THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2017
VOL. 70, NO. 4, 832–857
http://dx.doi.org/10.1080/17470218.2016.1171886

Routledge
Taylor & Francis Group

Associative independence revisited: competition between conflicting associations can be resolved or even reversed in one trial

Rachel L. Burton, Isabel Lek and Jeremy B. Caplan

Psychology Department, University of Alberta, Edmonton, AB, Canada

ABSTRACT

In one type of association-memory paradigm, after studying pairs of the form AB, AC, participants must recall both B and C in response to A. Counterintuitively, yet often replicated, recall probabilities of B and C are typically uncorrelated ("associative independence"). This face-value independence is now understood to reflect a negative correlation due to AB and AC competing, approximately offset by a positive correlation produced by subject- and item-variability. The outcome might vary with stimulus material; for noun-pairs, and with a single study trial per pair, AB and AC have been found to be positively correlated. We replicated the positive correlation between AB and AC for noun-pairs, but this did not differ from the correlation expected for independent memory tests, suggesting that for noun pairs, AB and AC are independent on average. In Experiment 2, participants instructed to form separate images for AB and AC again produced an independence pattern, but participants instructed to combine AB and AC into an integrative image produced a facilitation pattern. Thus, the relationship between AB and AC varies, and can be influenced by study strategy. Association-memory models may need to accommodate a diverse range of AB–AC relationships, and studies that build on AB/AC learning may need to consider whether AB/AC start out with a competitive, facilitatory or independent relationship.

ARTICLE HISTORY
Received 1 April 2015
Accepted 3 February 2016
First Published Online 26 April 2016

KEYWORDS
Association memory; Verbal memory; Associative interference; Strategy Imagery

People frequently face challenges to memory from conflicting associations. For instance, Jennifer might first be married to Brad and then later to Justin. In such situations, one may need to effectively retrieve the later associate (Justin) and avoid erroneously retrieving the earlier associate (Brad), overcoming proactive interference. Conversely, one might need to retrieve the earlier associate (Brad) while ruling out the more recently learned associate (Justin), overcoming retroactive interference. A long-standing question is: what is the relationship between such conflicting association-memories? Do they compete with one another, due to the common item, or could they even facilitate one another, linked by their common item?

The AB/AC procedure operationalizes associative interference. Participants study two lists of pairs (classically, mixed materials such as trigrams paired with adjectives, but in our experiments, pairs of nouns). The first word in a pair, A, appears in list 1, as AB, and a second time, in list 2, with a different associate, as AC. One can then ask how memory for AB and memory for AC relate to one another, by measuring the correlation between accuracy of AB and AC, with three possible outcomes: (1) AB and AC could be negatively correlated, which would suggest that AB and AC compete in memory (either at study or at test, or both); (2) AB and AC could have a near-zero correlation (note that this is a null hypothesis, so this case is more challenging to confirm), which would

CONTACT Jeremy B. Caplan jcaplan@ualberta.ca

signify that AB and AC are remembered independently of one another; or (3) AB and AC could be positively correlated, which would signify a facilitatory relationship, with memory for AB somehow boosting the participant's ability to remember AC, or vice-versa.

Exactly how memory for AB and AC is tested is important. Decades ago, researchers settled on what was called the "modified modified free recall" (MMFR) procedure (Barnes & Underwood, 1959). First, the participant learns list 1, including the AB pairs (in the classic experiments, over multiple trials to a near-perfect recall-criterion) followed by list 2, including the AC pairs (e.g., Delprato, 1972; Martin, 1971). In each MMFR test, one of the cue items, A, is presented, and participants are asked to recall both associates, B and C, in any order. If available, participants can express their memory for both B and C. Because participants can recall zero, one or both responses, if a negative correlation were still present between recall of B and recall of C, this could be considered more clear-cut evidence for associative competition.

Competition was expected for several reasons. First, the assumption behind unlearning theory (Melton & Irwin, 1940) was that, in order to learn AC, one had to forget AB. Second, consider that in the MMFR test, each cue (A) item has two possible target items (B and C). Most current memory models assume that candidate items, often called the "response set", compete to be retrieved. In an MMFR test, the corresponding B and C items may be in competition with one another. Moreover, as soon as an item is recalled, its association may be re-encoded (e.g., if B is recalled, then AB is re-encoded). This "output encoding" could immediately increase the level of competition against the unrecalled pair (here, AC), making the second target (C) more difficult to recall.

To get a feel for the magnitude of negative correlation one might expect, we simulated, in MATLAB, a simple strength model that assumed all pairs were stored independently, and included response competition. The simulation included four cycles, paralleling our experimental design (see Methods), with four AB pairs and four DE (control) pairs in each list 1, and four AC and four FG (control) pairs in each list 2. Encoding strength was drawn independently for each individual pair, from a Gaussian distribution, $N(\mu,\sigma)$, where $\mu = 1$ and $\sigma = 1.0$ or .5. Probability of a control pair being retrieved during MMFR was based on a choice rule like that suggested by Luce (1959): Probability correct $= \frac{s_i}{s_i + s_{bg}}$, where s_i is the strength of the target pair, and $s_{bg} = \mu/2$ stands in for a baseline level of retrieval of other items, given the cue (D for DE, or F for FG). Thus, the greater s_i, the greater is the chance the response will be correct. The outcome (correct versus incorrect) was then drawn at (pseudo-)random for this probability. For AB and AC pairs, probability correct was $\frac{s_i}{s_i + s_j + s_{bg}}$, where s_i is, again, the strength of the target pair (the AB or AC pair, respectively), and s_j is the strength of the competing pair (the corresponding AC or AB pair, respectively). Figure 1 plots the cumulative distribution functions of the correlation (quantified with Yule's Q, explained below) values produced by the model when run for 10,000 iterations with $\sigma = 1.0$ (a) and $\sigma = .5$ (b). Note that the only difference between control and interference pairs in this minimal model is that the interference pairs face competition from one additional pair. This model does not include variability across study sets (discussed later), so the expectation of Q_{EG} is exactly zero. In both cases, the median Q_{BC} is negative; more encoding-variability seems to produce larger negative values of Q_{BC}.

Surprisingly, with the exception of noun-pairs, discussed below, numerous studies have reported stochastic independence, supporting the associative-independence hypothesis— that is, no evidence of competition between recall of B and recall of C in MMFR (e.g., Delprato, 1972; Greeno, James, & DaPolito, 1971; Martin, 1971). In mathematical terms, recalls of B and C (given A as the cue) were not significantly correlated.

The interpretation of null correlations as associative-independence has been challenged. One problem is that correlations (or conditional probabilities, a different, but mathematically equivalent, way of testing the independence hypothesis) we re computed on aggregate data, pooled across participants (the "pooled" approach). Hintzman (1972) commented that subject variability would be expected to produce a positive correlation even between what should be unrelated tests of memory, known as Simpson's Paradox. For example, some participants may perform well on *all* memory tests, while other participants may perform poorly on *all* memory tests. Because memory tests are always paired within subjects, this would produce a positive correlation between any two memory measures, simply reflecting that variability across participants. In this way, if there

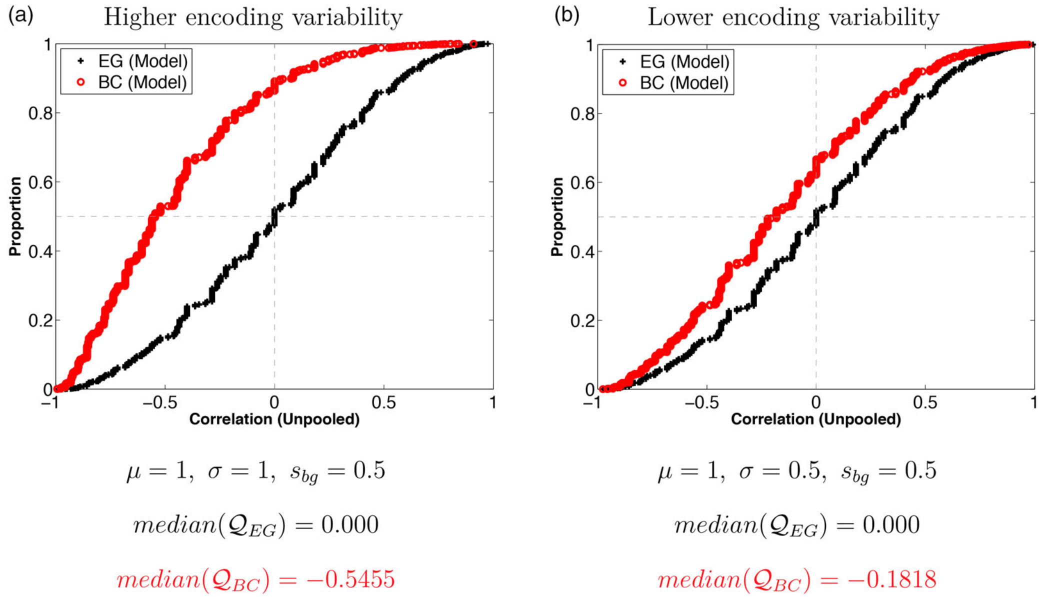

Figure 1. Strength model (see text), cumulative distribution functions of Yule's Q values; EG – Q_{EG} darker, black 'x' plot; BC – Q_{BC} (lighter, red 'o' plot). Each point represents one participant. Dashed grey lines denote the median (Proportion = .5) and numerical independence (Correlation, Q = 0). Two models are shown for illustration purposes, with σ = 1.0 (a) and σ = .5 (b).

were actually a negative correlation due to competition between AB and AC, it could be cancelled out by a positive correlation due to subject variability. Riefer and Batchelder (1988) found support for this kind of shifting of correlations toward positive values due to subject variability, in fits of a multinomial model to MMFR data. Thus, although many scholars may only remember the independence interpretation, the most up-to-date understanding of AB/AC learning is that with a subtle approach to data-analysis, AB and AC associations do compete in memory.

The AB/AC-learning results reported here deviate from competition, showing that a facilitatory relationship between AB and AC is possible (i.e., despite Simpson's Paradox). They also suggest that independence may not be such a poor characterization of AB/AC learning, at least for noun–noun pairs, and without specific strategy instructions. AB/AC learning is further understood by digging deeper into AB/AC learning, examining individual differences themselves. We address Simpson's Paradox—the confound due to artefactual sources of positive correlation—in two ways. First, control pairs are always included in both the AB and AC study sets ("DE" and "FG" pairs, respectively). The correlation between recall of E (given D) and recall of G (given F), pairs that should not interfere with one another, provides an estimate of the positive correlation due to subject variability (Caplan, 2005), to which the BC correlation can be compared. Second, the correlations are also calculated with subject variability removed, which we refer to as the "unpooled" approach (Caplan, Rehani, & Andrews, 2014; Rehani & Caplan, 2011). We do this by computing each correlation within participants before carrying out statistical tests across participants. By examining the outcomes of both approaches, we can effectively address Hintzman's critique of prior AB/AC learning studies and evaluate whether, and to what extent, the positive-correlation bias affects the conclusions one can draw from the observed pattern of correlations.

Further support for the idea that associations sharing an item compete in memory comes from a related associative-interference procedure. Caplan et al. (2014) had participants study double-function lists, which contained pairs of the form AB/BC. The chief difference from AB/AC learning is that the shared item (here, B) changes in position between the two pairs. This procedure also demanded that participants resolve interference within a single study set. Possibly for this reason, associative competition was more pronounced than is typically reported for AB/AC learning, and the correlation between memory for AB and BC was not only significantly more negative than a control for independent memory tests (which itself was positive), but was also significantly less than zero. In other words, the elusive face-value negative correlation, which had been sought after in AB/AC learning studies, was found using a double-function procedure with a single list. This suggests that when the challenge from associative interference is pronounced enough, competition can be plainly observed.

What is unknown is whether facilitation between pairs sharing a common item (like AB and AC) is possible. We found one report of a positive correlation between AB and AC (Tulving & Watkins, 1974), using noun-pairs, which might reflect a facilitatory relationship between AB and AC. Why did this study obtain a positive correlation where most of the classic studies found independence? Nearly all the classic studies did not use pairs of nouns. Rather, they typically used pairs for which the stimulus (A) and response (B, C) items were drawn from different stimulus types, such as consonant-vowel-consonant trigrams as stimuli and two-syllable adjectives as responses (e.g., Delprato, 1972; Martin, 1971; Wichawut & Martin, 1971). It could be that, whereas face-value independence (actually reflecting an underlying competitive relationship) between AB and AC is the norm for heterogeneous pairs, facilitation might be characteristic for homogeneous noun–noun pairs. Alternatively, pairs of nouns might be easier to integrate in a meaningful way, and may have greater ecological validity, than other materials that have been used in AB/AC research (including consonant-vowel-consonant trigrams and numerals), so there may be a good reason to expect different results for noun–noun pairs. However, because the face-value competition result reported by Caplan et al. (2014) was found with noun–noun pairs, this explanation alone seems unlikely. Moreover, Tulving and Watkins (1974) appropriately expressed caution in interpreting their positive-correlation result, for the reasons we just discussed— artefactual sources of positive correlation are hard to rule out. When we also obtained a positive correlation in a pilot experiment with verbal AB/AC learning, we wondered whether facilitation might in fact be the rule, at least for AB/AC learning with noun-pairs, or it in fact reflects independence or even competition when compared to an appropriate control. Thus, our main alternative hypothesis was that noun-pairs have even more pronounced sources of artefactually positive correlations (subject and item variability) than other materials, and the positive correlation between AB and AC in noun-pairs might still be significantly more negative than a control for independent memory tests.

