Intact strategic retrieval processes in older adults:

No evidence for age-related deficits in source-constrained retrieval

Louisa Salhi, & Zara M. Bergström*

University of Kent, UK

*Corresponding author: z.m.bergstrom@kent.ac.uk

Word count: Abstract: 199; Introduction: 1856; Methods: 3107; Results: 1554; Discussion: 2130. References: 1735; Total main body (including references, excluding abstract): 10382. Tables: 2; Figures: 2

Abstract

Aging is thought to involve impairments to cognitive control functions that support episodic memory, for example by enabling people to strategically constrain their retrieval search towards a specific context ("source") in order to facilitate retrieval of goal-relevant memories. The "memory-for-foils" paradigm investigates source-constrained retrieval by assessing whether incidental encoding of new foils during an old/new recognition test differs depending on the type of processing that was previously used during study of the old items in the test. If it does, it suggests that people process foils differently as a result of engaging in source-constrained retrieval attempts. Young adults typically show differences in incidental encoding foils, but such differences have not been found in older adults. Here, we compared source-constrained retrieval and reward effects on incidental foil encoding between younger and older adults, to assess if age-related reductions in strategic retrieval processing are accompanied by differences in responsiveness to external rewards. The results showed only minor effects of rewards on memory processing, in younger adults only. Contrary to prior findings, older adults had equivalent overall memory performance and spontaneously constrained retrieval to the same extent as the young group, showing that aging-related impairments to strategic retrieval processes are not inevitable.

Keywords: Aging; recognition; strategic retrieval; memory-for-foils; external rewards

Older adults often show deficits in memory retrieval tasks, in particular those tasks that require cognitive control processes to optimise memory search attempts or to evaluate the products of retrieval in relation to task goals (Morcom, 2016). In recognition tasks, aging is associated with an increased reliance on automatic, familiarity-driven recognition based on item memory strength rather than recollection of the context in which an item was previously encountered (Koen & Yonelinas, 2016; Yonelinas, 2002). In contrast, young people with intact control processes often self-initiate recollection to enhance the accuracy of their recognition decisions, for example by strategically orienting their memory search towards a particular context (Jacoby, Shimizu, Daniels, & Rhodes, 2005; Mecklinger, 2010; Rugg & Wilding, 2000). There is less consistent evidence regarding whether older adults also show lowered sensitivity to external factors that could potentially ameliorate memory difficulties, such as rewards, with different studies finding evidence for intact versus reduced reward effects on memory in older age (e.g. Spaniol, Schain, & Bowen, 2014; Geddes, Mattfeld, De los Angeles, Keshavan, & Gabrieli, 2018). Furthermore, recent research has only begun to assess the relationship between reward-induced motivation and strategic control processes in terms of their effects on memory (e.g. Cohen, Cheng, Paller & Reber, 2019), and how that relationship may differ across age groups, memory stages (e.g. encoding vs. retrieval), and memory processes (i.e. different types of encoding or retrieval operations, such as intentional vs. incidental encoding/retrieval processes). In this experiment, we attempted to replicate and extend on prior findings of age-related deficits in spontaneous use of controlled retrieval processes (Jacoby, Shimizu, Velanova, et al., 2005), by also investigating if such deficits were accompanied by age-related reductions in reward effects on memory.

Source-constrained retrieval refers to the idea that people can strategically enhance recognition by focusing their memory search towards a particular encoding context (or “source”). Such constraints on memory search may be achieved by re-implementing processes previously engaged at encoding during subsequent recognition attempts (Jacoby, Shimizu, Daniels, et al., 2005). For example, if an encoding task involved making judgements about the pleasantness of stimuli, then participants may in a subsequent old/new recognition test, focus on whether the stimuli are pleasant or not as a way to elicit recollection of the encoding context. Evidence that participants spontaneously constrain retrieval in this way comes from the "memory-for-foils" paradigm (Jacoby, Shimizu, Daniels, et al., 2005). In the most relevant version of this paradigm (Vogelsang, Bonnici, Bergström, Ranganath, & Simons, 2016; Vogelsang, Gruber, Bergström, Ranganath, & Simons, 2018), participants first encode two different word lists in two different encoding tasks. The first task encourages participants to focus on the meaning of the words by rating them for pleasantness (considered "deep", semantically-oriented processing in the Levels of Processing (LOP) framework; Craik & Lockhart, 1972). In contrast, the second task involves making judgements on a more perceptual, non-semantic basis, such as detecting if the words contain vowels or not (considered "shallow" processing in the LOP framework). Next, recognition of the two lists is tested in two different tests, one where the old words from the deep/semantic encoding task are intermixed with a set of new "foil" words (the "deep test"), and another test where the old words from the shallow/non-semantic encoding task are intermixed with a different set of new foils (the "shallow test"). Participants are informed of which test contains deep versus shallow old items, and it is thought that this knowledge encourages them to spontaneously focus more on semantic information (e.g. the pleasantness of the words) during the deep test than the shallow test, in order to constrain their memory search towards the appropriate source information.

Unsurprisingly, the encoding manipulation typically leads to higher recognition accuracy on the initial deep test than the shallow test, demonstrating a classic LOP effect (Craik & Lockhart, 1972). The novel aspect of the design involves adding a second subsequent surprise test where all the foils from the prior test phases are intermixed with completely new words, and participants are asked to indicate any words they recognise from the experiment. It is typically found that foils that were initially tested with deeply encoded old items are more likely to be later recognised in the second surprise test than foils that were first tested with shallowly encoded old items. This pattern suggests that participants spontaneously reinstate a more semantic processing mode during the deep recognition test than the shallow recognition test, and that doing so leads to enhanced incidental encoding of foils encountered during the deep test.

There is now strong evidence from the memory-for-foils paradigm that young participants often engage in source-constrained retrieval (Alban & Kelley, 2012; Bergström, Vogelsang, Benoit, & Simons, 2015; Danckert, MacLeod & Fernandes., 2011; Gray & Gallo, 2015; Halamish, Goldsmith, & Jacoby, 2012; Kantner & Lindsay, 2013; Marsh et al., 2009; Vogelsang, et al., 2016; 2018; Zawadzka, Hanczakowski, & Wilding, 2017). However, there is to our knowledge only one prior study that has investigated the effects of aging on source-constrained retrieval (Jacoby, Shimizu, Velanova, et al., 2005). In that study, Jacoby et al. found a deficit in spontaneous source-constrained retrieval in older adults. That is, older adults did not show any difference in subsequent recognition accuracy of foils that had previously been shown in the deep versus shallow recognition tests, whereas a young group of participants showed the standard pattern of enhanced memory for foils from the deep test compared to the shallow test. These findings thus converged with prior evidence showing that older adults are less likely to spontaneously use retrieval strategies in memory tasks (Cohn, Emrich, & Moscovitch, 2008; Kirchhoff, Anderson, Barch, & Jacoby, 2012).

In the current experiment, we aimed to replicate the above effect of aging on source-constrained retrieval, and to also extend on previous research by investigating whether monetary rewards would affect incidental encoding of foils in this paradigm, and whether the effects of reward would differ across younger and older people. Monetary rewards enhances both intentional and incidental encoding in younger adults (Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006; Spaniol, et al., 2014; Wittmann, Schott, Guderian et al., 2005). Some findings have indicated that reward effects on memory are reduced or absent in older adults (e.g. Geddes, et al., 2018), consistent with evidence for age-related impairments in dopaminergic neural systems linked with reward processing (Bäckman, Lindenberger, Li & Nyberg, 2010). However, other findings suggest that reward effects on incidental encoding may be intact in older adults (Mather & Schoeke, 2011) pointing towards a more complicated relationship between aging, reward processing, and different types of memory processes and stages. Furthermore, although some influences of reward on encoding are thought to depend on relatively automatic dopaminergic brain systems (e.g. Wittmann et al., 2005), reward effects on memory can also be mediated through strategic recruitment of cognitive control processes (Cohen, et al., 2019). For example, external rewards increase the use of strategic memory processes in younger samples who will prioritise encoding of rewarded information at the expense of non-rewarded information (Cohen, Rissman, Suthana, Castel, & Knowlton, 2014), with some evidence suggesting older adults will also strategically prioritise encoding in this way (Castel, Farb, & Craik, 2007; Castel et al., 2013; Cohen, Rissman, Suthana, Castel & Knowlton, 2016).

