Associative interference in older and younger adults

Rachel L. Burton

University of Alberta

Isabel Lek

University of Alberta

Roger A. Dixon

University of Alberta

Jeremy B. Caplan

University of Alberta

Author Note

Rachel L. Burton, Department of Psychology, University of Alberta; Isabel Lek, Department of Psychology, University of Alberta; Roger A. Dixon, Department of Psychology, University of Alberta; Jeremy B. Caplan, Department of Psychology, University of Alberta.

Correspondence concerning this article should be addressed to Jeremy B. Caplan, Department of Psychology, P217, Biological Sciences Building, University of Alberta, Edmonton, Alberta T6G 2E9, Canada. Email: jcaplan@ualberta.ca. We thank Yang Liu for help implementing the experiment in PyEPL and Jill Friesen for her help writing questionnaires, recruiting older participants and advice for ensuring smooth testing sessions. Supported in part by the Alberta Centre on Aging, the Natural Sciences and Engineering Research Council of Canada and the Alberta Ingenuity Fund. R.A.D. is supported by a grant from the National Institutes of Health (National Institute on Aging; R01 AG008235). Portions of the current data were presented at the International Neuropsychological Society and the Alberta Centre on Aging conferences in 2011.

Abstract

Healthy older adults are more challenged by associative interference than younger adults, but prior results could have been due to differences in list discrimination ability. We used a procedure that assessed interference without requiring knowledge of list membership to test the hypothesis that older adults (60–74 years old) would show more pronounced effects of associative interference in AB/AC learning. Despite our use of a self-paced, rather than timed, study procedure, older adults performed at lower levels of accuracy than younger adults, replicating the well established associative deficit in aging (Naveh-Benjamin & Mayr, 2018). Older participants also displayed more proactive interference on average. Older participants' memory for AB and AC showed statistical independence, resembling earlier data from younger participants with a timed study procedure (Burton, Lek, & Caplan, 2017). However, younger participants, with the current self-paced procedure, produced a facilitating relationship between memory for AB and AC. Thus, younger participants not only resolved, but reversed associative interference. List discrimination could not explain these age differences. Taken together, these results extend the associative deficit in aging, finding increased susceptibility to associative proactive interference and less resolution of associative interference in older than younger participants, even when given the opportunity to compensate during self-paced study.

Keywords: Associative memory, interference, AB/AC learning, verbal memory

Associative interference in older and younger adults

Introduction

Healthy aging is accompanied by a well established deficit in association memory,[1] even while memory for items (such as words) is intact (Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008; Overman & Becker, 2009), explored in depth in a special issue of *Psychology and Aging* (Naveh-Benjamin & Mayr, 2018). Association memory can be further challenged by interference. In associative interference conditions, knowledge is paired with one associate, and then re-paired with a different associate. For example, in locating one's keys, one may need to differentiate where they were today (the pair keys–pocket) versus yesterday (keys–coffee table). In many cases, it is valuable to be able to recall both associates (Reagan–actor, Reagan–president; or when a person shifts from one relationship to another: Brad–Jennifer became Brad–Angelina). Because they share a common item, such pairs present a challenge to memory. In AB/AC learning, an experimental model of associative interference (Martin, 1971a, 1971b), participants first learn List 1, composed of a set of "A" items paired with "B" items, AB (sometimes along with control pairs, DE). In List 2, participants must learn pairs composed of the earlier "A" items, re-paired with new, "C" items, AC (sometimes along with control pairs, FG). Proactive interference (PI) occurs when accuracy is worse (typically tested with cued recall; given a cue item, A, participants are asked to recall its associate), for AC than for FG pairs. When later tested for list 1, retroactive interference (RI) occurs when accuracy is worse for AB than DE pairs.

In the simple AB/AC procedure, reduced recall of C (but not B) could be due to AC being encoded more weakly than control pairs. Alternatively, memory for both C and B could be available, but the participant may simply have trouble determining which of the

[1] In place of "associative memory," we use the verbal-memory term "association memory" to avoid confusion with the influential Search of Associative Memory model (Raaijmakers & Shiffrin, 1981), and to maintain grammatical consistency with "item memory."

two candidate words is from list 2. The same logic can be applied to tests of retroactive interference. This conflates competition between memory for AB and AC with reduced source memory (e.g., whether B or C was paired with A first), and source-memory has also been found to be reduced in healthy aging (e.g., Bissig & Lustig, 2007; Carpenter & Schacter, 2018; Johnson, Hashtroudi, & Lindsay, 1993; Naveh-Benjamin & Mayr, 2018; Spencer & Raz, 1995). Consequently, apparent PI or RI in standard AB/AC learning may not, in fact, originate from a competitive relationship between memory for a pair of associations, but rather, difficulty in judging their relative time of presentation. To obtain a pure estimate of associative interference, unconfounded by source memory, Barnes and Underwood (1959) designed the so-called "modified modified free recall" (MMFR) procedure, in which participants are given an A item as a cue, and asked to recall both associated items, if possible. With the MMFR procedure, participants can show that they can remember both associates, even without knowledge of list-membership.

Our first aim was thus to compare both PI and RI in AB/AC learning between healthy older adults and younger adults. We expected older participants would perform worse than younger adults at association memory in general (Naveh-Benjamin & Mayr, 2018), quite apart from associative interference in particular. In addition, other evidence has mostly suggested that older adults are more challenged by interference in memory than younger adults (e.g., Biss, Rowe, Weeks, Hasher, & Murphy, 2018; Campbell & Hasher, 2018; Ebert & Anderson, 2009; Healey, Hasher, & Campbell, 2013; Lustig, May, & Hasher, 2001; Shimamura, 1994; Winocur & Moscovitch, 1983). Umanath and Marsh (2014) review how increased sensitivity to PI and RI can be either an advantage or a disadvantage, depending on task goals; Biss et al. (2018) even designed a clever paradigm to mobilize older adults' distraction toward task-irrelevant stimuli to their participants' benefit. In the case of associative interference, some evidence suggests that older adults experience a greater cost due to the challenge of PI (Ebert & Anderson, 2009; Hanseeuw, Seron, & Ivanoiu, 2012; Shimamura, 1994; Winocur & Moscovitch, 1983). However, we were unable to find published studies directly testing the presence of RI and PI in AB/AC learning.

Though AB/AC learning has been studied in samples of healthy older participants, they have only rarely been compared to young controls (rather, they are more frequently control groups for patient populations). PI is often measured (e.g., Guez & Naveh-Benjamin, 2016), but RI has only rarely been measured in older participants. Control pairs are usually not included, making it impossible to determine whether interference pairs are impaired or not (apart from comparisons between groups). And finally, MMFR is rarely used, leaving open the possibility that PI or RI effects, when found, could be due to list-discrimination or response competition or both.[2] The most relevant study was conducted by Siegel (2014), who compared younger and older adults with measures of PI and RI using MMFR, and found more PI in older than younger participants, but no difference in RI. However, the pairs were semantically related. Similarity might influence the relationship between AB and AC in memory, and it is hard to compare related AB/AC associations to semantically related control pairs that are not, in turn, similar to other studied pairs.

We compared healthy older participants to healthy younger participants, and tested memory for AB and AC compared to control pairs, with standard cued recall and MMFR. This enabled us to quantify RI and PI relative to control pairs, and test response competition or list-discrimination mechanisms.

Our second aim was to test the relationship between recall of B and C. With MMFR, one can measure the correlation between recall of B and recall of C in response to A. Surprisingly, when younger adults study word–word pairs, this correlation is near-zero, suggesting participants overcome associative interference and can retrieve B and C independently of one another (Burton et al., 2017; Tulving & Watkins, 1974).

Because general slowing is well established in healthy aging (Deary, Johnson, & Starr,

---

[2] but see Overman and Becker (2009), who found no interaction, in older participants, between association memory and list discrimination, albeit with an incidental study procedure.

2010; Salthouse, 1996) we deviated from the Burton and colleagues (2017) procedure and presented pairs self-paced (i.e., participants controlled the length of time each pair was presented) at study. Thus, a third aim was to test whether the pattern of mean performance and correlations between recall of B and C would generalize to conditions in which study of pairs was unconstrained, for both age groups.

Our fourth aim was to test for a possible relationship between associative interference and response competition. Compared to younger adults, older adults are well established to exhibit larger interference effects in response-selection in the color-Stroop (Stroop, 1935) task (e.g., Comalli, Wapner, & Werner, 1962; Ludwig, Borella, Tettamanti, & de Ribaupierre, 2010; Mayas, Fuentes, & Ballesteros., 2012; Spieler, Balota, & Faust, 1996), although this has been explained as an effect of general slowing rather than an age-difference in susceptibility to interference (Little & Hartley, 2000; Verhaeghen & De Meersman, 1998). Our older participants also performed the color-Stroop task between cycles of the AB/AC task, to enable us to test whether associative interference effects in AB/AC learning might have a common cause as interference due to response-competition in the Stroop task.

In sum, older and younger participants studied, self-paced, lists ("List 1") of word-pairs, AB, plus control pairs, DE, followed by cued recall of those List-1 pairs.[3] After each List 1, they studied a List 2, comprised of AC plus new control pairs (FG), followed

[3] We also included a strategy manipulation modelled on Burton et al. (2017), which modulated the relationship between AB and AC: Participants within each age group were randomly assigned to three Conditions: 1) Imagery-Only: instructed to study pairs using visual imagery. 2) Separation-Imagery: instructed to ensure images of pairs sharing an item were kept distinct. 3) Integration-Imagery: instructed to combine all items of pairs sharing an item into a single, integrated image. In part due to low participant counts in some Condition sub-groups, the strategy manipulation produced null or weak effects, and we could not clearly verify that participants were applying the strategies as instructed. We present the results with Condition included as a factor, and check that our key findings generalize across Condition, but caution the reader against confidently interpreting effects of Condition.

by cued recall of List 2. MMFR tests followed. We expected older participants would perform worse on cued recall, even on List 1, due to well established age-related reductions in association memory. Our primary interest was whether older participants would exhibit more RI or PI on average, and more competition between B and C than younger participants. Finally, older participants performed a Stroop task, to test whether Stroop interference bears any relationship to verbal-memory associative interference.

