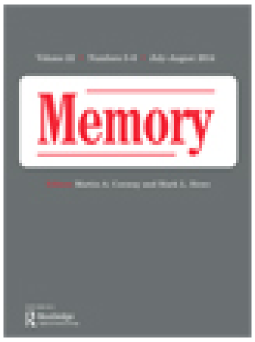

Memory

ISSN: 0965-8211 (Print) 1464-0686 (Online) Journal homepage: http://www.tandfonline.com/loi/pmem20

Word frequency influences on the list length effect and associative memory in young and older adults

Stephen P. Badham, Cora Whitney, Sumeet Sanghera & Elizabeth A. Maylor

MEMORY, 2016
http://dx.doi.org/10.1080/09658211.2016.1224358

Routledge
Taylor & Francis Group

Word frequency influences on the list length effect and associative memory in young and older adults

Stephen P. Badham[a], Cora Whitney[b], Sumeet Sanghera[b] and Elizabeth A. Maylor[b]

[a]College of Business Law & Social Sciences, Nottingham Trent University, Burton Street, Nottingham, UK; [b]Department of Psychology, University of Warwick, Coventry, UK

ABSTRACT
Many studies show that age deficits in memory are smaller for information supported by pre-experimental experience. Many studies also find dissociations in memory tasks between words that occur with high and low frequencies in language, but the literature is mixed regarding the extent of word frequency effects in normal ageing. We examined whether age deficits in episodic memory could be influenced by manipulations of word frequency. In Experiment 1, young and older adults studied short and long lists of high- and low-frequency words for free recall. The list length effect (the drop in proportion recalled for longer lists) was larger in young compared to older adults and for high- compared to low-frequency words. In Experiment 2, young and older adults completed item and associative recognition memory tests with high- and low-frequency words. Age deficits were greater for associative memory than for item memory, demonstrating an age-related associative deficit. High-frequency words led to better associative memory performance whilst low-frequency words resulted in better item memory performance. In neither experiment was there any evidence for age deficits to be smaller for high- relative to low-frequency words, suggesting that word frequency effects on memory operate independently from effects due to cognitive ageing.

ARTICLE HISTORY
Received 24 March 2016
Accepted 9 August 2016

KEYWORDS
Ageing; memory; word frequency; list length effect; associative deficits

The frequency at which a word is encountered in language has a strong influence on an individual's ability to name, identify, recognise and recall that word, and measures of word frequency are considered to be one of the most important variables in word processing and memory (Brysbaert & New, 2009). High-frequency words – words that occur often in language – are perceived and produced more rapidly than low-frequency words (Balota & Chumbley, 1984; Jescheniak & Levelt, 1994; Johnston & Barry, 2006). In free recall tests of memory, participants often perform better with high-frequency words than with low-frequency words (Balota & Neely, 1980; Gregg, 1976). In contrast, when memory is tested via recognition, low-frequency words benefit from higher hit rates to old items and lower false alarm rates to new items compared to high-frequency words (Glanzer & Adams, 1985; Yonelinas, 2002). High-frequency words are typically learned earlier in life than low-frequency words, with a negative correlation between age of acquisition and word frequency (Johnston & Barry, 2006).

Word frequency effects can vary across different groups of individuals. For example, patients with amnesia (Ellis, Miller, & Sin, 1983), schizophrenia (Huron et al., 1995), and Alzheimer's disease (Balota, Burgess, Cortese, & Adams, 2002) have responded differently to manipulations of word frequency. More extensive research has investigated word frequency effects with young and healthy older adults, which is the focus of the current study. The high-frequency word advantage in perception and production can be more extreme in older adults compared to young adults (Rayner, Reichle, Stroud, & Williams, 2006; Spieler & Balota, 2000), but not always (Allen, Madden, Weber, & Groth, 1993). In recognition memory tests, the low-frequency advantage has been shown to be smaller in older adults compared to young adults (Balota et al., 2002). Research has also shown a larger high-frequency word advantage using cumulative learning (tested by free recall) for older adults compared to young adults (Almond, Morrison, & Moulin, 2013; see also Almond & Morrison, 2014, for a qualitatively similar pattern with respect to age of acquisition). Therefore, it can be seen that increasing age can magnify word frequency effects, attenuate them, or leave them unchanged.

The current article focuses on word frequency effects on memory with young and older adults, and considers how theories explaining word frequency effects may relate to theories explaining age-related memory deficits. The key difference between high- and low-frequency words is the

CONTACT Stephen P. Badham stephen.badham@ntu.ac.uk
This article was originally published with errors. This version has been corrected. Please see Erratum (http://dx.doi.org/10.1080/09658211.2016.1232934).

degree of exposure to those words in language.[1] There is much evidence in the literature that older adults' memory is improved more than young adults' memory by the use of material that has been experienced more in pre-experimental environments (Umanath & Marsh, 2014). This has been shown across a variety of paradigms, largely in the domain of associative memory – for example, memory for plausible compared to implausible grocery prices (Castel, 2005), memory for typical compared to atypical actions in scripts (Hess, 1985), and memory for associations between semantically related compared to unrelated words (Badham, Estes, & Maylor, 2012). However, recent research from our laboratory (Badham, Hay, Foxon, Kaur, & Maylor, 2016; Badham & Maylor, 2015) and others (Mohanty, Naveh-Benjamin, & Ratneshwar, 2016) has shown that prior knowledge does not alleviate age-related memory deficits under certain conditions. This poses a challenge to existing theory which is explored further in the current article through examination of age differences in established memory paradigms that are known to be influenced by word frequency.

To explain word frequency effects in free recall, it has been hypothesised that remembering order is important, and that processing items occurs to the detriment of processing order (DeLosh & McDaniel, 1996). Therefore, as low-frequency words take longer to process (see above) there are less cognitive resources available to process order information, leading to poorer free recall for low-frequency compared to high-frequency words (DeLosh & McDaniel, 1996; Schmidt, 2008). The low-frequency word advantage in recognition has been explained in terms of discrimination between pre-experimental knowledge and episodic memory for the recognition memory test stimuli (Clark & Burchett, 1994; Schmidt, 2008). The degree of pre-experimental exposure to low-frequency words is lower than for high-frequency words – therefore it is easier to discriminate between experimental and pre-experimental exposure for low-frequency words during recognition. These influences of word frequency on recall and recognition are explored across two experiments in the current study.

It has been hypothesised that prior knowledge is easy to access and can therefore support the formation of episodic memory in older adults (Umanath & Marsh, 2014). Prior knowledge can be seen as a form of environmental support during memory tasks (Bäckman & Herlitz, 1990; Naveh-Benjamin, Craik, Guez, & Kreuger, 2005): Craik (1986) argued that environmental cues can be utilised to support effective encoding and retrieval by minimising the amount of cognitive resources necessary in memory tasks, with greater environmental support leading to smaller age deficits in memory. These ideas are aligned with the hypothesis that high-frequency words result in better free recall because they are processed more easily during encoding (DeLosh & McDaniel, 1996). Therefore, Experiment 1 will aim to establish if word frequency effects can alleviate age deficits commonly found in free recall (Zacks, Hasher, & Li, 2000). In Experiment 2, the link between word frequency effects and environmental support will be explored in the context of associative memory, which is known to be especially impaired in older adults (Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008). We hypothesise that the increased associability of high-frequency words compared to low-frequency words (Clark, 1992; Clark & Burchett, 1994) may alleviate age deficits in associative memory as has been found in other experiments where associations have been facilitated by prior knowledge (Badham et al., 2012; Castel, 2005; Naveh-Benjamin, Hussain, Guez, & Bar-On, 2003).

Experiment 1

Experiment 1 measured the list length effect with high- and low-frequency words and with young and older adults. The list length effect refers to the observation that the probability of recalling an item on a list is reduced as the total number of items in that list increases (Shiffrin, 1970). This provides a paradigm that manipulates the amount of interference between items. High-frequency words are considered to be more interconnected in memory (Almond et al., 2013; Steyvers & Tenenbaum, 2005) and it should therefore be the case that the list length effect is exaggerated by interference between high-frequency words. That is, the free recall advantage for high-frequency words should be smaller for longer lists, where interference between items is increased (cf. Schmidt, 2008). Furthermore, the free recall advantage for high-frequency words has been hypothesised to be driven by organisational processing (DeLosh & McDaniel, 1996). It may also be the case that greater organisational processing of high-frequency words compared to low-frequency words will be particularly disrupted by longer lists.

To our knowledge, only five studies have varied list length in free recall experiments with young and older adults. Craik (1968), Craik and Masani (1967), and Kahana, Dolan, Sauder, and Wingfield (2005) found that older adults have significantly greater memory deficits compared to young adults for long lists compared to short lists, and Cohen, Sandler, and Schroeder (1987) found a numerical trend in the same direction. These results appear to be aligned with a resource deficit hypothesis of cognitive ageing (although see our General Discussion) where age deficits are larger for the more difficult condition, namely long lists (Cohen et al., 1987). The inhibitory deficit hypothesis (Hasher & Zacks, 1988) also states that older adults have difficulty inhibiting irrelevant information during cognitive tasks. For example, older adults show more cue overload in fan effect studies (Gerard, Zacks, Hasher, & Radvansky, 1991; Radvansky, Zacks, & Hasher, 1996), suggesting that they may experience more interference between items when encoding and retrieving long lists. It may therefore be the case that less interconnected, low-frequency words may alleviate age deficits in inhibition and particularly so for long lists.