However to our knowledge, the effects of reward on encoding have not been investigated when encoding occurs during retrieval attempts, as in the memory-for-foils paradigm. We thus aimed to test whether rewards might enhance incidental foil encoding in similar ways as other factors, such as semantic processing, imagery (Danckert, et al., 2011), self-referential thinking (Bergström, et al., 2015), survival processing (Nairne, Pandeirada, VanArsdall, & Blunt, 2015) and full attention (Dudukovic, DuBrow, & Wagner, 2009). Reward-induced enhancement of incidental foil encoding could potentially be mediated by automatic, dopaminergic mechanisms (Bäckman, et al., 2010), or through recruitment of strategic memory processes during the initial recognition tests. That is, we reasoned that the prospect of external rewards during a recognition test might encourage participants to use cognitive-control dependent strategies such as careful retrieval monitoring or effortful attempts at recollection (*cf*. Halsband, Ferdinand, Bridger, & Mecklinger, 2012), and such strategies would likely involve enhanced attention and "deeper" processing of foils when compared to processing of foils in a non-rewarded test. Therefore, increased use of retrieval strategies in response to rewards might facilitate incidental foil encoding in younger adults. If strategic retrieval processes are impaired in older age, older adults might show reduced reward effects on incidental foil encoding.

In the current study, we therefore directly compared both source-constrained retrieval effects and reward effects on incidental foil encoding across younger and older adults in the memory-for-foils paradigm. The experiment was conducted with one older group (mean age 73 years old) and one younger group (mean age 20 years old), and included two different manipulations of retrieval processing during recognition tests. The LOP phase was closely based on previous research (e.g. Jacoby, Shimizu, Velanova, et al., 2005; Vogelsang et al., 2016; 2018), and included two initial recognition tests that varied whether foils were shown either together with old items that had been deeply/semantically encoded or shallowly/perceptually encoded, in order to manipulate source-constrained retrieval processing. An additional Reward phase was also added where we conducted two initial recognition tests that varied whether recognition performance was rewarded or not rewarded, in order to manipulate reward-related motivation. Subsequently, recognition memory for all foils from the four prior tests (deep/shallow/rewarded/not rewarded) was tested in a surprise final test.

We predicted that in line with prior findings (e.g. (Jacoby, Shimizu, Daniels, et al., 2005; Vogelsang et al., 2016; 2018), final recognition would be enhanced for foils previously shown in the deep compared to the shallow test in the younger group, however there would be no difference in final recognition of foils previously shown in the deep vs. shallow test in the older group (Jacoby, Shimizu, Velanova, et al., 2005). We also expected that surprise test recognition would be enhanced for foils from the rewarded compared to not rewarded test in young adults, in line with prior evidence for reward-related enhancement of incidental encoding (e.g. Mather & Schoeke, 2011; Wittmann, et al., 2005). Based on evidence that reward sensitivity (e.g. Geddes et al., 2018) and spontaneous use of strategic memory processes (e.g. Cohn, et al., 2008; Kirchhoff, et al., 2012) is reduced in aging, we predicted a smaller difference between reward and no reward foils in the older group compared to the younger group. In contrast, if reward-related motivational influences on memory are intact in older age (e.g. Mather & Schoeke, 2011), then both older and younger adults would be expected to show similar reward-related enhancements of foil recognition. To pre-empt the results, rewards did not modulate incidental foil recognition in either group. Foils from the deep test were however more accurately recognised than foils from the shallow test, supporting the notion that participants engaged in source-constrained retrieval. However, we surprisingly found equivalent levels of enhanced recognition of deep versus shallow foils across older and younger group, despite an average age difference of over 50 years between the groups. This finding conflicts with prior literature (Jacoby, Shimizu, Velanova, et al., 2005) and suggests intact source-constrained retrieval in our sample of older adults.

Methods

Participants

We aimed to recruit 48 participants per group in order to achieve >0.9 power to detect an effect size of Cohen's *d*=0.7 at two-tailed α=.05, and to fully rotate and cross all counterbalancing factors (task orders, stimuli assignment to conditions, etc.). This estimated effect size was based on previous research with a similar design and using similar stimuli, which found large effects of source-constrained retrieval on foil recognition in young adults (Cohen's *d*=0.75 and *d*=0.89 for recognition accuracy differences between Deep vs. Shallow foils on the surprise test, in Vogelsang et al., 2016; 2018). The final sample was reduced to 41 per group partly due to practical recruitment constraints and partly due to participant exclusions, resulting in >0.9 power to detect an effect size of Cohen's *d*=0.75 at two-tailed α=.05. Specifically, an additional nine young adults were replaced or excluded; three due to technical errors and six due to scoring below threshold on the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005). An additional six older adults were also excluded; three due to technical errors, two due to MoCA scores below threshold, and one due to a failure to follow task instructions. Six of the younger adults and three of the older adults in the final sample reported that they anticipated a final test of some kind, but these participants were not excluded as the pattern of results was not affected by their removal. Thus, the presented results are based on data from 41 young adults (*M* age = 19.6 years, *SD*=1.1, range = 18-24) and 41 older adults (*M* age = 72.6 years, *SD*=5.2, range= 64-85; based on *N*=40 due to one participant missing demographic data). Note however that our sample size is still larger than that used by Jacoby, Shimizu, Velanova, et al. (2005) in their most similar experiment (Experiment 1), which included 32 younger (mean age 19.6, range 18-26) and 32 older (mean age 75.8, range 61-87) adults in a memory-for-foils test that manipulated level of processing as a between subjects factor (thus testing 16 young and 16 older participants in each condition).

Both our age groups had scores within the normal range on the MoCA (Nasreddine et al., 2005; MoCA; cut off >= 25[1]; young adults *M* = 28.0, *SD*= 1.3; older adults *M* = 27.7, *SD*= 1.8). Older adults were highly educated (mean number of years in education = 17.7, *SD*= 4.4; based on *N*=40 due to one participant missing demographic data) and young adults were all currently in higher education (mean number of years in education = 14.5, *SD*= 1.4). The young adults were recruited from the University of Kent Psychology Department research participation scheme, and the older adults were recruited from the Canterbury District of the University of the Third Age. All participants were native English speakers, had normal or corrected to normal vision, reported not having depression, anxiety or dyslexia, and were not taking any psychoactive medication. All participants provided full written informed consent and 'won' the same monetary reward of £5 as part of the experimental manipulation. Young participants also received course credits in addition to the reward. The procedures were approved by the University of Kent, School of Psychology Ethics Board.

Materials

The experiment was programmed and run in PsychoPy (Peirce et al., 2019; Peirce, 2007) on a tablet computer. In total 288 English words from the MRC psycholinguistic database (Wilson, 1988) were used as stimuli (3-11 characters, Kucera-Francis frequency range 0-591), drawn from a larger set used by Vogelsang et al. (2016, 2018). The words were randomly split into twelve lists of 24 words each, with assignment to experimental conditions counterbalanced across participants (see procedure for details).

[1]The MoCA threshold of >=25 was used instead of the standard 26, in order to be less conservative with regards to inclusions. Since the main results showed a lack of memory difference between the age groups, we did not want to restrict group differences by being overly conservative. For the record, all key conclusions from analyses were the same when a MoCA cut off of >=26 was used.

A set of 160 emotionally neutral photographs from the IAPS database (Lang, Bradley, & Cuthbert, 2008) were used in a visual search filler task between encoding and retrieval tasks.

Design and procedure

The study was conducted as a “citizen science” project, where we collaborated with University of the Third Age (U3A) research group members who were older adults themselves and acted as advisors and experimenters. The aim of this collaboration was to mitigate potential confounding effects of experimenter and participant expectations for stereotyped age differences in memory (McDaniel, Einstein, & Jacoby, 2008). The U3A research group was consulted during the study design to make the task more suitable for older adults, and were extensively trained in order to independently conduct data collection, with assistance if necessary from the first author. Thus, we took great care to ensure that all tasks were suitable for both age groups based on consultation with the U3A research group members who were of similar age as the older sample. As a result of this consultation, all instructions were given both verbally and in writing, multiple breaks were provided at specific points during the task, and computerised response requirements were simplified and consistent across all tasks. All materials, task designs and instructions were identical across both groups, and instructions were delivered without mentioning that memory performance would be compared across age groups, to avoid expectancy effects. The experimenters’ ages were consistent with the age groups that they tested; the older adults were tested by an older adult experimenter from the U3A research group (together with the first author), whereas the younger adults were tested by a younger adult experimenter (either the first author or an undergraduate student assistant).

Main encoding and recognition tasks

The memory-for-foils procedure (Jacoby, Shimizu, Daniels, et al., 2005) was modified to incorporate two different encoding manipulations; Level of Processing (LOP) and Reward, which were presented to the participants in separate blocks and were followed by a joint surprise test for all foils. The key variable that changed across these blocks was what type of encoding task participants were asked to complete. Within each block, participants first completed two encoding phases (either Deep and Shallow, or Rewarded and Non-Rewarded), then went on to complete two separate recognition tests for old items drawn from each of the encoding tasks, together with randomly intermixed new foils. The recognition tests were always conducted in the same order as the encoding tasks, and participants were told which encoding task the old items had previously been shown in. The manipulation order (LOP/Reward or Reward/LOP) was counterbalanced across participants, as was the specific condition order within manipulation block (i.e. Deep/Shallow or Shallow/Deep within the LOP block, and Rewarded/Non-Rewarded or Non-Rewarded/Rewarded within the Reward block). So for example, a specific participant would have started with a Deep encoding task followed by a Shallow encoding task, then had their recognition of Deeply old items tested when intermixed with foils, followed by a recognition test for Shallowly encoded old items intermixed with foils. Next, they would have completed the Rewarded encoding task, followed by the Non-Rewarded encoding task. Then, they would have competed a recognition test for the Rewarded old items intermixed with foils, then another recognition test for the Non-Rewarded old items intermixed with foils. Other participants completed these blocks and conditions in counterbalanced orders so that there were no order confounds in the design, with the constraint that LOP conditions were always presented together in one block and Reward conditions in the other, and the initial recognition tests were always conducted in the same order as the encoding tasks, so that the time delay between encoding and recognition was always constant. Finally, all participants completed a Surprise recognition test at the end where foils from all four recognition tests were randomly intermixed with completely new words.