Methods

The materials and procedure were adapted from those used by Burton et al. (2017), illustrated in Figure 1. The protocol was reviewed and approved by University of Alberta research ethics board #2. All participants read and signed an informed consent form.

Participants. Younger participants: A total of 138 undergraduate students enrolled in a first-year introductory psychology course at the University of Alberta participated for course credit. These participants were assigned at random to the Imagery-Only, Separation-Imagery and Integration-Imagery groups, with subgroup sample sizes, sex and age reported in Table 1. Younger participants were given very general instructions in small groups, and then entered individual testing cubicles for the main task.

Older participants: Sixty older adults from the ages of 60 to 74 were recruited from the greater Edmonton area through ads. Older participants were administered a Personal Data Sheet and the Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975). We had planned to exclude older adults who scored less than 26 on the Mini Mental State Examination (MMSE), but no participants met this criterion. These 60 participants were assigned at random to the Imagery-Only, Separation-Imagery and Integration-Imagery groups, respectively. A larger sample size for the Imagery-Only group was acquired with the intent to compare this group to another data set, and sample sizes for Younger participants were greater because of ease of recruitment. Subgroup sample sizes, sex, age, MMSE scores and years of education are reported in Table 1. Older participants were also asked to rate their health in relation to other people their own age, on a five-point scale (Very good, Good Fair, Poor, Very poor). Responses were mostly Very Good or Good (Table 1). Older participants were tested individually, with the experimenter present throughout the practice phase, available to answer clarification questions and to ensure that the participant understood the instructions.

Materials. Materials were identical to those used by Burton and colleagues (2017). Lists of eight noun-pairs each were constructed. Stimuli were the 180 most concrete nouns with imagery rating $\geq$6 from the Toronto Word Pool (Friendly, Franklin, Hoffman, & Rubin, 1982). Each noun-pair was presented centrally on a computer screen with the two words side by side. For each participant, 112 words were pseudo-randomly drawn from the pool of 180 stimuli and were then assigned pseudo-randomly to pairs and pair types.

Procedure. Participants were tested in individual testing rooms. Each of the 8 noun-pairs in a list was presented self-paced, displayed until the participant pressed a key to continue, with an additional 150-ms interval between pairs (Figure 1). Participants completed a 20-s block of the distractor task following study of a List 1 (AB/DE) and were then tested on List 1 with cued recall. The left-hand words were cues, presented in random order, individually, in the centre of the screen, with one blank line (underline) below the probe. Younger participants responded by typing, and the letters appeared on the blank line as they typed. Backspacing was permitted, and the response was submitted by pressing the Enter key. Older participants responded vocally, and the experimenter typed their responses into the same interface. Following cued recall of List 1, participants similarly studied List 2, followed by another 20-s block of the distractor task and then a set of cued recall questions following the same procedure as for List 1, but specifying that participants were to retrieve the associate from the most recent list. Following one more 20-s block of the distractor task, participants were tested with MMFR, where they were given two response lines for each cue word, and asked to respond with the one or two words that had been presented alongside the cue word, in any order.

In each of the three experimental conditions, participants were asked to learn pairs by creating imagery relating paired items. In the Imagery-Only condition, no further strategic instruction was given. Participants in the Integration-Imagery group were asked to "create an image that incorporates all three words from both pairs" when studying the overlapping pairs. In the Separation-Imagery group, participants were asked to "create a separate image for each pair of words." No further training in these conditions was administered.

Testing-order of pairs during cued recall and MMFR tests was also random, with each pair tested once in cued recall, and each A, D and F word as a probe once during MMFR. During testing, participants typed (younger) or spoke (older, with the experimenter again typing) responses in blank lines presented below the cue word, followed by the Enter key.

The 20-s distractor task entailed adding three numerals between 2–8 inclusive, with typed (younger) or spoken, then typed by the experimenter (older), responses. Participants had 4 s to respond to each equation, and the program progressed from one question to the next automatically for a total of five math problems during each distractor block.

The experiment consisted of four study sets. Each set included two lists of noun-pairs presented on a computer screen: a List 1, comprised of AB (interference) and DE (control) pairs, and a List 2, comprised of AC (interference) and FG (control) pairs (Figure 1).

Following each of the practice and experimental AB/AC learning blocks, older participants performed a block of 45 trials of an adaptation of the Stroop task. In each trial, participants were asked to identify the font-color of each word stimulus, displayed centrally on the computer screen. Stimuli were color names 'RED,' 'GREEN' and 'BLUE,' as well as control words, 'CUP,' 'HOUSE' and 'PACK.' Font color could be red, green or blue, pseudo-randomly assigned, but counterbalanced to produce an equal number of trials of each condition in each block of Stroop trials. The three conditions were: (a) Congruent, color matching word; (b) Incongruent, color mismatching word; and (c) Control, word was not a color name. Participants responded with the number keys on the keypad, 1, 2 and 3, were labelled with color stickers with the colors red, green and blue, respectively. The stimulus remained on the screen until a response was made.

Statistical Analysis. To test for age differences in susceptibility to PI and RI (Aim 1), we analyzed accuracy in MMFR and cued recall, with analyses of variance for each list separately. RI was quantified on List 1 as the difference in MMFR accuracy between control (DE) and interference (AB) pairs. PI was quantified on List 2 as the difference in MMFR accuracy between control (FG) and interference (AC) pairs. Our interest was in whether Pair Type interacted with Age.

To quantify the relationship between recall of B and C in response to their shared A item (Aim 2), we computed Yule's Q, a measure of correlation suitable for dichotomous data and a special case of the gamma correlation (Kahana, 2002; Warrens, 2008). First, a 2×2 contingency table is constructed, tallying the number of pairs for which B and C were both recalled (quadrant 'a'), B recalled and C not recalled (quadrant 'b'), B not recalled and C recalled (quadrant 'c') and both not recalled (quadrant 'd'). $Q_{BC} = (ad - bc)/(ad + bc)$, and thus ranges from –1 to +1. $Q = 1$ signifies perfect, positive correlation (whenever B is recalled, so is C; whenever B is forgotten, so is C). $Q = 0$ signifies statistical independence (recall of B carries no information about recall of C). $Q < 0$ signifies a competitive relationship between B and C; that is, that recall of B and C are somewhat mutually exclusive. To avoid undefined quantities, one half an observation, 0.5, is added to each cell of the contingency table before computing Q. Q values are log-odds transformed (converted into logits) before conducting parametric analyses (Bishop, Fienberg, & Holland, 1975; Hayman & Tulving, 1989).

Hintzman (1972) noted that correlations will tend to be inflated due to Simpson's Paradox, explained as follows. In classic studies, correlations[4] were computed by first collapsing responses across all participants. However, consider that for some "high-performing" participants, nearly all tests of memory will be correct, whereas for other, "low-performing" participants, nearly all tests of memory will be incorrect. When

[4] actually, conditional probabilities, which are directly related to Q

the contingency table is collapsed across participants, this results in relatively more counts in the 'a' and 'd' cells, producing a more positive correlation, even for tests of memory of entirely unrelated pairs. Accordingly, we compute Yule's Q separately for each participant, which avoids the positive-valued contribution to the measure of correlation due to subject variability (Caplan, Rehani, & Andrews, 2014; Kahana, 2002; Rehani & Caplan, 2011; Rizzuto & Kahana, 2001). Still, the same logic can be applied to study sets. Because our participants perform multiple iterations of the entire AB/AC procedure, for some AB/AC sets, the participant might have performed well, and worse for other study sets. Because the contingency table is always computed from paired tests from within the same study set, one still expects a modest positive influence on the value of Q due to variability across study sets, as was confirmed by Burton et al. (2017). Thus, we also perform a bootstrap-based "control" correlation (Burton et al., 2017; Caplan et al., 2014; Caplan, 2005), computed from the contingency table relating the one AB with a different AC (A_iB_i and A_jC_j, $i \neq j$). The control correlation was computed using all eligible pairwise combinations of A_iB_i and A_jC_j, but then the contingency table was normalized to have the same number of total counts as for the Q_{BC} calculation, to match the effect of the 0.5 correction.

Finally, to test for a possible relationship between associative interference and Stroop interference (Aim 4), we computed Pearson correlations of measures of interference in the two tasks across older participants.

Complementing classical statistics, we present Bayesian analogues run in JASP (JASP Team, 2018) to test the robustness of important null effects, and the robustness of important positive effects. Bayes Factors, $\mathrm{BF}_{\mathrm{inclusion}}$, are reported for ANOVAs. Bayes Factors are ratios of evidence for the effect versus evidence for the null (i.e., the corresponding effect being absent from the data-model). We follow the convention that a Bayes Factor > 3 is considered "some" support for the effect and a Bayes Factor $< 1/3$ is considered "some" support for the null; BF> 5 or $< 1/5$ are considered moderate-to-strong support for the effect or the null, respectively.

Results

Self-pacing times and their relationship to MMFR are reported in the Appendix.