In apparent contrast to the above, Smith (1979) found a significantly larger list length effect in young adults compared to older adults (although see Experiment 1 discussion below for a review of the inconsistencies). He argued that longer lists disrupted organisational processes in young adults, which lowered their performance for long lists, whereas older adults may have used fewer organisational strategies overall, limiting the amount of disruption caused by long lists. DeLosh and McDaniel (1996) demonstrated greater organisational memory with high-frequency words, so it may be the case that a greater list length effect in young adults would only occur for high-frequency words where organisation is more likely. Alternatively, high-frequency words may facilitate organisational processing in older adults, aligning their list length effect with young adults for high- but not low-frequency words.

Method

Design

Young and older adults memorised word lists containing high- or low-frequency words; lists were either long (30 words) or short (15 words). The factors were age (young, older; between participants), word frequency (high, low; within participants), and list length (long, short; within participants).

Participants

Thirty young adults (17 female) aged 20–29 years ($M = 22.4$, $SD = 2.2$) and 30 healthy older adults (18 female) aged 65–82 years ($M = 72.5$, $SD = 4.8$) took part in the experiment. Young and older participants were recruited from the local community and received no incentives for participation. All participants were native English speakers. Both age groups reported mean levels of self-rated health equivalent to "good" on a five-point scale from 1 = "very poor" to 5 = "very good" ($M_{young} = 4.30$, $SD_{young} = 0.75$; $M_{older} = 3.93$, $SD_{older} = 0.79$; $t(58) = 1.85$, $p = .069$). Young and older participants differed significantly in their years of education, $t(48.85) = 5.11$, $p < .001$ ($M_{young} = 17.4$, $SD_{young} = 2.4$; $M_{older} = 13.2$, $SD_{older} = 3.8$). To assess cognitive functioning, participants completed the Digit Symbol Substitution test from the Wechsler Adult Intelligence Scale – Revised (Wechsler, 1981) as a measure of processing speed,[2] and the multiple choice part of the Mill Hill vocabulary test (Raven, Raven, & Court, 1988) as a measure of crystallised intelligence. The results were consistent with the literature (e.g., Hoyer, Stawski, Wasylyshyn, & Verhaeghen, 2004; Salthouse, 2010; Verhaeghen, 2003): young adults performed better than older adults at the speed task, $t(58) = 8.57$, $p < .001$ ($M_{young} = 73.90$, $SD_{young} = 13.05$; $M_{older} = 48.35$, $SD_{older} = 9.80$), and older adults performed better than young adults at the vocabulary task, $t(45.10) = 4.99$, $p < .001$ ($M_{young} = 18.23$, $SD_{young} = 2.49$; $M_{older} = 22.93$, $SD_{older} = 4.52$).

Materials

The English Lexicon Project (Balota et al., 2007) was used to select 150 high-frequency and 150 low-frequency nouns. Initially, all two-syllable words with 5–15 letters were obtained (15,523 words) and the top and bottom 999 words based on log SUBTL frequency (Brysbaert & New, 2009) were used to produce 150 high- and 150 low-frequency words, respectively. The final 300 chosen words were all two syllables long, 7–9 letters in length and none of them had any orthographic or phonographic neighbours.

From this set of 300 words, 90 high- and 90 low-frequency words were selected such that there were equal numbers of seven-letter (33 per group), eight-letter (35 per group), and nine-letter (22 per group) words. For each of the three word lengths, there was also the same amount of plurals for high- and low-frequency groups (word lengths seven, eight, and nine letters each had just one plural per frequency group). A non-significant *t*-test indicated that the number of phonemes remained matched between the high- and low-frequency word groups, $t(178) = 1.38$, $p = .17$ (high frequency, $M = 6.28$, $SD = 1.00$; low frequency, $M = 6.50$, $SD = 1.14$).

The word frequency measures showed significant differences between high- and low-frequency words after the exclusions. For log HAL frequency (Lund & Burgess, 1996), $t(178) = 22.9$, $p < .001$ (high frequency, $M = 8.91$, $SD = 1.50$; low frequency, $M = 3.73$, $SD = 1.54$). For log SUBTL frequency (Brysbaert & New, 2009), $t(116.11) = 46.69$, $p < .001$ (high frequency, $M = 2.84$, $SD = 0.44$; low frequency, $M = 0.57$, $SD = 0.17$). Finally, for log SUBTL contextual diversity, which accounts for multiple uses of a word within a single context (Brysbaert & New, 2009), $t(126.31) = 50.29$, $p < .001$ (high frequency, $M = 2.63$, $SD = 0.36$; low frequency, $M = 0.53$, $SD = 0.17$).

In addition to the above measures, the mean age of acquisition in years for high- and low-frequency words was calculated using data from Kuperman, Stadthagen-Gonzalez, and Brysbaert (2012). These statistics were generated on a post hoc basis and were not used for data selection. For high-frequency words, 86 out of 90 words existed on the database, $M = 8.13$, $SD = 2.21$; for low-frequency words, 82 out of 90 words existed on the database, $M = 11.03$, $SD = 2.05$. Age of acquisition significantly differed between high- and low-frequency words, $t(166) = 8.80$, $p < .001$.

The 300 words were also rated by 16 independent volunteers (eight undergraduate students and eight older adults) on seven-point scales for imageability (from 1 = low imagery to 7 = high imagery) and valence (from 1 = negative to 7 = positive). Cronbach's alpha was used to measure the internal consistency of the ratings given by the different participants. For imageability ratings, Cronbach's alpha was 0.93, and for valence ratings, Cronbach's alpha was 0.92, indicating high concordance between the different raters (Table A1 of the appendix shows all 300 words from which stimuli for Experiments 1 and 2 were selected).

Mean imageability across the 16 raters was calculated for each of the chosen 90 high-frequency and 90 low-frequency words. No significant difference was found between the high- ($M = 4.04$, $SD = 1.34$) and low-frequency words ($M = 4.14$, $SD = 1.23$), $t < 1$. For valence, again the overall means did not differ significantly between the high- ($M = 4.15$, $SD = 1.01$) and low-frequency words ($M = 3.94$, $SD = 0.61$), $t(178) = 1.66$, $p > .05$. However, with regard to valence, perhaps of more relevance is whether the words were neutral or valenced (either positively or negatively). The vast majority of the 180 words (79%) were neutral (with mean ratings between 3 and 5); nevertheless, calculating each word's absolute distance from the mid-point (4) revealed a significant difference between high- ($M = 0.82$, $SD = 0.61$) and low-frequency words ($M = 0.45$, $SD = 0.41$), $t(154.92) = 4.75$, $p < .001$, such that high-frequency words were more valenced than were low-frequency words. Crucially, however, there appears to be no evidence of any consistent age-related differences – for either recall or recognition – in the benefit to memory of valenced over neutral material (e.g., Kensinger, 2008; Mather & Knight, 2005). Thus, any age by frequency interactions observed here cannot be attributable to this valence difference between word sets.

During the experiment, words were displayed using E-Prime 2.0 (Psychology Software Tools, Pittsburgh, PA) on a laptop computer, using lowercase in a size 40 font that corresponded to a letter height of approximately 1 degree of viewing angle.

Procedure

During encoding, participants were shown a list of words presented sequentially at a rate of one word every 3 seconds. After presentation there was a 30-second delay where participants were asked to count backwards in threes from a three-digit number. Following this, participants completed a free-recall memory test where they were given unlimited time to recall verbally any words they could remember from the list whilst the experimenter wrote down their responses. There were four conditions that were completed by each participant twice (the speed and vocabulary tests were administered in the middle of the session after the participant had completed one of each of the four conditions). Word frequency of the stimuli (high- and low-frequency words) was crossed with list length (either 15 or 30 words presented at encoding). Words were selected randomly from the high- and low-frequency lists for each participant and no individual was shown the same word twice such that all 180 words were used. The order of the conditions was counterbalanced.

Results

Throughout the article, standard null hypothesis tests are accompanied by an estimated Bayes Factor implemented through JASP computer software (Love et al., 2015). The Bayes Factor (BF_{10}) provides an odds ratio for the alternative/null hypotheses (values < 1 favour the null hypothesis and values > 1 favour the alternative hypothesis). For example, a BF_{10} of 0.40 would indicate that the null hypothesis is 2.5 times more likely than the alternative hypothesis (see Jarosz & Wiley, 2014).