We further speculated that individual differences might be obscuring the picture one gets from aggregate data. Memory for AB/AC pairs might be independent, competitive, or facilitative, depending on the participant. In other words, aggregating participants together may also have the consequence of positive-correlation participants cancelling out negative-correlation participants. The objectives of Experiment 1 were to test whether the correlation between word pairs of the form AB/AC is positive compared to a control for independent memory tests, and to acquire enough data per participant to examine individual differences.

In addition, previous AB/AC studies, such as Wichawut and Martin (1971), had participants study the AB pairs multiple times until the pairs were learned to a criterion. At that time, unlearning theory, which assumes that previously well learned AB pairs are unlearned following study of AC pairs, was a theoretical focus in the literature and this design allowed that theory to be studied (Martin, 1971). Unlearning theory has since been seriously challenged; for example, Verde's (2004) findings, testing AB/AC pairs with associative recognition, suggested that AB and AC can co-exist in memory. Because of the interest in unlearning theory, nearly all the apparent findings of independence in AB/AC learning have been obtained with procedures wherein participants learned list 1 (which includes the AB pairs) to a perfect, or near-perfect, initial-recall criterion (Greeno et al., 1971; Martin, 1971; Wichawut & Martin, 1971). As Mensink and Raaijmakers (1988) suggested, such overlearning might also be blurring the interpretation of the correlations. Overlearning may reduce variability in memory-strength across the AB set; if there is less meaningful variability in AB strengths, then correlations between AB and AC accuracies would be more driven by noise, which would be expected to nudge correlations toward zero. We therefore asked whether the classic finding of independence would be replicated if participants had only one study trial per AB (and AC) pair (as did Tulving & Watkins, 1974). This overlearning hypothesis was tested with a follow-up data analysis, confined only to AB/AC sets for which the AB pair was initially recalled correctly. The prediction was that this resampled data set, screened for accurate cued recall of AB pairs, would produce a correlation closer to zero.

In sum, the AB/AC design presented here is adapted in two major ways: first, list 1, which included the AB pairs as well as control pairs with no repeated items, DE, was studied only once, to simplify the procedure and avoid potential over-learning problems. List 2, which included the AC set as well as control pairs with no repeated items, FG, was also studied only once. Second, the one-trial procedure freed up experimental session-time, which we used to acquire several cycles of AB/AC learning (each with a completely new randomly selected set of nouns) from each participant. This provided enough data to compute correlations for each participant. The chief objective of Experiment 2 was to test whether an explicit manipulation of study strategy via instruction could influence the relationship (correlation) between AB and AC. In two experiments we tested four hypotheses:

Hypothesis 1. The correlation between recall of AB and AC is positive for noun–noun pairs. This would replicate Tulving and Watkins (1974) but would not be conclusive about whether this reflects facilitation or independence (or even competition). A control for independent memory tests specifies the interpretation of the aggregate behavioural pattern.

Hypothesis 2. A near-zero correlation can be a result of positive correlations in some participants cancelling negative correlations in others. The distributions of correlation values speak to this possibility.

Hypothesis 3. Variability in strategy for handling the repeated items (A) can produce different values of the BC correlation. We tested this hypothesis, first by examining subjective strategy reports in Experiment 1, and second by manipulating strategy instructions to participants, in an attempt to influence the correlation, in Experiment 2.

Hypothesis 4. The typical procedure of requiring a high degree of initial AB learning can remove important sources of variability (i.e., degree of learning of individual pairs) that would otherwise have correlated with retrieval of the AC associations. If a major source of common variability is removed, then the remaining variability may be primarily noise, producing a zero or near-zero correlation between recall of B and recall of C for trivial reasons. This was investigated by conditionalizing the correlation between AB and AC on accurate initial cued-recall.

Experiment 1

Noun-pairs were studied once each, then tested initially with cued recall, to reinforce to the participants that they were supposed to learn each list in its own right. This was followed by MMFR tests. Four full cycles of the design (list 1 with cued recall, list 2 with cued recall, and MMFR) were performed by each participant, to produce reasonably reliable estimates of the correlation between AB and AC. No instruction about strategy was given to participants in this experiment. Rather, participants described their strategy at the end of the main experiment.

Methods

Participants

A total of 101 undergraduate students enrolled in a first-year introductory psychology course at the University of Alberta participated for course credit. Of those, 12 were eliminated from further analysis because their MMFR accuracy was at ceiling (>95%) or floor (<5% correct) or had perfect or zero accuracy in any one MMFR condition. This left 89 participants.

Materials

Lists of eight noun-pairs each were constructed from the nouns from the Toronto Word Pool (TNPnorms) (Friendly, Franklin, Hoffman, & Rubin, 1982). Nouns with imagery rating ≥6 were retained, and of those, the 180 most concrete nouns were used. Thus, all of the words used in the experiment had an imagery rating ≥6 and a concreteness rating ≥4. Noun-pairs were presented in the centre of the computer screen with one word slightly to the left (A, D and F words) and the other slightly to the right (B, C, E and G words). Assignment of words to pairs and pair types was randomized.

Testing-order of the pairs during the cued recall and MMFR tests was also random, with each pair being tested exactly once in cued recall, and each A, D and F item being used as a probe exactly once during MMFR. During testing, participants typed their responses in blank lines presented below the cue word, followed by the ENTER key.

The 20-s distractor task consisted of adding three integers between two and eight (inclusive), with typed 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 mathematics problems during each distractor block.

The experiment consisted of four study sets, where each set included two lists of noun-pairs presented on

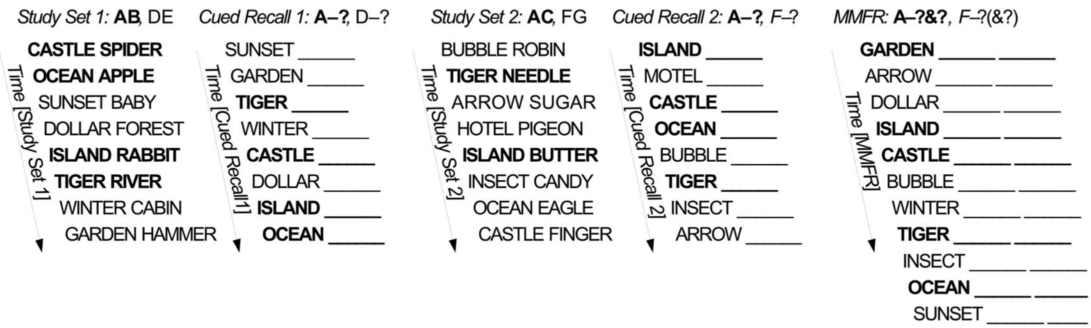

Figure 2. Experimental design for Experiments 1 and 2. Participants studied two lists of noun-pairs. In each list, half the pairs were interference pairs (**AB** or **AC**, denoted in boldface in the figure for illustration only) and half of the pairs are no-interference pairs (DE or FG). Each list of pairs was tested with cued recall, and the set concluded with an MMFR test. Underlines denote response fields and are depicted to the right of the cue word for compactness only; in the experiment, response fields were displayed underneath the cue word. Participants completed a distractor task (not shown) that separated all study and test phases.

a computer screen: a list 1, composed of AB (interference) and DE (control) pairs, and a list 2, composed of AC (interference) and FG (control) pairs, reusing the A items from list 1 (Figure 2).

Procedure

Participants were tested in individual testing rooms. Each of the eight noun-pairs in a list was presented for 3000 ms with an additional 150-ms interval between pairs. Participants completed a 20-s block of the distractor task following study of a list 1 (AB/DE) and were then tested on that 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. 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.

Following a second 20-s block of the distractor task, participants studied the list 2 (AC/FG), followed by another 20-s distractor task and cued recall of list 2. In this cued-recall block, participants were instructed to respond only with the associates from the most recently studied list. This was followed by a 20-s distractor task and the MMFR test, which concluded each cycle. In MMFR, participants were presented with the left-hand item (A, D and F) as a cue, and were asked to type the one or two words that were paired with the word on the screen in any order. Initially, a single line below the cue word was presented on the screen, as in the cued-recall procedure. When a first word was entered, a blank line replaced the first word. Participants were told to type "pass" if the cue word had only one associate. Participants were told that, if the cue word had two associates, they should type the other word on the second line. If a participant could not recall any words after typing "pass" on the first line, the next cue word appeared.

At the end of the session, participants answered a free-form question: "When you were learning the pairs and you noticed that a word had been used more than once, what strategies did you use to handle this potentially confusing situation?" Then they completed a multiple-choice strategy questionnaire that asked how often they repeated pairs, visualized or imagined the words, created a sentence that included both words or looked for similarities between the two words in a pair. Participants also answered questions about how they handled the interference generated by the overlapping AB/AC pairs. They were asked about how often they tried to learn the two pairs that contained the same word together as unit, or tried to link all three of these words together. They were also asked how often they tried to learn the overlapping pairs as two independent units. Participants circled "Never", "Sometimes", "Mostly" or "Always".

Correlation analyses

The correlations for MMFR responses were calculated using Yule's Q, a measure of correlation appropriate for dichotomous data (for reviews, see Kahana, 2002; Warrens, 2008). Yule's Q is calculated by creating a 2 $\times$ 2 contingency table. The 'a' quadrant contains the number of times both items were recalled, the 'b' quadrant contains the number of time the first item but not the second was recalled, the 'c' quadrant contains the number of times the second item but not the first was recalled, and the 'd' quadrant contains the number of times neither item was recalled. Yule's $Q = (ad - bc) / (ad + bc)$. We report two different Yule's Q calculations. First, we computed Q in the conventional manner, to enable a direct comparison with most of the published work on AB/AC learning. In the pooled approach, the data from all participants were pooled into a single contingency table, and a single Yule's Q was calculated. Significance and pairwise comparisons were done by transforming Q into log-odds ratios, also known as "logits": $\log((Q - 1)/(1 - Q))$ and calculating standard errors in log-odds units (Bishop, Fienberg, & Holland, 1975; Hayman & Tulving, 1989). Note that with the pooled method, Q values potentially include sources of positive correlation due to subject variability. Second, in the "unpooled" approach, Q values were calculated for each participant first, and these values were analysed after a log-odds transform. By computing Q values for individual participants, subject variability cannot inflate the correlations. The trade-off is that fewer data points go into each Q value. Because there are often empty cells in the contingency table, we apply a correction, adding one-half of an observation (0.5) to each contingency table quadrant. Because each approach to computing Q has its advantages and disadvantages, we report both methods, and check whether or not the patterns of findings depend on the approach to calculating Q (pooled versus unpooled). We have not found precedents for the unpooled approach applied to AB/AC learning but it has been used in other association-memory paradigms (Caplan et al., 2014; Kahana, 2002; Rehani & Caplan, 2011; Rizzuto & Kahana, 2001). We denote the correlation between recall of B and C, given A as a cue, by Q_{BC}.

As a second way to address the correlation due to subject variability, we also obtained a separate measure of this correlation, by computing a "Control" correlation between recall of E and G items, denoted Q_{EG}. Because DE and FG pairs have no relationship with one another, for the pooled correlation analyses and cumulative distribution function analyses(unpooled approach), we chose a random shuffle, assigning each DE pair to a single FG pair within each study set. In this way, the same number of counts went into the calculations of Q_{EG} as of Q_{BC}. For the unpooled correlation analyses, a different approach was needed to obtain a more reliable estimate of the correlation induced by variability across study sets. Making maximum use of the data, we used all combinations of DE with FG pairs within a given two-list study set. Computed in this manner, the contingency tables for Q_{EG} would have a much larger number of counts (16 combinations per study set × 4 study sets = 64) than the contingency tables for Q_{BC} (4 AB/AC per study set × 4 study sets = 16). The Q_{EG} contingency table were normalized to have the same total number of counts as the Q_{BC} contingency table (4 AB/AC per study set × 4 study sets = 16 counts) by multiplying the contingency by this ratio (16/64).

Accuracy values and Yule's Q values were always log-odds transformed prior to parametric statistical tests (t tests and ANOVAs) to better satisfy the condition of normality.