Accuracy in cued recall. In a 2×3×2 mixed, repeated-measures ANOVA for list 1, with design Age[Younger/Older]×Condition[Imagery-Only/Separation-Imagery/Integration-Imagery]×Pair Type[AB/DE], only the main effect of Age was significant, $F(1, 193) = 18.8,\ MSE = 0.10,\ p < 0.0001,\ \eta_p^2 = 0.89,\ \mathrm{BF_{inclusion}} = 614$ (Figure 2). Greater accuracy was observed for younger ($M = 0.80$) than older ($M = 0.65$) participants, replicating prior findings of lower association memory in healthy aging. All other effects were non-significant ($p > 0.4,\ \mathrm{BF_{inclusion}} \leq 0.23$). These results raise no concern about sampling bias, which is expected, due to the randomization of materials and assignment of participants to groups.

An ANOVA with the same design, for list 2 [AC/FG], again found a main effect of Age, $F(1, 193) = 29.5,\ MSE = 0.12,\ p < 0.0001, \eta_p^2 = 0.13,\ \mathrm{BF_{inclusion}} > 1000$, with greater accuracy for younger ($M = 0.74$) than for older ($M = 0.52$) participants. Condition was not a significant main effect, nor did it interact significantly with any factor ($p > 0.05$ but Bayes Factors were inconclusive, $0.3 < \mathrm{BF_{inclusion}} < 3$). However, the main effect of Pair Type was significant, $F(1, 193) = 16.9,\ MSE = 0.01,\ p < 0.0001,\ \eta_p^2 = 0.08,$ $\mathrm{BF_{inclusion}} > 1000$. This was qualified by a significant Age×Pair Type interaction, $F(1, 193) = 24.2,\ MSE = 0.010,\ p < 0.0001,\ \eta_p^2 = 0.11,\ \mathrm{BF_{inclusion}} > 1000$. This was a cross-over interaction where AC pairs were less accurate than FG pairs for older participants, $t(60) = -5.22,\ p < 0.0001$, paired samples, indicating PI for this group. The reverse was observed (non-significant) for the younger participants, $t(137) = 0.87,\ p = 0.39$.

MMFR. To test Aim 1, we analyzed MMFR (Figure 3) in the same manner as cued recall. The ANOVA for list 1 revealed a significant main effect of Age, $F(1, 193) = 31.0,\ MSE = 0.11,\ p < 0.0001,\ \eta_p^2 = 0.14,\ \mathrm{BF_{inclusion}} > 1000$, with greater accuracy for younger ($M = 0.77$) than older ($M = 0.56$) participants, showing that the associative deficit in aging persists with MMFR testing. All other effects were non-significant ($p > 0.2$, $\mathrm{BF_{inclusion}} < 0.18$), thus indicating no RI.

For list 2, Age was again a significant main effect, $F(1, 193) = 36.7$, $MSE = 0.12$, $p < 0.0001$, $\eta_p^2 = 0.16$, $\mathrm{BF_{inclusion}} > 1000$, with greater accuracy for younger ($M = 0.74$) than older ($M = 0.50$) participants. Pair Type was a significant main effect, $F(1, 193) = 14.7$, $MSE = 0.01$, $p < 0.001$, $\eta_p^2 = 0.07$, $\mathrm{BF_{inclusion}} > 1000$, but qualified by an interaction, Age×Pair Type, $F(1, 193) = 25.4$, $MSE = 0.01$, $p < 0.0001$, $\eta_p^2 = 0.12$, $\mathrm{BF_{inclusion}} > 1000$. As with cued recall, older participants had PI, significantly worse recall of AC than FG pairs, $t(60) = -4.77$, $p < 0.0001$, whereas younger participants had slightly greater accuracy for AC than for FG (n.s.; $p = 0.20$).

To test if this PI effect in MMFR was either amplified or attenuated in MMFR compared to cued recall, we repeated this last ANOVA on the difference between MMFR and cued-recall accuracy (older participants). This produced a non-significant effect of Pair Type, with very small effect size, $F(1, 58) = 0.15$, $p = 0.70$, $MSE = 0.006$, $\eta_p^2 = 0.003$. This raises one of two possibilities: One the one hand, the MMFR PI effect could be reiterating the same effect as in cued recall. Alternatively, the MMFR PI effect might be entirely due to output encoding during cued recall. If the latter were true, and participants, challenged by list discrimination, remembered and re-encoded AB pairs during cued-recall tests of AC, at the expense of AC, then one would expect a negative correlation between recall of AB and recall of AC when measured with MMFR. To anticipate, the correlation results revealed no such negative correlation, suggesting that PI during MMFR test is not caused by a list-discrimination deficit.

Testing for mediation by general association memory accuracy. Unsurprisingly, older participants performed worse even in List 1, than younger participants. This raises the question whether the effect of Age on PI might be mediated by individual differences in association memory ability, rather than a direct consequence of age. To test this, we reran the List-2 ANOVAs as ANCOVAs, with mean list-1 cued-recall accuracy as a covariate. For cued recall, the covariate was significant, $F(1,192) = 505.0,\ MSE = 0.03,\ p < 0.0001\ \eta_p^2 = 0.73$, but did not interact with Pair Type ($p > 0.7,\ \eta_p^2 = 0.001$). The main effect of Pair Type became non-significant, $F(1,192) = 2.55,\ MSE = 0.01,\ p = 0.112\eta_p^2 = 0.01$, suggesting that to a small degree, PI may be associated with low performance levels. However, the main effect of Age remained significant, $F(1,192) = 10.01,\ MSE = 0.33,\ p < 0.01,\ \eta_p^2 = 0.05$, as did Age×Pair Type, $F(1,191) = 21.1,\ MSE = 0.01,\ p < 0.0001,\ \eta_p^2 = 0.10$. The remaining effects were not significant ($p \geq 0.099$).

For MMFR, the outcome was similar. The covariate was a significant main effect, $F(1,192) = 513.1,\ MSE = 0.03,\ p < 0.0001,\ \eta_p^2 = 0.73$. The main effect of Pair Type was no longer significant, $F(1,192) = 1.81,\ MSE = 0.01,\ p = 0.18$. The main effect of Age remained significant, $F(1,192) = 18.6,\ MSE = 0.03,\ p < 0.0001,\ \eta_p^2 = 0.09$, as did Age×Pair Type, $F(1,192) = 22.7,\ MSE = 0.01,\ p < 0.0001,\ \eta_p^2 = 0.11$ (and see Figure 5a–c). All other effects were not significant ($p > 0.29$).

MMFR correlations. To test aims 2 and 3, we turn to the the correlation between AB and AC. We excluded participants who had zero accuracy in one or more of the four pair types in MMFR, because Yule's Q is undefined when all responses are correct or all responses are incorrect (excluded: older participants, N=3, 0 and 5; younger, N=8, 13 and 15, from Imagery-Only, Separation-Imagery and Integration-Imagery, conditions, respectively). A mixed, 2×3×2 ANOVA[5] on the log-odds transformed correlation, with design Age[2]×Condition[3]×Correlation Type[BC, Control] revealed a significant, and very robust, main effect of Correlation Type, $F(1,149) = 11.7$, $MSE = 0.57$, $p < 0.01$, $\eta_p^2 = 0.073$, $\mathrm{BF}_{\mathrm{inclusion}} > 1000$, reflecting a more positive value for the correlation between recall of B and C than the control correlation (i.e., associative facilitation overall). Age was

[5] Condition is included to check whether this factor might alter the overall interpretation, but effects involving Condition should be interpreted with caution, as noted, and are included in the Appendix.

not a significant main effect ($\text{BF}_{\text{inclusion}} = 0.32$), nor did it interact significantly with any factor ($p > 0.2,\ \eta_p^2 < 0.02$), although the two-way interaction, Age×Correlation Type was associated with an inconclusive Bayes Factor ($\text{BF}_{\text{inclusion}} = 0.42$).

Considering the diversity of instructed strategies across groups (although those effects were not robust), we followed up, analyzing the Imagery-Only participants. A mixed, 2×2 ANOVA on the log-odds transformed correlation, with design Age[2]×Correlation Type[BC, Control], revealed a significant main effect of Correlation Type, $F(1, 63) = 20.2,\ MSE = 0.42,\ p < 0.001,\ \eta_p^2 = 0.24,\ \text{BF}_{\text{inclusion}} > 1000$, and although the effect of Age was not significant, $F(1, 63) = 2.07$, $MSE = 1.37$, $p = 0.16$, $\eta_p^2 = 0.03$, $\text{BF}_{\text{inclusion}} = 1.33$, the interaction was significant, $F(1, 63) = 4.62$, $MSE = 0.42$, $p = 0.035$, $\eta_p^2 = 0.07$, $\text{BF}_{\text{inclusion}} = 3.03$. A Bayesian, one-tailed, paired-samples t test found support for the difference in correlations (BC minus Control) being significantly greater for younger than older participants, $\text{BF}_{-0} = 3.41$. we next asked whether direct, pair-to-pair associative interference, could be confidently ruled out for the older participants (Imagery-Only condition). A Bayesian, one-tailed, paired-samples t test rejected the prediction log-odds(Q_{BC})<log-odds($Q_{Control}$), $\text{BF}_{-0} = 0.078,\ \text{df} = 30$. Thus, approximate statistical independence of recall of AB and AC, as found previously for younger participants in a *timed* study procedure, when they were not instructed as to how to handle associative ambiguity (Burton et al., 2017), also extends to older participants with a *self-paced* procedure. However, for the younger participants in the Imagery-Only condition, not only was the prediction, log-odds(Q_{BC}) <log-odds($Q_{Control}$), confidently rejected, $\text{BF}_{-0} = 0.034,\ \text{df} = 44$, but the reverse, log-odds(Q_{BC}) <log-odds($Q_{Control}$), was strongly supported, $\text{BF}_{+0} = 997$. In other words, younger participants, when self-paced, spontaneously produced associative facilitation between AB and AC. In sum, with the self-paced study procedure, older participants show associative independence, but viewed alongside the younger participants' data, this can be understood as *less associative facilitation* than younger participants.