The LOP phase was very similar to prior research with the memory-for-foils paradigm (e.g. Jacoby, Shimizu, Velanova, et al., 2005). In this phase, participants encoded two different lists of 24 words during separate Deep vs. Shallow encoding tasks (order counterbalanced). In the 'Deep' encoding task, they were asked to judge whether each word was pleasant, whereas in the 'Shallow' encoding task, they were asked to judge whether each word contained an 'O' or a 'U'. Words were presented for 3 seconds and participants pressed 'z' or 'm' keys on the keyboard within this time to indicate 'Yes' or 'No' to the relevant question. Each word was preceded by a 600ms fixation cross.

Participants then took part in two recognition tests (the Test 1 stage), in the same order as the encoding tasks (i.e. if the Deep encoding task was conducted before the Shallow encoding task, then Deep old items were tested before Shallow old items, and vice versa). Each test contained 24 old target words and 24 new foil words. As in previous research (Jacoby, Shimizu, Daniels, et al., 2005; Jacoby, Shimizu, Velanova, et al., 2005; Vogelsang et al., 2016; 2018) participants were explicitly told that the 'old' words in these separate recognition tests were drawn from only one of the earlier judgment tasks (Deep or Shallow) and that they were not intermixed. Their task was to say whether they recognised each word as having been shown in the relevant earlier judgement task. These instructions thus aimed to hint to participants that it could be useful for them to strategically reinstate the processing mode they had used during the corresponding study task, however participants were not explicitly asked to do so. Therefore, we consider any reinstatement of Deep vs. Shallow processing during the recognition test to have been spontaneous. Participants again used the 'z' or 'm' keys to respond 'Yes' or 'No', this time to the question: 'Do you recognise the word?'. Each word was preceded by a 600ms fixation cross, and remained on the screen until participants gave a response. Participants were asked to respond quickly and accurately.

In the Reward phase, participants encoded two different lists of 24 words during separate Rewarded vs. Non-rewarded encoding tasks (order counterbalanced). They were told that for one list (the Reward condition), they would receive a monetary reward for correctly recognising those words later (10p per word, max £5), whereas there would be no monetary reward for the other list. As in the LOP encoding tasks, all words were presented for 3 seconds, preceded by fixation cross for 600 milliseconds, but no judgments or responses were required during the Reward phase encoding tasks. Participants studied both lists before moving onto the corresponding recognition test phases. As in the LOP phase, word lists were tested in two separate "Test 1" tests by intermixing old words from the encoding tasks with new foils (24 old and 24 new words in each test), in the order of study presentation (i.e. if the Rewarded encoding task was conducted before the Non-rewarded encoding task, then Rewarded old items were tested before Non-rewarded old items, and vice versa). Participants were once again explicitly told which test corresponded to which study list so that they knew whether they would be receiving rewards or not, based on their recognition performance. Participants were informed that they would be rewarded for correct responses to both old and new words. Rewards were balanced across old and new items to avoid inducing response biases, which could have emerged if only correct "old" responses were rewarded (such a design could encourage participants to adopt a "lax" response bias and respond 'old' to the majority of items in an effort to maximise their winnings). For ethical reasons, but unknown to participants, at the end of the experiment all participants received the full £5 win. As in the LOP recognition tests, each word was preceded by a 600ms fixation cross, and remained on the screen until participants gave a response. Participants were asked to respond quickly and accurately.

After all the LOP and Reward encoding phases and initial recognition test phases had been completed, participants received a surprise final recognition task for all the 96 foil words that had appeared in the four tests (24 Deep, 24 Shallow, 24 Rewarded, 24 Non-Rewarded foils), intermixed with an equal number of new words (96 new words). Participants were instructed that in this test, words that had been presented previously in any of the blocks could appear, intermixed with completely new words that had never been presented during the study. It was ensured that participants understood that this included words from all previous phases, including the ‘foil’ words that had been presented in the previous test phases, and that their task was to say whether they recognised each word from any of the previous phases. As in previous tests, participants again used the ‘z’ or ‘m’ keys to respond ‘Yes’ or ‘No’ to the question: ‘Do you recognise the word?’ as quickly and accurately as possible and the task was self-paced, with each word preceded by a 600ms fixation cross.

Visual Search filler tasks

A visual search task was used as a filler to prevent rehearsal between each experimental task (eight times in total). Participants were presented with a series of photographs and were asked to search for a white triangle that had been superimposed onto a subset of photographs. As in previous tasks, participants again used the ‘z’ or ‘m’ keys to respond ‘Yes’ or ‘No’, this time to the question, ‘Is there a white triangle on the image?’. This filler task always lasted 1 minute for all blocks and all participants, regardless of the number of trials completed (i.e. faster participants were given more trials, and slower participants were given fewer trials, in order to ensure a constant delay between tasks).

Post- Experiment tasks and debrief

After the participants had completed the experimental tasks they completed a short questionnaire, assessing if they had anticipated the final foil test as well as questions about effort, motivation and strategies used during the reward phases (see Supplemental file for the self-report questionnaire and analysis of responses). They also completed the MoCA (Nasreddine et al., 2005), and were given a verbal as well as a written debrief alongside their reward phase earnings of £5.00.

Data Analysis

For the Test 1 stage, we calculated the proportion of correct responses for each test block (Deep, Shallow, Reward, No Reward), separately for Old words (i.e. hit rates) and New words (i.e. correct rejection rates) within each test block. Mean reaction times (RTs) were also calculated for Old and New words split by test block type. The raw accuracy scores above were used to derive estimates of discrimination ability and response biases during each recognition test, in order to provide more meaningful measures of recognition performance for testing potential differences between the conditions and age groups. *Pr* was calculated to measure participants' ability to discriminate between old and new items regardless or response biases. *Pr* represents how much more likely participants are to accurately identify previously seen items as old, compared to mistakenly "false alarm" by identifying new items as old (*Pr* = hits- false alarms; see Snodgrass & Corwin, 1988). The *Br* measure of response bias was also calculated (*Br* = false alarm / (1- *Pr*); Snodgrass & Corwin, 1988) to estimate participants' general tendency to respond 'old' versus 'new' when uncertain. *Br* values can fall between 0-1, where values around 0.5 indicate a neutral response bias (equally likely to guess old or new), values >0.5 indicate a "liberal" response bias (guessing "old" more often than "new") and values <0.5 indicate a "conservative" response bias (guessing "new" more often than "old").

On the Surprise foil test, proportion accurate responses were calculated separately for each foil condition (Deep foils, Shallow foils, Reward foils, Standard foils), and for New items (resulting in hit rates for foils since these items are "old" on the final test, and correct Rejection rates for New items). Mean RTs were also calculated for these conditions. As there was only one set of New items and all Foils were intermixed, one measure of Surprise test response bias (*Br*) was calculated for each participant, using a summary measure of *Pr* that was not split according to foil condition (*Br* = false alarm / (1- summary *Pr*)). Condition-specific *Pr* scores were then calculated for each foil type by subtracting the common false alarm rate from the separate foil hit rates.

Pr was used as the key measure of recognition discrimination ability for inferential statistical analyses of both Test 1 and Surprise test data, in order to make the results comparable across phases and to compensate for individual differences in response biases. The *Br* measure was also included as a key measure to investigate potential age and condition differences in response biases. Raw proportion accurate measures (hit and correct rejection rates) were also analysed, but since these measures are only complementary to the key *Pr* and *Br* measures, they are presented descriptively in the main article with inferential tests available on the Open Source Framework website (OSF, https://osf.io/aejxm/). Likewise, differences in RTs across conditions were also analysed, but since there were not relevant to the key hypotheses and did not reveal any results of interests, inferential analyses of RTs are available on the OSF site, and RT data are only shown descriptively in the main paper. All analyses compared measures across conditions and groups within the LOP and Reward manipulation types, but did not compare measures across LOP and Reward manipulations, since such comparisons would not be meaningful.