Results and discussion

The results start with cued-recall and MMFR accuracy, to characterize any proactive or retroactive interference on average. We then report correlation analyses, the main measure of interest. The conventional pooled correlations are followed by the novel, unpooled correlations, testing whether associative independence characterizes the relationship between AB and AC in memory (Hypothesis 1). We then examine the potential effects of subject variability (Hypothesis 2) and a high degree of learning of the AB pairs (Hypothesis 4). Finally, the self-report strategy responses are examined to test whether individual difference in strategy adoption could explain individual variability in the correlation (Hypothesis 3).

Cued recall accuracy

At the cued-recall stage of the procedure (Figure 3a), the interfering AC pairs have not been presented to the participants; thus, the AB and DE pairs should not be differentiated, which was confirmed, $t(88) = -1.39$, n.s., paired-samples, two-tailed. Proactive interference was evident in initial cued recall; participants recalled more FG than AC pairs, $t(88) = -2.56$, $p < .05$.

MMFR accuracy

In MMFR (Figure 3b), again, AB and DE pairs did not differ in accuracy, $t(88) = 0.31$, n.s.; thus, no evidence of retroactive interference on average was found. As with cued recall, probably inherited from output encoding during the cued-recall phase, proactive

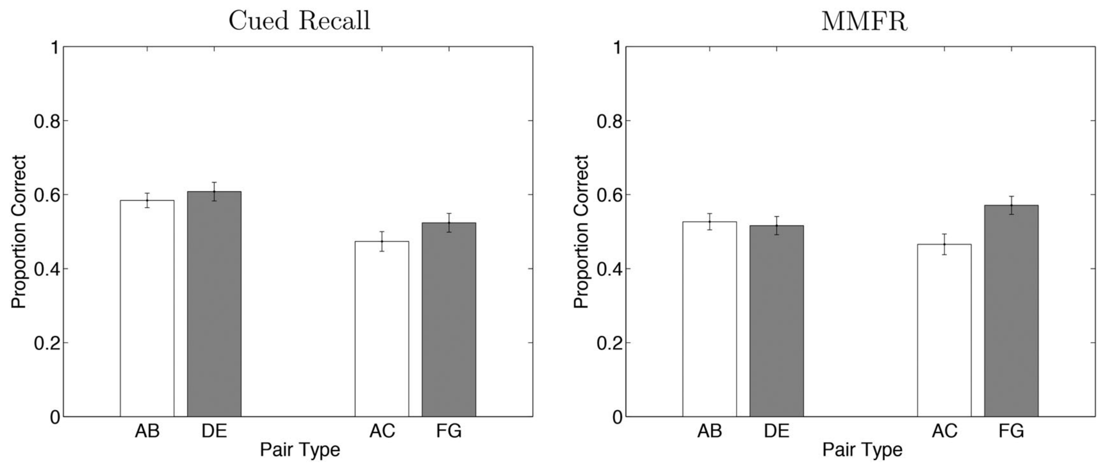

Figure 3. Experiment 1, cued recall (a) and MMFR (b), proportion correct as a function of pair type. Error bars plot 95% confidence intervals corrected for subject variability *(Loftus and Masson, 1994)*.

interference was present: fewer AC than FG pairs were recalled on average, $t(88) = -4.45$, $p < .0001$.

MMFR correlations

We now turn to the measure that can be used to test our central hypotheses: the correlation between accuracy in MMFR of AB and AC pairs, starting with the pooled approach to computing correlations (Figure 4a), because prior studies have computed correlations in this way.

Pooled correlations. There was a significantly positive correlation between the overlapping AB and AC pairs (Figure 4a), a face-value facilitation result, $Q_{BC} = .29$; $Z_{log-OR} = 5.60$, $p < .001$. This is similar to Tulving and Watkins (1974), who obtained Q_{BC} values of .16 and .39 using noun-pairs, in their conditions that were most comparable to ours. However, as explained in the introduction, due to variability across subjects and study/test cycles, it is not clear whether this positive correlation is truly due to the relationship between memory for the AB and AC associations. This is because if some participants perform better

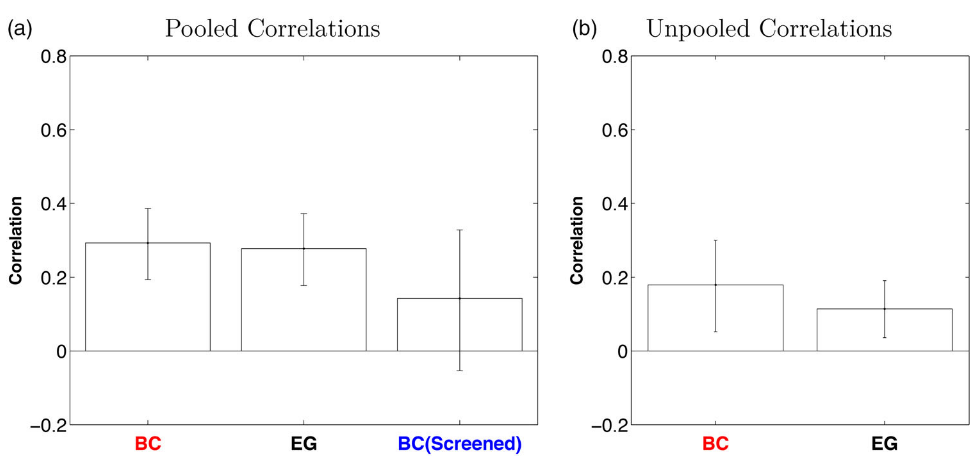

Figure 4. Experiment 1, MMFR correlations. (a**)** Pooled approach (see text). **BC**, Q_{BC}, the correlation of interest, measuring the relationship between recall of AB and AC. **EG**, Q_{EG}, the control correlation, between DE and EG pairs, computed via a bootstrap, estimating the correlation expected between independent MMFR measures paired within study sets. Error bars plot 95% confidence intervals computed via the log-odds transform. **BC(Screened)**, Q_{BC} computed only using AB/AC pairs for which the AB pair was initially correctly recalled in cued recall, approximating the effect of overlearning of the AB list as done in prior studies. (b) Unpooled approach (see text).

overall than other participants, all measures would covary together, resulting in a moderately positive correlation. The same reasoning applies to variability across study/test cycles (e.g., due to fatigue or learning-to-learn effects). Rather than testing correlations against zero, Q_{BC} must be compared to Q_{EG}, the control correlation for what should be independent memory tests, between arbitrarily paired DE and FG pairs. The DE and FG pairs do not share any words, so there is no reason to expect that the recall of one DE pair would be related to the recall of an FG pair. Consistent with Hintzman (1972, 1980) and Riefer and Batchelder (1988), Q_{EG} was significantly greater than zero, Q_{EG} = .27; Z_{log-OR} = 5.27, p < .001. Thus, subject /cycle variability does indeed result in a substantial, positive correlation even for independent memory tests. Q_{BC} and Q_{EG} had very similar values, and were not significantly different, z = 0.22, p = .83. Thus, when viewed alongside an appropriate control for independence, this face-value positive correlation seems, instead, to be consistent with independence.

Effect of high degree of initial learning of AB pairs. Hypothesis 4 was that the typical procedure, initially requiring a high-accuracy criterion for the AB pairs, may have reduced the amount of meaningful variance relative to noise variance in memory for the AB pairs, which should have shifted Q_{BC} values toward zero. Consistent with this, Tulving and Watkins (1974), who also obtained a positive Q_{BC}, also did not have an initial learning-to-criterion phase. To understand those effects, we re-analysed our data in a way that would simulate the effect of having had only a high degree of initial learning of AB pairs. To this end, Q_{BC}*(Screened)* was computed using only the MMFR data for AB/AC pairs for which the AB pair was initially correctly recalled in initial *cued recall* (Figure 4a, rightmost bar). Thus, the Q_{BC}*(Screened)* (still during *MMFR*) was recalculated including only those pairs A_iB_i and A_iC_i for which the participant had responded correctly with B_i when initially given A_i as a cued-recall probe. The AB-screened correlation was no longer significantly different from zero, Q_{BC}*(Screened)* = .14, z = 1.42, p > .1, in line with Hypothesis 4. Thus, when the MMFR correlation was confined to well learned AB associations, AB and AC appeared more consistent with face-value independence (failure to reject a correlation of zero). This may be one reason the face-value correlations in prior AB/AC learning studies have been near-zero. However, it should also be noted that fewer data points went into the computation of the screened correlations; this may also have influenced their values.

Unpooled correlations. When correlations were computed for each participant (Figure 4b), the mean correlation values dropped as expected, indicating that both the control and interference pooled correlations were inflated by subject variability. Although Q_{BC} was greater than zero (p < .01), so was Q_{EG} (p < .01), because of study-test-cycle effects which are not controlled for here. With four AB and four AC pairs in a single study set, there are not enough pairs to calculate an AB/AC correlation for each individual study set. This is why Q_{EG} is still necessary as a control for (otherwise) independent memory tests. A paired-samples t test on the log-odds transformed correlations found no significant difference between Q_{BC} and Q_{EG}, $t(88)$ = 0.92, p = .36. This is consistent, once again, with independence of AB and AC, at least on average.

In summary, the pooled and unpooled approaches produced convergent results: EG and BC correlations in MMFR were inflated by subject variability. When the BC correlation is compared to an appropriate control, and when the effects of subject variability are removed, Q_{BC} is not more negative than what one expects for independent memory tests (Q_{EG}), as is understood for prior AB/AC learning studies using other materials (Hintzman, 1972; Riefer & Batchelder, 1988) and found for noun-pairs using a different kind of associative interference (Caplan et al., 2014). Q_{BC} is in fact slightly more positive than the control for independence, but not significantly so, compatible with associative independence.

Individual variability in the sign and magnitude of the BC correlation

Hypothesis 2 was that individual differences in the sign (not to mention magnitude) of the BC correlation may have effectively cancelled one another in prior studies to produce near-zero correlation. This variability can be visualized by plotting the cumulative distribution function (CDF) of the BC correlations for individual participants (red circles in Figure 5). There appears to be a broad spread in individual-participant correlation values, with correlations ranging from very negative to very positive. However, each Q value was based on a small number of values (16 in this experiment), so this variability could be due to chance. The black crosses in Figure 5 plot the same CDF for the control for independence, Q_{EG}. Importantly, for

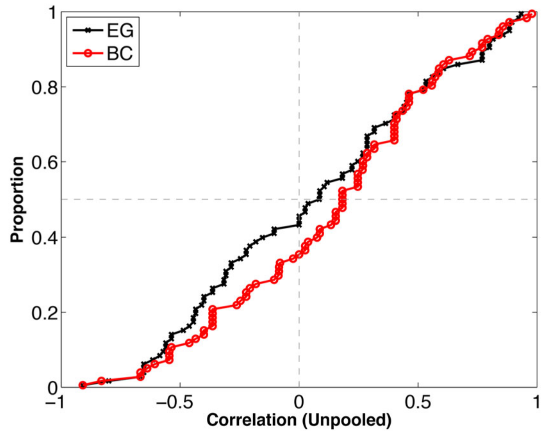

Figure 5. Cumulative distribution functions of Yule's Q values; EG – Q_{EG} (darker, black 'x' plot); BC –Q_{BC} (lighter, red 'o' plot). Each point represents one participant. Dashed grey lines denote the median (Proportion = .5) and numerical independence (Correlation, Q = 0).

the cumulative distribution analyses, we computed *Q* values by randomly assigning each DE to a different FG pair (from the same list-1/list-2 cycle), without replacement for each participant. This ensured that the number of data points entering into the Yule's *Q* calculation was equal for the EG and the BC correlations, avoiding a potential bias in variability. Indeed, the distributions nearly overlapped, and were not significantly different by a Kolmogorov-Smirnoff test, $d = -.13$, $p = .39$. Thus, for this data set, we failed to find reliable support for Hypothesis 2. Nevertheless, the distribution of Q_{BC} appears slightly right-shifted relative to the distribution of Q_{EG}. Although non-significant, this raised the possibility that, at least in some experimental conditions, variability in strategy might modulate Q_{BC}.

Self-reported strategy use

A pilot experiment (not reported here) had also produced an apparent facilitation effect ($Q_{BC} > 0$). This led us to hypothesize that participants might actively attempt to link the AB and AC items into a single representation, or the "Integration" strategy, which could cause the facilitation pattern, a more positive Q_{BC}. Other participants who attempt to keep the two associations distinct, the "Separation" strategy, may produce a competitive relationship between AB and AC, and thus a more negative Q_{BC}. To test this, we asked participants to self-report their strategy use with an open-ended strategy questionnaire followed by a forced-choice questionnaire including questions about strategy use. On the free-form questionnaires participants were asked, "When you were learning the pairs and you noticed that a word had been used more than once, what strategies did you use to handle this potentially confusing situation?" Two of the co-authors (IL and RB) independently categorized these free-form answers into three categories: Integration-like, Separation-like and Other. Inter-rater reliability was Kappa = .84, 95% CI (.75, .94). Among the responses to this free-form question, 44 participants reported strategies that were judged by both raters to resemble the Integration strategy (examples of participants' free-form question responses that resembled an Integration strategy are presented in Appendix A). A set of 18 participants reported strategies that were judged by both raters to resemble the Separation strategy.