Relationship of associative interference to Stroop interference. Our fourth aim was to compare interference in the Stroop task to associative interference. The Stroop task, administered only with the older participants, produced very robust effects of interference, measured with both accuracy and response time. However, the hypothesized correlation of Stroop interference with associative interference (both mean-accuracy and correlation measures) were not supported. Full details are reported in the Appendix.

Discussion

With self-paced study, older participants exhibited PI on average, with similar magnitudes for both list-specific cued recall and MMFR, wherein participants are asked to recall both associates of the cue in any order. Younger participants showed no PI on average. Both groups showed no evidence of RI, consistent with prior results in younger participants, with a timed procedure (Burton et al., 2017). Associative independence, previously reported for younger participants with a timed study procedure (Burton et al., 2017), was found here for older participants with a self-paced study phase, suggesting that at least in the absence of time pressure, older participants are also able to resolve associative interference. On the other hand, the self-paced study procedure apparently gave younger participants the ability not only to neutralize associative interference, but to produce a facilitating relationship between AB and AC. Such a facilitating outcome was found by Burton et al. (2017) in a group of younger participants who were explicitly instructed to apply integration imagery, combining A, B and C into a single image. This suggests either, with the self-paced procedure, younger participants might spontaneously construct integration imagery, or else integration imagery is not strictly necessary to produce a facilitating relationship between AB and AC. Correlations across older participants suggest no common cause of Stroop interference and associative interference in this task. We discuss the implications of these results below.

Aim 1: Retroactive and proactive interference in mean-accuracy measures. The most striking age effect was the overall presence of PI for older, in contrast to no net PI for the younger participants. This held both for the (list-specific) cued-recall measure and for the MMFR measure, with similar magnitudes, suggesting PI in older participants is not due to a list-discrimination deficit. These findings echo those of Siegel (2014), who also found PI in older but not younger participants. Our findings suggest this pattern was not limited to semantically similar pairs.

Because cued recall preceded MMFR, an objection could be raised. Namely, it is conceivable that the PI effect in cued recall is entirely due to confusion about list-source. List-discrimination difficulties would sometimes result in B being recalled in place of C during cued recall. If one additionally assumes one form of output encoding, that during cued recall, the participant re-encodes the cue word paired with their response, whether or not the response were correct, then if AB were re-encoded during such error trials, one would fully expect MMFR also to exhibit PI: erroneous recalls of B would displace AC from being re-encoded. This would not occur for control pairs (FG pairs). In other words, one could argue that the PI (on average) we observed for MMFR responses may in fact be entirely attributable to list-discrimination, inherited from cued recall through output encoding. However, a telltale signature of this output-encoding effect would be a negative correlation between recall of B and recall of C. Inconsistent with that, recall of B and C were independent in older participants' responses. In sum, it seems implausible that the PI observed in the average MMFR accuracy measure reflects list-discrimination difficulties.

In our reading of research on the AB/AC paradigm with healthy older adult participants, we could find no prior study that provided a complete picture about retroactive interference and proactive interference in healthy aging. Healthy older samples are often included as age-matched controls for patients, in which case, no comparison to younger participants is typically included (e.g., Mayes, Pickering, & Fairbairn, 1987; Shimamura, Jurica, Mangels, Gershberg, & Knight, 1995; Van der Linden, Bruyer, Roland,

& Schils, 1993). Studies often also omit control pairs (e.g., Mayes et al., 1987; Shimamura et al., 1995; Van der Linden et al., 1993; Winocur, Moscovitch, & Bruni, 1996). Finally, most studies used list-specific cued-recall, often with no test of RI (but see Siegel, 2014). Without MMFR, one cannot distinguish whether PI effects are due to response competition or list-discrimination. Shimamura et al. (1995) administered a variant of AB/AC learning with initial cued recall followed by MMFR in frontal-lesioned amnesic patients and controls. Demographics were not reported, but if age-matched, the controls would have ranged from 62–75 years old. Although the results were suggestive of the presence of PI, they were ambiguous, because only AB and AC pairs were included. We also note that Winocur and Moscovitch (1983) did not include control pairs within the same lists, but included a separate control group of participants who studied only control pairs. They found clear and large-magnitude PI, larger for older than younger participants, compatible with our findings. However, they did not use MMFR, which left open the possibility that the these differences could have been due to differences in list-discrimination ability. Our findings lend clarity, suggesting that Winocur and Moscovitch's results, indeed, probably would not have been explained away with a list-discrimination account. Guez and Naveh-Benjamin (2016) found age-related deficits in associative recognition, a task that is related to cued recall, but age did not interact with their measure of PI. Thus, our study is, to our knowledge, the first to quantify the presence/absence of both RI and PI in AB/AC learning of unrelated word-pairs in healthy older, compared to younger participants. Moreover, with MMFR, we are able to isolate competition and facilitation effects that are not explainable based on list-discrimination ability.

Limitations. Baseline associative memory performance. One complicating factor is that overall accuracy was reduced in the older compared to younger participants, replicating a well known, robust finding (e.g., Castel & Craik, 2003; Naveh-Benjamin, 2000; Naveh-Benjamin, Guez, & Marom, 2003). This was the case even despite older participants taking far longer during self-paced study. Some researchers have argued that it is necessary to equate association memory before one can compare associative interference between age groups (e.g., Ebert & Anderson, 2009). Alternatively, in doing so, one might actually increase the strengths of the AB pairs, exacerbating the challenge from PI for older participants.[6]

There may be no perfect design. We checked whether our finding of an age difference in PI might be due to differences in basic association memory. Figure 5 shows that the age differences in both proactive facilitation/interference (a–c) and correlation between recall of B and C (Q_{BC}, d–f) are not likely mediated by general association memory ability.

Source of proactive interference. PI, as measured with mean accuracy, could have several causes. As already argued, list-discrimination is an unlikely explanation. Next, consider that compared to control pairs, the "A" words are seen twice. Item-repetition might plausibly enhance cued-recall. Moreover, there could be a cognitive "overhead" cost to integrating each new word into a visual image. AC pairs, thus, may benefit from reduced overhead-cost because the participant has already figured out how to imagine A. If these putative effects enhance cued recall of interference pairs, perhaps Younger participants do experience PI, but this interference is roughly cancelled out by those putative enhancing effects. For older participants, then, the enhancement due to A items being repeated may not completely compensate for their PI. Thus, we cannot conclusively interpret the PI effect, as measured by mean accuracy, as indicating that older participants are more susceptible to PI than younger participants. The correlation measure of associative interference, Q_{BC}, was not negative for either age group, so it is plausible that the age difference in the mean-accuracy PI measure is not due to differences in PI, but a difference in the advantage due to item-repetition (cf. Fine, Shing, & Naveh-Benjamin, 2018).

Pair-specific versus response-set mechanisms. Finally, although reduced recall of AC compared to FG pairs could be due to pair-specific competition (A_iB_i competing with A_iC_i), our correlation measure suggests otherwise. Rather than

[6] but note more PI was found for our older than younger participants, opposite the concern this raises.

pair-specific competition, the entire set of response words ("C" items) is inhibited, an account that was proposed to accommodate classic findings of associative independence in younger participants (Postman, Stark, & Fraser, 1968; Postman & Underwood, 1973; Thune & Underwood, 1943; Underwood, 1945, 1949; Wang, 1980; Winograd, 1968).

Aims 2 and 3: Coupling of recall of B and C, effects of age and self-pacing. Our findings replicate and extend the boundary conditions of Burton et al.'s (2017) results. First, with younger participants, Burton and colleagues (2017) found associative independence, and associative facilitation for participants asked to use Integration Imagery, using a timed study procedure. Here, we used a self-paced study procedure and also found independence for older participants, and facilitation (positive Q_{BC}) for younger participants given only general instructions to form visual imagery. Thus, in contrast to mixed-material pairs, which produced associative competition (Burton et al., 2017; Caplan et al., 2014; Hintzman, 1972; Riefer & Batchelder, 1988), we replicate lack of competition (independence, and even associative facilitation) for word–word pairs reported by Burton et al. (2017); Tulving and Watkins (1974), and extend the lack of associative competition to healthy older participants.

Limitations.

Imagery. Although intended to test effects of instructed strategy, sample sizes Condition were modest for older participants, and clear effects that might have validated the instructional manipulation did not emerge. Future studies could better assess the operation of instructed or spontaneous strategies in aging, as investigations of strategy usage in aging, including visual imagery mediators, has often found similarities, rather than differences across age groups, without explaining aging deficits (e.g., Bailey & Hertzog, 2014; Dunlosky & Hertzog, 1998; Kuhlmann & Touron, 2017).

Self-pacing. In addition, it is unclear how the data for older adults would look with a timed procedure. Younger participants exhibited associative independence with a timed procedure in data reported by Burton et al. (2017), and facilitation with the self-paced procedure here. Extrapolating, it is thus conceivable that while older participants showed independence of AB and AC with self-paced study, they would show competition (a negative correlation between recall of AB and AC) if the procedure had been timed (see Appendix). Our main concern in selecting the self-paced procedure was to prevent older participants' accuracy (due to the well established associative memory deficit in aging) lurking near floor levels, which would have caused problems for the correlation analyses. Future work could investigate whether self-pacing is a necessary precondition for healthy older adults to resolve associative interference.