The key analyses used frequentist inferential statistical tests from the General Linear Model (ANOVA, t-tests). Effect sizes were estimated using partial eta squared

(ANOVA) or Cohen's *d* (t-tests). Cohen's *d* for both paired and independent t-tests was calculated as the difference between means divided by the pooled standard deviation to avoid inflating effect size estimates for paired t-tests (Dunlap, Cortina, Vaslow, & Burke, 1996). Frequentist statistical tests of key hypotheses were supplemented with Bayesian statistics in order to assess if evidence supported the null hypothesis, which is not possible to determine with frequentist statistics alone. Therefore, Bayes factors (BF_{10}) were calculated to estimate the relative support for the alternative (H_1) versus the null hypothesis (H_0). The Bayes Factor is a ratio that contrasts the likelihood that the data would occur under the alternative (H_1) versus null (H_0) hypotheses, with values over 1 indicating support for H_1 and values below 1/3 indicating support for the H_0. Values close to 1 are only considered weakly/anecdotally supportive of one hypothesis over the other, whereas BF_{10} >3 are typically interpreted as substantial evidence in support of H^1 over H^0, and BF_{10} <1/3 are interpreted as substantial evidence in support of H^0 over H^1 (see Wagenmakers, Wetzels, Borshboom & Van der Mass, 2011). All Bayes factors were calculated with two-tailed tests, thus the alternative hypothesis was a statistical difference between groups/association between variables in either direction, and the null hypothesis was no difference between groups/no association between variables. Bayes Factors were calculated using JASP (JASP Team, 2017) with default priors (a Cauchy distribution with centre = 0, r = 0.707).

Results

Both group level and individual participant's data from all recognition tests in all phases can be found in an anonymised format on the Open Science Framework website along with other supplemental materials and additional analyses, such as Bayesian analysis output including robustness checks (https://osf.io/aejxm/).

Recognition Test 1 phase

LOP manipulation

Performance on the first recognition tests from the LOP phase is shown in Table 1. As can be seen here, both age groups were highly accurate on the Deep recognition test, and less accurate on the Shallow recognition test, as expected. The older adults responded slower than younger adults, but discrimination and response biases were similar across groups.

[Table 1 around here]

A mixed ANOVA with the within subjects factors LOP (Deep vs. Shallow test) and the between subjects factor Age Group (Young vs. Old) showed only a main effect of LOP on *Pr* ($F(1,80)=389.78$, $p<.001$, $n^2p=.83$, *EMM*; Deep = 0.82, *SE*=0.02, Shallow = 0.37, *SE*= .02) with no main effect of Age Group on discrimination ($F(1,80)= 0.39$, $p=.532$, $n^2p<.01$, *EMM*; Older = 0.58, *SE*=0.02; Younger =0.60, *SE*=0.02), and no interaction between LOP and Age Group ($F(1,80)=0.02$, $p=.90$, $n^2p<.01$). Thus, both younger and older adults were better at discriminating between old and new words in the Deep compared to Shallow test, with no evidence for a difference between the groups. To confirm that the effects of the LOP manipulation on discrimination performance was indeed equal across Age Groups, a Bayes factor was calculated for an independent t-test that compared the difference in *Pr* between LOP conditions (Deep *Pr* minus Shallow *Pr*) across the two Age Groups. The Bayes factor provided substantial evidence in support for the null hypothesis of no difference in discrimination between the Age Groups (BF_{10}= 1/4.32).

Participants showed a more conservative response bias (i.e. a reduced tendency to guess “old” when uncertain) in the Shallow compared to the Deep test (main effect of LOP: $F(1,80)=46.90$, $p<.001$, $n^2p=.37$, *EMM*; Deep= 0.50, *SE=0*.03, Shallow= 0.30, *SE*= 0.02). There was no main effect of Age Group (F(1,80=<.001, $p=.994$, $n^2p<.01$, *EMM*; Older = 0.40, *SE=0*.03; Younger =0.40, *SE=0*.03) nor interaction with age group ($F(1,80)=0.43$, $p=.516$, $n^2p<.01$). To confirm that the LOP effect on response bias was equal across Age Groups, a Bayes factor was calculated for an independent t-test that compared the difference in *Br* between LOP conditions (Deep *Br* minus Shallow *Br*) across the two Age Groups. The Bayes factor provided substantial evidence in support for the null hypothesis of no difference in LOP effect on response bias between the Age Groups (BF_{10} = 1/3.61).

So to summarise, discrimination and response bias on the first tests in the LOP phase did not differ across Age Groups, who both showed increased discrimination and a less conservative response bias on the Deep compared to the Shallow test.

Reward manipulation

Performance on the first recognition tests from the Reward phase is shown in Table 2.

[Table 2 around here]

The *Pr* measure of discrimination showed no main effect of Reward condition ($F(1,80)=1.71$, $p=.195$, $n^2p=.02$, *EMM*; Reward=0.60, *SE*=0.03, No Reward=0.57, *SE*=0.03) or Age Group ($F(1,80)= 2.28$, $p=.135$, $n^2p=.03$, *EMM*; Older=0.62, *SE*=0.03; Younger=0.55, *SE*=0.03) on discrimination ability, but there was a significant interaction between Age Group and Reward condition ($F(1,80)=7.88$, $p=.006$, $n^2p=.09$). The Reward manipulation had no significant effect on Older adults' ability to discriminate between Old and New Words ($t(40)= 1.36$, $p=.182$, $d=0.20$) whereas Reward did lead to higher discrimination performance for Younger adults compared to No Reward in the same group ($t(40)= 2.47$ $p=.018$, $d= 0.41$). To provide complementary evidence that the Reward manipulation did indeed affect discrimination performance differently across Age Groups, a Bayes factor was calculated for an independent t-test that compared the difference in *Pr* between Reward conditions (Reward *Pr* minus No Reward *Pr*) across the two Age Groups. The Bayes factor provided substantial evidence in support for the alternative hypothesis that the effect of Reward on discrimination was different between the Age Groups ($BF_{10} = 6.503$).

The response bias index (*Br*) indicated that both Age Groups responded fairly conservatively when uncertain, with a slightly more conservative response bias in the Non-Rewarded test than the Rewarded test ($F(1,80)=4.38$, $p=.04$, $n^2p=.05$, *EMM*; Reward = 0.41, *SE*=0.03, No Reward =0.37, *SE*=0.03). There was however no main effect of Age Group ($F(1,80)= 0.02$, $p=.889$, $n^2p<.01$, *EMM*; Older=0.40, *SE*=0.03; Younger=0.40, *SE*=0.03), and no interaction between these factors ($F(1,80)=1.27$, $p=.263$, $n^2p=.02$). A Bayes factor was calculated for an independent t-test that compared the difference in *Br* between Reward conditions (Reward *Br* minus No Reward *Br*) across the two Age Groups. The Bayes factor provided anecdotal evidence in support for the null hypothesis that response bias was equal across the Age Groups ($BF_{10} = 1/2.501$).

So to summarise, the Reward manipulation enhanced performance on the first tests only in the Younger adults, who showed improved discrimination between Old and New Words when rewarded compared to when they were not rewarded.

Surprise Foil Test phase

On the final foil test, Older and Younger adults not differ significantly in their ability to correctly reject New Words ($t(80)=1.27$, $p=.21$, $d=0.28$; Younger=0.76, *SD*= 0.17; Older= 0.80 *SD*=0.14; $BF_{10} = 1/2.17$), nor did they differ in general response bias (*Br* = false alarm / (1- summary *Pr)*; $t(80)= 1.21$, $p=.229$, $d= 0.27$; mean *Br*: Older= 0.38, *SE*=0.04; Younger=0.49, *SE*=0.02; $BF_{10} = 1/2.30$).

LOP manipulation effects on foils

First, we investigated whether recognition of foils on the final test differed depending on whether they had been shown intermixed with deeply encoded versus shallowly encoded old words on the first recognition test, and whether this effect was reduced in older adults, as had been previously found (Jacoby, Shimizu, Velanova, et al., 2005). Discrimination performance for the two Age Groups and LOP conditions is shown in Figure 1.

[Figure 1 around here**]**

A mixed ANOVA with the within subjects factors LOP (Deep vs. Shallow) and the between subjects factor Age Group (Young vs. Old) showed a main effect of LOP on foil *Pr* ($F(1,80)=35.87$, $p<.001$, $n^2p=.31$, *EMM*; Deep = 0.51, *SE*=0.02, Shallow = 0.42, *SE*= .02) with better discrimination for Deep compared to Shallow foils. There was no main effect of Age Group ($F(1,80)= 0.40$, $p=.531$, $n^2p<.01$, *EMM*; Older = 0.48, *SE*=0.03; Younger =0.46, *SE*=0.03), nor an interaction between LOP and Age Group ($F(1,80)=0.02$, $p=.892$, $n^2p<.01$). To confirm that the effect of LOP on discrimination performance was indeed equal across Age Groups, a Bayes factor was calculated for an independent t-test that compared the difference in *Pr* between LOP conditions (Deep foil *Pr* minus Shallow foil *Pr*) across the two Age Groups. The Bayes factor provided substantial evidence in support for the null hypothesis of no difference in the effect of LOP on discrimination between the Age Groups (BF_{10} = 1/4.315). Thus, paralleling the first test, both Age Groups were equally affected by the LOP manipulation in terms of their ability to discriminate between old and new items.