The data were analysed in terms of accuracy, measured as the proportion of words that were correctly recalled from a given list (see Figure 1 for overall means). The data from the two tests for each condition were averaged together.[3] A 2 (Age: young, older) × 2 (Word Frequency: high, low) × 2 (List Length: long, short) repeated measures ANOVA showed all main effects to be significant: Young adults performed better than older adults, $F(1, 58) = 39.29$, $MSE = 0.04$, $p < .001$, $\eta_p^2 = .404$, $BF_{10} = 5.226 \times 10^5$. High-frequency words were recalled better than low-frequency words, $F(1, 58) = 26.17$, $MSE = 0.01$, $p < .001$, $\eta_p^2 = .311$, $BF_{10} = 1.882 \times 10^5$. A higher proportion of words were recalled for short than for long lists, $F(1, 58) = 141.34$, $MSE = 0.01$, $p < .001$, $\eta_p^2 = .709$, $BF_{10} > 10^{15}$. There was an interaction between age and list length, $F(1, 58) = 6.38$, $MSE = 0.01$, $p < .05$, $\eta_p^2 = .099$, $BF_{10} = 10.15$, with a smaller list length effect in older adults. The simple effects of list length (Bonferroni corrected here and throughout the article) were significant for both young ($p < .001$) and older ($p < .001$) adults. And the simple effects of age were significant for both short ($p < .001$) and long ($p < .001$) lists. There was also an interaction between word frequency and list length, $F(1, 58) = 11.02$, $MSE = 0.003$, $p < .005$, $\eta_p^2 = .160$, $BF_{10} = 15.20$, with a larger list length effect for high-frequency words. Simple effects of word frequency were significant for both short ($p < .001$) and long ($p = .022$) lists. And the simple effects of list length were significant for both high- and low-frequency word lists (both $ps < .001$).There was no interaction between age and word frequency, $F < 1$, $BF_{10} = 0.730$. Crucially, age differences in the list length effect were not influenced by word frequency as the triple interaction between age, word frequency, and list length was nonsignificant, $F < 1$, $BF_{10} = 0.867$.

Discussion

Despite main effects of age, word frequency, and list length, there were no significant interactions between age and word frequency. The data therefore indicate that for free recall, effects due to cognitive ageing may act independently from effects due to word frequency. The interaction between word frequency and list length was present in the predicted direction – high-frequency words suffered more from increased list length than did low-frequency words. As high-frequency words are more interconnected (Steyvers & Tenenbaum, 2005) they might interfere with each other more (Schmidt, 2008), which would be particularly detrimental to memory when interference is also higher due to longer lists. This view is

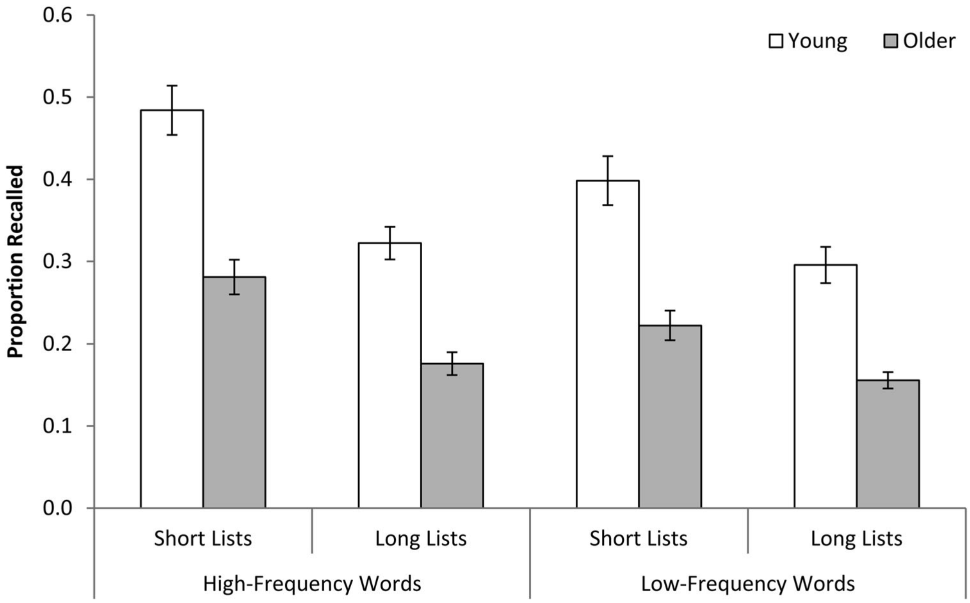

Figure 1. Proportion of words recalled by young and older adults, for short and long word lists, and for high- and low-frequency words in Experiment 1. Error bars are ± 1*SE*.

aligned with conclusions from Smith (1979) who argued that organisational processing may be disrupted more in long lists, and the work of DeLosh and McDaniel (1996) who showed that organisational processing is greater with high-frequency words (i.e., for high-frequency words, there is more organisational processing to be disrupted by long lists).

DeLosh and McDaniel (1996) also showed a different pattern of results in free recall using mixed lists of high- and low-frequency words – in mixed lists, a low-frequency advantage was found. They hypothesised that order memory in this case was "equivalent across item type" (p. 1144), which disrupted the usual high-frequency advantage in free recall in favour of an advantage for the more distinctive/richly encoded low-frequency words. In terms of the list length effect with mixed lists, we would expect the low-frequency advantage to be smaller for short lists than for long lists. This is because all items would be more distinctly encoded in short lists, minimising the distinctiveness advantage available to low-frequency words. Given the equivalent effects of frequency across age in the current study, we would expect similar effects of mixed lists in young and older adults.

The current data showed a smaller list length effect in older adults compared to young adults. This finding is consistent with that of Smith (1979). Other studies appeared to show the opposite, with larger age deficits for longer lists (Cohen et al., 1987; Craik, 1968; Craik & Masani, 1967; Kahana et al., 2005). Smith compared his result to Craik's study and argued that the opposite age by list length effects may be due to the method of analysis. Like the current study, Smith assessed performance in terms of the proportion correct, whereas many of the other studies analysed raw scores (i.e., absolute numbers of items correctly recalled). It has been argued that proportion correct is the preferable measure for such tasks (Shiffrin, 1971; Smith, 1979). If we consider the actual data from these studies, the results are less contradictory. All of the studies showed a larger difference in absolute recall performance between short and long lists for young adults which resulted in larger age deficits for long lists (and the same is true of the current data as presented in Table A2 of the appendix – for short lists older adults recalled 2.8 fewer words on average compared to young adults but for long lists older adults recalled 4.3 fewer words on average compared to young adults). This means that the age deficits are larger for long lists in absolute terms but this typically translates into a smaller list length effect (which is examined in proportional terms) for older adults.

Experiment 2

Experiment 2 tested item and associative recognition of high- and low-frequency words with young and older adults. In Experiment 1, the free-recall data showed that young and older adults responded similarly to manipulations of word frequency. The same may not be true for recognition memory tests where word frequency effects are reversed, with better recognition performance for low- than for high-frequency words (Gregg, 1976). The introduction showed how word frequency effects are hypothesised to work differently for recognition and for recall. Therefore, recognition offers another paradigm by which to observe potential age differences in word frequency effects.

For item recognition, low-frequency words are remembered better than high-frequency words; however, for associative recognition, high-frequency word pairings are remembered better than low-frequency word pairings (Clark, 1992; Clark & Burchett, 1994). Item and relational processing are considered to make independent contributions to retrieval (Hunt & McDaniel, 1993). For recognition of items, item-specific processing is important with low-frequency words encouraging discriminative processing of items, leading to a low-frequency word advantage. However, for associative recognition, relational processing is more important, which leads to a high-frequency word advantage as high-frequency words are easier to associate (cf. Clark & Burchett, 1994; Yonelinas, 2002).

Item and relational/associative memory performance is reliably dissociated for different age groups (Old & Naveh-Benjamin, 2008). Naveh-Benjamin (2000) proposed an associative deficit hypothesis whereby older adults are seen to have larger deficits for associative than for item memory compared to young adults. Of particular relevance to the current study is the common finding that the age-related associative deficit can be alleviated under conditions that draw upon pre-experimental knowledge: Age deficits in associative memory become smaller when knowledge and experience can support the formation of associations (Badham et al., 2012; Naveh-Benjamin et al., 2003). We therefore predict that the high-frequency word advantage in associative memory should be particularly beneficial to older adults as high-frequency words facilitate relational memory (Clark & Burchett, 1994). Additionally, we manipulated encoding time, under the assumption that longer encoding would benefit low-frequency words more than high-frequency words as the former take longer to process.

Method

Design

Young and older adults encoded word pairs comprising high- or low-frequency words, encoding speed was manipulated, and separate memory tests were conducted to measure recognition of items and associations between items. The factors were age (young, older; between participants), word frequency (high, low; within participants), presentation rate (fast, slow; within participants), and test type (item, associative; within participants).

