 75 word pairs were used as new pairs in the CR task for both younger and older adults.

**Procedure**

Each participant was tested in three consecutive sessions distributed across 2–3 weeks. The first and the last sessions were used to assess a battery of behavioural covariates tasks. In the second session the participants were first instructed and practised the CR task (Schnider et al., 1996). Then, while lying in the MRI scanner, each participant performed the familiarisation task immediately followed by three consecutive runs of the CR task (see Fandakova, Lindenberger, & Shing, 2012, for neuroimaging data with this paradigm).

*Familiarisation task.* During the familiarisation task each participant viewed all word pairs that were later used for the main task (132 pairs for younger adults, 102 pairs for older adults). Each pair was presented in black against a white background for 3 seconds, followed by a fixation cross for 500 ms. The participants were instructed to indicate by button press whether none, one, or both of the words depicted a living object. The familiarisation task was introduced in order to familiarise the participants with the lure pairs that were later used in the main task. Thus, in the first run of the subsequent CR task, the lure pairs were already familiar to the participants, making the run more comparable to the remaining runs of the task in terms of monitoring demands (instead of relying solely on detecting word pair repetitions among novel word pairs).

*Modified continuous recognition task (CR task).* The CR task started immediately after the familiarisation task and consisted of three runs (see Figure 1). The same set of stimuli was presented in all runs with different quasi-randomised orders. The number of trials varied between younger and older adults. In each run the task set of word pairs was presented once (termed *lure pairs*), resulting in 132 trials for younger adults and 102 pairs for older adults. Due to the

<table>
<thead>
<tr>
<th>Measure</th>
<th>Younger adults</th>
<th>Older adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N$</td>
<td>28</td>
<td>30</td>
</tr>
<tr>
<td>Age</td>
<td>25.92 (1.84)</td>
<td>72.28 (2.01)</td>
</tr>
<tr>
<td>Spot-a-word</td>
<td>25.54 (4.05)</td>
<td>28.87 (3.9)</td>
</tr>
<tr>
<td>Digit Symbol</td>
<td>68.06 (10.18)</td>
<td>48.83 (10.61)</td>
</tr>
<tr>
<td>Face–name task (Face Recognition)</td>
<td>.48 (.18)</td>
<td>.43 (.19)</td>
</tr>
<tr>
<td>Face–name task (Associative Recognition)</td>
<td>.04 (.17)</td>
<td>.15 (.13)</td>
</tr>
<tr>
<td>WCST</td>
<td>10.76 (4.83)</td>
<td>19.83 (8.17)</td>
</tr>
</tbody>
</table>
familiarisation task, the lure pairs were already familiar to the participants in run 1. One third of the lure pairs were selected to reappear once throughout the run (termed target pairs), resulting in 44 pairs for younger adults and 34 pairs for older adults. The lag between the first presentation of a lure pair and its corresponding reoccurrence as target varied between 13 and 49 trials ($M = 31$ trials). The target pairs were never repeated within and across runs. Another third of the lure pairs (44 for younger adults, 34 pairs for older adults) were selected to reappear as recombined pairs, that is, the left word of a given lure pair was presented with the right word from another lure pair (termed rearranged pairs). The rearranged pairs were unique; that is, although the single words were familiar to the participants, the recombination of words was never repeated within and across runs. Each word was also selected only once for recombination across the three runs. Finally, for both age groups 25 word pairs were used in each run that were not included in the familiarisation task and were therefore never seen by the participants before or during the task (termed new pairs). The order of presentation of the lure, target, rearranged, and new pairs was randomised within runs. Each word pair was presented for 3 seconds. Across the three runs of the task the participants were instructed to pair words as presented for the first time were used in each run that were not included in the familiarisation task and were therefore never seen by the participants before or during the task (termed new pairs). The order of presentation of the lure, target, rearranged, and new pairs was randomised within runs. Each word pair was presented for 3 seconds. Across the three runs of the task the participants were instructed to indicate reoccurrences of the same word pair within the ongoing run by using four different buttons—“sure new”, “unsure new”, “unsure old”, and “sure old”. Participants were told to be as accurate as possible in their response and had up to 3 seconds to respond on each trial.