Reward manipulation effect on foils

The next analysis was conducted to investigate whether recognition of foils on the final test differed depending on whether they had previously been shown in a recognition test where performance was rewarded or not rewarded, and whether this potential reward effect was affected by aging. Discrimination performance for the two Age Groups and Reward conditions is shown in Figure 2.

[Figure 2 around here]

A Reward condition x Age Group ANOVA showed no main effect of Reward condition on *Pr* ($F(1,80)=0.08$, $p=.778$, $n^2p<.01$, *EMM*; Reward=0.48, *SE*=0.02, No Reward=0.47, *SE*=0.02) and no effect of Age Group ($F(1,80)= 0.54$, $p=.466$, $n^2p<.01$, *EMM*; Older=0.49, *SE*=0.03; Younger=0.46, *SE*=0.03), nor an interaction between these factors ($F(1,80)=0.24$, $p=.624$, $n^2p<.01$). To provide complementary evidence that discrimination performance was indeed similar across Age Groups, a Bayes factor was calculated for an independent t-test that compared the difference in *Pr* between Reward conditions (Reward *Pr* minus No Reward *Pr*) across the two Age Groups. The Bayes factor provided substantial evidence in support for the null hypothesis that the effect of Reward on discrimination was not different between the Age Groups (BF_{10}=1/ 3.914).

Correlations between older adults' demographics and the LOP foil effect

In a final analysis of the Surprise test data, we investigated whether the source-constrained retrieval effect (i.e. the effect of LOP on foil recognition) was related to age or level of education within the older sample. Since we had not found an age-related impairment in source-constrained retrieval as described in prior research (Jacoby, Shimizu, Velanova, et al., 2005), we wanted to assess whether this might be related to demographic variables of relevance (for example, perhaps only the oldest or less educated adults in our sample showed a reduced source-constrained retrieval effect). These issues were examined by calculating the difference in foil discrimination (*Pr*) on the final test between the Deep and Shallow conditions, and correlating this difference with age and numbers of years in education within the Older adults only. There was no significant relationships between age (r_s =.045, p=.783; $BF_{10, (r \text{ width}=1)}$ = 1/4.76) or years in education (r_s =.178, p=.271; $B_{10, (r \text{ width}=1)}$ = 1/4.98; N=40 for both due to one participant missing demographic data) and the LOP source-constrained retrieval effect, and the Bayes Factor indicated support for the null hypothesis of no association between variables in both cases. However these results should be interpreted cautiously since our study design and sample size was not optimised for detecting such correlations.

Discussion

The current study investigated whether older adults would show deficits in strategic retrieval processes, as has been found in previous research (Cohn, et al. 2008; Kirchhoff, et al., 2012; Morcom, 2016). Specifically, we aimed to replicate previous findings that older adults are less likely than young adults to spontaneously re-implement processes from a prior encoding task during a subsequent recognition test in order to constrain their retrieval search towards a particular source (Jacoby, Shimizu, Velanova, et al., 2005). We also extended on previous research by exploring the effects of monetary rewards on incidental encoding processes engaged during retrieval attempts, in order to investigate if age deficits in strategic retrieval processing were accompanied by reduced reward influences on memory within the same sample. The main findings showed no influence of rewards on incidental encoding during retrieval attempts in either age groups, but suggested that both age groups engaged in source-constrained retrieval to a similar extent. Thus, we unexpectedly found that the older adults appeared to have intact ability to strategically reinstate encoding processes during retrieval, conflicting with prior findings (Jacoby, Shimizu, Velanova, et al., 2005).

Our older and younger groups were comparable on general memory performance and showed equivalent effects of the encoding LOP manipulation (Craik & Lockhart, 1972) during the initial encoding task, as assessed with a first old/new recognition test that showed that both groups benefited equally from a ‘deep’, semantically oriented task compared to a ‘shallow’, perceptually oriented tasks. Such equivalent effects of LOP encoding task manipulations on recognition accuracy across age groups have been found previously (e.g. Grady, McIntosh, Rajah, Beig & Craik, 1999; Jacoby, Shimizu, Velanova, et al., 2005), even though older adults typically also show a general reduction in recognition discrimination ability (Fraundorf, Hourihan, Peters, & Benjamin, 2019), which was not found in our older group. Furthermore, in the second, surprise foil test that followed, both groups showed equivalent enhanced encoding of foils that had been tested together with old items from the deep encoding task compared to foils that had been tested together with old items from the shallow encoding task. This pattern suggests that both age groups spontaneously and strategically altered their retrieval processing, producing differential incidental encoding of foils. That is, in the memory-for-foils paradigm, enhanced subsequent recognition for the deep over the shallow foil words is thought to be caused by participants spontaneously reinstating encoding processes during the initial tests. This strategy thus involves focusing on semantic aspects of stimuli (“deep” processing) in a recognition test if the initial encoding task involved semantic judgements, versus focusing on perceptual aspects of stimuli (“shallow” processing) if the initial encoding task involved perceptual judgements. By reinstating such different processing modes, the foils become more deeply encoded in the deep test than the shallow test (see Danckert, et al., 2011; Gray & Gallo, 2015; Jacoby, Shimizu, Daniels, et al., 2005; Vogelsang et al., 2016, 2018).

Prior research has suggested that intact recognition test performance in older adults may be due to a reliance on automatic familiarity over more controlled recollection of context (Koen & Yonelinas, 2016; Yonelinas, 2002). However, if the older adults had relied on automatic memory processes during the first recognition test, we would not have expected to see any difference between foils in terms of subsequent memory in this group. Instead, it appeared that they were just as likely as the young group to engage strategic retrieval processes to facilitate recollection of the study context (Jacoby, Shimizu, Velanova, et al., 2005). Our results therefore conflict with prior findings from a very similar experimental design that also compared young university students against older members of the community and found no differences between deep and shallow foils in the older participants (Jacoby, Shimizu, Velanova, et al., 2005), and are also inconsistent with other findings of age-related impairments to strategic retrieval processing (reviewed in Morcom, 2016). It is unclear exactly what caused this inconsistency. Our older sample was of similar age (mean age 73, range 64-85) to those in prior relevant studies (e.g. Jacoby, Shimizu, Velanova, et al., 2005, where the mean age was 75 and range 61-87). Furthermore, there was no indication of a reduction in the source-constrained retrieval effect (i.e. the difference between ‘deep’ and ‘shallow’ foil recognition discrimination) as a function of age within the older sample, despite the relatively large range between our youngest ‘old’ and oldest ‘old’ participant.

However, it is well known that aging-related memory problems can be highly variable across individuals of the same age, and there are many people who have intact memory function even at a very old age (Nyberg, Lövdén, Riklund, Lindenberger, & Bäckman, 2012). One possibility is that our older sample was unusually high performing because of demographic reasons – they were mostly highly educated, and all were socially active members of the University of the Third Age, which is an organisation for lifelong learning that encourages retired or semi-retired people to share their skills and knowledge with each other. High levels of education is typically associated with intact cognitive performance in old age, but we did not observe a correlation between years in formal education and the source constrained retrieval effect in our sample (although we had a restricted range of values for education, so this lack of association should be interpreted cautiously). Education is not the only important factor for healthy cognitive aging however, since being socially active and engaging in mentally challenging activities have also been found to independently predict high cognitive functioning in old age (Chan et al., 2018; Nyberg & Pudas, 2019). Members of social/educational groups such as the U3A may be particularly likely to have a lifestyle that promotes brain health and intact cognitive function in old age. Although we did not intentionally aim to sample older adults with particularly high memory functioning, we may nevertheless have done so by recruiting only from the U3A. Consistent with this account, our participants did not show a general deficit in memory performance as found in the older sample in Jacoby, Shimizu, Velanova, et al. (2005), where older participants’ recognition accuracy was reduced overall across initial and final recognition tests.

It should also be noted that reductions in strategic retrieval processing in older age have not always been found in prior research. As reviewed in Morcom (2016), older adults sometimes engage strategic retrieval control when required to do so by the task (Duverne Motamedinia & Rugg, 2009). This finding suggests that aging may not lead to permanent impairments in strategic retrieval processes, but rather that older people tend not to engage such processes spontaneously in item recognition tasks (see also Morcom & Rugg, 2004). It is possible that subtle aspects of our procedure could have encouraged older participants to use source-constrained retrieval strategies more effectively than in prior studies. For example, it is known that experimenter and participant expectations can have strong effects on cognitive test results, such as in the case of stereotype threat where older adults perform worse when it is implied that their memory is likely to be impaired (Krendl, Ambady, & Kensinger, 2015; Lamont, Swift, & Abrams, 2015). Such expectancy effects may be more likely when a younger experimenter tests older adults (McDaniel et al., 2008). In our study, U3A members who were of similar age as the older group collected data from that group, which may have reduced expectancy effects. We also took great care to ensure that the computerised task requirements and instructions would be suitable for the older sample, which may have enabled this group to perform well. Thus, the equivalent performance of our two age groups may be due to individual cognitive and/or situational factors.