These self-reports are proof-of-principle that at least some participants, some of the time opted to use an Integration-like strategy and others applied a Separation-like strategy. The next question was whether these self-reported strategy measures relate to the effects of associative interference on performance. We compared a Subjective-Integration group with a Subjective-Separation group, combining the free-form questionnaire with responses to the forced-choice questionnaire (Table 1). The Subjective-Integration group comprised 25 participants who reported an Integration-like strategy in the free-form questionnaire, and in the forced-choice questionnaire confirmed this (reporting "Mostly" or "Always" to the Integration-strategy question and then not reporting "Mostly" or "Always" to the Separation-like strategy). The Subjective-Separation group, likewise, comprised 17 participants who reported a Separation-like strategy in the free-form questionnaire, confirmed in their forced-choice questionnaire responses ("Mostly" or "Always" responses to the Separation-strategy question and "Never" or "Sometimes" to the Integration-strategy question).

First, we asked whether the pattern of mean accuracy in MMFR differed between groups. If participants

Table 1. Tallies of response rates to the forced-choice questions at the end of the session in Experiment 1. The "Integration Question" asked participants if they used a strategy resembling our Integration strategy, and likewise, the "Separation Question" asked about the Separation strategy

	Never	Sometimes	Mostly	Always
Separation question	24	34	19	12
Integration question	24	28	22	15

in the Subjective-Integration group are retrieving AB during study of AC, one would expect them to have increased probability of recalling AB, because of this additional study. One expects a cost for AC, since retrieving AB may displace study-time from study of AC. Two repeated-measures ANOVAs tested these predictions (Figure 6a,b). Analysing list 1, with Subjective Group as a between-subjects factor and Control/Interference (DE versus AB pairs) as a within-subjects factor (and log-odds transformed MMFR accuracy as the measure), the main effect of Subjective Strategy was not significant, $F(1,34) = 1.14$, showing no evidence that these self-report-based groups differed in overall performance. We therefore have no reason to suspect that the groups self-selected based on overall memory ability. The main effect of Control/Interference was non-significant, $F(1,34) = 0.27$, showing no evidence of an overall presence of retroactive interference or facilitation. The interaction, however, was significant, $F(1,34) = 5.00$, $MSE = 0.515$, $p < .05$. A post hoc *t* test found a significant difference between groups on the accuracy of AB–DE (the difference), $t(34) = 2.12$, $p < .05$, with the effect in the expected direction (more facilitation of AB in the Subjective-Integration group compared to more inhibition of AB in the Subjective-Separation group). Analysing list 2 (AC and FG) in the same way, again, the main effect of Subjective Group was not significant, $F(1,34) = 0.615$. The main effect of Control/Interference was significant, $F(1,34) = 8.61$, $MSE = 5.58$, $p < .01$. This main effect was qualified by a significant interaction, $F(1,34) = 6.37$, $MSE = 0.648$, $p < .05$. A post hoc comparison found that the groups differed significantly, $t(34) = -2.93$, $p < .01$, again with the effect in the expected direction (more inhibition of AC in the Subjective-Integration group than the Subjective-Separation group). These findings suggest that the subjectively reported strategy responses reflect a difference in strategy (not general performance level), and that these strategy reports resemble the two strategies we had in mind.

Turning to the correlations, the hypothesis was that the Integration strategy should produce a facilitatory relationship between AB and AC, whereas the Separation strategy should produce a competitive relationship, or at least independence, between AB and AC. Descriptively, the results fit this hypothesis; the median Q_{BC} was .04 for the Subjective-Integration strategy and −.08 for the Subjective-Separation group. The cumulative distribution function was more positive at nearly all percentiles for the Subjective-Integration group compared to the Subjective-Separation group (Figure 6c). However, this difference did not reach significance with a Kolmogorov-Smirnoff test, $d = -.28$, $p = .59$, thus inconclusive. This is unsurprising, given the low power in these sub-samples.

In sum, these results suggest that, without receiving specific strategy instructions, some participants adopted an Integration-like strategy and some adopted a Separation-like strategy. However, the central measure of interest failed to show a statistically robust difference in the facilitatory or competitive relationship between AB and AC. Self-reported strategy use measures have three limitations. First, participants are not necessarily accurate judges of the strategy they used, particularly retrospectively. Second, the differences between groups may be due to self-selection effects rather than the effects of the strategy per se. Third, participants may have been inconsistent in application of their self-reported strategies, and may have used multiple different strategies. This kind of mix of strategies could have produced both positive and negative influences on the correlation measures, particularly by acting on study-set variability. Finally, the sample sizes for the Subjective Strategy groups were relatively low. The predicted effect of strategy on Q_{BC} might be more robust if we could use most of our participants, as would be possible with a manipulation of strategy rather than rating strategy post hoc. Therefore, Experiment 2 was a manipulation of instructed strategy to test whether these effects could reach statistical significance if strategy adoption were not left to the participants' spontaneous choice.

Experiment 2

Integrative strategies have long been thought to hold potential for participants in overcoming associative interference. In some integrative strategies, participants combine items into a single representation, such as an image (the variant of integrative mediation we pursue here). In others, participants identify a concept that is common to both AB and AC. Integrative strategies have been found to reduce associative interference (both retroactive and proactive) in AB/AC learning of adjective pairs when the B and C response terms were related (Kanungo, 1967; Postman, 1964), in AB/AC learning of nonsense syllables paired with words where mediators were available via free association during study (Martin & Dean, 1964), in AB/AC learning involving sentences (Smith, Adams, & Schorr, 1978). Integrative strategies have also reduced associative interference in retrieval-induced forgetting (Anderson & Bell, 2001; Anderson & McCulloch, 1999; Smith & Hunt, 2000) and when elaborative strategies, like the peg list method and the method of loci, are used (Bower & Reitman, 1972). These findings suggest that integrative strategies could, under certain conditions, overcome associative interference based on average-accuracy measures, but do not speak to whether integrative strategies might affect the relationship (i.e., correlation) between the competing AB and AC pairs.

The chief aim of Experiment 2 was to directly test whether the two different kinds of strategy, Separation and Integration, could produce different degrees of correlation between recall of AB and recall of AC.

Methods

Participants

There were 193 participants, from the same population as in Experiment 1. Three failed to complete the experiment due to equipment failure. Of those remaining, 25 were eliminated from further analysis

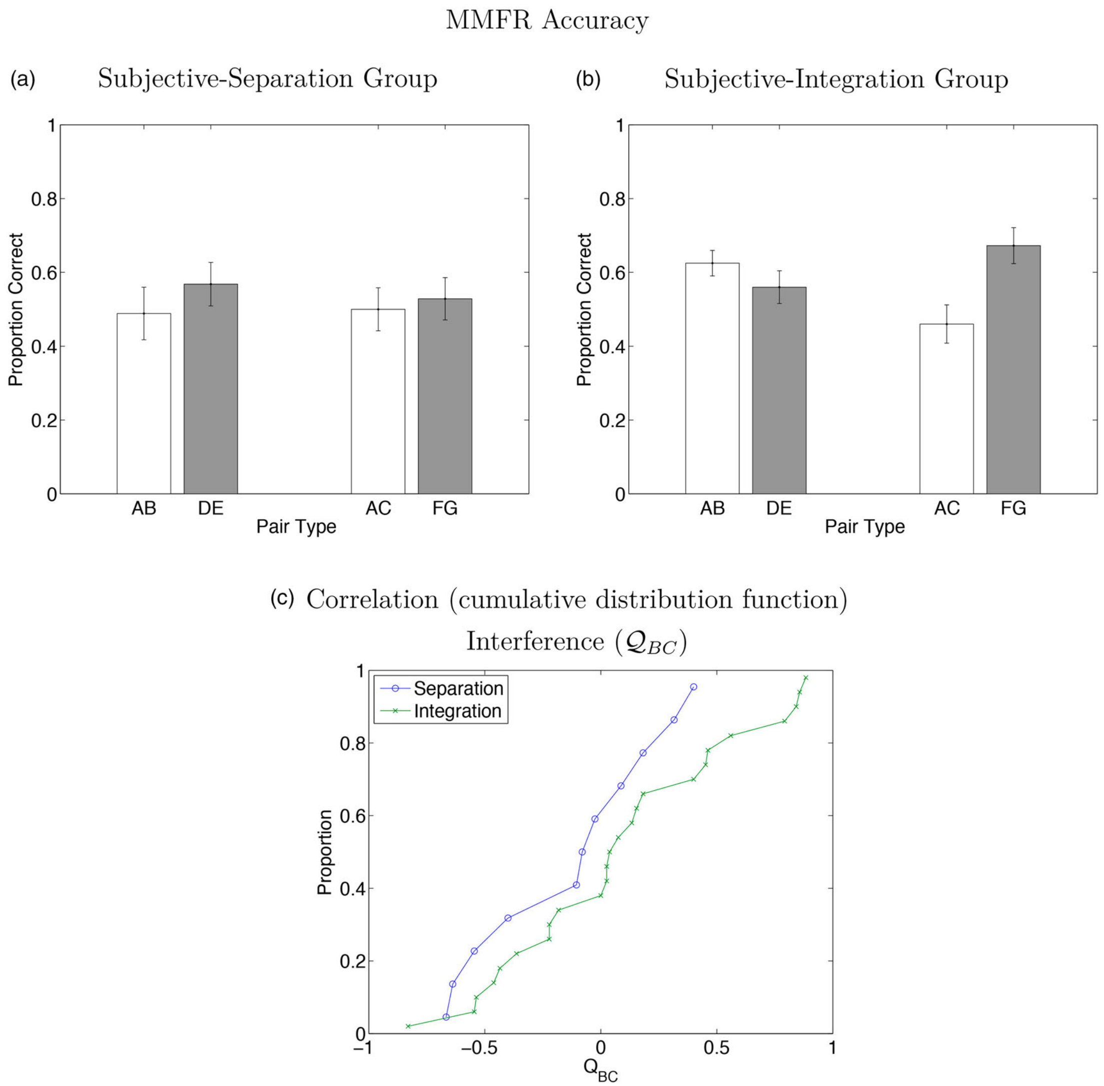

Figure 6. Experiment 1, MMFR, proportion correct, for participants classified as "Subjective-Separation" (a) and "Subjective-Integration" (b) (see text for details). Error bars plot 95% confidence intervals corrected for subject variability (Loftus and Masson, 1994). Cumulative distribution functions by group for Q_{BC} (c).

because their MMFR accuracy was at ceiling (> 95%) or floor (<5%) in at least one MMFR condition. This left 81 in the Integrative group and 87 in the Separation group.

Materials

The same word pool and distractor task from Experiment 1 were used in Experiment 2. The same method of pairing and presenting the words from the pool was also used.

Procedure

Two groups were asked to learn pairs by creating imagery relating paired items, using one of the two strategies to learn AB/AC pairs. Participants in the Integration group were asked to "create an image that incorporates all three words from both pairs" when studying the overlapping pairs, whereas participants in the Separation group were asked to "create a separate image for each pair of words".

The study and test procedures were identical to those used in Experiment 1, except that following each MMFR test set, participants were shown each pair from both sets in random order. They were asked whether they had incorporated any other words from the experiment into the image they had created of the pair on the screen. This was immediately followed by a second question asking participants to rate the quality of the image they created from 1 to 7. Each pair was presented once and for each pair participants completed the two ratings one after the other.[1] Analyses of the two post hoc rating tasks provided some additional information about compliance and implementation of the strategies, and are presented in Appendix B.

Results and discussion

Accuracy in cued recall

In a 2 × 2 repeated-measures ANOVA for list 1, with design Group [Separation/Integration] × Pair Type [AB/DE], the main effects and interaction were all non-significant, all $Fs < 1$, $ps > .4$ (Figure 7a,b). These results raise no concern about sampling bias, which should be the case due to the randomization of materials. It also suggests one strategy was not more effective than the other prior to the introduction of the overlapping (AC) pairs, at least given our sensitivity.

In a 2 × 2 ANOVA for list 2 [AC/FG], there was no main effect of Pair Type, $F(1,166) = 1.71$, $p > .1$, but the main effect of Group was significant, $F(1,166) = 10.80$, $MSE = 2.06$, $p < .01$, with the Separation group outperforming the Integration group. The Pair Type × Group interaction was not significant, $F(1,166) = 0.091$, $p > .5$. Thus, in cued recall, there was no reliable evidence of proactive interference or facilitation on average, and the Integration group performed worse on the second list.