Aim 4: Color-Stroop interference and associative interference. The Stroop task measures interference due to response conflict. Although robust Stroop effects were found in the older sample, they failed to show robust relationships to measures of either PI or RI, nor to the correlation between recall of B and C. Little and Hartley (2000) found a positive correlation between the standard measure of Stroop interference and a measure of PI in the Stroop task. Intriguingly, our analyses also revealed a robust correlation, across participants, between the Stroop-PI measure and PI in the AB/AC task. However, the sign of the correlation indicated that more Stroop-PI was associated with less PI in AB/AC learning. It is possible this complex pattern is due to the fact that in the Stroop task, the participant has no need to remember prior trials; the best way to cope with the current trial is to suppress or forget prior trials, in stark contrast to AB/AC learning. This result leaves us with two possible conclusions: either the two sources of interference are moderately mutually exclusive (anti-correlated across participants) or the cause of PI in the Stroop task is, actually, protective against PI (measured with average accuracy) in AB/AC learning.

Conclusion

Proactive interference was observed for older, but not younger participants on average. Despite this, and the typical reduction in association memory with age, older participants apparently resolved direct associative competition between memory of AB and AC, exhibiting statistical independence between recall of B and C. As elegantly shown by Healey and Kahana (2016) for age effects on free recall, age differences like this may originate from differences in more than one underlying cognitive mechanism. At least in the absence of time pressure, this suggests that older participants may resolve associative interference in AB/AC learning situations, even without any particular instruction or training on how to handle associative ambiguity. Intriguingly, the challenge from list-discrimination did not produce direct competition between memory for AB and AC; to the contrary, it may be that studying AB and AC pairs in distinct lists or contexts enables resolution of associative interference, as suggested by Burton et al. (2017). It remains to be seen whether associative interference that materializes within a single study set, which produced clear associative competition effects in younger participants (Caplan et al., 2014), could be strategically overcome by either younger or older participants. As well, for the younger participants, the self-paced procedure shifted what was previously an independence result with timed study, to facilitation, even without instructions to integrate AB and AC into a single representation. This suggests either that participants can spontaneously discover integrative strategies, or that integration is not strictly necessary to reverse associative competition.

References

Bailey, D. J., Heather R., & Hertzog, C. (2014). Does strategy training reduce age-related deficits in working memory? *Gerontology*, *60*(4), 346-356.

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology*, *58*(2), 97-105.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). *Discrete multivariate analysis: theory and practice*. Cambridge, MA: MIT Press.

Biss, R. K., Rowe, G., Weeks, J. C., Hasher, L., & Murphy, K. J. (2018). Leveraging older adults' susceptibility to distraction to improve memory for face-name associations. *Psychology and Aging*, *33*(1), 158-164.

Bissig, D., & Lustig, C. (2007). Who benefits from memory training? *Psychological Science*, *18*(8), 720-726.

Burton, R. L., Lek, I., & Caplan, J. B. (2017). Associative independence revisited: competition between conflicting associations can be resolved or even reversed in one trial. *Quarterly Journal of Experimental Psychology*, *70*(4), 832-857.

Campbell, K. L., & Hasher, L. (2018). Hyper-binding only apparent under fully implicit test conditions. *Psychology and Aging*, *33*(1), 176-181.

Caplan, J. B. (2005). Associative isolation: unifying associative and order paradigms. *Journal of Mathematical Psychology*, *49*(5), 383-402.

Caplan, J. B., Rehani, M., & Andrews, J. C. (2014). Associations compete directly in memory. *Quarterly Journal of Experimental Psychology*, *67*(5), 955-978.

Carpenter, A. C., & Schacter, D. L. (2018). Flexible retrieval mechanisms supporting successful inference produce false memories in younger but not older adults. *Psychology and Aging*, *33*(1), 134-143.

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*(4), 873-885.

Comalli, P. E. J., Wapner, S., & Werner, H. (1962). Interference effects of Stroop color-word test in childhood, adulthood, and aging. *Journal of Genetic Psychology*, *100*(1), 47-53.

Deary, I. J., Johnson, W., & Starr, J. M. (2010). Are processing speed tasks biomarkers of cognitive aging? *Psychology and Aging*, *25*(1), 219-228.

Dunlosky, J., & Hertzog, C. (1998). Aging and deficits in associative memory: what is the role of strategy production? *Psychology and Aging*, *13*(4), 597-607.

Ebert, P. L., & Anderson, N. D. (2009). Proactive and retroactive interference in young adults, healthy older adults, and older adults with amnestic mild cognitive impairment. *Journal of the International Neuropsychological Society*, *15*, 83-93.

Fine, H. C., Shing, Y. L., & Naveh-Benjamin, M. (2018). Effects of changes in schematic support and of item repetition on age-related associative memory deficits: theoretically-driven empirical attempts to reduce older adults' high false alarm rate. *Psychology and Aging*, *33*(1), 57-73.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189-198.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods and Instrumentation*, *14*, 375-399.

Guez, J., & Naveh-Benjamin, M. (2016). Proactive interference and concurrent inhibitory processes do not differentially affect item and associative recognition: implication for the age-related associative memory deficit. *Memory*, *24*(8), 1091-1107.

Hanseeuw, B. J., Seron, X., & Ivanoiu, A. (2012). Increased sensitivity to proactive and retroactive interference in amnestic mild cognitive impairment: new insights. *Brain and Cognition*, *80*(1), 104-110.

Hayman, C. A. G., & Tulving, E. (1989). Contingent dissociation between recognition and fragment completion: the method of triangulation. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, *15*(2), 228-240.

Healey, M. K., Hasher, L., & Campbell, K. L. (2013). The role of suppression in resolving interference: evidence for an age-related deficit. *Psychology and Aging*, *28*(3), 721-728.

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review*, *123*(1), 23-69.

Hintzman, D. L. (1972). On testing the independence of associations. *Psychological Review*, *79*(3), 261-264.

JASP Team. (2018). *JASP (Version 0.9)[Computer software].* Retrieved from `https://jasp-stats.org/`

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993, 3-28). Source monitoring. *Psychological Bulletin*, *114*, 1.

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition*, *30*(6), 823-840.

Kuhlmann, B. G., & Touron, D. R. (2017). Relate it! objective and subjective evaluation of mediator-based strategies for improving source memory in younger and older adults. *Cortex*, *91*, 25-39.

Little, D. M., & Hartley, A. A. (2000). Further evidence that negative priming in the Stroop color-word task is equivalent in older and younger adults. *Psychology and Aging*, *15*(1), 9-17.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*(4), 476-490.

Ludwig, C., Borella, E., Tettamanti, M., & de Ribaupierre, A. (2010). Adult age differences in the Color Stroop Test: a comparison between an item-by-item and a blocked version. *Archives of Gerontology and Geriatrics*, *51*(2), 135-142.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*(2), 199-207.

Martin, E. (1971a). Stimulus component independence. *Journal of Verbal Learning and Verbal Behavior*, *10*(6), 715-721.

Martin, E. (1971b). Verbal learning theory and independent retrieval phenomena. *Psychological Review*, *78*(4), 314-332.

Mayas, J., Fuentes, L. J., & Ballesteros., S. (2012). Stroop interference and negative priming (NP) suppression in normal aging. *Archives of Gerontology and Geriatrics*, *54*(2), 333-338.

Mayes, A. R., Pickering, A., & Fairbairn, A. (1987). Amnesic sensitivity to proactive interference: its relationship to priming and the causes of amnesia. *Neuropsychologia*, *25*(1B), 211-220.

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*(5), 1170-1187.

Naveh-Benjamin, M., Guez, J., & Marom, M. (2003). The effects of divided attention at encoding on item and associative memory. *Memory & Cognition*, *31*(7), 1021-1035.

Naveh-Benjamin, M., & Mayr, U. (2018). Age-related differences in associative memory: empirical evidence and theoretical perspectives. *Psychology and Aging*, *33*(1), 1-6.

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*(1), 104-118.

Overman, A. A., & Becker, J. T. (2009). The associative deficit in older adult memory: recognition of pairs is not improved by repetition. *Psychology and Aging*, *24*(2), 501-506.

Postman, L., Stark, K., & Fraser, J. (1968). Temporal changes in interference. *Journal of Verbal Learning and Verbal Behavior*, *7*(3), 672-694.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition*, *1*, 19-40.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93-134.

Rehani, M., & Caplan, J. B. (2011). Interference and the representation of order within associations. *Quarterly Journal of Experimental Psychology*, *64*(7), 1409-1429.

Riefer, D. M., & Batchelder, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, *95*(3), 318-339.

Rizzuto, D. S., & Kahana, M. J. (2001). An autoassociative neural network model of paired-associate learning. *Neural Computation*, *13*, 2075-2092.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, *103*(3), 403-428.

Shimamura, A. P. (1994). Neuropsychological perspectives on memory and cognitive decline in normal human aging. *Seminars in the Neurosciences*, *5*(6), 387-394.

Shimamura, A. P., Jurica, P. J., Mangels, J. A., Gershberg, F. B., & Knight, R. T. (1995). Susceptibility to memory interference effects following frontal lobe damage: findings from tests of paired-associate learning. *Journal of Cognitive Neuroscience*, *7*(2), 144-152.

Siegel, J. (2014). *Brain volume differences and proactive interference: a VBM investigation of healthy cognitive aging* (Unpublished master's thesis). University of Pittsburgh.

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: a meta-analysis. *Psychology and Aging*, *10*(4), 527-539.

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 461-479.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643-662.

Thune, L. E., & Underwood, B. J. (1943). Retroactive inhibition as a function of degree of interpolated learning. *Journal of Experimental Psychology*, *32*(3), 185-200.

Tulving, E., & Watkins, M. J. (1974). On negative transfer: effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior*, *13*(2), 181-193.

Umanath, S., & Marsh, E. J. (2014). Understanding how prior knowledge influences memory in older adults. *Psychological Science*, *9*(4), 408-426.

Underwood, B. J. (1945). The effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs*, *59*(3), 1-33.