We expected that as lure pairs are presented in each run of the CR task, correctly rejecting those pairs as presented for the first time would put higher demand on monitoring processes with increasing run. We hypothesised that older adults’ false alarms to lure pairs would increase across runs due to deficits in the ability to monitor currently relevant memories and withstand interference from previous presentations of the lure pairs. In line with the associative binding deficit in old age, we expected that older adults would benefit less than younger adults from the across-run repetitions of the lure pairs to build distinctive representations of these word pairs and use them to correctly reject novel rearrangements. Hence, the decrease in false alarm rates for rearranged pairs across runs would be less expressed in older adults than in younger adults.

**Face–Name task.** The materials and procedure of the face–name task were similar to the task reported by Naveh-Benjamin and colleagues (2009), Experiment 2. Participants studied a total of 48 face–name pairs for impending memory tests while judging how well the face and the name fitted together. Immediately after study, participants performed recognition tests for the face–name associations followed by a test for the faces only. The data of one older adult was excluded due to floor performance on both the associative and face test ($Pr < .01$). We found no age differences in the face task, $t(55) = .94, p > .05$, but lower performance of older adults on the associative test, $t(55) = 6.37, p < .05$ (see Table 1). For further analysis only the performance of the associative test was analysed, because this test required binding of the face and the name into a cohesive representation, and therefore reflects individual differences in binding (Naveh-Benjamin et al., 2004).

**Wisconsin card sorting task (WCST).** We used a computerised version of the standard WCST (Heaton et al., 1993; see Nagel et al., 2008). Participants were instructed to sort a response card presented at the bottom of a computer screen to one of four key cards presented in the upper part of the screen. Following each response, a feedback of the correct answer was briefly displayed on the screen. Whenever a person reached 10 correct consecutive sorting orders, the sorting rule changed without warning. The WCST automatically ended after six categories were completed or after all 128 cards were sorted. Performance was evaluated by applying the standard WCST scoring (Heaton et al., 1993). Due to technical problems the data of four younger adults could not be analysed. For the present analysis the percentage of perseverative errors was used as an index of performance. In line with previous studies (Nagel et al., 2008) we found evidence for age-related increase in the percentage of perseverative errors, $t(52) = 4.80, p < .05$ (see Table 1). Perseverative errors in WCST are a sensitive marker of age differences in executive functioning (Rhodes, 2004) and are related to adult age differences in the structural integrity of the frontal lobes (Gunning-Dixon & Raz, 2003).

WCST and Face–Name performance were not correlated in either age group ($r = -.14$ and $r = -.24$ for younger and older adults, respectively, $p > .05$). This finding supports our assumption that the covariate tasks measured relatively different processes.
RESULTS

Overview of analysis

Proportions of hit ("old" response to target pairs) minus false alarm (FA, "old" response to lure and rearranged pairs, respectively) were computed for each participant and each run of the CR task (Pr-values; Snodgrass & Corwin, 1988). For analysis of confidence judgements, the proportions of high-confidence ratings from hit and false alarm responses were computed, respectively. No outliers were found for any of the measures at \( p < .001 \) (two-tailed test). Skewness and kurtosis of each variable were within acceptable ranges (Kline, 2005; Tabachnik & Fidell, 2006).