Accuracy in MMFR

As in cued recall, first a 2 × 2 ANOVA on Pair Type [AB/DE] × Group [Separation/Integration] was carried out (Figure 7). The main effect of Group was not significant, $F(1,166) = 0.53$, $p > .4$. The main effect of Pair Type was significant, $F(1,166) = 9.62$, $MSE = 0.327$, $p < .01$, with interference pairs (AB) recalled worse than control pairs (DE). The Pair Type × Group interaction was not significant, $F(1,166) = 2.44$, $p > .1$. In a 2 × 2 ANOVA for list two, on Pair Type [AC/ FG] × Group (Separation/Integration), the main effect of Group was significant, $F(1,166) = 8.71$, $MSE = 2.05$, $p < .01$, with the Separation group outperforming the Integration group. The main effect of Pair Type and interaction were both not significant, $Fs < 1$, $p > .4$. This differs from Experiment 1, where both the aggregate data and the Subjective-Integration group showed proactive interference on average. Although speculative, it is possible that the proactive interference effect was not due to the integrative strategy itself, but rather that those participants who experienced proactive interference were those who were motivated to adopt an integrative strategy. In Experiment 2, participants were told which strategy to use, so this selection effect is not possible. Regarding retroactive interference, although a main effect in Experiment 2, the effect is nominally larger for the Separation group, which is consistent with what was found for the subjectively reported strategy groups in Experiment 1.

[1]These two ratings were included to check that participants were able to use the Separation and Integration strategies as instructed. However, we later discovered a possible ambiguity in this wording: asking participants to indicate "whether there were any other words incorporated into the image of the pair" does not make it clear whether to indicate if they had incorporated another word into the image or to indicate if they knew another word was also presented.

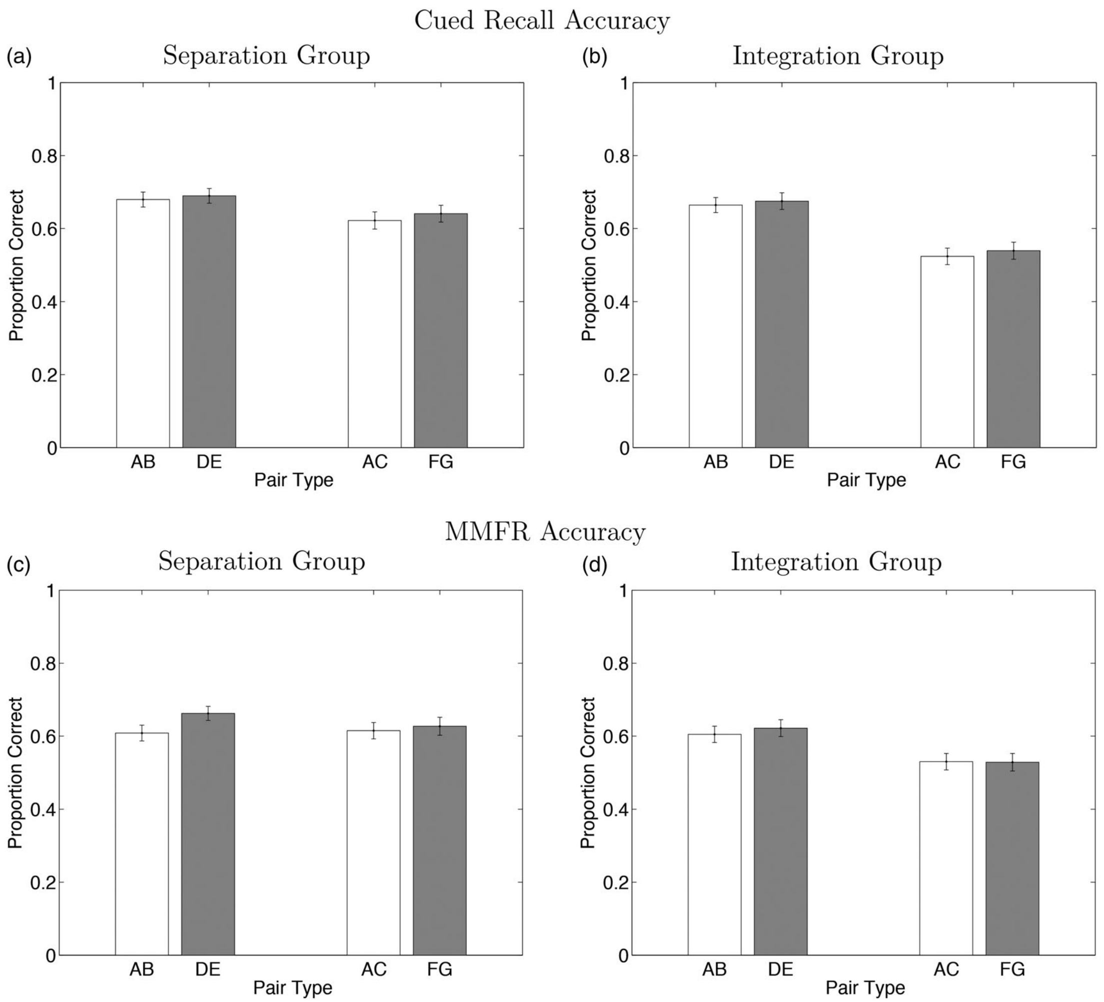

Figure 7. Experiment 2, cued recall (a, b) and MMFR (c, d) proportion correct, for participants in the Separation group (a, c) and the Integration group (b, d). Error bars plot 95% confidence intervals corrected for subject variability (Loftus and Masson, 1994).

MMFR correlations

Pooled correlations. For the Separation group, Q_{BC} (0.34) was not significantly greater than Q_{EG} (0.28, $Z_{log-OR} = 0.82$, $p = .41$), failing to reject independence on average (Figure 8a). For the Integration group, Q_{BC} (0.53) was significantly more positive than Q_{EG} (0.28), $Z_{log-OR} = 3.66$, $p < .001$, consistent with facilitation. The control correlation, Q_{EG}, did not differ significantly between groups, $Z_{log-OR} = 0.10$, $p > .5$, but the critical correlation, Q_{BC}, was significantly greater for the Integration group than the Separation group, $Z_{log-OR} = 2.99$, $p < .01$. The difference, Q_{BC}–Q_{EG} was in the same direction, and significant, $Z_{log-OR} = 2.04$, $p < .05$. These results are in line with the idea that when participants use the Integration strategy, associative facilitation occurs, whereas when they use the Separation strategy, overlapping associations are recalled independently, supporting Hypotheses 2 and 3.

Unpooled correlations. As in Experiment 1, removing the effects of subject variability (Figure 8b) decreased all correlations.

For the Separation group, Q_{EG} (median = .032) did not reach significance, $t(86) = 1.82$, $p = .07$, but Q_{BC} (median = .087) was just barely significantly positive, $t(86) = 2.01$, $p = .048$. Q_{BC} and Q_{EG} were not significantly different, $t(86) = 1.07$, $p > .2$. As with the pooled correlations, this suggests independence, on average, for the Separation group.

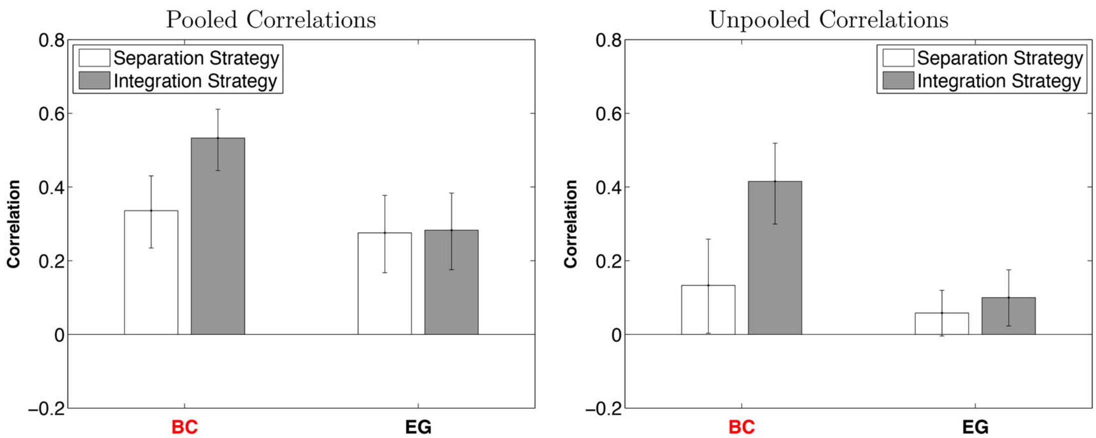

Figure 8. Experiment 2, MMFR correlations. (a) Pooled approach (see text). **BC**, Q_{BC}, the correlation of interest, measuring the relationship between recall of AB and AC. **EG**, Q_{EG}, the control correlation, between DE and EG pairs, computed via a bootstrap, estimating the correlation expected between independent MMFR measures paired within study sets. Error bars plot 95% confidence intervals computed via the log-odds transform. (b) Unpooled approach (see text).

For the Integration group, Q_{EG} (median = .087) was significant, $t(80) = 2.55$, $p < .05$, and Q_{BC} (median = .40) was significantly positive, and robustly so, $t(80) = 6.52$, $p < 10^{-7}$, and Q_{BC} was greater than Q_{EG}, $t(80) = 4.56$, $p < .0001$. Consistent with the pooled correlation analyses, this indicates facilitation, on average, for the Integration group.

Finally, the control correlations were not significantly different between groups, $t(166) = 0.83$, $p > .4$, but Q_{BC} was significantly more positive for the Integration group than the Separation group, $t(166) = 3.24$, $p < .01$, as was the difference, $Q_{BC} - Q_{EG}$, $t(166) = 2.58$, $p < .05$. Thus, manipulating instructed strategy influenced the relationship between AB and AC, from apparent independence (Separation group, consistent with Experiment 1) to facilitation (Integration group).

Individual-differences analysis

As in Experiment 1, CDFs of Q_{BC} were plotted with their corresponding control for independence, Q_{EG} (Figure 9, panels a and b, for the Separation and Integration groups, respectively). To directly compare the groups with one another, as well as to the distribution of correlation values from Experiment 1, all three distributions are plotted in panels c and d (Q_{EG} and Q_{BC}, respectively). First, reassuringly, the control-correlation distributions are quite similar across the three group (panel c), and did not differ from one another significantly by Kolmogorov-Smirnoff tests, $|d| < .12$, $p > .5$. Second, note that the distribution of Q_{BC} values for the Separation group resembles that from Experiment 1, and a Kolmogorov-Smirnoff test was not significant, $d = -.13$, $p > .4$. For the Integration group, the distribution of Q_{BC} was shifted to the right from the distribution of Q_{EG} values (nearly a relationship of stochastic dominance); as already reported, the medians differed significantly, and not surprisingly the Kolmogorov-Smirnoff test was also significant, $d = .30$, $p < .01$. The relationship between the distribution of Q_{BC} for the Integration group compared to the Separation group and Experiment 1 Q_{BC} values (panel d) was one of approximate stochastic dominance (i.e., at nearly every percentile value, the correlation value is more positive for the Integration group than for each of the other samples). Correspondingly, Kolmogorov-Smirnoff tests were significant comparing the Integration group with the Separation group, $d = .29$, $p < .01$ and with Experiment 1, $d = .25$, $p < .05$.

In sum, the pattern of cumulative distribution functions further reinforces the idea that strategy shifts the relationship between AB and AC from more likely to be independent (with the Separation strategy), to more likely to be facilitatory (with the Integration strategy). Experiment 2 thus provided further support for the hypothesis that variation in strategy adoption can produce wide variability in the correlation between memory for AB and AC.