Underwood, B. J. (1949). Proactive inhibition as a function of time and degree of prior learning. *Journal of Experimental Psychology*, *39*(1), 24-34.

Van der Linden, M., Bruyer, R., Roland, J., & Schils, J. P. (1993). Proactive interference in patients with amnesia resulting from anterior communicating artery aneurysm. *Journal of Clinical and Experimental Neuropsychology*, *15*(4), 525-536.

Verhaeghen, P., & De Meersman, L. (1998). Aging and the Stroop effect: a meta-analysis. *Psychology and Aging*, *13*(1), 120-126.

Wang, A. Y. (1980). Evidence for a "selective" selector mechanism. *American Journal of Psychology*, *93*(4), 643-656.

Warrens, M. J. (2008). On association coefficients for 2 × 2 tables and properties that do not depend on the marginal distributions. *Psychometrika*, *73*(4), 777-789.

Winocur, G., & Moscovitch, M. (1983). Paired-associate learning in institutionalized and noninstitutionalized old people: an analysis of interference and context effects. *Journal of Gerontology*, *38*(4), 455-464.

Winocur, G., Moscovitch, M., & Bruni, J. (1996). Heightened interference on implicit, but not explicit, tests of negative transfer: evidence from patients with unilateral temporal lobe lesions and normal old people. *Brain and Cognition*, *30*(1), 44-58.

Winograd, E. (1968). List differentiation as a function of frequency and retention interval. *Journal of Experimental Psychology*, *76*(2, Suppl. 2), 1-18.

Condition	Total	Sex M/F/NR*	Age M±SD (NR*)	Education**	MMSE†	Health‡
	Younger Participants					
Imagery-Only	45	17/13/4	19.50±1.61 (4)	–	–	–
Separation-Imagery	46	19/17/1	19.11±1.57 (2)	–	–	–
Integration-Imagery	47	11/23/3	19.00±1.14 (2)	–	–	–
	Older Participants					
Imagery-Only	30	8/22/0	67.53±4.38 (0)	14.84±3.11	29.11±1.24	11/15/3/1/0
Separation-Imagery	15	5/9/1	67.53±5.03 (0)	15.73±2.76	29.33±0.72	10/4/1/0/0
Integration-Imagery	15	4/11/0	66.87±3.89 (0)	14.60±2.92	28.69±1.20	10/5/1/0/0

Table 1

*Sample sizes, sex breakdown and ages (Mean±SD) for all participants. *NR=not reported. For younger participants, note that due to a technical failure, age and sex responses were not requested for 11, 10 and 9 remaining participants, respectively, but these participants were recruited from the same research participation pool. For older participants: **years (Mean±SD). † Mini Mental State Examination (Mean±SD). ‡Self-reported on a five-point scale (Mean±SD): Very good, Good Fair, Poor, Very poor (see text). Condition: Strategy instruction subgroup (see text).*

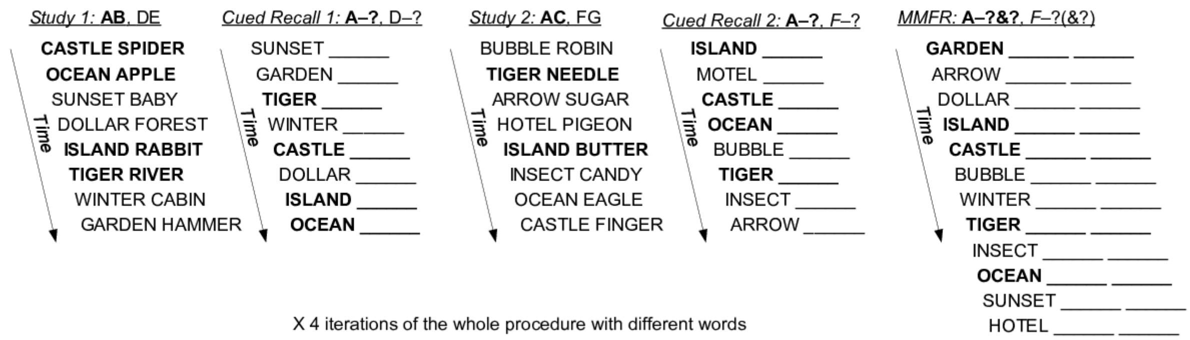

Figure 1. Experimental design. Participants studied two noun-pair lists, each containing half interference pairs (**AB** or **AC**, respectively, set in boldface here for illustration only) and half no-interference pairs (DE or FG, respectively). Following study, each list was tested with cued recall. After both lists had been studied and tested, both lists were tested with modified modified free recall (MMFR). Underlines denoting response fields are depicted to the right of the cue word for illustration only; participants saw response fields were underneath the corresponding cue word. A mathematical distractor task (not shown) separated all study and test phases.

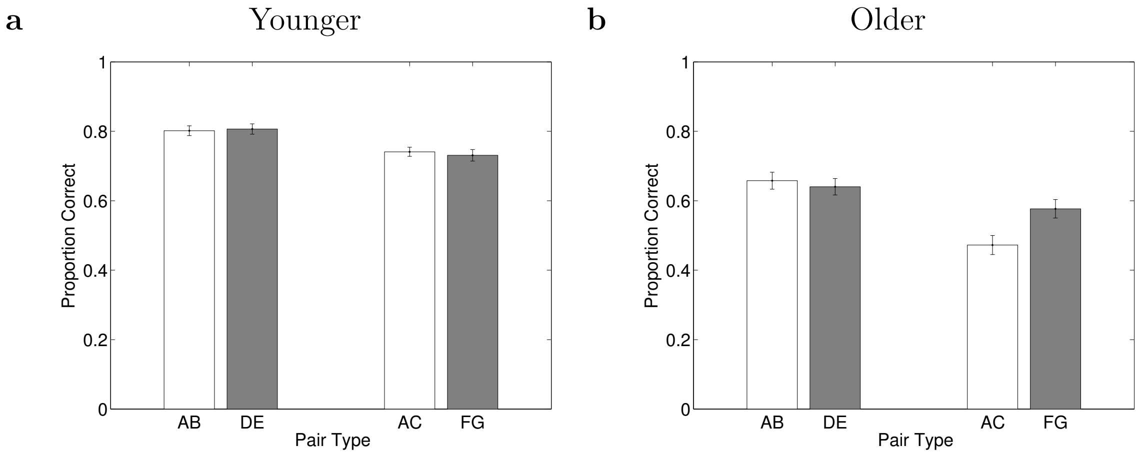

Figure 2. Cued recall accuracy. Error bars plot 95% confidence intervals corrected for subject variability (Loftus & Masson, 1994).

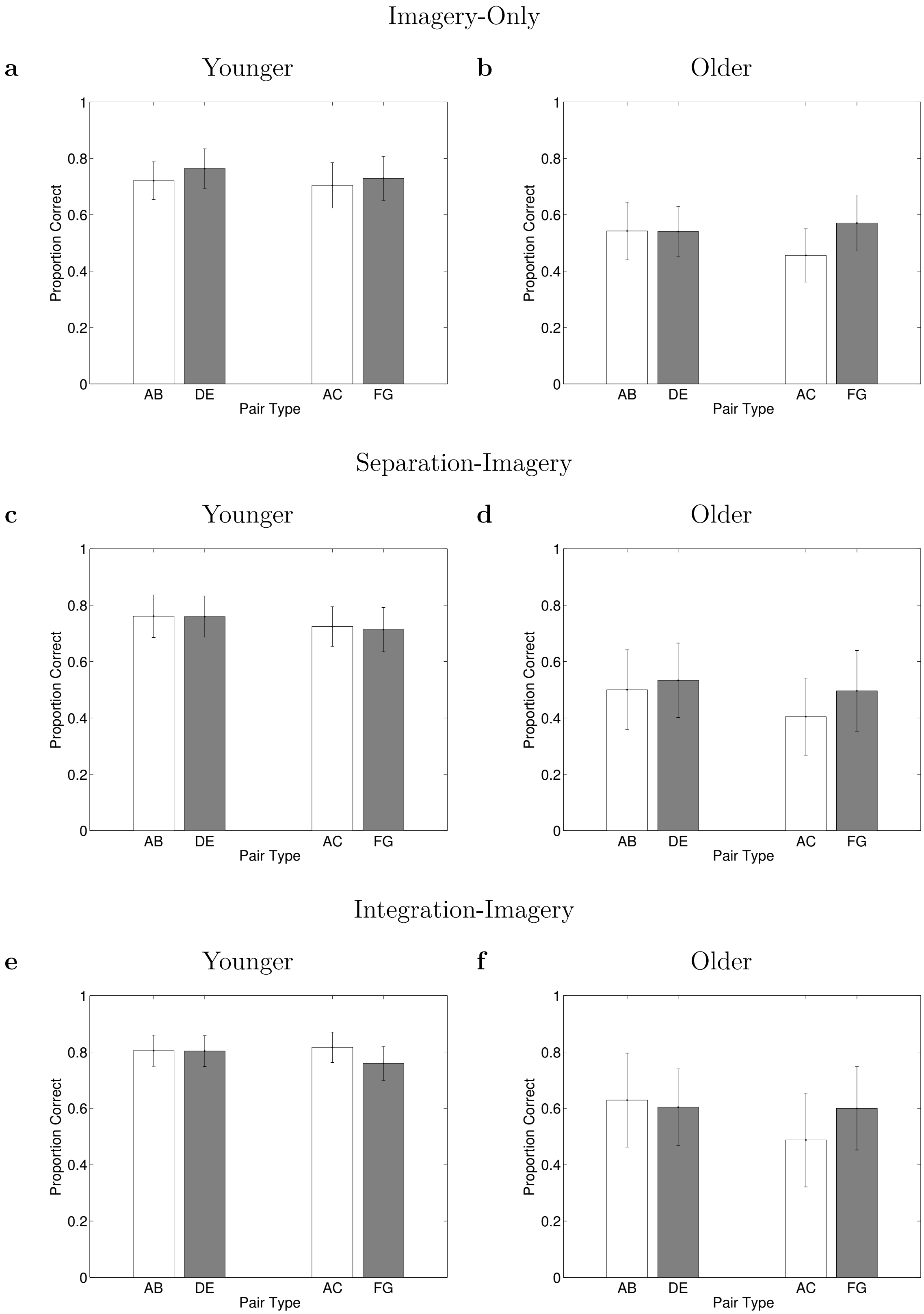

Figure 3. Modified modified free recall (MMFR) accuracy. Error bars plot 95% confidence intervals corrected for subject variability (Loftus & Masson, 1994).