The main statistical analyses were conducted using Mplus (Version 5.2; Muthén & Muthén, 1998–2010). Given our interest in testing age differences in initial performance as well as change in performance across runs, we chose to utilise linear growth models in the multiple-group framework of structural equation modelling (Bollen & Curran, 2006; McArdle, 2005) for analysing the data. A linear growth model approach was preferred over an analysis of variance (ANOVA) approach, because it is more amenable for the analysis of relationships between initial performance, change and independent covariate measures that were of main interest in the present study (Bollen & Curran, 2006). However, the overall pattern of results reported in this section was replicated in a control set of analyses using an ANOVA approach. The loadings of all observed variables on the intercept were specified to 1. Linear change was incorporated by specifying loadings of the observed variables on the slope to 0 for run 1, 1 for run 2, and 2 for run 3. Centring change in this way allows the intercept to represent estimated initial level of performance, or the expected value of the outcome variable when time is 0 (i.e., at run 1), while the slope represents the expected amount of change in performance per time point; that is, with increasing runs. As the assessments were equally and identically spaced for all individuals, the models reported in the current study are equivalent to corresponding model specifications in the framework of multilevel linear models (Bauer, 2003; Curran, 2003; Mehta & Neale, 2005; for a comparison of the two approaches, see Ghisletta & Lindenberger, 2004). All models were also tested for quadratic trends by specifying a saturated model with a combination of linear and quadratic trends. Unless otherwise reported, the quadratic trend was not significant for the reported outcome variables in any of the age groups.

All models were estimated via maximum likelihood. Models were considered to have acceptable fit with \( \chi^2/df \) of less than 2, comparative fit index of more than .95 (CFI; Bentler, 1990), as well as a root-mean-square error of approximation of less than .08 that included .05 within its 90% confidence interval (RMSEA; Browne & Cudeck, 1993; Steiger, 1990). To compare age differences, age-equality constraints were imposed on the parameters of interest of the previously freely estimated model. The free and restricted models are nested models, which can be formally compared by testing the significance of the difference with \( \chi^2 \), with the degree of freedom (df) equal to the difference in the number of free parameters between the two models. An alpha level of .05 was used for all statistical tests.

Prior to the main analysis we examined performance on new pairs within each run of the CR task as a manipulation check that the participants understood and followed the instructions of the task. As the new pairs were always novel, we expected that participants in either age group would easily detect these stimuli. As expected, the overall number of FA to new pairs was low (Figure 2, Panel B). A two-way analysis of variance (ANOVA) revealed no significant effect of run, \( F(2, 112) = 2.54, p > .05 \), age, \( F(1, 56) = .01, p > .05 \), or run \( \times \) age interaction, \( F(2, 112) = .41, p > .05 \).

Age differences in Pr-values

First, we examined the Pr-values with FA for lure pairs (Figure 1, Panel A). Here we expected that participants would increasingly endorse lure pairs as targets due to increasing interference from previous presentations of the word pairs. Moreover, we expected that due to age-related deficits in memory monitoring processes this tendency would be more strongly expressed in the older adults’ group. The freely estimated model showed excellent fit, \( \chi^2 = 6.41, df = 6, CFI = .995 \) and RMSEA = 0.05 [90% CI .00 – .25]. In the subsequent model comparisons older adults showed a trend for lower initial performance compared to younger adults, \( \Delta \chi^2 = 3.46, df = 1, p = .06 \), \( d = 0.55 \). Older, but not younger, adults
showed a significant decrease in performance across runs of the task (Table 2).

However, the difference between the two groups (i.e., the group by time interaction) only approached statistical significance, $\chi^2 = 3.39$, $df = 1$, $p = .07$, $d = .60$.

Next we compared the Pr-values computed with FA for rearranged pairs (Figure 1, Panel B). Here we expected that seeing the same set of lure pairs across the three runs of the task would strengthen their memory representations, making it increasingly easier to reject the rearranged pairs. However, given previous findings of age-related deficits in associative binding, we expected that younger adults would benefit more from the repeated presentation of the lure pairs than the older adults. In addition to the linear slope, the quadratic slope was also significant in the group of younger adults (Table 2), reflecting the levelling off of performance between runs 2 and 3 in this group, $\chi^2 = .00$, $df = 6$, CFI = 1.00 and RMSEA = .00 [90% CI .00 – .00]. In the subsequent model comparisons older adults showed lower performance in the first run of the task compared to younger adults, $\Delta \chi^2 = 9.26$, $df = 1$, $p < .05$, $d = .86$ as well as a significantly smaller linear increase in performance across runs of the task than younger adults (i.e., a group by time interaction), $\Delta \chi^2 = 5.15$, $df = 1$, $p < .05$, $d = .61$. There were no age differences in the mean of the quadratic trend, $\Delta \chi^2 = 2.73$, $df = 1$, $p > .05$, $d = .44$.