General discussion

Taken together, the results suggest a range of possible ways in which memory for AB and AC could relate to one another. We first address several ways in which

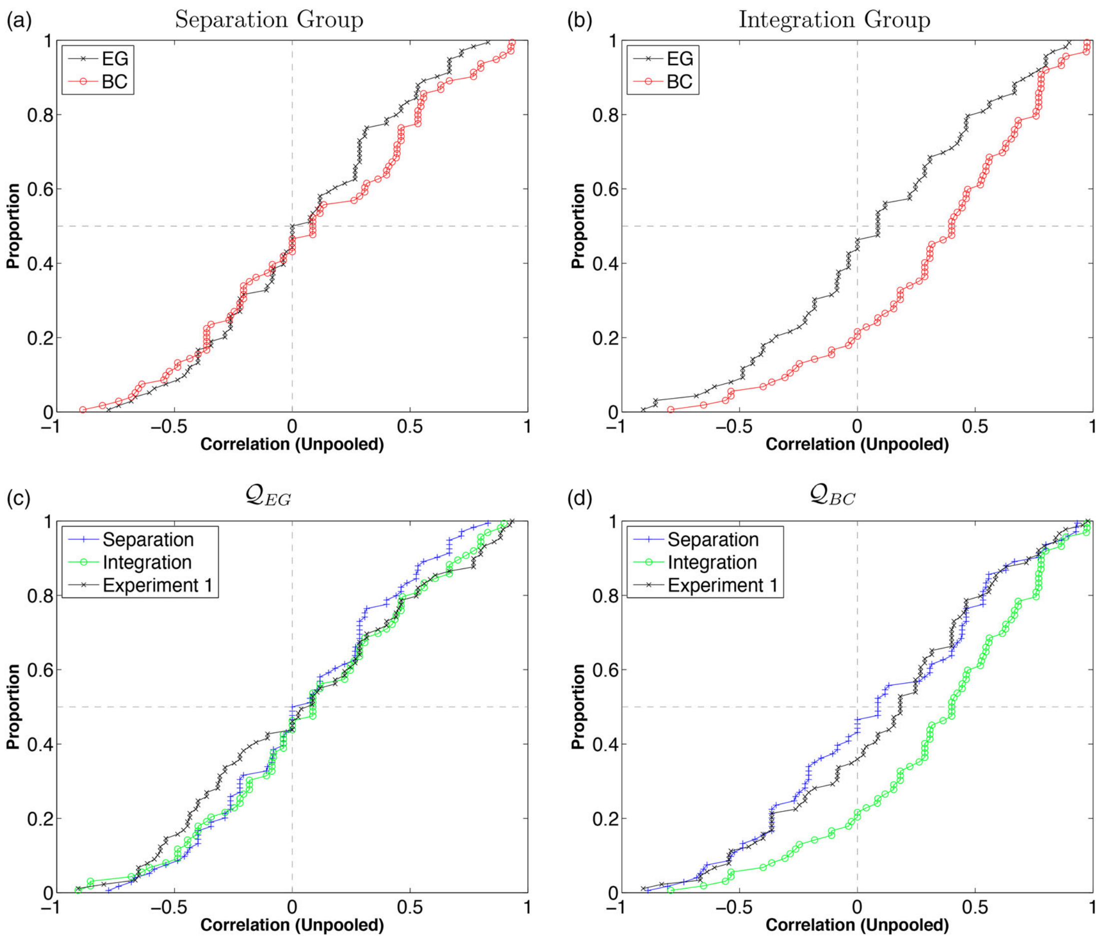

Figure 9. Cumulative distribution functions of Yule's Q values; EG – Q_{EG}; BC – Q_{BC}. Each point represents one participant. Dashed grey lines denote the median (Proportion = .5) and numerical independence (Correlation, Q = 0).

the classic interpretation (associative independence) and the modified interpretation (underlying competition) may need to be reconsidered. We then discuss implications for related memory paradigms, and conclude by discussing the implications of our results for models of association-memory.

The aggregate findings. First, Experiment 1 replicated Tulving and Watkins (1974), finding that the face-value correlation between AB and AC for noun-pairs, with a single study trial for each pair, is positive. This supports Hypothesis 1, but the interpretation was specified only by additional analyses. Compared to a control for independence, the aggregate pattern for strategy-uninstructed participants (Experiment 1) and participants instructed to use a Separation strategy (one group in Experiment 2) was consistent with independence; for participants instructed to use an Integration strategy (the other group in Experiment 2), the aggregate pattern supported facilitation.

Individual variability and its effects on the correlation. Our second hypothesis was that there may be a mix of AB/AC relationships across participants. The cumulative distribution functions were broad, but the distributions of values of the independence-control correlation were similarly broad. With the current sensitivity the two distributions were not significantly different. Thus, although a mix of spontaneous strategies may be present, this does not appear to affect the general finding of independence when no strategy instructions are given; our findings suggest that independence under these conditions not only characterizes the aggregate pattern, but is sufficient to explain individual variability as well.

Strategy can influence the correlation between AB and AC. Our third hypothesis was that strategy could influence the correlation between AB and AC. First, the existence of multiple strategies for handling associative interference was reinforced by an analysis of post-experimental subjective reports of spontaneous strategy adoption in Experiment 1. Although suggestive, self-reported strategy could not robustly differentiate the correlation between AB and AC. However, the results of Experiment 2 show that even a gentle strategy manipulation (without detailed instruction and without extensive training) can create between-group differences in Q_{BC}. The relationship between AB and AC for participants using the Integration strategy was more likely to be all-or-none —the A, B and C words may all be associated into an ABC unit.

The distributions for participants in the Separation strategy group in Experiment 2 are similar to those for the strategy-uninstructed participants of Experiment 1 (Figure 9). This suggests either that participants have a strong tendency to spontaneously adopt Separation-like strategies, or that the spontaneously adopted strategies, at least in this respect, resemble the Separation strategy. It may be that participants in the uninstructed condition guess that, to avoid confusing AB and AC, the better strategy would be to avoid thinking about AB while studying AC. They would not be so wrong with respect to these task parameters, in that accuracy was not better for the Integration group (Figure 7). In contrast, the Integration strategy seems to deviate more from spontaneously adopted strategies. An important difference between the two experiments is that participants in Experiment 2 were asked to use visual imagery, whereas participants in Experiment 1 were not. We can assume that Experiment-2 participants applied visual imagery more frequently than Experiment-1 participants, who were not given instructions to use imagery (cf. the Imagery versus No-Set groups of Paivio & Yuille, 1969). However, the similarity of the control correlations across the three groups, and the similarity of the interference correlation, Q_{BC}, between the Separation group of Experiment 2 and Experiment 1 participants, suggests that visual imagery did not play a major role in determining the relationship between AB and AC. It is more plausible that the positive (facilitatory) outcome in the Integration group in Experiment 2 was caused by actual linking of A, B and C into a single concept.

Initial overlearning might distort the correlations. Our fourth hypothesis, in line with Postman and Underwood (1973), was that the conventional procedure of ensuring a high degree of initial AB learning might remove important sources of variability in behaviour, and thereby produce an artefactual independence result. A number of studies have reported independence in terms of the correlation between responses and the degree to which the AB and AC associations were learned, and a number of these examples of independent recall come from the re-analysis of previous research (e.g., Greeno et al., 1971; Martin, 1971). In the Wichawut and Martin (1971) study as well as re-analyses by Martin (1971), the AB pairs were learned to a perfect or nearly perfect (e.g., 7/8 pairs recalled correctly) criterion. In each of the six data sets that are discussed in these three articles, independence was reported. Our re-analysis exercise (Figure 4) showed that confining the calculation to highly learned AB pairs can indeed shift the correlation toward zero; in our case, the correlation started positive, and the AB-screened analysis reduced in magnitude, but in other data sets it is conceivable that an underlying negative correlation was made more positive due to over-learning. Mensink and Raaijmakers (1988) suggested this and demonstrated it in their model. Thus, past results may have been biased toward independence, not due to the lack of a relationship between recall of B and C, but rather a lack of meaningful variability due to the high degree of learning. To the best of our knowledge, there is only one reported zero BC correlation where AB associations were not initially learned to a high criterion: Da Polito's (1966) doctoral dissertation (results summarized in Greeno et al., 1971), which was the first reported independence finding.

Facilitation between AB and AC and sources of spurious positive correlations. Previously, positive correlations between the recall of overlapping associations in MMFR have been reported, but have not been interpreted as associative facilitation (e.g., Postman & Gray, 1977). In fact, Tulving and Watkins (1974) stated that their positive correlation between recall of B and C might not be an important finding (p. 188). They rightly referred to Hintzman's observation that subject-variability artefacts (and a related problem, item-variability artefacts, discussed next) make it difficult to interpret conditional-recall data (Hintzman, 1972, 1980), due to Simpson's Paradox. Postman and Underwood (1973) suggested that the relationship between individual B and C items was not theoretically relevant, or at least could not be assessed by aggregate contingency tables generated from AB/AC learning when either the AB association or the AC association (or both) has been learned to a criterion.

The experiments and analyses discussed here solve these problems in three ways. First, neither the AB set nor the AC set is learned to a criterion. Therefore, the number of trials required to learn AC does not affect the strength of AB. Second, the Q_{EG} correlation is included as a control; this control correlation is also influenced by subject differences (pooled correlation) and study-set differences (both pooled and unpooled correlations), to a similar degree to Q_{BC}. Independence can be tested by comparing the interference and control correlations rather than by the value of the interference correlation alone. Third, subject variability was excluded in the unpooled correlations. Consequently, we found evidence that recall of AB/AC associations can be either independent (aggregate pattern in Experiment 1 and the Separation group of Experiment 2) or even mutually facilitatory (Integration group of Experiment 2).

The issue of item variability is subtler. Hintzman (1972) argued that because AB and AC have a common item A, it is possible that some A items may be easier to form associations with than other A items. If so, this would be expected to inflate the correlations in a positive direction, which might explain away the facilitation result. The control for independence, Q_{EG}, the correlation between the control pairs (which have no such shared variability due to items), would not control for this inflation. However, the effect of the strategy manipulation (Experiment 2) argues against variability in the general associability of the shared A item producing the facilitation effect. Both the Separation and Integration strategies should be susceptible to this artefact. Instructions for both strategies involved the same basic kind of mediation: imagery. If characteristics relevant to mediators, such as imageability, were inflating the correlations, this too would be controlled. The same stimuli were used for both strategies. Thus, the only difference between the two groups was the instruction to link (Integration) or keep distinct (Separation) the AB and AC pairs. This strategy manipulation shifted the entire distribution of BC correlation values more positive for the Integration group (Figure 9c,d), and was very significant as well as sizeable. This implies that the facilitation result is unlikely to be due to shared variance due to the common A items. Finally, with a double-function procedure, Caplan et al. (2014) found large negative correlations between pairs sharing an item (in double-function pairs, the shared item is in a different position in each pair). In that data set, a face-value negative correlation suggestive of associative competition would have had to overcome any such item effects, which suggests that, if present, item-variability effects may be small.

Effects of materials. Prior studies have found face-value independence but, as previously explained, this is better interpreted as a competitive effect offset by artefactual sources of positive correlation between memory tests (Hintzman, 1972, 1980; Riefer & Batchelder, 1988). Bearing this subtlety in mind, our findings stand in contrast to the classic results. Our pattern, in the aggregate, was quite close to independence in Experiment 1 and in the Separation group of Experiment 2. One important difference between our experiment and most prior studies may be the materials. Classic studies have typically used heterogeneous pairs; a common choice of materials was non-words (consonant-vowel-consonant "nonsense" syllables) paired with two-syllable adjectives. These materials might be more difficult to link together into a meaningful concept to support later memory. In contrast, noun–noun pairs, such as the stimuli we used here, and used by Tulving and Watkins (1974), may be easier to associate together, including, perhaps, into ABC triads, as participants using Integration-like strategies may do. Thus, our materials may have been more amenable to strategies that overcome associative interference, compared to materials used in prior AB/AC learning studies. If this is the case, one would need to explain why Caplan et al. (2014) found starkly negative correlations, reflecting competition, even in the aggregate data, in double-function (AB/BC) pairs learned in a single study list. At the very least, there may be a limit to how much meaningful or associable materials can overcome associative competition.

Relative effectiveness of the strategies. One might wonder to what degree spontaneously adopted strategies really resemble our instructed Integration and Separation strategies, which could be viewed as extreme. Figure 9d shows the three CDFs from both experiments plotted together. The Separation distribution resembles the distribution of Experiment 1, whereas the Integration distribution deviates from the other two, in the positive direction. This suggests that, not given specific strategy instructions, our Experiment-1 participants adopted strategies that were less integrative and thus produced an aggregate pattern that was closer to independence. Possibly, in these experimental conditions, the Integration strategy was not optimal for handling associative interference in terms of maximizing overall accuracy, which was suggested in the analyses of accuracy in list 2 (AC and FG pairs). If so, this may explain why not all participants in Experiment 1 reported spontaneously adopting an Integration-like strategy. In different conditions (perhaps with longer study time or more instruction and training on how to use the strategy), the Integration strategy might outperform the Separation strategy. It is also possible that the Integration strategy requires more effort, and possibly more time, so it may require more practice or additional incentives to be able to implement effectively. This kind of effect has been found, for example, with the peg list method applied to serial learning (e.g., Bugelski, Kidd, & Segmen, 1968).

Stimulus encoding theory. According to stimulus encoding theory, the A item takes on different properties depending on its context—either the B or the C item (Delprato, 1972; Martin, 1971). Delprato (1972) contended that stimulus encoding theory predicts independent recall of B and C in MMFR because, as Wichawut and Martin (1971) showed, the different encodings of A (A_B and A_C) are not interdependent and therefore there is no reason to expect that their associates would be. The results of the Separation strategy are consistent with stimulus encoding theory: B and C associates were recalled independently of each other and this may be because AB became independent of AC. It is hard to imagine that it would be useful to an Integration-adopting participant to learn two different forms of A, since the three items are to be encoded together. Thus, stimulus encoding theory may be incompatible with integrative strategies, and this is consistent with positive statistical dependence, rather than independence, between AB and AC for the Integration group.