Moreover, in this study, we added a novel manipulation to investigate whether external rewards would modulate incidental encoding of foils during a recognition test, and whether such effects would differ between young and older adults, as might be expected based on prior findings of reduced reward-sensitivity in older age (e.g. Geddes, et al., 2018; but see Mather & Schoeke, 2011). However, the monetary reward manipulation only affected recognition of words that had been intentionally encoded, and only in the young group. The younger adults showed enhanced recognition on the first test for old words that had been intentionally encoded in a study phase with reward compared to those encoded in a standard, non-rewarded study phase, whereas the older group showed no difference between reward and non-rewarded conditions on the first test. Consistent with these performance differences, younger adults reported higher levels of motivation for the rewarded study and test phases, coupled with lower effort to learn and retrieve non-rewarded words compared to older adults (see supplementary file for self-reports). Thus when compared to older adults, younger adults did not only seem to increase their encoding efforts because of a strong motivation to win a monetary reward, but they also reduced their retrieval efforts when there was no reward, perhaps to minimise interference (Hennessee, Castel, & Knowlton, 2017), or simply due to finding the tasks intrinsically less motivating than the older group (Murayama & Kuhbandner, 2011). Older adults however showed high recognition performance for both the rewarded and non-rewarded conditions, indicating that their memory ability was equal to the younger adults, but they were not incentivised by the monetary reward. Instead they reported being less motivated by monetary reward than the young adults, in line with the literature on the ‘undermining effect’ of monetary rewards (for review see Deci, Koestner, & Richard, 2001). The older adults were instead perhaps more motivated by an internal drive to succeed and perform well, as indicated by their high effort levels in the non-rewarded phase.

In contrast to the reward effects on the first test, neither group showed a difference in recognition accuracy on the final surprise test for foils that had been shown in a rewarded versus non-rewarded recognition test. We expected to find enhanced incidental encoding of rewarded compared to non-rewarded foils, since previous research has shown that incidental encoding is enhanced for stimuli associated with reward (Adcock et al., 2006; Mather & Schoeke, 2011; Spaniol, Bowen, Wegier & Grady, 2015). More generally, prior studies have found that factors that enhance intentional encoding also enhance incidental encoding of foils in a recognition test, for example self-referential processing, imagery, full attention, and survival processing (Bergström et al., 2015; Danckert, et al., 2011; Dudukovic et al., 2009; Nairne et al., 2015). We hypothesised that rewards during a recognition test could enhance encoding of foils either directly through automatic, dopamine-mediated brain mechanisms (e.g. Bäckman, et al., 2010) or more indirectly by increasing participants' use of strategic retrieval processes (e.g. Halsband et al., 2012), which could have resulted in more attentive/deeper processing of foils and thereby enhanced encoding. However, the results showed no support for either of these accounts. Since the effect of our reward manipulation on intentional encoding was rather subtle, the manipulation may not have been strong enough to produce differential encoding of foils. For example, even if participants sometimes used "deep" strategies to enhance encoding of words in the rewarded compared to the non-rewarded study phase, they may have used such strategies inconsistently (e.g. switching strategies across items within a study phase, or using similar strategies in both rewarded and non-rewarded study phases). Therefore, the reward study manipulation may not have provided a clear source context that participants could use to constrain retrieval towards during the subsequent test, thus leading to similar retrieval processing in both the rewarded and non-rewarded test, and no difference in incidental foil encoding. Furthermore, reward effects on memory tend to be larger when memory is tested after a delay of at least 24-hours, which indicates that reward influences memory retention/consolidation over time (Murayama & Kitagami, 2014; Spaniol, et al., 2014). Future research should test if rewarded foils are better remembered than non-rewarded foils after a longer delay, and using stronger reward manipulations. Based on the current null results of rewards on foil encoding in both groups, we cannot draw firm conclusions regarding the relationship between reward-induced motivation and strategic control processing as modulators of memory, and how these factors and their relationship are influenced by aging.

In sum, the main finding in this experiment was fully intact source-constrained retrieval processing in a group of socially active, mostly highly educated older adults who are members of a lifelong learning organisation. Our older group had equal memory accuracy to a group of university students who were on average 50 years younger, and showed equivalent engagement of complex retrieval strategies. Although we cannot provide a conclusive explanation for why our older sample of adults had intact memory performance when so many prior studies have found aging-related impairments on similar memory tasks, we believe this is an important demonstration that aging-related memory decline is far from inevitable and universal.

Acknowledgements: This research was part funded by a Public Engagement with Research Grant from the University of Kent, and part funded by a small grant from the Experimental Psychology Society. The research was conducted in collaboration with the University of the Third Age (U3A) Canterbury Branch research group, which consisted of Rona Hodges, Jennifer Harrop, Diane Billam, Jacky Moore, Karen Evans, Lynda Bonard. We thank Tara Lalani and Jen Huynh Nguyen for help with data collection. We have no conflict of interest to declare.

References

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. E. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. *Neuron*, *50*(3), 507–517. https://doi.org/10.1016/j.neuron.2006.03.036

Alban, M. W., & Kelley, C. M. (2012). Variations in constrained retrieval. *Memory & Cognition*, *40*(5), 681–692. https://doi.org/10.3758/s13421-012-0185-5

Bäckman, L., Lindenberger, U., Li, S.-C., & Nyberg, L. (2010). Linking cognitive aging to alterations in dopamine neurotransmitter functioning: recent data and future avenues *Neuroscience and Biobehavioural Reviews*, *34,* 670-677. https://doi.org/10.1016/j.neubiorev.2009.12.008

Bergström, Z. M., Vogelsang, D. A., Benoit, R. G., & Simons, J. S. (2015). Reflections of Oneself: Neurocognitive Evidence for Dissociable Forms of Self-Referential Recollection. *Cerebral Cortex, 25(9),* 2648–2657. https://doi.org/10.1093/cercor/bhu063

Castel, A. D., Farb, N. A. S., & Craik, F. I. M. (2007). Memory for general and specific value information in younger and older adults: Measuring the limits of strategic control. *Memory & Cognition*, *35*(4), 689–700. https://doi.org/10.3758/BF03193307

Castel, A. D., Murayama, K., Friedman, M. C., Mcgillivray, S., Link, I., Mcgil-Livray, S., … Sungkhasettee, V. (2013). Selecting Valuable Information to Remember: Age-Related Differences and Similarities in Self-Regulated Learning. *Psychology and Aging*, *28*(1), 232–242. https://doi.org/10.1037/a0030678

Chan, D., Shafto, M., Kievit, R., Matthews, F., Spink, M., Valenzuela, M., & Henson, R. N. (2018). Lifestyle activities in mid-life contribute to cognitive reserve in late-life, independent of education, occupation, and late-life activities. *Neurobiology of Aging*, *70*, 180–183. https://doi.org/10.1016/j.neurobiolaging.2018.06.012

Cohen, M.S., Cheng, L.Y., Paller, K.A., Reber, P.J. (2019). Separate memory-enhancing effects of reward and strategic encoding. *Journal of Cognitive Neuroscience, 31(11),* 1658-73. https://doi.org/10.1162/jocn_a_01438

Cohen, M. S., Rissman, J., Suthana, N. A., Castel, A. D., & Knowlton, B. J. (2014). Value-based modulation of memory encoding involves strategic engagement of fronto-temporal semantic processing regions. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(2), 578–592. doi: 10.3758/s13415-014-0275-x

Cohen, M. S., Rissman, J., Suthana, N. A., Castel, A. D., & Knowlton, B. J. (2016). Effects of aging on value-directed modulation of semantic network activity during verbal learning. *NeuroImage*, *125*, 1046–1062. https://doi.org/10.1016/j.neuroimage.2015.07.079

Cohn, M., Emrich, S. M., & Moscovitch, M. (2008). Age-Related Deficits in Associative Memory: The Influence of Impaired Strategic Retrieval. *Psychology and Aging*, *23*(1), 93–103. https://doi.org/10.1037/0882-7974.23.1.93

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Danckert, S. L., MacLeod, C. M., & Fernandes, M. A. (2011). Source-constrained retrieval influences the encoding of new information. *Memory & Cognition*, *39*(8), 1374–1386. https://doi.org/10.3758/s13421-011-0117-9

Deci, E. L., Koestner, R., & Richard, R. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research; Spring*, *71*(1), 1–

27. https://doi.org/10.3102/00346543071001001

Dudukovic, N. M., DuBrow, S., & Wagner, A. D. (2009). Attention during memory retrieval enhances future remembering. *Memory and Cognition*, *37,* 953-961. https://doi.org/10.3758/MC.37.7.953