### Age differences in hits and FA

Mean hits and FA rates for lure and rearranged pairs are presented in Figure 3. The mean estimated intercept and slope for each model are presented in Table 2.

**TABLE 2**

Mean model estimates for memory performance

<table>
<thead>
<tr>
<th>Measures</th>
<th>Younger adults</th>
<th></th>
<th>Older adults</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intercept</td>
<td>Linear slope</td>
<td>Intercept</td>
<td>Linear slope</td>
</tr>
<tr>
<td>Pr-values</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>with lure FA</td>
<td>.599 (.03)*</td>
<td>-.012 (.02)</td>
<td>.526 (.02)*</td>
<td>-.052 (.02)*</td>
</tr>
<tr>
<td>with rearranged FA</td>
<td>.314 (.05)*</td>
<td>.221 (.05)*</td>
<td>.148 (.02)*</td>
<td>.061 (.05)*</td>
</tr>
<tr>
<td></td>
<td>-.053 (.02)*</td>
<td>(quadratic)</td>
<td>.001 (.02)</td>
<td>(quadratic)</td>
</tr>
<tr>
<td>Hit &amp; FA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hits</td>
<td>.773 (.02)*</td>
<td>.003 (.01)</td>
<td>.805 (.03)*</td>
<td>.001 (.02)</td>
</tr>
<tr>
<td>Lure FA</td>
<td>.165 (.02)*</td>
<td>.016 (.01)</td>
<td>.267 (.02)*</td>
<td>.054 (.02)*</td>
</tr>
<tr>
<td>Rearranged FA</td>
<td>.433 (.04)*</td>
<td>-.115 (.01)*</td>
<td>.653 (.03)*</td>
<td>-.064 (.02)*</td>
</tr>
</tbody>
</table>

Estimates with significant age differences are in bold. * Estimate is significant at $p < .05$. 

Figure 2. Pr-values across runs of the CR task. (A) Pr-values computed with FA to lure pairs, (B) Pr-values computed with FA to rearranged pairs. Error bars represent standard error of the mean.
The model for hits fitted the data well, $\chi^2 = 6.58$, df = 6, CFI = .99 and RMSEA = 0.06 [90% CI .00 – .25]. Both age groups showed similar initial hit rates, which did not change across runs of the task, $\Delta \chi^2 = 0.80$, df = 1, $p > .05$, $d = 0.27$ and $\Delta \chi^2 = 0.01$, df = 1, $p > .05$, $d = 0.04$, respectively (Table 2).

The model for FA to lure pairs showed excellent fit to the data, $\chi^2 = 2.5$, df = 6, CFI = 1.00 and RMSEA = 0.00 [90% CI .00 – .12]. In the subsequent model comparisons we found that older adults committed more FA to lure pairs than did younger adults in the first run of the task, $\Delta \chi^2 = 10.38$, df = 1, $p < .05$, $d = 0.95$. Furthermore, older adults showed a greater increase in FA rates across runs than younger adults (i.e., a group by time interaction), $\Delta \chi^2 = 7.06$, df = 1, $p < .05$, $d = 0.64$.