Pooled versus unpooled calculations of correlations. We compared the conventional, pooled approach to computing correlations, collapsing data from all participants together, to the unpooled approach, calculating a correlation for each participant. One could ask, which approach is generally better? Both methods require a separate estimation of the correlation expected for independent memory tests to be properly interpreted. However, here, the two calculations produced qualitatively similar results. This convergence suggests that, for these particular data sets, there is no need to worry too much about our choice of pooled versus unpooled calculations. However, it is possible that in some experiments individual differences may act in different ways than correlations within-subjects, so in principle more caution is advised for pooled correlations. On the other hand, unpooled correlations, however, demand more data per participant, which may not be feasible for many experiments. Moreover, in procedures that are feasible within a standard hour-long testing session, a correction may need to be applied to avoid infinities and undefined (0/0) correlation values. This correction may result in a small distortion of the results.

Prior suggestions that integration produces facilitation in AB/AC learning. Facilitation in mean accuracy has been reported previously. Wahlheim and Jacoby (2013) trace this back to Barnes and Underwood (1959), who found that if the B and C items in AB/AC learning were similar to one another (termed the AB/AB′ procedure), proactive facilitation was found. Their participants reported using the B responses to mediate learning of the AC pairs, which might resemble what the Integration participants here were doing. Martin and Dean (1964) found facilitation in AB/AC learning only when participants reported using this mediation strategy. Wahlheim and Jacoby (2013) asked participants to judge whether a pair had been repeated (AB, AB) or repeated and the response item changed (AB, AC). They found proactive facilitation when participants were aware of the change and remembered the B item, and proactive interference otherwise. They suggested that, when they detect such a repetition, participants apply something like Hintzman's (2011) recursive reminding strategy; upon studying AC, they can retrieve AB, and then encode AC embedded within AB, in a manner that preserves their order (and see Jacoby & Wahlheim, 2013; Wahlheim, Maddox, & Jacoby, 2014, for extensions of this work to other paradigms). If this interpretation is correct, it may closely resemble the strategy the Integration participants engaged. Or, perhaps the Integration instructions urged participants to attend to repetitions more, giving them the opportunity to exhibit facilitation. In this case, our findings extend this line of study, suggesting that recursive reminding or integrative representations produce facilitation at the level of the AB/AC pair, in the correlation measure, an effect that can be completely mathematically independent of any changes in mean accuracy. Our Separation strategy might, in turn, lead participants to neglect or actively ignore repetitions of the A items; if so, our findings of independence in the Separation group, although not evidence of interference per se, may at least reflect a reduced number of opportunities to apply recursive reminding or an equivalent process. The parallels with Wahlheim and Jacoby's work also raise the intriguing possibility that the only necessary precondition for associative facilitation (even as quantified by the correlation between recall of B and C, as we do here) is simply retrieval of B while studying AC.

In knowledge assembly theory (Hayes-Roth, 1977), more elementary units of knowledge can become associated with one another, with experience. Additional experience can lead to what she called "unitization", meaning that the entire set of knowledge is accessed at once. It is possible that our Integration strategy resembles the kind of unitization that Hayes-Roth referred to, which could result in either (a) access of all three items at once or (b) access failure and retrieval of none of the three items. This unitization result might explain the positive correlations indicative of associative facilitation.

Relationship to other experimental paradigms. In studies of directed forgetting (the list method), Sahakyan and colleagues (Sahakyan & Delaney, 2003, 2005; Sahakyan & Kelley, 2002) have provided evidence that participants keep list 1 and list 2 separate by changing strategies from list 1 to list 2 (for example, applying shallow versus deep processing). Such a strategy shift, or, more generally, any kind of large contextual shift, could mean that ambiguous associates could be retrieved via fairly separate retrieval routes, namely, cueing with one strategy (or context) versus the other. This is reminiscent of "list discrimination" or "selector" mechanisms proposed by early AB/AC learning researchers (e.g., Postman, Stark, & Fraser, 1968; Postman & Underwood, 1973; Thune & Underwood, 1943; Underwood, 1945, 1949; Wang, 1980; Winograd, 1968) to account for what they interpreted as associative independence. However, when one considers this carefully, contextual shifts may in fact be more likely to be functioning for the Separation than the Integration group. The Separation strategy asks participants to keep AB and AC separate, so a contextual shift would be compatible with the Separation strategy. A contextual shift should moderate what would otherwise be a larger negative correlation between recall of B and C. The Integration strategy asks participants to relate the AB pairs to the AC pairs. This may make it challenging to shift contexts, particularly if context is retrieved whenever an item is retrieved, as in the Temporal Context Model (Howard & Kahana, 1999). One could alternatively argue that, because the integration step cannot occur during study of the first set (AB), but only during the second set (AC), that might function very much like the kind of strategy shift that Sahakyan and colleagues were investigating, and thus the Integration strategy might incorporate something like a contextual shift. This could contribute to the positive value of the correlation for Integration participants, by removing underlying effects of competition, but it is hard to see how a contextual shift could, on its own, produce a positive correlation. An even simpler way of differentiating AB from AC pairs would be to rely on recency, since all list-2 pairs were studied (and tested) more recently than list-1 pairs.

The class of list discrimination and strategy/contextual shift accounts of independence (or reduced competition), including the use of recency for list discrimination, leads to an interesting and testable prediction: these strategies should not be possible when the competing associations are presented within a single study set, as in the double-function list procedure (Primoff, 1938). Using a procedure that combines double-function pairs with the MMFR test procedure, Caplan et al.'s (2014) results support this prediction—namely, a negative correlation was found on average and nearly all single participants' correlations were less than zero.

A different, but related, associative-interference paradigm that has been the focus of a large number of studies in the last decade is termed retrieval-induced forgetting (RIF), developed by Anderson, Bjork, and Bjork (1994). Pairs of the form AB, AC are studied within a single study set. Typically, the A items are category labels and the B and C items are corresponding category members. Retrieval practice is given to one of the overlapping pairs (e.g., given A and a word-stem cue for B) but not the other (AC).

This reduces subsequent memory for the non-practised pair, AC. Note that the retrieval-induced forgetting procedure starts with stimuli that have an AB/AC relationship to one another. Given the evidence, here, that AB and AC can be independent or facilitatory (and in other conditions, possibly competitive), it may be that retrieval-induced forgetting depends critically on the starting conditions of the AB/AC pairs. Interestingly, integration-based strategies have been suggested to overcome interference in retrieval-induced forgetting as well (Anderson & McCulloch, 1999). If integrative strategies also induce a facilitatory relationship (i.e., a positive correlation, regardless of what mean-accuracy measures show) between AB and AC associations in retrieval-induced forgetting procedures, that may somehow counteract suppression during retrieval-practice. However, Smith and Hunt (2000) found that this effect may be undermined when participants are asked to seek similarities between items belonging to a common category cue, so there may be important boundary conditions for the effect of integration on associative interference.

When word-pairs are embedded within simple, meaningful sentences, sentences involving words that appear in more sentences can take longer and be more error-prone in a later recognition task (similar to associative recognition) than sentences containing words that appear fewer times (Anderson, 1974). This is called the "fan effect", where fan quantifies associative ambiguity; AB/AC pairs would have fan = 2, whereas the control pairs (DE and FG) would have fan = 1. The fan effect, therefore, can be viewed as a kind of associative interference, but it has boundary conditions. Radvansky and colleagues have argued that if high-fan sentences can be combined within certain types of "situation" models (e.g., Radvansky, 1999a, 1999b, 2005; Radvansky, Spieler, & Zacks, 1993), this form of associative interference can be overcome, but see Sohn, Anderson, Reder, and Goode (2004) for an alternative perspective. To our knowledge, fan-effect data have not been analysed at the pair level, as has been done in the AB/AC learning line of research; thus, the correlation between sentences sharing an item is not known. Still, it is possible that situation models also overcome associative interference, possibly even producing facilitation between sentences sharing an item, due to a mechanism similar to our Integration strategy in Experiment 2.

Model mechanisms for associative interference, independence and facilitation. Classic empirical findings are often considered benchmark data for models to account for—and rightly so. As Roediger (2008) argues and summarizes, citing Jenkins's (1979) tetrahedral model, strictly generalizable "laws" are rare in memory-behaviour research, and behavioural patterns nearly always depend on specific parameters of the study, from participant characteristics to encoding conditions, stimulus materials, to retrieval tests. When such findings are qualified by newly identified boundary conditions, modellers may want to change the way in which the benchmark is used to test or constrain the model. In this regard, our findings have important implications: association-memory models must now not only explain associative independence (Experiment 1 and Separation group in Experiment 2) in MMFR tests of AB/AC learning, but must also accommodate both associative competition (Caplan et al., 2014) and facilitation (Integration group of Experiment 2). While this is unlikely to constrain or rule out entire models, variability in the AB/AC correlation suggests that modellers should identify how a given model could range from negative to positive correlations. We first consider models that, to our knowledge, have addressed the apparent-independence result as a benchmark, and then consider other possible mechanisms including some that have been proposed to explain behaviour in different memory-interference paradigms.

Although independence (a numerical correlation value of zero, even without considering appropriate controls) was initially a surprising result (even in 1994 when Chappell and Humphreys reported it with an exclamation point!), it may be the least difficult outcome for a model to produce. Three sets of authors, Metcalfe Eich (1982), Mensink and Raaijmakers (1988) and Chappell and Humphreys (1994), found that independence naturally fell out of their respective models (the convolution-based Composite Holographic Associative Recall Model, CHARM; a concatenation-based association-memory model based on the Search of Associative Memory, SAM; and an auto-associative neural network model, respectively) without the need for any additional assumptions. In all cases, this was presented as a positive success of the model. Independence comes about easily when the model does not assume that prior knowledge of AB affects learning of AC; in this case, the random encoding strengths of each will be independent. If one further does not assume that competition at test drives response accuracy, recall of B and recall of C will remain independent. None of these modellers considered the controls for independence we emphasize here, so in some sense it is still not clear how these models stack up against their data. Metcalfe Eich (1982) assumed that associations were learned independently, but there was response competition at retrieval. Interestingly, although non-significant, her model produced Yule's *Q* values that were nominally negative (when one recalculates *Q* from her published contingency table), which, with sufficient power, might produce a result more consistent with associative competition, even relative to a control for independence.

Mensink and Raaijmakers (1988) adapted SAM (Raaijmakers & Shiffrin, 1981) such that each item "image" in SAM was replaced by an image of a concatenation of the A and B items, and only one such compound image occupied the short-term store at any given time. The authors commented that learning of AB and AC were independent by design. However, their careful follow-up modelling investigations showed that their model had several ways to produce positive and negative correlations as well. For example, negative correlations could be contributed by competition during item-sampling itself, and especially if sampling were not reset between the two responses. Interestingly, item-variability effects also produced a small negative correlation. Incrementing (output encoding) during MMFR also produced a small negative correlation. In contrast, subject variability produced a positive correlation, as expected. Thus, a model as complete as Mensink and Raaijmakers's may already possess enough degrees of freedom to produce a range of correlations; although small in their model, these mechanisms might be amplified to produce larger effects and enable such a model to explain AB/AC correlations over a broader range of experimental procedures.

Chappell and Humphreys (1994) also assumed independent learning of AB and AC, but response competition at test. When we recalculated *Q* from their published contingency table values, their model produced a (non-significant) Yule's *Q* value of .42. Notably, this is positive, but is presumably due to item differences. They assumed this would have to be offset, and they suggested offsetting it with the use of distinct contexts for the AB and AC associations (see discussion of Sahakyan's work below). However, taking into account our results, it is not obvious that their model would need to be modified to accommodate average behavioural data based on materials and the single-trial procedure we report here. In associative recognition of AB/AC pairs, Dyne, Humphreys, Bain, and Pike (1990) found that associative interference increased both hit and false-alarm rates, with no net change in sensitivity. Thus, AB and AC can co-exist in memory, at least when associative recognition is used as the memory test.

Associative facilitation might be plausibly implemented simply by assuming that when confronted with an A item, the model attempts to retrieve any prior association before encoding the new one (resembling a strategy investigated by Postman & Gray, 1977). If successful, the prior association is re-encoded. It is reasonable to assume that encoding variability is temporally autocorrelated. If so, some variability in encoding strength could be due to slow fluctuations in attention leading to subsequent recall of B and C being positively correlated. However, this would have to overcome the retrieval of B stealing study time away from AC, which would tend to push the correlation between AB and AC more negative. A more complicated, perhaps less parsimonious, means of modelling the Integration strategy would be to take it more literally and assume that participants construct short triple-item representations, ABC. If ABC were accessed during MMFR, then there should be some positive correlation between recall of B and recall of C simply due to a single random encoding strength applied to the entire ABC triple. This common variability would drive recall of both B and C.