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170–177. http://dx.doi.org/10.1037/1082-989X.1.2.170

Duverne, S., Motamedinia, S., & Rugg, M. D. (2009). Effects of age on the neural correlates of retrieval cue processing are modulated by task demands. *Journal of Cognitive Neuroscience, 21,* 1–17. https://doi.org/10.1162/jocn.2009.21001

Fraundorf, S. H., Hourihan, K. L., Peters, R. A., & Benjamin, A. S. (2019). Aging and recognition memory: *A meta-analysis. Psychological Bulletin, 145(4),* 339-371. http://dx.doi.org/10.1037/bul0000185

Geddes, M. R., Mattfeld, A. T., de los Angeles, C., Keshavan, A., & Gabrieli, J. D. E. (2018). Human aging reduces the neurobehavioral influence of motivation on episodic memory. *NeuroImage, 171,* 296-310. https://doi.org/10.1016/j.neuroimage.2017.12.053

Grady, C. L., McIntosh, A. R., Rajah M. N., Beig S., & Craik, F. I. (1999). The effects of age on the neural correlates of episodic encoding. *Cerebral Cortex, 9,* 805–14. https://doi.org/10.1093/cercor/9.8.805

Gray, S. J., & Gallo, D. A. (2015). Disregarding familiarity during recollection attempts: Content-specific recapitulation as a retrieval orientation strategy. *Journal of Experimental Psychology: Learning Memory and Cognition, 41*(1), 134–147. https://doi.org/10.1037/a0038363

Halamish, V., Goldsmith, M., & Jacoby, L. L. (2012). Source-constrained recall: Front-end and back-end control of retrieval quality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(1), 1-15. http://dx.doi.org/10.1037/a0025053

Halsband, T. M., Ferdinand, N. K., Bridger, E. K., & Mecklinger, A. (2012). Monetary rewards influence retrieval orientations. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(3), 430–445. https://doi.org/10.3758/s13415-012-0093-y

Hennessee, J. P., Castel, A. D., & Knowlton, B. J. (2017). Recognizing what matters: Value improves recognition by selectively enhancing recollection. *Journal of Memory and Language*, *94,* 195–205. https://doi.org/10.1016/j.jml.2016.12.004

Jacoby, L. L., Shimizu, Y., Daniels, K. A., & Rhodes, M. G. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review*, *12*(5), 852–857. https://doi.org/10.3758/BF03196776

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005). Age differences in depth of retrieval: Memory for foils. *Journal of Memory and Language*, *52*(4), 493–504. https://doi.org/10.1016/j.jml.2005.01.007

JASP Team (2019). JASP (Version 0.10.2). [Computer software]

Kantner, J., & Lindsay, D. S. (2013). Top-down constraint on recognition memory. *Memory & Cognition*, *41*(3), 465–479. https://doi.org/10.3758/s13421-012-0265-6

Kirchhoff, B. A., Anderson, B. A., Barch, D. M., & Jacoby, L. L. (2012). Cognitive and Neural Effects of Semantic Encoding Strategy Training in Older Adults. *NeuroImage*, *62*(3), 1956–1964. https://doi.org/10.1093/cercor/bhr129

Koen, J. D., & Yonelinas, A. P. (2016). Recollection, not familiarity, decreases in healthy ageing: Converging evidence from four estimation methods. *Memory*, *24(1)*, 75-88.

https://doi.org/10.1080/09658211.2014.985590

Krendl, A. C., Ambady, N., & Kensinger, E. A. (2015). The dissociable effects of stereotype threat on older adults' memory encoding and retrieval. *Journal of Applied Research in Memory and Cognition*, *4*(2), 103–109. https://doi.org/10.1016/J.JARMAC.2015.02.001

Lamont, R. A., Swift, H. J., & Abrams, D. (2015). A Review and Meta-Analysis of Age-Based Stereotype Threat: Negative Stereotypes, Not Facts, Do the Damage. *Psychology and Aging*, *30*(1), 180-193. https://doi.org/10.1037/a0038586

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* Gainesville, FL: University of Florida, The Center for Research in Psychophysiology.

Marsh, R. L., Meeks, J. T., Cook, G. I., Clark-Foos, A., Hicks, J. L., & Brewer, G. A. (2009). Retrieval constraints on the front end create differences in recollection on a subsequent test. *Journal of Memory and Language*, *61*(3), 470–479. https://doi.org/10.1016/j.jml.2009.06.005

Mather, M., & Schoeke, A. (2011). Positive outcomes enhance incidental learning for both younger and older adults. *Frontiers in Neuroscience*, *5*, 129. https://doi.org/10.3389/fnins.2011.00129 [doi]

McDaniel, M. A., Einstein, G. O., & Jacoby, L. L. (2008). New considerations in aging and memory: The glass may be half full. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 251-310). New York, NY, US: Psychology Press.

Mecklinger, A. (2010). The control of long-term memory: Brain systems and cognitive processes. *Neuroscience and Biobehavioral Reviews*, *34*, 1055–1065. https://doi.org/10.1016/j.neubiorev.2009.11.020

Morcom, A. M. (2016). Mind Over Memory. *Current Directions in Psychological Science*, *25*(3), 143–150. https://doi.org/10.1177/0963721416645536

Morcom, A. M., & Rugg, M. D. (2004). Effects of age on retrieval cue processing as revealed by ERPs. *Neuropsychologia*, *42*, 1525–1542. https://doi.org/10.1016/j.neuropsychologia.2004.03.009

Murayama, K., & Kitagami, S. (2014). Consolidation power of extrinsic rewards: reward cues enhance long-term memory for irrelevant past events. *Journal of Experimental Psychology: General*, *143*(1), 15. http://dx.doi.org/10.1037/a0031992

Murayama, K., & Kuhbandner, C. (2011). Money enhances memory consolidation–But only for boring material. *Cognition*, *119*(1), 120–124. https://doi.org/10.1016/j.cognition.2011.01.001

Nairne, J. S., Pandeirada, J. N. S., VanArsdall, J. E., & Blunt, J. R. (2015). Source-constrained retrieval and survival processing. *Memory & Cognition*, *43*(1), 1–13. https://doi.org/10.3758/s13421-014-0456-4

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Nyberg, L., Lövdén, M., Riklund, K., Lindenberger, U., & Bäckman, L. (2012). Memory aging and brain maintenance. *Trends in Cognitive Sciences*, *16(5)*, 292-305. https://doi.org/10.1016/j.tics.2012.04.005

Nyberg, L., & Pudas, S. (2019). Successful Memory Aging. *Annual Review of Psychology*, *70*, 219–243. https://doi.org/10.1146/annurev-psych-010418

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E & Lindeløv, J. K. (2019). *PsychoPy2: Experiments in behavior made easy*. *Behavior Research Methods* (Vol. 51). https://doi.org/10.3758/s13428-018-01193-y

Peirce, J. W. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1–2), 8–13. https://doi.org/10.1016/J.JNEUMETH.2006.11.017

Rugg, M. D., & Wilding, E. L. (2000). Retrieval processing and episodic memory. *Trends in Cognitive Sciences, 4(3),* 108-115. https://doi.org/10.1016/S1364-6613(00)01445-5

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117(1),* 34-50. http://dx.doi.org/10.1037/0096-3445.117.1.34

Spaniol, J., Bowen, H. J., Wegier, P., & Grady, C. (2015). Neural responses to monetary incentives in younger and older adults. *Brain Research*, *1612*, 70–82. https://doi.org/10.1016/j.brainres.2014.09.063

Spaniol, J., Schain, C., & Bowen, H. J. (2014). Reward-enhanced memory in younger and older adults. *The Journals of Gerontology.Series B, Psychological Sciences and Social Sciences*, *69*(5), 730–740. https://doi.org/10.1093/geronb/gbt044 [doi]

Vogelsang, D. A., Bonnici, H. M., Bergström, Z. M., Ranganath, C., & Simons, J. S. (2016). Goal-directed mechanisms that constrain retrieval predict subsequent memory for new "foil" information. *Neuropsychologia*, *89,* 356–363. https://doi.org/10.1016/J.NEUROPSYCHOLOGIA.2016.07.016

Vogelsang, D. A., Gruber, M., Bergström, Z. M., Ranganath, C., & Simons, J. S. (2018). Alpha Oscillations During Incidental Encoding Predict Subsequent Memory For New

"Foil" Information. *Journal of Cognitive Neuroscience, 30,* 667-679. https://doi.org/10.1162/jocn_a_01234

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology, 100(3),* 426-432. http://dx.doi.org/10.1037/a0022790

Wilson, M. (1988). MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*(1), 6–10. https://doi.org/10.3758/BF03202594

Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H. J., and Duzel, E. (2005). Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron, 45,* 459–467. https://doi.org/10.1016/j.neuron.2005.01.010

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46(3),* 441-517. http://dx.doi.org/10.1006/jmla.2002.2864

Zawadzka, K., Hanczakowski, M., & Wilding, E. L. (2017). Late consequences of early selection: When memory monitoring backfires. *Journal of Memory and Language*, *92*, 114–127. https://doi.org/10.1016/j.jml.2016.06.003

Tables

Table 1. Performance on the first recognition tests in the LOP phase. The table shows proportion accurate responses, discrimination (*Pr*), response bias (*Br*) and mean reaction times, and split by age group and type of preceding encoding task.