Participants

Thirty-two young adults (15 female) aged 20–21 years ($M = 20.5$, $SD = 0.5$) and 32 healthy older adults (15 female) aged 65–83 years ($M = 73.5$, $SD = 5.0$) took part in the experiment. None of the participants had taken part in Experiment 1. Young participants were undergraduate students at the University of Warwick. Older participants were recruited from the local community. They were offered no financial incentives for participation. All participants were native English speakers. Self-rated health was again equivalent to “good” in both age groups ($M_{young} = 4.63$, $SD_{young} = 0.49$; $M_{older} = 4.13$, $SD_{older} = 0.87$); young adults’ ratings were significantly higher, $t(62) = 2.83$, $p < .01$. Young and older participants differed significantly in their years of education, $t(37.33) = 3.26$, $p < .01$ ($M_{young} = 16.2$, $SD_{young} = 1.4$; $M_{older} = 13.8$, $SD_{older} = 3.8$). The same measures of cognitive speed and vocabulary were used as in Experiment 1: young adults performed better than older adults at the speed task, $t(62) = 11.07$, $p < .001$ ($M_{young} = 71.78$, $SD_{young} = 11.20$; $M_{older} = 42.19$, $SD_{older} = 10.17$), and older adults performed better than young adults at the vocabulary task, $t(62) = 4.40$, $p < .001$ ($M_{young} = 18.09$, $SD_{young} = 2.61$; $M_{older} = 23.00$, $SD_{older} = 5.74$).

Materials

The same 300 words from Experiment 1 (see earlier) were used as a pool from which to select 120 high-frequency and 120 low-frequency words. In these 240 words, across the high- and low-frequency groups there were equal numbers of seven-letter (45 per group), eight-letter (45 per group), and nine-letter (30 per group) words. For each of the three word lengths, there was also the same amount of plurals for high- and low-frequency groups (word lengths seven, eight, and nine letters: 13, 11, and 9 plurals per frequency group, respectively). A non-significant *t*-test indicated that the number of phonemes remained matched between the high- and low-frequency word groups, $t(238) = 1.33$, $p = .18$ (high frequency, $M = 6.35$, $SD = 1.06$; low frequency, $M = 6.54$, $SD = 1.17$).

All word frequency measures remained significantly different for the final set of high- and low-frequency words. For log HAL frequency, $t(238) = 25.02$, $p < .001$ (high frequency, $M = 8.97$, $SD = 1.44$; low frequency, $M = 3.94$, $SD = 1.67$). For log SUBTL frequency, $t(162.00) = 55.92$, $p < .001$ (high frequency, $M = 2.81$, $SD = 0.40$; low frequency, $M = 0.59$, $SD = 0.17$). Finally, for log SUBTL contextual diversity, $t(179.81) = 60.93$, $p < .001$ (high frequency, $M = 2.61$, $SD = 0.33$; low frequency, $M = 0.55$, $SD = 0.17$).

The mean age of acquisition in years for the words used was calculated as described in Experiment 1. For high-frequency words, 88 out of 120 words existed on the database, $M = 8.12$, $SD = 2.20$; for low-frequency words, 82 out of 120 words existed on the database, $M = 11.02$, $SD = 2.05$. Age of acquisition significantly differed between high- and low-frequency words, $t(168) = 8.90$, $p < .001$.

As in Experiment 1, there was no difference in imageability between the 120 high- ($M = 3.91$, $SD = 1.32$) and 120 low-frequency words ($M = 4.16$, $SD = 1.20$), $t(238) = 1.53$, $p > .05$. For valence, in this case the small overall difference between high- ($M = 4.16$, $SD = 0.97$) and low-frequency words ($M = 3.95$, $SD = 0.62$) just reached significance, $t(202.01) = 2.02$, $p < .05$. Note again, however, that 79% of the 240 words were neutral (mean ratings of 3–5); like Experiment 1, the high-frequency words were slightly but significantly more valenced (in terms of absolute distance from neutral) than were the low-frequency words (high, $M = 0.79$, $SD = 0.59$; low, $M = 0.46$, $SD = 0.42$), $t(214.63) = 4.96$, $p < .001$ (see earlier for why this difference is not critical in the current context).

During the experiment, words were displayed using E-Prime 2.0 on a laptop computer, using lowercase in a size 40 font that corresponded to a letter height of approximately 1 degree of viewing angle.

Procedure

Participants were asked to remember 23 pairs of words presented on a computer screen. There was then a 30-second delay period where participants were required to count backwards in threes from a number randomly generated between 80 and 100. Participants then completed item and associative memory tests. Memory was only tested for the middle 21 word pairs, with the first and 23rd pairs serving as buffers that were not seen again.

For the item memory test, participants viewed one word at a time, and they had to respond with a button press as to whether the word was seen earlier during encoding or not. Participants pressed "J" with their right index finger if they thought the word was seen earlier, or "F" with their left index finger if they did not remember seeing the word earlier. Half of the words were taken from the encoding phase and half were new words introduced for the first time during the test phase. Immediately after each button press the next test word was shown. In total there were 14 old and 14 new words in the item test, using a third of the non-buffer encoded stimuli for item testing.

For the associative test, participants viewed intact word pairs that were presented exactly the same as during encoding, or recombined word pairs that presented words taken from separate pairs. When words were recombined, words initially presented on the left were still presented on the left and similarly for words presented on the right (i.e., words never changed sides). The method for responding was similar to the item test – participants used their index fingers on the "J" and "F" keys to indicate intact and recombined pairs, respectively. Immediately after each button press the next test pair was shown. In total there were seven intact and seven recombined pairs using two-thirds of the non-buffer encoded stimuli for associative testing. Each encoded word was only tested once as all encoding stimuli were either presented in the item test or in the associative test.

Participants repeated the procedure for four separate conditions. The conditions crossed fast or slow presentation rate (2 seconds per pair versus 4 seconds per pair) with word type (high frequency versus low frequency). Lure stimuli (new stimuli not seen during encoding) were always of the same word type as the condition (i.e., with high-frequency words displayed at encoding, item lures were also high frequency, and similarly for low-frequency words). The selection of stimuli was as random as possible: Words used for encoding were selected entirely randomly for each participant from the set of 120 high-frequency and 120 low-frequency words without replacement such that a given word would only ever be seen once at encoding in the entire testing session. Words were combined randomly during encoding, lures were selected randomly for item tests, encoded words were selected randomly to be part of the item or associative tests, associative pairs were recombined randomly and presentation orders of old/new and intact/recombined pairs were random at test.

For counterbalancing, an individual would either always have the item test before the associative test or vice versa. An individual would also have either high-frequency, low-frequency, high-frequency then low-frequency word stimuli across the four tests or low-frequency, high-frequency, low-frequency then high-frequency word stimuli across the four tests. Finally, an individual would be presented with the fast encoding conditions before the slow encoding conditions or vice versa. This produced a $2 \times 2 \times 2$ design with eight possible order combinations, which were used four times for each age group.

Before any of the main tests took place, each participant completed a short practice version with three word pairs at encoding, followed by two associative and four item test trials. The practice words were not used in the main memory tests. Participants were therefore aware of the requirements of the test and encoded the information intentionally (note that intentional encoding produces a larger age-related associative deficit than does incidental encoding; see Old & Naveh-Benjamin, 2008).

Results

A corrected recognition measure of memory performance was used, that is, hit rates minus false alarm rates (e.g., as used by Naveh-Benjamin, 2000).[4] A 2 (Age: young, older) × 2 (Word Frequency: high, low) × 2 (Presentation Rate: fast, slow) × 2 (Test Type: item, associative) repeated measures ANOVA was conducted on the data (see Figure 2 for means). All main effects were significant: Young adults performed better than older adults, $F(1, 62) = 26.27$, $MSE = 0.23$, $p < .001$, $\eta_p^2 = .298$, $BF_{10} = 1.429 \times 10^4$. Low-frequency words were remembered better than high-frequency words, $F(1, 62) = 4.40$, $MSE = 0.04$, $p < .05$, $\eta_p^2 = .066$, $BF_{10} = 190.9$. Slow presentation rates resulted in better memory than fast presentation rates, $F(1, 62) = 4.84$, $MSE = 0.06$, $p < .05$, $\eta_p^2 = .072$, $BF_{10} = 0.264$. Item memory tests resulted in higher memory performance than associative memory tests, $F(1, 62) = 71.61$, $MSE = 0.07$, $p < .001$, $\eta_p^2 = .536$, $BF_{10} = 3.854 \times 10^{14}$.