With regard to rearranged pairs, the initial freely estimated linear model resulted in negative elements in the covariance matrix in younger adults despite acceptable fit to the data, $\chi^2 = 9.95$, df = 6, CFI = .96 and RMSEA = 0.15 [90% CI .00 – .31]. In this model the slope variance of younger adults did not differ reliably from zero. Hence the slope variance in younger adults and the slope-intercept covariance in both age groups were fixed to zero (Dillon, Kumar, & Mulani, 1987; Ghisletta & Lindenberger, 2005). The resulting model showed acceptable fit to the data, $\chi^2 = 11.59$, df = 9, CFI = .98 and RMSEA = 0.10 [90% CI .00 – .24] and did not differ from the initial unrestricted model, $\Delta \chi^2 = 1.64$, df = 3, $p > .05$. In the subsequent model comparisons it was found that older adults committed more FA to rearranged pairs at the intercept than younger adults, $\Delta \chi^2 = 17.55$, df = 1, $p < .05$, $d = 1.34$. Importantly, both groups had a significant negative slope, indicating decrease in the amount of FA to rearranged pairs across runs. However, the increase was more pronounced for younger than for older adults, as indicated by a steeper slope in younger adults (i.e., a group by time interaction), $\Delta \chi^2 = 5.17$, df = 1, $p < .05$, $d = 0.93$.

Finally, we included the associative test from the face–name task as a measure of binding and the WCST task as a measure of executive functioning in the models for hits, lure pair FA, and rearranged pair FA reported above (Table 4).
There were no relations between the covariate measures and the estimates for hits. For lure pair FA, in younger adults the intercept was positively related to WCST perseverative errors, \( r = .35, p < .05 \). For rearranged pair FA, in younger adults the intercept was positively related to WCST perseverative errors, \( r = .50, p < .05 \), and negatively related to performance in the face-name task, \( r = -.40, p < .05 \). Due to the fixed variance of the slope in younger adults, we could not estimate any relationships of the slope with the covariates. For older adults, we found a significant negative correlation between the slope for rearranged pair FA and the face-name task, \( r = -.55, p < .05 \), suggesting that those participants who showed higher performance on the face-name task also showed more decrease in the FA rate to rearranged pairs across runs of the task. Bootstrapping analyses confirmed those results, as the 95% confidence intervals for the reported correlations did not include zero.

### Age differences in confidence judgements

Mean proportions of “sure” response for hits, FA to lure pairs, and FA to rearranged pairs are presented in Figure 4. The mean estimated intercept and slope for each model are presented in Table 3.

With this analysis we sought to examine to what extent misrecollection, memory monitoring deficits, or both underlie older adults’ high-confidence false memories. We expected that if older adults commit high-confidence errors for rearranged pairs only, this would speak for the misrecollection account, as this condition places high demand on binding. However, if additional deficits in memory monitoring contribute to older adults’ tendency to commit errors with high confidence, we expected to see this pattern of results for both the lure and rearranged pairs.

First, the model for high-confidence responses for FA to lure pairs showed acceptable fit, \( \chi^2 = 8.36, df = 6, CFI = .97 \) and \( RMSEA = .12 \) [90% CI \( .00 - .29 \)]. Initially older adults committed more high-confidence FA to lure pairs compared to younger adults, \( \Delta \chi^2 = 9.45, df = 1, p < .05, d = 0.94 \), and there were no age differences in change across the runs of the task (i.e., a group by time interaction), \( \Delta \chi^2 = 1.28, df = 1, p > .05, d = 0.40 \).

Next we examined high-confidence responses FA to rearranged pairs. The quadratic slope was significant in older adults, due to the slight pattern of inverted U-shape present in the data (\( \chi^2 = .00, df = 6, CFI = 1.00 \) and \( RMSEA = .00 \) [90% CI \( .00 - .00 \)]). Older adults exhibited higher subjective confidence in FA to rearranged pairs, \( \Delta \chi^2 = 6.71, df = 1, p < .05, d = 0.70 \). There was no difference in linear (\( \Delta \chi^2 = 0.80, df = 1, p > .05, d = 0.24 \)) or quadratic (\( \Delta \chi^2 = 2.64, df = 1, p > .05, d = 0.45 \)) change across runs.