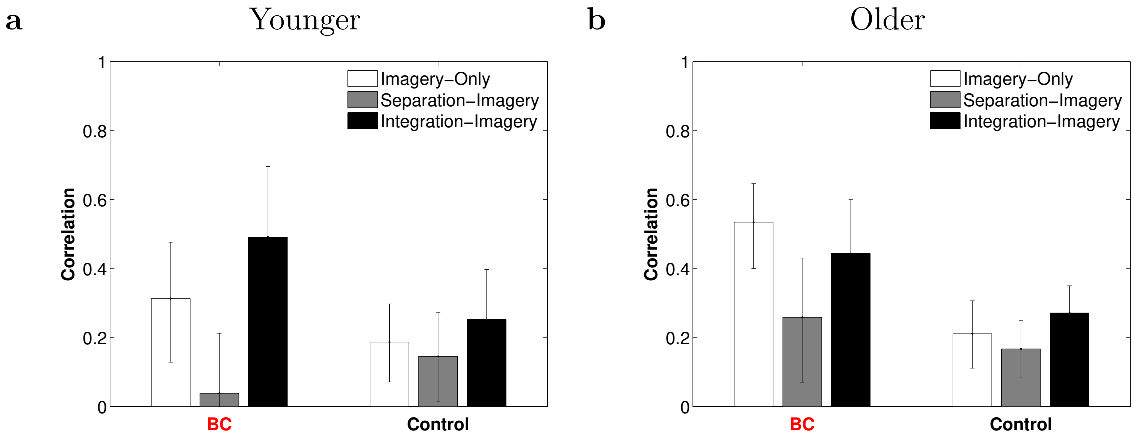

Figure 4. Correlation between recall of B and C, given A as the cue ("BC") and a control correlation, computed between pairs of control pairs (see text), to estimate the expected positive correlation due to variability across study/test cycles ("Control"), for younger (a) and older (b) participants. Error bars plot 95% confidence intervals computed via the log-odds transform.

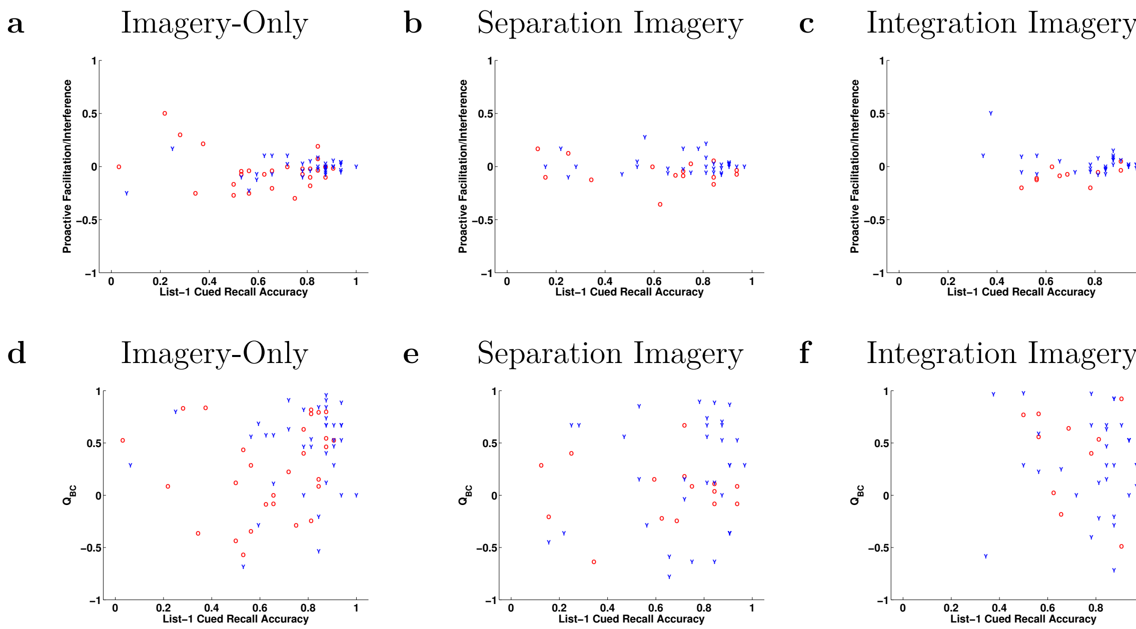

Figure 5. Lack of strong dependence of proactive interference and Q_{BC} on basic association memory skill. Basic association memory is quantified as the average accuracy in cued recall of List 1 (both AB and DE pairs), thus, before any associative-interference effects appear for a given list; this is plotted on the horizontal axis. a–c, a measure of proactive facilitation/interference is plotted on the vertical axis. This is a normalized measure, quantified as MMFR Accuracy 0.5×(AC–FG)/(AC+FG). Thus, positive values indicate proactive facilitation, and negative values, proactive interference. d–f, The correlation between recall of B and recall of C during MMFR (Q_{BC}) is plotted on the vertical axis. Each point represents a single participant; older participants are denoted with the letter 'O' and younger participants, 'Y.'

Appendix A

Self-pacing times

We conducted a 2×3×2, mixed, repeated-measures ANOVA on self-pacing time per pair, with design Age[Younger/Older]×List[1/2]×Condition[Imagery-Only/Separation-Imagery/Integration-Imagery]. The significant main effect of List, $F(1, 193) = 51.7,\ MSE = 268, 101\ \text{ms}^2,\ p < 0.0001,\ \eta_p^2 = 0.21$, indicated faster self-pacing on second, AC/FG lists ($M = 9260$ ms) than the first, AB/DE lists ($M = 9339$ ms). The significant main effect of Age, $F(1, 193) = 48.3$, $MSE = 26, 801, 137\ \text{ms}^2$, $p < 0.0001$, $\eta_p^2 = 0.20$, was due to Older participants taking longer to study ($M = 11, 587$ ms) than Younger participants ($M = 7, 512$ ms). Qualifying these main effects, the List×Age interaction, $F(1, 193) = 22.2,\ MSE = 268, 101\ \text{ms}^2$, $p < 0.0001$, $\eta_p^2 = 0.10$, List×Condition interaction, $F(2, 193) = 8.33,\ MSE = 268, 101\ \text{ms}^2,\ p < 0.0001$, $\eta_p^2 = 0.08$, and three-way, List×Condition×Age interaction, $F(2, 193) = 7.79,\ MSE = 268, 101\ \text{ms}^2$, $p < 0.001$, $\eta_p^2 = 0.08$, were significant. The main effect of Condition and Age×Condition interaction did not reach significance ($F < 2.2,\ p > 0.11,\ \eta_p^2 \leq 0.023$). To understand the significant interactions, simple effects were conducted, analyzing Younger and Older participants separately. For Older participants, the main effect of Condition was not significant, $F(1, 58) = 1.30,\ MSE = 56, 041, 320\ \text{ms}^2,\ p = 0.28,\ \eta_p^2 = 0.04$. The main effect of List was significant, $F(1, 58) = 17.0,\ MSE = 780, 483\ \text{ms}^2,\ p < 0.001,\ \eta_p^2 = 0.23$. This was qualified by a significant List×Condition interaction, $F(2, 58) = 3.74,\ MSE = 780, 483\ \text{ms}^2$, $p < 0.05$, $\eta_p^2 = 0.11$. Pairwise, paired-samples t tests found, for Older Imagery-Only participants, significantly slower self-pacing times on List 1 ($M = 10, 364$ ms) than List 2 ($M = 9, 920$ ms), $t(30) = 2.40,\ p < 0.023$ and likewise for the Older Integration-Imagery participants ($M = 12, 971$ ms and $11, 600$ ms on Lists 1 and 2, respectively), $t(14) = 3.17,\ p < 0.01$, but no significant difference for Separation-Imagery participants ($M = 12, 206$ ms and $11, 943$ ms, for Lists 1 and 2, respectively), $t(14) = 0.88,\ p = 0.39$. For Younger participants, the main effect of Condition was also non-significant,

$F(1,135) = 0.24,\ MSE = 1,423,8687\ \text{ms}^2,\ p = 0.79,\ \eta_p^2 = 0.004.$ The main effect of List was significant, $F(1,135) = 30.2,\ MSE = 47,966\ \text{ms}^2,\ p < 0.0001,\ \eta_p^2 = 0.18$, with longer self-pacing times on List 1 ($M = 7,584$ ms) than List 2 ($M = 7439$ ms). Unlike the Older participants, the List×Condition interaction was not significant for the Younger participants, $(2,135) = 0.10,\ MSE = 47,966\ \text{ms}^2,\ p = 0.90,\ \eta_p^2 = 0.001.$ In sum, older participants took more advantage of the self-paced procedure, taking longer to study. In general, list 2 was studied slightly faster than list 1 by all participants. As noted, Condition only minimally influenced self-pacing times, suggesting that subsequent effects of Condition should be viewed as inconclusive.