There was an interaction between age and test type, $F(1, 62) = 8.19$, $MSE = 0.07$, $p < .01$, $\eta_p^2 = .117$, $BF_{10} = 16.54$, showing the expected age-related associative deficits with larger age deficits for associative memory tests than for item memory tests. The simple effects of age were significant for both item and associative memory (both $ps < .001$), and the simple effects of test type were significant for both young and older adults (both $ps < .001$). There was also an interaction between word frequency and test type, $F(1, 62) = 23.07$, $MSE = 0.05$, $p < .001$, $\eta_p^2 = .271$, $BF_{10} = 911.3$, with item memory better for low- than for high-frequency words, but associative memory better for high- than for low-frequency

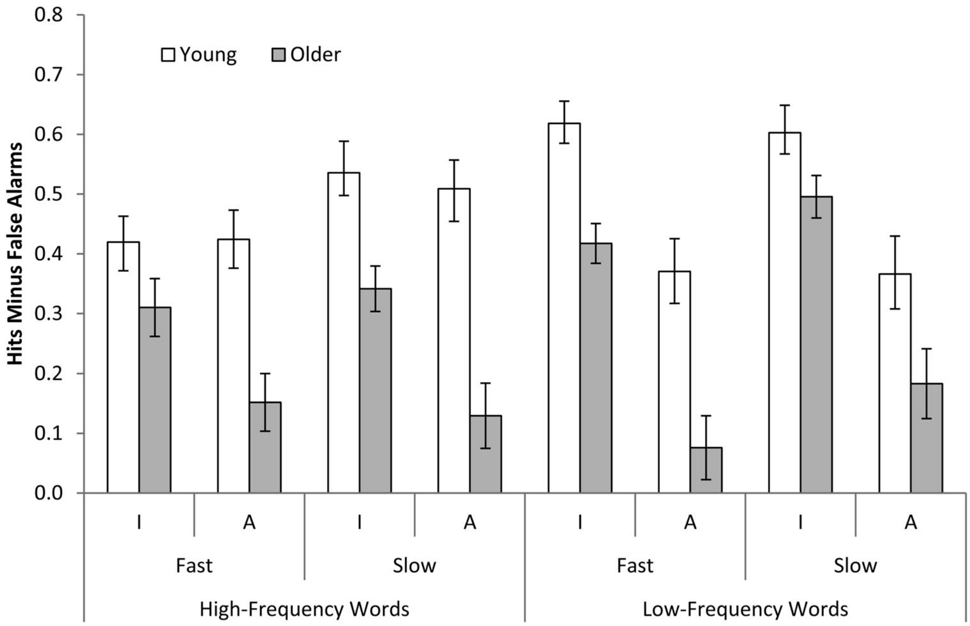

Figure 2. Hits minus false alarms memory performance data for young and older adults, high- and low-frequency words, fast and slow presentation rates, and for item (I) and associative (A) recognition memory tests in Experiment 2. Error bars are ± 1*SE*.

words. The simple effects of word frequency were significant for item ($p < .001$) but only marginal for associative ($p = .083$) memory; the simple effects of test type were significant for both high ($p = .001$) and low ($p < .001$) frequency lists. Finally, there was a triple interaction between age, word frequency and presentation rate, $F(1, 62) = 5.66$, $MSE = 0.07$, $p < .05$, $\eta_p{}^2 = .084$, $BF_{10} = 0.048$, which is explored further below. Crucially, there was no significant interaction between age, test type, and word frequency, $F(1, 62) = 1.32$, $MSE = 0.05$, *ns*, $\eta_p{}^2 = .021$, $BF_{10} = 0.417$, indicating that word frequency did not influence the age-related associative deficit. There were no other significant interactions ($Fs < 1.34$): Age × Word Frequency, $BF_{10} = 0.280$; Age × Presentation Rate, $BF_{10} = 0.097$; Word Frequency × Presentation Rate, $BF_{10} = 0.102$; Presentation Rate × Test Type, $BF_{10} = 0.088$; Word Frequency × Presentation Rate × Test Type, $BF_{10} = 0.021$; and Age × Word Frequency × Presentation Rate × Test Type, $BF_{10} = 2.974 \times 10^{-4}$.

Following up the Age × Word Frequency × Presentation Rate interaction, although slow presentation generally benefitted memory relative to fast presentation, tests of simple effects of presentation rate revealed that for young adults, the benefit was evident for high-frequency words ($p = .012$) but not for low-frequency words ($p = .825$), whereas for older adults the benefit was evident for low-frequency words ($p = .045$) but not for high-frequency words ($p = .909$). Separate three-way ANOVAs on the data for the fast and slow presentation rates both revealed significant effects of age, test type, Age × Test Type, and Word Frequency × Test Type as described above. For the slow presentation rate only, there was also a significant interaction between age and word frequency, $F(1, 62) = 6.85$, $MSE = 0.047$, $p = .01$, $\eta_p{}^2 = .099$, $BF_{10} = 4.520$, with smaller age differences for low- than for high-frequency words.

Some studies show that age-related associative deficits are driven by false alarms in the associative recognition test (Castel & Craik, 2003; Healy, Light, & Chung, 2005). It is possible that the increased familiarity of high-frequency words relative to low-frequency words may lead to an age-related increase in endorsing lures in the associative memory test, especially given that older adults can show increased reliance on familiarity during associative memory tests (Cohn, Emrich, & Moscovitch, 2008). We therefore explored our data further:[5] Separate analyses (Age × Word Frequency × Presentation Rate × Test Type) of hit rates and false alarm rates (see Table A3 for means) revealed that age effects were mostly driven by false alarms rather than by hits. For example, age was significant for false alarms, $p < .001$, $BF_{10} = 3.08 \times 10^8$, but not for hits, $p = .169$, $BF_{10} = 0.079$. Interestingly, there was a significant Age × Word Frequency interaction for false alarms only, $F(1, 62) = 11.28$, $MSE = 0.02$, $p = .001$, $\eta_p^2 = .154$, $BF_{10} = 15.508$, with a greater age-related increase in false alarms for high- than for low-frequency words, but this did not interact with test type and was therefore similar in both item and associative memory tests, $F < 1$, $BF_{10} = 1.440$.

Discussion

As in Experiment 1, all main effects were significant, indicating that our manipulations were effective. We replicated prior literature in terms of word frequency effects, with the word frequency by test type interaction showing superior item memory for low-frequency words alongside superior associative memory for high-frequency words (Clark, 1992; Clark & Burchett, 1994). The data also showed a clear age-related associative deficit, with larger deficits for the associative memory test compared to the item memory test (Old & Naveh-Benjamin, 2008).

Longer encoding time aided memory but presentation rate did not significantly interact with word frequency as predicted. It was supposed that longer encoding would particularly aid memory for low-frequency words, which take longer to process (e.g., Balota & Chumbley, 1984; Jescheniak & Levelt, 1994; Johnston & Barry, 2006). A triple interaction showed that a slow presentation rate helped older adults' memory particularly for low-frequency words as predicted, but young adults showed the opposite effect, with a slow presentation rate benefitting memory more for high-frequency words than for low-frequency words. It may be the case that young and older adults made use of the extra encoding time differently as studies show age deficits in strategic processing during associative memory tasks (Dunlosky & Hertzog, 2001; Naveh-Benjamin, Brav, & Levy, 2007).

Additionally, when analysing hits and false alarms separately, we found that age effects were largely driven by more false alarms in older than in young adults. This is consistent with the notion that false alarm rates could be responsible for the age-related associative deficit (Cohn et al., 2008). We also found an interaction between age and word frequency for false alarms, with older adults' increased endorsement of lures occurring more for high- than for low-frequency words. Whilst this result is consistent with the view that prior knowledge can lead older adults astray by enhancing the familiarity of lures (Umanath & Marsh, 2014), this outcome was similar across item and associative memory.

Our main prediction, that the age-related associative deficit would be alleviated with high- (relative to low-) frequency words, was not supported by the current data. Note that an earlier study found the same when comparing words with nonwords (Badham & Maylor, 2011). Recent research from our laboratory indicates that age differences in the use of prior knowledge depend on the experimental paradigm and this is explored further in the General Discussion in the context of both Experiments 1 and 2.

General discussion

Both of the experiments reported here showed overall word frequency effects that were as predicted from the literature. Crucially, young and older adults responded similarly to word frequency manipulations, providing a dissociation between processes related to word frequency effects and processes related to cognitive ageing.

In Experiment 1, the list length effect in free recall was larger for high-frequency words, indicating that interference between items was greater for the less distinctive high-frequency words than for low-frequency words (Schmidt, 2008). Young adults showed a greater list length effect than did older adults and we showed how this is generally consistent with the existing literature. Resource deficit accounts of cognitive ageing would predict a larger list length effect in older adults due to their greater susceptibility to interference which is higher for long lists (cf. Hasher & Zacks, 1988). Similarly, age deficits tend to be larger as tasks become more demanding so it could be expected that this would cause a greater list length effect in older adults (Cohen et al., 1987). However, it appeared to be the case that young adults had "more to lose" with long lists possibly disrupting their organisational processing (Smith, 1979). Age differences in the list length effect have been interpreted in terms of task difficulty (Cohen et al., 1987; Kahana et al., 2005). These authors used raw scores instead of proportion correct which led them to conclude that larger age deficits for long lists were due to increasing task difficulty. On inspection of reported data, studies of age differences in the list length effect generally showed a greater difference in proportions correct between short and long lists for young than for older adults (Cohen et al., 1987; Craik, 1968; Craik & Masani, 1967; Kahana et al., 2005; Smith, 1979), in line with the current data.

In Experiment 2, word frequency influenced the difference in performance between item and associative recognition memory tests. High-frequency words were easier to associate and low-frequency words resulted in better item recognition, in line with prior research (Clark, 1992; Clark & Burchett, 1994). Consistent with the ageing literature (Old & Naveh-Benjamin, 2008), older adults showed an age-related associative deficit, with significantly larger age deficits for associative compared to item recognition tests.