To confirm that the age differences reported above were not driven by a general tendency of older adults to report high confidence, we examined age differences in the proportion of high-confidence hit responses. The model showed
acceptable fit, $\chi^2 = 9.41$, $df = 6$, $CFI = .97$ and $RMSEA = 0.14$ [90% CI .00 — .31]. Compared to older adults, younger adults reported higher confidence when making a hit, $\Delta \chi^2 = 4.78$, $df = 1$, $p < .05$, $d = 0.66$, as well as significantly greater decrease in confidence across runs, $\Delta \chi^2 = 6.99$, $df = 1$, $p < .05$, $d = 1.06$.

Similar to the performance data, we correlated the intercepts and the slopes of the models for confidence ratings described above with the face-name and WCST tasks (Table 4). We found that subjective confidence for FA to both lure and rearranged pairs was related to WCST perseverative errors in the older adults, $r = .37$, $p < .05$ and $r = .43$, $p < .05$, respectively. In younger adults high-confidence FA to lure pairs was also positively related to WCST perseverative errors, $r = .39$, $p < .05$. All statistical inferences were confirmed in bootstrapping analyses where the 95% confidence intervals for the reported correlations did not include zero. Finally we found no reliable associations between the covariates and high-confidence hits.

### DISCUSSION

#### Summary of findings

In the present study we examined age differences in false memories for highly familiar events. We hypothesised that, due to the age-related decrease in memory monitoring processes accompanied by deficits in associative binding, older adults would...
have more difficulties to withstand interference from previous exposure to the same events. Furthermore, we examined monitoring deficits and misrecollection as possible mechanisms underlying older adults’ increased susceptibility to commit memory errors with high subjective confidence.

In the modified CR task we found no age differences in hits and FA to new pairs. That is, both younger and older adults were equally able to identify the repetition of a pair within a run as well as to correctly reject pairs that were not seen before. With respect to FA to lure pairs, the intercept was higher in older adults, indicating that they committed more FA than younger adults. Moreover, older adults showed a significant increase in FA to lure pairs as familiarity of the pairs increased across runs of the task, indicating that control over episodic memory declines with age. With regard to rearranged pairs, we found that older adults committed more FA than younger adults. Across the three runs, both groups reduced their false responses to rearranged pairs, but the decrease was significantly more pronounced in younger than in older adults. Moreover, individual differences in initial performance of rearranged pairs were positively related to perseverative errors in the WCST and negatively related to performance in the face-name task in younger adults, indicating that better executive and binding functioning are related to committing less FA. In older adults, on the other hand, more decrease in FA across runs was related to higher performance in the face-name task, indicating that older adults with better associative binding ability also benefited more from the repetition of the original pairs to facilitate correct rejection of the rearranged pairs.

With respect to subjective confidence in memory judgements, younger adults exhibited greater subjective confidence when making a hit. In contrast, older adults showed higher subjective confidence when making FA to both lure and rearranged pairs. Committing errors with higher subjective confidence was positively related to lower performance on the WCST task in both age groups, possibly indicating the important role of memory monitoring mechanisms implemented by frontal functioning. Taken together, these results suggest that younger adults are more accurate than older adults when introspecting the validity of their memory representations.

**False memory in old age: The role of strategic and associative components**

The CR task used in the present study was a modified version of a continuous recognition task used to investigate monitoring deficits in confabulating patients (Schnider et al., 1996). According to the reality-monitoring account (Dalla Barba, Nedjam, & Dubois, 1999; Schnider, 2003; Schnider & Ptak, 1999), confabulating patients’ increased FA across runs results from a deficit in the ability to distinguish events that pertain to the present situation from events that have occurred at another time in the past. Furthermore, deficits in strategic memory processes might also contribute to confabulators’ performance (i.e., the strategic retrieval account; Gilboa et al., 2006; Moscovitch & Melo, 1997). Namely, by presenting the same stimuli across runs, their familiarity as well as the interference from previous presentations of the stimuli increases. As direct retrieval cues are presented during the task, the demands placed on monitoring and evaluation processes are high (Burgess & Shallice, 1996; Moscovitch & Winocur, 2002).