Relationship between self-pacing times and MMFR accuracy

We next asked if participants' self-pacing could have been responsible for the associative interference effects. Collapsed across Condition, but separately for Older and Younger participants, we computed correlations between three measures of self-pacing time and three measures of associative interference. Self-pacing measures were: 1) SP1 = average self-pacing time on List 1, 2) SP2 = average self-pacing time on List 2 and 3) SPdiff = SP2–SP1. The associative interference measures were: 4) Q = log-odds transformed Q_{BC}, 5) RI=MMFR Accuracy(AB)–MMFR Accuracy(DE) and 6) PI=MMFR Accuracy(AC)–MMFR Accuracy(FG). The resulting 18 correlations were all non-significant when Bonferroni-corrected ($\alpha \simeq 0.0028$, smallest $p = 0.015$). With this caveat in mind, effects that were significant when uncorrected may give us clues as to possible relationships between self-pacing and interference effects.

For the Younger participants, uncorrected, all $p > 0.1$ apart from the correlation between the RI with SP1, $r(136) = 0.17,\ p = 0.044$ and with SP2, $r(136) = 0.17, p = 0.050$. Because the correlation between RI and SPdiff was far from significant, this suggests an overall self-pacing speed effect: participants who were faster to study both lists tended to produce a disadvantage for recall of AB, compared to DE, whereas participants who took longer during study of both lists tended to produce greater recall of AB than DE.

For the Older participants, there was a a positive correlation of Q with SP1, $r(59) = 0.31,\ p = 0.015$ and with SP2, $r(59) = 0.29,\ p = 0.023$ (but not with SPdiff). Thus, older participants may, indeed, have been taking more time to study, with the effect of resolving associative interference; or put the opposite way, older participants who took less time during study produced more negative Q values. This is in line with our suspicion that with a timed procedure, our older sample may very well have produced associative competition (somewhat mutually exclusive relationship between recall of B and recall of C). SP1 and SP2 were also positively correlated with PI; $r(59) = 0.22,\ p = 0.88$ and $r(59) = 0.20,\ p = 0.13$, respectively, which may suggest that a timed procedure would have amplified the aging PI effect even further, but even uncorrected, these correlations failed to reach significance. The remaining comparisons were quite far from significant ($p > 0.4$, uncorrected).

In sum, the analyses of self-pacing times suggest that self-pacing was unlikely to have produced the age differences we observed in MMFR; if anything trends suggest the opposite, that some older participants may have taken advantage of additional study time to reduce what would otherwise have been more negative Q_{BC} values and more pronounced proactive interference.

Appendix B

MMFR correlations: effects involving Condition

Here we report the effects resulting from the ANOVA on MMFR correlations, that involve Condition. The main effect of Condition was significant, $F(2, 149) = 3.28,\ MSE = 1.07,\ p = 0.041,\ \eta^2_p = 0.046$, but no post-hoc tests reached significance, $p > 0.05$, and the corresponding Bayes Factor was inconclusive ($\mathrm{BF}_{\mathrm{inclusion}} = 0.81$). This was qualified by a significant Condition×Correlation Type interaction, $F(2, 149) = 3.39,\ MSE = 0.57,\ p = 0.037,\ \eta^2_p = 0.043$.

Bayesian ANOVA found strong support for the null interactions Pair Type×Age×Condition, $\mathrm{BF}_{\mathrm{inclusion}} = 0.053$ and Age×Condition, $\mathrm{BF}_{\mathrm{inclusion}} = 0.13$, but was inconclusive for Pair Type×Condition, $\mathrm{BF}_{\mathrm{inclusion}} = 1.1$ and Pair Type×Age, $\mathrm{BF}_{\mathrm{inclusion}} = 0.52$).

To follow up on the interaction, simple effects analyzing the control correlation found no significant main effect of Condition ($p = 0.35,\ \eta^2_p = 0.014$), but for Q_{BC}, the main effect of Condition was significant, $F(1, 149) = 3.88,\ MSE = 1.58,\ p < 0.05,\ \eta^2_p = 0.049$, although the Bayes Factor was inconclusive, $\mathrm{BF}_{\mathrm{inclusion}} = 1.6$. Bonferroni-corrected post-hoc t tests indicated that the Imagery-Only and Interactive-Imagery groups did not differ significantly from each other, but the Separation-Imagery participants had a lower correlation than the Imagery-Only group ($p = 0.045$). The Separation-Imagery group had a non-significant trend toward a lower correlation than the Integration-Imagery group ($p = 0.057$). Because the latter was a planned comparison, we note that uncorrected, the Separation-Imagery and Integration-Imagery conditions were significantly different, $p = 0.019$. Both the main effect of Age and the interaction Age×Condition were non-significant ($p = 0.20$ and $p = 0.47$, respectively, although here, the Bayes Factor suggests a meaningful main effect of Age, $\mathrm{BF}_{\mathrm{inclusion}} = 4.08$; the interaction had an inconclusive, $\mathrm{BF}_{\mathrm{inclusion}} = 0.65$). This suggests that both the younger and older participants may have produced more associative facilitation when asked to form Integration Imagery than when asked to form Separation Imagery, but due to the statistical fragility of the results, effects of Condition should be considered with caution.

Appendix C

Relationship of associative interference to Stroop interference

First we checked whether there was reliable evidence of Stroop interference in our sample. A 3×3, mixed, repeated-measures ANOVA on Stroop accuracy, with the design Condition[3]×Stroop Type[3] (Congruent/Incongruent/Control) revealed only a very robust main effect of Stroop Type, $F(2, 116) = 137.1$, $MSE = 0.002$, $p < 0.0001$, $\mathrm{BF_{inclusion}} > 1000$. Post-hoc t tests attributed this effect to Control > Congruent > Incongruent (all comparisons significant, $p < 0.0001$, Bonferroni-corrected; $M = 0.990$, 0.928 and 0.864, respectively). The remaining effects were not significant ($p > 0.3$, $BFi \leq 0.14$). Similarly, for Stroop response time (correct responses only), the only significant effect was a robust main effect of Stroop Type, $F(2, 116) = 114.9$, $MSE = 9936$, $p < 0.0001$, $\mathrm{BF_{inclusion}} > 1000$ (Greenhouse-Geisser corrected for violation of non-sphericity). Post-hoc pairwise comparisons found no difference between Control and Congruent trials, but the other comparisons were significantly different ($p < 0.0001$). This supported the rank ordering Congruent = Control < Incongruent ($M = 1116$, 1139 and 1354 ms, respectively). Thus, the Stroop effect was present and pronounced: Incongruent Stroop responses were both less accurate and slower than both Congruent and Control responses. The latter two were close to one another in accuracy and response time.

Next, we asked if Stroop interference might covary with associative interference. The two measures of interference in the Stroop task are the difference in accuracy, and correct-response times, for Incongruent minus Congruent Stroop trials, normalized by the mean of Incongruent and Congruent trials. We correlated these Stroop interference measures with the measure of retroactive interference (AB–DE, normalized by mean accuracy of AB and DE pairs) and proactive interference (AC–FG, normalized) in the associative interference task, without regard to Condition. All four pair-wise correlations were non-significant (greatest $r(59) = -0.147$, $p = 0.258$; all BF< 0.3). Excluding participants with undefined Q_{BC} values, both Stroop measures (accuracy and response time) were also not significantly correlated with (log-odds-transformed) Q_{BC} (accuracy: $r(51) = 0.004,\ p = 0.98$, response time: $r(51) = 0.076,\ p = 0.59$). Broken down by Condition, no correlations were significant.

Finally, we examined a measure of proactive interference within the Stroop task. Response times are slowed in so-called "negative priming" trials, when the current congruent trial has the same response as the (incorrect) color-label (word) from the prior incongruent trial. The typical contrast is with trial sequences for which there is no relation between the prior and current trial ("no-relation" condition). Little and Hartley (2000) found that this measure was positively correlated with the standard Stroop-interference response-time measure, and both were unaffected by Age. We wondered if additional variance in the Stroop-PI measure might share variance with our associative-interference effects. To quantify Stroop-PI, we computed the difference in response times (correct trials only) for negative-priming trials minus congruent trials preceded by incongruent trials, but where the previous color and word were not the current color (no-relation). First, a mixed, repeated-measures ANOVA with design Condition[3]$\times$Relation[negative priming, no relation], revealed no main effect nor interaction with Condition (both $p > 0.7$, BF< 0.3), but as expected, a main effect of Relation, $F(1, 58) = 8.00$, $MSE = 29416$, $p < 0.01$, $\mathrm{BF}_{\mathrm{inclusion}} = 7.7$. This main effect was due to longer response times for negative priming trials ($M = 1217$ ms) than no relation trials ($M = 1124$ ms; difference=93 ms), indicating the presence of proactive interference in the Stroop task. Next, we asked if the size of the Stroop-PI effect (negative priming minus no relation response time, normalized by their mean) correlated with the measures of retroactive interference and proactive associative interference, without regard to Condition. The Stroop-PI measure correlated positively with associative-interference PI, $r(59) = 0.273,\ p < 0.05$ (although the Bayesian correlation was inconclusive, BF=1.45), indicating that the more PI in the Stroop task (lengthened response times), the weaker the PI in the AB/AC task (less of a reduction in AC compared to FG memory). For associative interference RI, the correlation was not significant, $r(59) = -0.013,\ p = 0.92$, BF$< 0.3$. Broken down by Condition, no individual correlation reached significance. The Stroop-PI measure was not significantly correlated with the log-odds-transformed Q_{BC}, $r(51) = 0.073$, $p = 0.58$, $BF = 0.19$. This was also the case when computed separately for each Condition (all $r < 0.1,\ p > 0.7$), suggesting no relationship between proactive interference in the Stroop task and the relationship between memory for AB and AC.