Finally, Hayes-Roth (1977) reported data in a variant of AB/AC learning (study of AB and AC followed by associative recognition of AB and AC with confidence ratings), suggesting that the relationship between retrieval of AB and retrieval of AC is highly non-monotonic. For low AB strengths, the correlation was positive, for moderate AB strengths the relation became negative, and for high AB strengths the relation became more gently positive. If a model were designed to incorporate this kind of non-monotonicity, it might very well be sufficient to account for the kind of variability we observed in participants as a result of the strategy manipulation, in both the sign and magnitude of the BC correlation.

Conclusion. In sum, our findings suggest that independence was obtained in prior studies due to numerical coincidence. A zero correlation is better thought of as a combination of positive correlations, signalling associative facilitation, and negative correlations, due to unresolved competition. With certain procedures (noun-pairs, with each pair studied once), spurious sources of positive correlation can become even more pronounced; in these conditions, independence appears to explain the aggregate pattern and is even sufficient to account for the spread of correlation values, but with certain strategies facilitation is even possible. Individual differences in handling of associative interference, including variability in strategy, may hold the key to understanding the range of ways in which conflicting associations are handled.

References

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 6, 451–474.

Anderson, M. C., & Bell, T. (2001). Forgetting our facts: the role of inhibitory processes in the loss of propositional knowledge. *Journal of Experimental Psychology: General*, 130(3), 544–570.

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1063–1087.

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(3), 608–629.

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.

Bower, G. H., & Reitman, J. S. (1972). Mnemonic elaboration in multilist learning. *Journal of Verbal Learning and Verbal Behavior*, 11, 478–485.

Bugelski, B. R., Kidd, E., & Segmen, J. (1968). Image as a mediator in one-trial paired-associate learning. *Journal of Experimental Psychology*, 76(1), 69–73.

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.

Chappell, M., & Humphreys, M. S. (1994). An auto-associative neural network for sparse representations: analysis and application to models of recognition and cued recall. *Psychological Review*, 101(1), 103–128.

Delprato, D. J. (1972). Pair-specific effects in retroactive inhibition. *Journal of Verbal Learning and Verbal Behavior*, 11, 566–572.

Dyne, A. M., Humphreys, M. S., Bain, J. D., & Pike, R. (1990). Associative interference effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(5), 813–824.

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.

Greeno, J. G., James, C. T., & DaPolito, F. J. (1971). A cognitive interpretation of negative transfer and forgetting of paired associates. *Journal of Verbal Learning and Verbal Behavior*, 10 (4), 331–345.

Hayes-Roth, B. (1977). Evolution of cognitive structures and processes. *Psychological Review*, 84(3), 260–278.

Hayman, C. A. G., & Tulving, E. (1989). Contingent dissociation between recognition and fragment completion: the method of triangulation. *Journal of Experimental Psychology: Learnings, Memory, and Cognition*, 15(2), 228–240.

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

Hintzman, D. L. (1980). Simpson's paradox and the analysis of memory retrieval. *Psychological Review*, 87(4), 398–410.

Hintzman, D. L. (2011). Research strategy in the study of memory: fads, fallacies, and the search for the "coordinates of truth". *Psychological Science*, 6(3), 253–271.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4), 923–941.

Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: the role of recursive remindings in recency judgments and cued recall. *Memory & Cognition*, 41(5), 625–637.

Jenkins, J. J.(1979). Four points to remember: a tetrahedral model of memory experiments. In L. S. Cermak & F. I. M. craik (Eds.), Levels of processing in human memory (p. 429–446). Hillsdale, NJ: Erlbaum.

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

Kanungo, R. (1967). Meaning mediation in verbal transfer. *British Journal of Psychology*, 58(3/4), 205–212.

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

Luce, R. D. (1959). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (pp. 103–189). New York: Wiley.

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

Martin, R. B., & Dean, S. J. (1964). Implicit and explicit mediation in paired-associate learning. *Journal of Experimental Psychology*, 68(1), 21–27.

Melton, A. W., & Irwin, J. M. (1940). The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology*, 53(2), 173–203.

Mensink, G.-J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, 95(4), 434–455.

Metcalfe Eich, J. (1982). A composite holographic associative recall model. *Psychological Review*, 89(6), 627–661.

Paivio, A., & Yuille, J. C. (1969). Changes in associative strategies and paired-associate learning over trials as a function of word imagery and type of learning set. *Journal of Experimental Psychology*, 79(3), 458–463.

Postman, L. (1964). Studies of learning to learn II. *Changes in transfer as a function of practice. Journal of Verbal Learning and Verbal Behavior*, 3, 437–447.

Postman, L., & Gray, W. D. (1977). Maintenance of prior associations and proactive inhibition. *Journal of Experimental Psychology: Human Learning and Memory*, 3(3), 255–263.

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.

Primoff, E. (1938). Backward and forward associations as an organizing act in serial and in paired-associate learning. *Journal of Psychology*, 5, 375–395.

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

Radvansky, G. A. (1999a). The fan effect: a tale of two theories. *Journal of Experimental Psychology: General*, 128 (2), 198–206.

Radvansky, G. A. (1999b). Memory retrieval and suppression: the inhibition of situational models. *Journal of Experimental Psychology: General*, 128(4), 563–579.

Radvansky, G. A. (2005). Situation models, propositions, and the fan effect. *Psychonomic Bulletin & Review*, 12(3), 478–483.

Radvansky, G. A., Spieler, D. H., & Zacks, R. T. (1993). Mental model organization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(1), 95–114.

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.

Roediger, H. L. I. (2008). Relativity of remembering: why the laws of memory vanished. *Annual Review of Psychology*, 59, 225–254.

Sahakyan, L., & Delaney, P. F. (2003). Can encoding differences explain the benefits of directed forgetting in the list method paradigm? *Journal of Memory and Language*, 48, 195–206.

Sahakyan, L., & Delaney, P. F. (2005). Directed forgetting in incidental learning and recognition testing: support for a two-factor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(4), 789–801.

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(6), 1064–1072.

Smith, E. E., Adams, N., & Schorr, D. (1978). Fact retrieval and the paradox of interference. *Cognitive Psychology*, 10, 438–464.

Smith, R. E., & Hunt, R. R. (2000). The influence of distinctive processing on retrieval-induced forgetting. *Memory & Cognition*, 28(4), 503–508.

Sohn, M.-H., Anderson, J. R., Reder, L. M., & Goode, A. (2004). Differential fan effect and attentional focus. *Psychonomic Bulletin & Review*, 11(4), 729–734.

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, 181–193.

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.

Verde, M. F. (2004). Associative interference in recognition memory: a dual-process account. *Memory & Cognition*, 32(8), 1273–1283.

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: the critical role of recursive remindings in proactive effects of memory. *Memory & Cognition*, 41(1), 1–15.

Wahlheim, C. N., Maddox, G. B., & Jacoby, L. L. (2014). The role of reminding in the effects of spaced repetitions on cued recall: sufficient but not necessary. *Memory & Cognition*, 40(1), 94–105.

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.

Wichawut, C., & Martin, E. (1971). Independence of A-B and A-C associations in retroaction. *Journal of Verbal Learning and Verbal Behavior*, 10, 316–321.

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

Appendix A

Examples of Integration-like responses to the open-ended questionnaire

- Recalled the word it was originally paired with and tried to remember all three.
- Tried to recall the word it was previously paired with, then repeated all three in my head until the screen went onto the next pair. It was a lot easier to recall the previous word pair when I had linked the words in some meaningful way.
- Picture the object that is the same but change the object that is different. Or apply the same situation but just with different objects. Or try to relate them together as much as possible. Sometimes I just remembered two different situations. The ones that I found more amusing were easier to remember along with pairs that related to each other.
- I connect the two different words (with either meaning, picture …) and then related with that one word. Somehow related all three words together.

- Tried to find a connection that relates the words logically, or try to list all three together.
- When I noticed overlapping words I would incorporate the new word paired with it and the old word paired with it all into one sentence. Ex. First round: TREE & APPLE, Second round: TREE & MONKEY. I would think: The MONKEY picked an APPLE from the TREE.
- By associating all words that appeared twice, I made a mental image of all three items which I was able to call upon if only one of the images was mentioned. It cues the whole image and I just have to decipher what the picture was.
- Try to recall the one from before to remember both it was hard because there wasn't much time.
- I would try and connect the two different (separate) words together. In example, for DAYLIGHT → MOISTURE, BOTTOM. I connected BOTTOM to MOISTURE by thinking of a grungy basement or something of the like. I figured if I could connect the two separate words it would be easy to remember the common variable in both. Connecting them to things that were somehow part of my life also was a strategy of mine. For instance, COLOR and CLOTHING was easy to remember because I wear lots of COLORful CLOTHING.
- Tried to remember which one I had seen first or which strategy I most recently used to remember the word. I also tried to memorize groups of three words if a word was used twice. If none of those worked I tried to choose the word I remembered less well as the second word I had seen paired with the original.
- I would try and find some similarities between the words that had been used more than once. I would try and relate the word which had already been used to the other word by trying to find some connection whether it was in meaning or similarities in the way they words sounded. I also tried to think of that word in a sentence with the other one.
- Tried to visualize the words as a mental picture, e.g., CASTLE/LAWYER = a LAWYER in a CASTLE and when it came up again as LAWYER/APPLE, I would visualize a LAWYER in a CASTLE holding an APPLE.

Appendix B

Post hoc ratings of pairs

In Experiment 2, following the MMFR tests, both groups were asked to rate each pair for the quality of the image they produced and to judge whether they incorporated additional items within the pair. We report the corresponding analyses, because they provide some additional clues as to how the strategies may have functioned, but then note that the interpretation of these findings must be viewed in light of several important caveats.

A 2 × 2 × 2 mixed, repeated-measures ANOVA, with the design Group [Separation/Integration] × List[1/2] × Pair Type [Interference/Control], was conducted on answers to the incorporation question. The main effect of Group was not significant, $F(1,166) = 0.65$, $MSE = 0.214$, n.s. List was a significant main effect, $F(1,166) = 4.53$, $MSE = 0.012$, $p < .05$, with more "yes" responses to list-1 pairs than to list-2 pairs (M = 0.40 and 0.38, respectively). The main effect of Pair Type was also significant, $F(1,166) = 23.83$, $MSE = 0.044$, $p < .001$, with more "yes" responses to AB/AC pairs than to DE/FG pairs (M = 0.43 and 0.35, respectively), as one would expect. Pair Type additionally interacted with Group, $F(1,166) = 8.07$, $MSE = 0.044$, $p < .01$, with a bigger effect of Pair Type for the Integration group (M = 0.44 and 0.31, respectively) than for the Separation group (M = 0.42 and 0.39, respectively), which, although small in magnitude, is the form of interaction one expects if the Integration group were indeed successful in integrating AB and AC together compared to the Separation group. No other effects approached significance.

An ANOVA with the same design was then conducted on the image-quality question. Again, the main effect of Group was not significant, $F(1,166) = 0.076$, $MSE = 3.54$, n.s. The main effect of List was significant, $F(1,166) = 53.1$, $MSE = 3.56$, $p < .001$, with higher image-quality ratings for list-1 pairs than for list-2 pairs (M = 4.85 and 4.47, respectively). Pair Type interacted significantly with Group, $F(1,166) = 4.85$, $MSE = 0.419$, $p < .05$, describing an apparent crossover interaction, with higher image-quality ratings for control pairs (M = 4.72) than interference pairs (M = 4.64) for the Separation group, but higher image-quality ratings for the interference pairs (M = 4.71) than control pairs (M = 4.57) for the Integration group. The interaction between List and Group was also significant, $F(1,166) = 7.87$, $MSE = 0.452$, $p < .01$, with both groups producing greater image-quality ratings for list-1 than list-2 pairs, but to a lesser degree for the Separation group (M = 4.79 and 4.56, respectively) than for the Integration group (M =

4.90 and 4.38, respectively). No other effects approached significance.

The first question was intended to check compliance with the strategy instructions. However, in retrospect, it seems difficult to know how participants in the Separation group understood their task. These participants had been asked not to combine AB and AC into a single image. In one sense, then, we were asking participants to admit non-compliance, and it is hard to know how truthful they were. Furthermore, if participants in the Separation condition were attempting to keep images of AB and AC distinct from one another, they may have nonetheless incorporated additional (extra-experimental) items in the images to enrich them. It is plausible that Separation participants were answering this question in several ways other than what we had intended. Nonetheless, the outcome of the analysis offers some support for the idea that the Integration participants were indeed attempting to combine AB and AC pairs into one image.

The second question also suffers from some ambiguity, as nothing prevents participants from producing new imagery ad hoc, and rating that instead. Still, the groups seemed matched for overall self-reported image quality (no main effect of Group). The Integration group reported better-quality imagery for interference than control pairs, and the reverse for the Separation group, so image quality, at least assessed in this way, does not appear to be able to explain away the accuracy effects (cf. Figure 7).