	Recognition accuracy				Discrimination (Pr)		Response Bias (Br)		Reaction Times (ms)			
	Old Words		New Words						Old Words		New Words	
	Deep	Shallow	Deep	Shallow	Deep	Shallow	Deep	Shallow	Deep	Shallow	Deep	Shallow
Younger Adults	.94 (.07)	.57 (.19)	.92 (.13)	.83 (.14)	.83 (.14)	.38 (.20)	.49 (.27)	.31 (.19)	871 (146)	1131 (317)	984 (185)	1110 (264)
Older Adults	.93 (.09)	.55 (.18)	.92 (.11)	.82 (.14)	.81 (.14)	.36 (.18)	.51 (.24)	.29 (.17)	1165 (359)	1567 (413)	1399 (437)	1454 (421)

Note: Values shown are Means (SD).

Table 2. Performance on the first recognition tests in the Reward phase. The table shows proportion accurate responses, discrimination (*Pr*), response bias (*Br*) and mean reaction times, and split by age group and type of preceding encoding task.

	Recognition accuracy				Discrimination (Pr)		Response Bias (Br)		Reaction Times (ms)			
	Old Words		New Words						Old Words		New Words	
	Reward	No Reward	Reward	No Reward	Reward	No Reward	Reward	No Reward	Reward	No Reward	Reward	No Reward
Younger Adults	.78 (.21)	.69 (.22)	.86 (.16)	.83 (.16)	.60 (.27)	.50 (.27)	.42 (.23)	.35 (.23)	1014 (239)	1020 (256)	1076 (255)	1005 (163)
Older Adults	.76 (.16)	.79 (.14)	.86 (.11)	.87 (.12)	.60 (.19)	.64 (.21)	.41 (.24)	.39 (.18)	1350 (413)	1348 (372)	1390 (397)	1433 (442)

Note: Values shown are Means (SD).

Figures

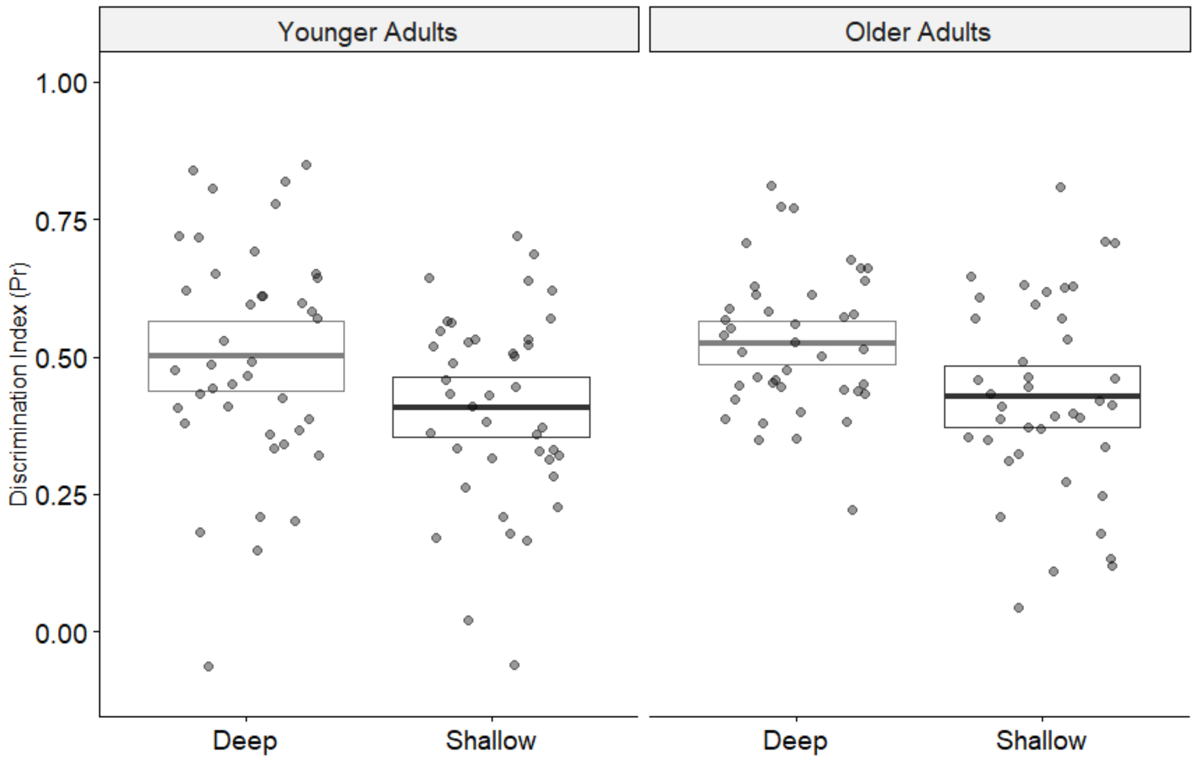

Figure 1. Discrimination scores (*Pr*) for Deep and Shallow foil word recognition on the final surprise test, split by Age Group. The dot plot shows each individual score, with the mean value and confidence interval plotted as a centre line and box respectively. Scores within groups have been randomly scattered across the x-axis for visualisation purposes.

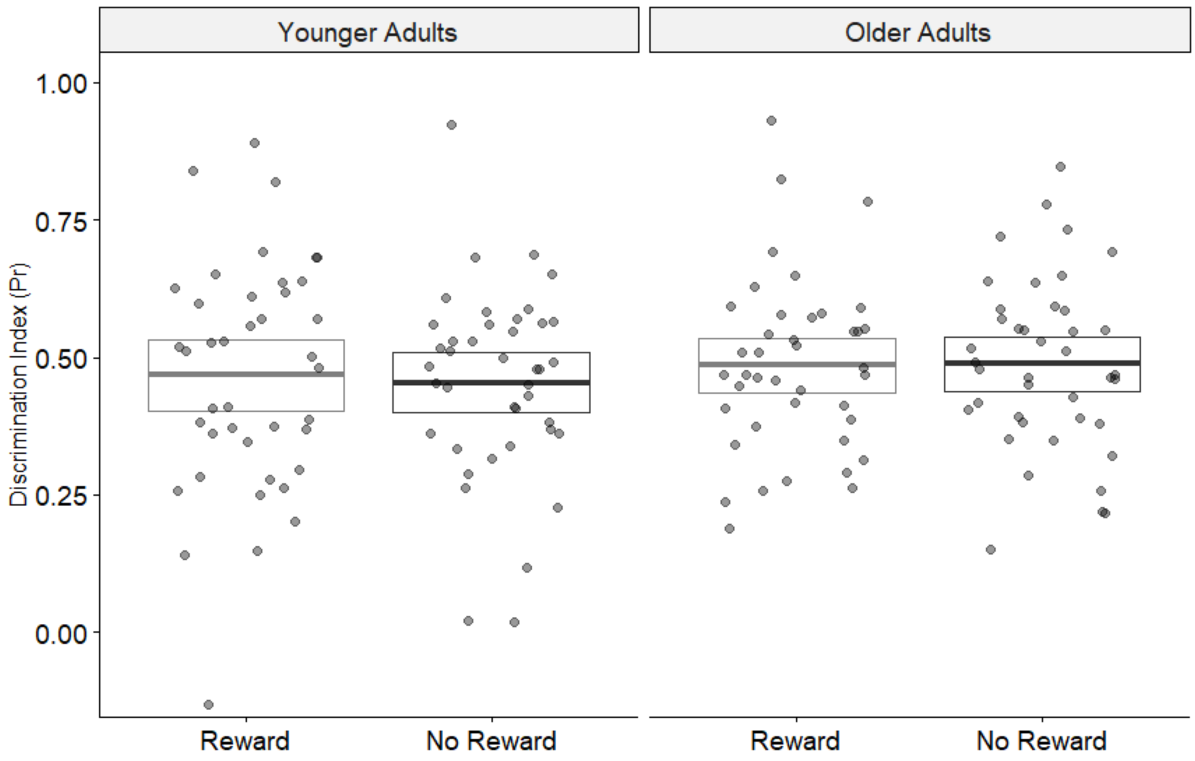

Figure 2. Discrimination scores (*Pr*) for Reward and No Reward foil word recognition on the final surprise test, split by Age Group. The dot plot shows each individual score, with the mean value and confidence interval plotted as a centre line and box respectively.