With the interesting exceptions of somewhat larger age deficits in overall recognition memory for high- compared to low-frequency words at the slow presentation rate of Experiment 2, and the generally larger age increase in false alarms for high- than for low-frequency words (both perhaps indicative of older adults responding more on the basis of familiarity), in both experiments there was age invariance in word frequency effects. Thus, young and older adults responded similarly to our frequency manipulations, which contrasts with studies showing greater benefits from prior knowledge in older adults. Although much research indicates that the influence of prior experience is greater for older adults (see Umanath & Marsh, 2014, for a review), particularly for associative memory (e.g., Naveh-Benjamin et al., 2003), it might not always be the case. In recent work from our laboratory, we identified boundary conditions on this effect (Badham et al., 2016): Our data and review of the literature showed that there are many circumstances under which young and older adults make similar use of pre-experimental knowledge and the same is true of the current word frequency data. In Badham et al. (2016), age deficits in associative memory were reduced by using semantically related word pairs (e.g., spear-pistol, horn-trombone) compared to unrelated word pairs (e.g., whiskey-jacket, hawkvolcano) but only if the relations were unique to each pair. When participants studied a list of word pairs with similar relations (e.g., banker–fireman, engineer–cook, athlete–teacher), young and older adults showed no difference in use of those relations. We argued that prior knowledge disproportionately benefits older adults if it provides environmental support *independent* from the episodic memory itself. For example, knowing that the word pairs were related was useless when all the relations were the same, but that same knowledge could be used to narrow down the search process when all the relations were unique, and in this latter case, particularly so for older adults. Therefore, in terms of the current data, word frequency effects were influential on memory for both age groups, but knowing that the list just studied contained common or uncommon words offered little extra information for orienting memory processes, and offered no independent information from which older adults could disproportionally benefit.

In the current article, we have explored free recall (Experiment 1) and recognition (Experiment 2) but there is also literature on word frequency effects with cued recall. Several studies have investigated memory using paired associates comprising high- and low-frequency words (Clark & Burchett, 1994; Criss, Aue, & Smith, 2011; Madan, Glaholt, & Caplan, 2010), where participants study pairs of words and are later cued with one word of each pair and asked to recall the other. Generally, cued recall is greater for pairs of high-frequency words compared to pairs of low-frequency words. It has been suggested here that high-frequency words are easier to associate, but data from some of these cued recall studies suggest that the high-frequency word advantage is mainly driven by target words and not cues (Criss et al., 2011; Madan et al., 2010). Therefore, the high-frequency word advantage in cued recall may operate by different mechanisms to the advantage typically seen for cued recall using related compared to unrelated words (e.g., Badham et al., 2012; Naveh-Benjamin et al., 2005). Future word frequency research may show different patterns of age differences to the current study if it used cued recall rather than recognition, even though for other types of word pairs (related vs. unrelated), age patterns are similar for cued recall (Badham et al., 2012) and recognition (Naveh-Benjamin et al., 2003).

A potential issue with the current data is the fact that there are systematic differences in the words that young and older adults experience. Worden and Sherman-Brown (1983) showed that there are age differences in memory for contemporary compared to dated words. The current data showed equivalent word frequency effects in young and older adults, suggesting that our stimuli were appropriate for both groups. Our predictions were for greater word frequency effects in older adults, and it is possible that cohort differences may have minimised age differences in frequency effects. This is unlikely, however, as cohort effects would need to have cancelled out developmental age differences perfectly in both experiments.

A further point to consider is the degree to which the effects can be attributed to word frequency alone. As already noted, other factors such as age of acquisition and contextual diversity are known to correlate highly with word frequency. Indeed, our high- and low-frequency words differed significantly in terms of both of these factors, as expected. Our primary concern here was to explore age-related differences in memory for word-based stimuli with a manipulation of pre-existing knowledge, operationalised in this case using word frequency as a robust objective measure of prior exposure. To isolate effects due to word frequency alone would require the matching of stimuli on both age of acquisition and contextual diversity. Given that these two measures correlate highly with word frequency, such matching alongside all our other controls was not attempted here. Nonetheless, our overall conclusion across the two experiments is that factors contributing to differences in pre-existing knowledge of verbal stimuli appear to act independently from factors influencing age differences in memory.

Notes

1. High-frequency words tend to differ from low-frequency words not only in terms of their frequency of occurrence in the language but also in terms of their duration of exposure (i.e., earlier age of acquisition; Johnston & Barry, 2006) and variety of exposure (i.e., greater contextual diversity; Brysbaert & New, 2009). Note, however, that our focus here is on manipulations of pre-existing knowledge of, or prior exposure to, verbal stimuli, broadly defined; hence, our high- and low-frequency words differed significantly on both of these measures, as detailed in the Materials sections.
2. Due to experimenter error, the task was conducted for 60 seconds rather than 90 seconds in Experiment 1 so the data were multiplied by 1.5 here to aid comparison with other experiments.
3. An initial correlation between exact age and overall recall averaged across all conditions was highly significant, $r(58) = -.618$, $p < .001$. This remained highly significant when education was partialled out, $r(57) = -.446$, $p < .001$, when vocabulary was partialled out, $r(57) = -.634$, $p < .001$, but not when processing speed was partialled out, $r(57) = -.196$, $p = .137$, consistent with the reduced processing speed hypothesis of cognitive ageing (Salthouse, 1996).
4. There was a highly significant correlation between exact age and overall recognition averaged across all conditions, $r(62) = -.556$, $p < .001$, which remained highly significant when education was partialled out, $r(60) = -.479$, $p < .001$, when vocabulary was partialled out, $r(61) = -.654$, $p < .001$, but not when processing speed was partialled out, $r(61) = -.138$, $p = .281$, again as expected (see Salthouse, 1996).
5. We thank an anonymous reviewer for suggesting these analyses.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the UK's Economic and Social Research Council under Grant [ES/K002732/1] awarded to SPB and EAM.

References

Allen, P. A., Madden, D. J., Weber, T. A., & Groth, K. E. (1993). Influence of age and processing stage on visual word recognition. *Psychology and Aging*, *8*, 274–282.

Almond, N. M., & Morrison, C. M. (2014). Episodic intertrial learning of younger and older participants: Effects of age of acquisition. *Aging, Neuropsychology, and Cognition*, *21*, 606–632.

Almond, N. M., Morrison, C. M., & Moulin, C. J. (2013). Episodic intertrial learning of younger and older adults: Effects of word frequency. *Aging, Neuropsychology, and Cognition*, *20*, 174–194. doi:10.1080/13825585.2012.679914

Bäckman, L., & Herlitz, A. (1990). The relationship between prior knowledge and face recognition memory in normal aging and Alzheimer's Disease. *Journal of Gerontology*, *45*, P94–P100.

Badham, S. P., Estes, Z., & Maylor, E. A. (2012). Integrative and semantic relations equally alleviate age-related associative memory deficits. *Psychology and Aging*, *27*, 141–152. doi:10.1037/a0023924

Badham, S. P., Hay, M., Foxon, N., Kaur, K., & Maylor, E. A. (2016). When does prior knowledge disproportionately benefit older adults' memory? *Aging, Neuropsychology, and Cognition*, *23*, 338–365.

Badham, S. P., & Maylor, E. A. (2011). Age-related associative deficits are absent with nonwords. *Psychology and Aging*, *26*, 689–694. doi:2011-06348-001

Badham, S. P., & Maylor, E. A. (2015). Antimnemonic effects of schemas in young and older adults. *Aging, Neuropsychology, and Cognition*, *18*, 1–25.

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language*, *46*, 199–226.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 340–357.

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 576–587.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., … Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*, 445–459.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990. doi:10.3758/BRM.41.4.977

Castel, A. D. (2005). Memory for grocery prices in younger and older adults: The role of schematic support. *Psychology and Aging*, *20*, 718–721.

Castel, A. D., & Craik, F. I. M. (2003). The effects of aging and divided attention on memory for item and associative information. *Psychology and Aging*, *18*, 873–885. doi:10.1037/0882-7974.18.4.873

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition*, *20*, 231–243.

Clark, S. E., & Burchett, R. E. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition*, *22*, 55–62.

Cohen, R. L., Sandler, S. P., & Schroeder, K. (1987). Aging and memory for words and action events: Effects of item repetition and list length. *Psychology and Aging*, *2*, 280–285.

Cohn, M., Emrich, S. M., & Moscovitch, M. (2008). Age-related deficits in associative memory: The influence of impaired strategic retrieval. *Psychology and Aging*, *23*, 93–103. doi:2008-02853-012 [pii] 10.1037/0882-7974.23.1.93

Craik, F. I. M. (1968). Short-term memory and the aging process. In G. A. Talland (Ed.), *Human aging and behavior* (pp. 131–168). New York, NY: Academic Press.

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

Craik, F. I. M., & Masani, P. A. (1967). Age differences in the temporal integration of language. *British Journal of Psychology*, *58*, 291–299.

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language*, *64*, 119–132.

DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1136–1146.

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition*, *29*, 247–253.

Ellis, A. W., Miller, D., & Sin, G. (1983). Wernicke's aphasia and normal language processing: A case study in cognitive neuropsychology. *Cognition*, *15*, 111–144.

Gerard, L., Zacks, R. T., Hasher, L., & Radvansky, G. A. (1991). Age deficits in retrieval: The fan effect. *Journal of Gerontology*, *46*, P131–P136.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*, 8–20.