The present modified version of the CR task differs from the original version in a number of ways. Specifically, the modified version had a larger number of stimuli, shorter time intervals between runs, and consisted of word pairs instead of pictures, which allowed us to introduce rearranged pairs as a new type of lures. Under these supposedly more difficult conditions, we observed that older adults’ false recognition increased across runs of the task, in line with an age-related decline in source monitoring, in general (Chalfonte & Johnson, 1996; Spencer & Raz, 1995), and in monitoring and evaluation processes, in particular. Such monitoring deficits may result in less effective memory control by means of inhibiting irrelevant information or enhancing relevant information (Gazzaley, Cooney, Rissman, & D’Esposito, 2005).

Taken together, the present findings indicate that memory monitoring processes become less efficient in old age, thus magnifying the effects of age-related decline in recollection and leading to heightened false memory (Jennings & Jacoby, 1997; Schmiedek, Li, & Lindenberger, 2009).

In line with previous studies of associative recognition, we found that initially older adults committed more FA to rearranged pairs than younger adults (Castel & Craik, 2003; Old &
Naveh-Benjamin, 2008). Given that individual words were familiar to the participants, the correct rejection of the rearranged pairs required recollection of the binding from the original pairs, which points to a recall-to-reject monitoring process that declines in old age (Gallo, Kensinger, & Schacter, 2006; Jacoby, 1999; Jones & Jacoby, 2005; Multhaup, 1995).

Furthermore, the goal of the present study was to examine the extent to which binding deficits (Naveh-Benjamin, 2000) also contribute to false memories in old age. We expected that under conditions of optimal inter-item binding, seeing the same set of lure pairs across the three runs of the task would strengthen a coherent memory representation of the two items that constitute the pair. Hence, to the extent that binding deficits also contribute to false alarms, older adults would not be able to benefit to the same degree as younger adults from the repeated presentation of the lure pairs (Jacoby, 1999; Light et al., 2004). In agreement with this prediction, we found that the decrease in FA to rearranged pairs across runs of the task was almost twice as large in younger adults as in older adults. The observed associations between rearranged pair FA and the covariate tasks strengthen this interpretation. For younger adults, individual differences in initial rearranged pair FA rate were related to performance in the WCST as well as the face–name task. These findings indicate that, in line with our hypotheses, to reject rearranged pairs entails both efficient memory monitoring and binding. For older adults, we found that the ability to benefit from repeated presentation of the pairs was related to the face–name task, suggesting that older adults with better preserved binding benefited more from the pair repetition, which may support building up coherent memory representations that in turn facilitate recall-to-reject processes.

Finally, while younger adults exhibited greater subjective confidence when making a hit, older adults showed higher subjective confidence when committing FA to both lure and rearranged pairs. In both age groups the amount of high-confidence errors was positively related to more WCST perseverative errors. Taken together, these results indicate that age-related deficits in executive functioning contribute to older adults’ difficulties to calibrate subjective confidence according to memory performance (Baldo & Shimamura, 2002). Interestingly, in younger adults both the overall FA rate and the high-confidence errors to rearranged pairs decreased significantly across runs of the task (Figures 2 and 3). In contrast, while for older adults the slope for rearranged FA was significant, indicating that the amount of false recognitions for those pairs decreased across runs (Figure 2), the level of high-confidence memory errors remained relatively constant across runs (Figure 3). Taken together, these findings offer the intriguing possibility of a differential effect of repeated lure pair presentation in ageing, such that for older adults strengthening memory representations may have a beneficial effect on false recognition, but not necessarily on confidence calibration. Future studies are needed to examine the distinct effects of associative binding manipulations on memory performance and confidence calibration in old age (cf. Hertzog, Dunlosky, & Sinclair, 2010).