Gregg, V. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183–215). Oxford: John Wiley & Sons.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). New York: Academic Press.

Healy, M. R., Light, L. L., & Chung, C. (2005). Dual-process models of associative recognition in young and older adults: Evidence from receiver operating characteristics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 768–788. doi:10.1037/0278-7393.31.4.768

Hess, T. M. (1985). Aging and context influences on recognition memory for typical and atypical script actions. *Developmental Psychology*, *21*, 1139–1151.

Hoyer, W. J., Stawski, R. S., Wasylyshyn, C., & Verhaeghen, P. (2004). Adult age and digit symbol substitution performance: A meta-analysis. *Psychology and Aging*, *19*, 211–214. doi:10.1037/0882-7974.19.1.211

Hunt, R. R., & McDaniel, M. A. (1993). The enigma of organization and distictiveness. *Journal of Memory and Language*, *32*, 421–445.

Huron, C., Danion, J.-M., Giacomoni, F., Grangé, D., Robert, P., & Rizzo, L. (1995). Impairment of recognition memory with, but not without, conscious recollection in schizophrenia. *American Journal of Psychiatry*, *152*, 1737–1742.

Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *Journal of Problem Solving*, *7*, 2–9.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 824–843.

Johnston, R. A., & Barry, C. (2006). Age of acquisition and lexical processing. *Visual Cognition*, *13*, 789–845.

Kahana, M. J., Dolan, E. D., Sauder, C. L., & Wingfield, A. (2005). Intrusions in episodic recall: Age differences in editing of overt responses. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *60*, P92–P97.

Kensinger, E. A. (2008). Age differences in memory for arousing and nonarousing emotional words. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *63*, P13–P18.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990.

Love, J., Selker, R., Marsman, M., Jamil, T., Verhagen, A. J., Ly, A., … Wagenmakers, E.-J. (2015). JASP (Version 0.6.5). Retrieved from https://jasp-stats.org/

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*, 203–208.

Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language*, *63*, 46–63.

Mather, M., & Knight, M. (2005). Goal-directed memory: The role of cognitive control in older adults' emotional memory. *Psychology and Aging*, *20*, 554–570.

Mohanty, P. P., Naveh-Benjamin, M., & Ratneshwar, S. (2016). Beneficial effects of semantic memory support on older adults' episodic memory: Differential patterns of support of item and associative information. *Psychology and Aging*, *31*, 25–36.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1170–1187.

Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging*, *22*, 202–208. doi:2007-03358-020 [pii] 10.1037/0882-7974.22.1.202

Naveh-Benjamin, M., Craik, F. I. M., Guez, J., & Kreuger, S. (2005). Divided attention in younger and older adults: Effects of strategy and relatedness on memory performance and secondary task costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 520–537. doi:2005-05101-010 [pii] 10.1037/0278-7393.31.3.520

Naveh-Benjamin, M., Hussain, Z., Guez, J., & Bar-On, M. (2003). Adult age differences in episodic memory: Further support for an associative-deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 826–837.

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, *23*, 104–118. doi:s10.1037/0882-7974.23.1.104

Radvansky, G. A., Zacks, R. T., & Hasher, L. (1996). Fact retrieval in younger and older adults: The role of mental models. *Psychology and Aging*, *11*, 258–271.

Raven, J. C., Raven, J., & Court, J. H. (1988). *The Mill Hill vocabulary scale*. London: H. K. Lewis.

Rayner, K., Reichle, E. D., Stroud, M. J., & Williams, C. C. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*, 448–465.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.

Salthouse, T. A. (2010). *Major issues in cognitive aging*. New York, NY: Oxford University Press.

Schmidt, S. R. (2008). Distinctiveness and memory: A theoretical and empirical review. In H. L. Roediger (Ed.), *Learning and memory: A comprehensive reference: Volume 2: Cognitive psychology of memory* (pp. 125–144). Oxford: Elsevier.

Shiffrin, R. M. (1970). Forgetting: Trace erosion or retrieval failure? *Science*, *168*, 1601–1603.

Shiffrin, R. M. (1971). Retrieval failure or selective attention? *Science*, *173*, 1040–1041.

Smith, A. D. (1979). The interaction between age and list length in free recall. *Journal of Gerontology*, *34*, 381–387.

Spieler, D. H., & Balota, D. A. (2000). Factors influencing word naming in younger and older adults. *Psychology and Aging*, *15*, 225–231.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*, 41–78.

Umanath, S., & Marsh, E. J. (2014). Understanding how prior knowledge influences memory in older adults. *Perspectives on Psychological Science*, *9*, 408–426. doi:10.1177/1745691614535933

Verhaeghen, P. (2003). Aging and vocabulary scores: A meta-analysis. *Psychology and Aging*, *18*, 332–339.

Wechsler, D. (1981). *Manual for the Wechsler adult intelligence scale — revised*. New York, NY: Psychological Corporation.

Worden, P. E., & Sherman-Brown, S. (1983). A word-frequency cohort effect in young versus elderly adults' memory for words. *Developmental Psychology*, *19*, 521–530.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In T. A. Salthouse, & F. I. M. Craik (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 293–357). Mahwah, NJ: Lawrence Erlbaum Associates.

Appendix

Table A1. List of 300 words from which stimuli for Experiments 1 and 2 were selected along with imageability and valence ratings.

Low-Frequency Words								
Word	Imageability	Valence	Word	Imageability	Valence	Word	Imageability	Valence
alcoves[2]	4.57	3.94	fineness[12]	2.25	4.82	puffball[12]	3.87	4.00
anthems[2]	3.69	5.41	flatness[12]	3.81	3.38	pullback	2.06	3.29
awnings[2]	3.93	3.88	flutist[12]	5.19	5.12	ravines[2]	4.81	4.12
bathers[2]	4.75	4.29	foodstuff[12]	5.50	5.47	rhomboid	3.00	3.53
beetroot[12]	5.67	3.76	forebear	1.69	4.12	ripcord[12]	4.00	3.53
bequests[2]	2.62	4.06	funfair[12]	6.25	4.94	rosettes[2]	4.63	3.69
billionth[12]	1.63	3.76	geysers[2]	4.88	4.56	roundness[12]	4.75	4.18
biplane[12]	4.75	3.88	gherkin[12]	5.88	3.47	roundups[12]	2.88	3.59
birthrate[12]	2.50	3.81	goatskin[12]	4.38	3.53	rulings[2]	2.63	3.41
boaster	2.50	2.53	goldfinch[12]	4.69	5.00	scabies	4.31	1.94
breadline	2.81	2.65	gourmets[2]	2.88	4.53	senders[12]	2.00	4.18
brickyard[12]	4.94	3.35	granule[12]	4.31	3.71	sewerage[12]	5.25	2.65
briquette[12]	3.67	3.76	griffon[12]	4.00	3.71	shallot[12]	5.19	4.12
casement	3.00	3.82	grinders[2]	3.88	3.47	shortfall	1.63	2.59
casework	2.69	3.18	grounder	1.88	3.24	showpiece[12]	3.25	4.31
chairmen[2]	4.75	3.71	groupings[2]	2.88	3.82	sightseer[12]	4.56	4.06
chanter[12]	2.81	3.94	gullies	3.40	3.35	signposts[2]	5.56	4.47
chisels[2]	4.50	4.00	hackles	2.94	2.76	silkworm[12]	4.69	4.29
claimants[2]	2.25	3.24	handbooks[2]	4.81	4.18	slowness	2.88	2.76
clamshell[12]	4.19	3.82	harpist[12]	5.50	5.29	snowshoe[12]	5.31	3.82
clansman[12]	3.50	3.76	haunches[2]	3.75	3.35	soundness[12]	1.75	4.47
clothier[12]	2.63	4.00	headstand[12]	5.19	4.06	splotches[2]	4.56	3.18
cockpits[2]	4.88	3.65	hearers[2]	2.07	3.82	stairways[2]	6.00	4.24
coinage[12]	3.69	4.29	hilltops[2]	5.75	5.35	starlings[2]	5.44	4.71
courtier[12]	3.44	3.82	hyacinths[2]	4.63	4.76	stoneware[12]	4.38	4.41
crafter[12]	2.81	4.71	ironwork[12]	4.31	4.18	streaker[12]	5.44	3.47
crassness	1.88	2.12	issuance[12]	1.44	3.71	streamer[12]	4.81	4.35
crevasse[12]	4.19	3.53	justness[12]	2.25	5.82	strivings[12]	1.81	4.35
croquette[12]	4.44	4.41	lattice[12]	4.38	4.24	subgroups	2.13	3.47
currant[12]	5.13	4.31	leaflet[12]	6.06	3.47	subtype	1.88	3.59
deftness[12]	2.27	4.71	lectern	3.73	4.06	sultans[2]	5.13	4.00
despots[2]	2.53	1.88	lenience[12]	2.13	4.25	sweatband[12]	5.50	3.47
deviance	2.00	2.24	lifebelt[12]	5.50	4.88	tankard[12]	4.81	3.88
disquiet	2.06	2.35	longship[12]	4.44	4.12	tempter	2.38	3.35
dockyard[12]	5.13	3.59	lounger[12]	4.75	3.94	tendril[12]	3.75	4.24
doctrines	1.94	3.29	marquees[2]	5.88	4.24	tiredness[12]	3.75	2.53
doubter	2.47	2.50	midpoint[12]	3.44	3.76	tollgate[12]	4.69	3.24
downpour[12]	5.94	2.88	midstream[12]	3.25	3.88	tracksuit[12]	5.75	3.71
downturn	2.13	2.41	mischance	1.56	2.94	travail[12]	2.07	3.71
droplet[12]	5.38	4.06	nocturne[12]	2.63	4.53	triplet[12]	4.81	3.88
ductwork	3.25	3.35	nosebag[12]	4.63	3.59	tripods[12]	5.44	3.81
duellist[12]	3.69	3.12	outgrowth[12]	2.88	3.35	tweezer[12]	6.40	3.82
earldom[12]	2.13	3.53	paleness[12]	4.69	2.76	untruth	1.94	1.88
effluent[12]	3.19	3.18	parsnip[12]	5.94	3.76	vestment[12]	3.33	4.06
eightieth[12]	3.13	5.06	pipette[12]	4.13	3.82	viscount	3.13	3.29
enclave	2.19	3.53	pitfall	2.75	2.35	wastage[12]	3.69	2.29
entrant[12]	3.19	3.94	plantain[12]	4.13	3.65	waxwork[12]	4.75	3.59
facades[2]	3.88	3.82	plateful[12]	5.13	4.12	whippet[12]	5.33	4.06
fastener[12]	3.81	3.76	playsuit[12]	5.38	3.94	wielder[12]	3.19	3.35
fielders[2]	4.07	3.82	pleader	2.81	3.00	wolfhound[12]	5.44	4.12

High-Frequency Words								
Word	Imageability	Valence	Word	Imageability	Valence	Word	Imageability	Valence
actions[12]	3.25	4.76	graveyard[12]	6.38	2.76	pursuit[12]	3.13	4.38
actress	5.94	5.06	greetings[2]	3.38	5.76	quantum[12]	2.50	4.18
address[12]	3.87	4.53	guardian[12]	4.06	4.76	rainbow	6.69	6.12
airplane	6.63	4.76	guidance[12]	2.44	4.81	response[12]	2.63	4.41
alliance[12]	2.20	5.24	hardware[12]	4.56	4.12	roommate[12]	5.25	4.88
applause	4.56	6.24	headaches[2]	3.25	2.12	sausage[12]	6.38	4.82
aspirin[12]	4.94	3.35	heartbeat[12]	4.06	5.00	scientist[12]	5.38	5.06
audience[12]	5.50	4.76	highness[12]	3.00	4.29	sequence[12]	2.31	3.94
bathtub	6.38	4.81	homework[12]	5.00	2.94	sergeant[12]	5.44	4.06
bedtime[12]	4.07	5.82	hundreds[2]	3.13	4.06	servant[12]	5.38	3.20
birthday	4.75	6.06	instincts[2]	2.06	4.71	shipment[12]	3.75	3.75
blanket	5.81	4.76	jewelry	6.00	4.76	sickness[12]	4.63	2.06
bourbon	5.19	4.18	judgment[12]	2.69	4.65	someone[12]	3.56	4.41
boyfriend	4.88	4.76	kindness[12]	3.25	6.47	something[12]	2.56	4.24
bracelet	6.19	4.53	knowledge[12]	2.75	6.29	speeches[2]	3.56	3.75

(*Continued*)

Table A1. Continued.

High-Frequency Words								
Word	Imageability	Valence	Word	Imageability	Valence	Word	Imageability	Valence
breakdown[12]	4.07	1.82	landlord[12]	4.69	3.00	stadium[12]	5.81	4.24
briefcase[12]	6.13	3.82	laundry[12]	5.40	3.53	standards[2]	2.19	4.65
buildings[12]	6.00	4.24	lawsuit[12]	2.81	2.12	statement[12]	3.13	3.82
chambers[12]	4.19	3.88	lawyers[2]	5.06	3.06	storage[12]	3.69	4.00
champagne	6.13	5.35	lifetime[12]	3.25	5.06	strangers[2]	3.50	3.53
childhood[12]	3.69	5.82	lightning[12]	6.50	4.18	students	6.06	5.00
children	6.19	5.47	luggage[12]	6.19	4.41	subjects[2]	2.94	4.12
chocolate	6.44	6.12	mankind[12]	3.63	5.00	substance[12]	2.88	3.82
circuit[12]	4.63	3.71	mansion	5.81	4.94	suitcase	6.63	4.65
clearance[12]	2.75	3.76	mattress[12]	6.25	4.41	sunrise	6.25	6.18
cocktail	5.81	5.00	medicine[12]	5.06	4.41	supreme[12]	2.00	5.19
colleague[12]	4.94	5.35	meetings[2]	4.31	3.47	suspects[2]	3.44	2.59
conflict[12]	3.19	2.24	members[2]	4.06	4.38	sweetie[12]	4.19	4.59
costume[12]	5.00	4.53	methods[2]	2.50	4.06	symptoms[2]	3.00	2.82
cottage	6.44	5.24	moments[2]	2.19	4.53	syndrome[12]	1.94	2.53
countries	5.00	5.00	moonlight	6.06	5.82	systems[2]	1.69	3.82
courtroom[12]	5.69	3.35	movement[12]	4.25	4.65	teachers[2]	5.93	5.29
cousins[2]	5.33	4.94	necklace	6.25	4.65	technique[12]	2.69	4.53
creatures[12]	5.38	5.12	nightmare[12]	4.13	1.94	theater[12]	5.44	5.41
crystal	5.38	5.24	nothing[12]	2.56	2.76	theories[2]	1.94	5.06
customs[2]	3.13	4.00	objects[2]	4.56	3.76	thousands[2]	3.44	4.41
darkness[12]	4.38	3.24	options[2]	1.94	5.06	tourists[2]	5.56	4.06
daughters[2]	5.06	5.00	package[12]	5.69	4.50	transfer[12]	2.38	3.65
daylight	5.31	6.06	paintings	6.31	5.47	transport[12]	4.31	4.59
dentist[12]	6.19	3.71	passion	3.63	5.65	treatment[12]	3.13	4.65
dessert	5.44	5.47	passport	6.19	5.06	trousers[12]	6.50	4.41
disgrace[12]	2.13	1.76	patience[12]	2.19	5.59	upstairs[12]	5.06	4.47
efforts[2]	2.50	5.12	payment[12]	3.07	4.47	vengeance[12]	2.38	2.12
engines[2]	5.00	4.47	percent[12]	3.81	3.59	visions[2]	3.00	4.47
essence[12]	2.31	4.59	perfume[12]	5.06	4.53	waitress[12]	5.50	4.29
evening[12]	4.63	5.13	physics[12]	3.31	4.65	wardrobe	6.63	4.71
farewell[12]	3.50	2.59	platform[12]	5.56	3.94	warehouse[12]	5.63	3.65
footage[12]	3.31	4.00	precinct[12]	4.06	3.18	weakness[12]	2.50	2.24
footsteps[2]	4.94	4.06	progress[12]	2.25	5.18	weekends[2]	3.50	5.53
fortune[12]	3.19	5.00	purpose[12]	1.75	5.41	welfare[12]	2.63	4.82

[1]Words used in Experiment 1.
[2]Words used in Experiment 2.

Table A2. Mean numbers and standard deviations of words correctly recalled by young and older adults, for high- and low-frequency words, and for short (15-item) and long (30-item) word lists in Experiment 1.

	High-frequency words				Low-frequency words			
	Short lists		Long lists		Short lists		Long lists	
Age Group	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Young	7.25	2.45	9.68	3.28	5.97	2.44	8.87	3.62
Older	4.22	1.74	5.30	2.27	3.33	1.46	4.65	1.64

Note: In a repeated measures ANOVA on numbers correctly recalled, there were highly significant main effects of age, word frequency, and list length, together with an interaction between age and list length.

Table A3. Means (and standard deviations) for Hit (H) and False Alarm (FA) rates in young and older adults for high- and low-frequency words, fast and slow presentation rates, and for item and associative recognition memory tests in Experiment 2.

			Young		Older	
Frequency	Rate	Test	H	FA	H	FA
High						
	Fast					
		Item	.652 (.158)	.232 (.183)	.692 (.192)	.382 (.215)
		Associative	.616 (.197)	.192 (.191)	.629 (.195)	.478 (.220)
	Slow					
		Item	.730 (.206)	.194 (.178)	.645 (.212)	.304 (.195)
		Associative	.741 (.183)	.232 (.177)	.674 (.277)	.545 (.273)
Low						
	Fast					
		Item	.775 (.155)	.156 (.145)	.616 (.230)	.199 (.180)
		Associative	.696 (.198)	.326 (.247)	.621 (.245)	.545 (.275)
	Slow					
		Item	.710 (.205)	.107 (.144)	.676 (.223)	.181 (.138)
		Associative	.714 (.208)	.348 (.232)	.705 (.226)	.522 (.273)