Our findings are in line with findings from the repetition-lag procedure (Jennings & Jacoby, 1997). In this procedure participants read aloud a list of words that they have to remember. In a subsequent recognition memory test, along with studied words, the participants see new words that are repeated after 0, 3, or 12 intervening items, and part of their task is to reject these stimuli. With this procedure older adults are more likely to recognise repeated new items as studied before already at a lag as small as 3, thus showing an effect of familiarity in the absence of recollection. Crucially, our task differed from the repetition-lag procedure in that the source of the memory about the word pairs is not as clearly distinctive as in the repetition-lag procedure due to the high similarity and short intervals among the runs. Under these conditions the demand on memory monitoring processes increases (Burgess & Shallice, 1996; Henson, Shallice, & Dolan, 1999; Moscovitch & Winocur, 2002). Furthermore, a number of studies have shown that older adults can be trained with the repetition-lag procedure by implementing a gradual increase in the repetition lag and the corresponding demand on recollection-based processing (Bissig & Lustig, 2007; Jennings & Jacoby, 2003; Jennings, Webster, Kleykamp, & Dagenbach, 2005). Our results speak to these findings by demonstrating that the slope for rearranged pairs in older adults was reliably different from zero. Hence older adults benefited from the across-run repetition of the lure pairs, although not to the same degree as younger adults. An intriguing question for future research would be to identify possible interventions that may enhance older adults’ memory performance by strengthening the
ability to withstand interference from familiar information. Given the substantial inter-individual variability in training gain from the repetition-lag procedure (Bissig & Lustig, 2007) as well as the present finding that across-run repetition benefits were greater for older adults with relatively preserved binding abilities, investigating training procedures that are individually adapted to the available memory resources may be of particular importance (Fandakova, Shing, & Lindenberger, 2012; Lindenberger et al., 2008).

In conclusion, our findings demonstrate that deficits in memory monitoring mechanisms underlie older adults’ difficulties to withstand interference from increasing familiarity of memory representations. At the same time, age-related decline in binding mechanisms also contribute to older adults’ difficulties to reject highly familiar information. The two mechanisms interact with each other such that deficits in binding may lead to weaker and less distinctive memory traces, which call for higher need for strategic intervention. At the same time, deficits in memory monitoring may preclude critical evaluation of retrieved memories and limit the effectiveness of recall-to-reject processes, hence leading to higher false memory and deficient confidence calibration in old age.

The present findings are in good agreement with the two-component framework of episodic memory development across the lifespan (Shing et al., 2008, 2010). According to this framework, episodic memory performance is supported by two interacting associative and strategic components (see also Miller & Cohen, 2001; Prull, Gabrieli, & Bunge, 2000; Simons & Spiers, 2003). The strategic component refers to PFC-dependent cognitive control processes that aid and regulate memory functions at encoding and retrieval, whereas the associative component refers to MTL-dependent mechanisms that bind different features of memory events into coherent representations. Importantly, according to the framework both components undergo age-related changes, suggesting that older adults exhibit deficits both in strategic and associative aspects of memory.

Our results not only provide general support for the two-component model, but also help to further specify the interacting contributions of strategic and associative components to adult age changes in episodic memory. First, we examined monitoring processes, an aspect of the strategic component that has not been investigated in previous studies of the two-component framework (Shing et al., 2010). Second, we support the framework by demonstrating that increased false memory in old age is related to memory monitoring deficits in interaction with age-related decline in binding. Finally, our results with respect to high-confidence errors in old age suggest that those errors might be more related to ageing deficits in memory monitoring than to binding deficits alone. An important next step towards a more complete understanding of episodic memory decline in ageing is to investigate how processing of PFC and MTL regions and their interactions underlie older adults’ susceptibility to high-confidence memory errors.

References


Manuscript received 2 July 2012
Manuscript accepted 30 November 2012
First published online 10 January 2013


